EMERGING TECHNOLOGIES ACCELERATOR (ETA) PROGRAMS

ATTACHMENT 1

AMENDMENT 0003

1. ARCADE – ARTIFICIAL REASONING FOR CIRCUIT AUTOMATION AND DESIGN ENGINEERING (FY26) - ACTIVE

Program Overview

ARCADE transforms and accelerates electrical circuit design within the Intelligence Community (IC) by leveraging Artificial Intelligence (AI) to develop an intelligent, comprehensive AI-knowledge assistant that extracts key information from vast technical documentation and enables engineers to perform quick, intuitive queries to meet the speed of mission.

The technical information ingested by ARCADE will include text, datasheets, diagrams, schematics, tables, and graphs. ARCADE extracts crucial details about electrical components, their specifications and interfaces to compile them into a comprehensive searchable knowledge platform.

Going beyond a simple search engine, ARCADE aims to suggest and recommend optimal components based on design requirements, constraints, and performance parameters as specified through user-defined prompts. This capability enables engineers to perform swift, targeted queries, dramatically reducing the time it takes to find, compare, and select suitable parts. It also aims to help identify optimal components and potential alternatives more efficiently by providing comparison tables and explanations for why specific components were selected, effectively saving engineers from manually sifting through thousands of documents. Ultimately, ARCADE will lead to faster, more accurate, and more robust circuit designs, ensuring Government missions can be deployed with unprecedented speed and reliability.

Objectives

Successful ARCADE solutions must culminate in the delivery of a fully functional software prototype capable of executing all three task areas below:

INGEST: Intelligent Data Ingestion and Interpretation

ARCADE solutions must be capable of quickly ingesting and interpreting large corpora of complex technical documents and datasheets, which will include specifications, capabilities, tables, formulas, graphs, and reference schematics in primarily PDF format. This involves

interpreting and encoding complex specifications (e.g., voltage ratings, efficiency curves, package size, pinouts, assembly information) into a knowledge base with minimum human involvement. Interpolation of data that expands beyond what is explicitly documented is encouraged (e.g., efficiency curves of a voltage regulator). Engineers must be able to incrementally update the resulting knowledge base (e.g., to add new versions of a datasheet or remove obsolete information).

DIGEST: Optimized Knowledge Base

Transform data from INGEST into an indexed format optimized for intelligent, context-aware queries. This task may include the design of an Application Programming Interface (API) to ensure rapid, low-latency access to the centralized knowledge base. This serves as the information retrieval backend to the SUGGEST interface, where engineers enter design intent and specifications. Inputs are quickly mapped to component performance parameters, specifications, capabilities and paired with direct citations and links to the original manufacturer datasheets for verification. Performers are encouraged to design the knowledge base to support complex relationships between components.

SUGGEST: Intelligent Assistant for Component Selection

Develop an intuitive standalone interface (e.g., chat) that enables engineers to identify optimal components based on specific application requirements. By processing both natural language and technical design parameters, the UI will rapidly provide optimal component suggestions with reasoning to provide rapid, data-driven solutions from the DIGEST knowledge base. The interface must support text-based input, but Performers can propose additional input modalities. The goal of SUGGEST is to allow engineers to easily define design intent and constraints, execute queries, and review results. Follow-up questions and/or suggestions to the engineers to refine their search is allowed. The interface does not need to be integrated with any electronic design automation (EDA) tools.

Program Scope and Limitations

Proposals shall explicitly address the following:

- Approaches that intelligently process technical documents to extract key information about electrical components (including capabilities), while minimizing compute-related costs.
- Approaches which recognize and digitize parameters from both text and graphical representations.
- Approaches which implement a scalable backend to support complex relationships between components and their specifications, enabling sophisticated querying.

- Research which accelerates design workflow for electrical component selection.
- Software solutions which provide an intuitive interface for engineers to leverage.
- Approaches which automate workflow for updating the knowledge base (e.g., ingestion of additional technical documents, removing/modifying outdated documents).
- Performers are encouraged to utilize state-of-the-art LLMs during development; however, final architectures should remain adaptable for customers requiring self-hosted, closed network deployments.
- Underlying theory: Proposed strategies to meet program-specified metrics must have firm theoretical bases that are described with enough detail that reviewers will be able to assess the viability of the approaches. Proposals shall properly describe and reference previous work upon which their approach is founded.
- Research & Development approach: Proposals shall describe the technical approach to meeting program metrics.
- Technical risks: Proposals shall identify technical risks and proposed mitigation strategies for each identified risk.
- Software development: Proposals shall describe the approach to software architecture, modularization, and integration. If the Performer leverages backbone language models, the ability to "switch out" and retrain using future language model releases is highly desirable.

The following areas of research are out of scope for the ARCADE program:

- Research that does not have strong theoretical and experimental foundations.
- Approaches that require significant manual effort.
- Approaches that address component placement, and Printed Circuit Board (PCB) layout.
- EDA tool connectivity and integration is not required.
- Development of hardware.
- Research that utilizes proprietary data that restricts the Government's ability to transfer the technology or data.
- Approaches that consist merely of integrating currently existing software.

Program Specific Terms/Definitions

Datasheet: In the context of electronic design, a datasheet for a specific component (like a resistor or an integrated circuit) provides critical information such as its electrical ratings, pin configuration, and operating conditions, which are essential for its proper use in a circuit.

Human or Animal Subject Research applicable to this Program?

N/A

Government Furnished Equipment/Information?

The Government will provide an initial component library and technical documentation to support data ingestion and extraction. This library is expected to expand throughout the program. Performers are encouraged to contribute additional data to enhance the collective knowledge base.

Program Metrics

Performers shall deliver fully functional prototype software to be evaluated against program metrics. All submissions encompassing algorithms, subcomponent modules, and complete end-to-end systems, must be executable codebases that an independent Test and Evaluation (T&E) team can directly run and evaluate using sequestered datasets in their own test environment.

The metrics below are preliminary and may be modified by the Government. Proposers are encouraged to suggest additional metrics and evaluation constructs based on their approach and expertise. Metric scores will be calculated by the T&E Team. The test set (component library) will contain a variety of passive, active, and electromechanical components.

a) **Data Interpretation Accuracy:** The percentage of key component attributes (e.g., electrical specifications, performance parameters, pinout, capabilities, formulas, assembly guidelines) correctly extracted from datasheets and diagrams.
b) **Data Interpolation Error:** For test queries that require numerical reasoning (and/or interpolation) between known information to derive a new answer, the relative error between the true answer (per source of truth or human SME assessment) and the generated answer.
c) **Semantic Constraint Accuracy:** Percentage of test queries whose specifications and design intents (e.g., "output voltage should be 5V," "low power consumption”) are correctly interpreted and applied to retrieve components.
d) **Normalized Response Time:** System response time from user input to response, relative to baseline system of equivalent compute capacity.
e) **Recommendation Accuracy:** Component recommendation accuracy across test queries, calculating the percentage of optimal component retrieval (per source of truth or human SME assessment).

Program Waypoints, Milestones, Deliverables

ARCADE is a 12-month program comprised of two 6-month phases to evaluate performers' progress to meet the program metrics. The program will culminate in the delivery of a fully functional and thoroughly documented software prototype, optimized for deployment by Government engineers within a cloud-based environment.

Table 1 shows a timeline for the program with Government-defined milestones and deliverables.

Event	Months after Kick-off		Deliverables
	Phase 1	Phase 2	
Program Wide Milestone Kickoff	1	6	Read-ahead package is due from Performers to the Government 7 days before meeting. If required by the PM, updates after the meeting are due 15 days after the meeting date.
Technical Review Meetings	Every 2 Weeks	Every 2 Weeks	Read-ahead package is due from Performer to the Government 1 day before meeting. If required by the PM, updates after the meeting are due 7 days after the meeting date.
Site Visits and Waypoint Review	3, 5	8, 10	Site visits (to be held concurrently with Technical Review Meetings)
Software Deliverables	5	11	Source code, build scripts and documentation for facilitating reproduction of system operation by an independent Test and Evaluation Team. Deliverables shall be received at the T&E site specified by the Government no later than the final day of the listed month. Performers shall also provide documentation and training to the Government for proper usage.
Final Technical Report	6	12	Performers shall provide a final technical report at the end of each phase outlining their technical progress, methodology, challenges, successes, and recommendations for future efforts. A draft final report is to be delivered 2-4 weeks prior to the deadline.
Independent T&E	6	12	Upon receipt of the deliverables, T&E will be conducted. Performers may

Event	Months after Kick-off		Deliverables
	Phase 1	Phase 2	
			expect test results within one month of submission.
Financial and Technical Reports	Monthly	Monthly	Monthly financial and technical reports are due by the 10th day of the following month.
Collaborative Transition Workshop		12	A Collaborative Transition Workshop held in the DC area at the end of Program
End of Phase	6	12	Phase Period of Performance ends

ble 1:

Software Deliverable Formatting

Performers will be required to provide algorithms, software deliverables, design documentation, and operating instructions to execute the automated software in a manner that conforms to a standardized industrial method or methods that will be provided at program Kickoff. To facilitate planning, Proposers may assume that the standardized configuration will require the use of software containerization technology (e.g., Docker or Kubernetes). This means that the entirety of a Performer's system, including pre- and post-processing, must be included within the delivered software container. The deliverable package must ensure that a third-party evaluation team can fully reconstruct, and execute the system with the exact model configurations, weights, and dependencies utilized during deployment. Performers shall also provide source code to the Government and T&E team to enable modification of Performer containers to suit deployment constraints. Proposer teams that do not include the requisite expertise to conduct such software development shall include costs in their proposal to obtain software development support.

Each team is required to include among their key personnel a Lead System Integrator (LSI) who shall be responsible for preparing software Deliverable subcomponents, modules, and systems, performing quality control of Deliverables, and integrating key components into the primary ARCADE system(s). The LSI will also oversee communication and coordination across a Performer's research teams including subcontractors, if applicable, to ensure research products are functional and following software coding best practices (e.g., inline comments, documentation). Additional team members and roles are dependent on the proposed research; there is no predetermined or required skill mix.

Place of Performance

Performance will be conducted at the Performers' recommended sites with Government's concurrence.

Test & Evaluation

T&E will be performed by a team or teams selected by the IARPA ARCADE Program Manager and may be modified throughout the program.

T&E will be conducted by an independent team of Government and contractor staff carrying out evaluation and analyses of Performer Deliverables using program test datasets and protocols. In addition to independent T&E, the program will regularly gauge interim progress of Performer research activities towards ARCADE objectives and target metrics using T&E results measured and reported by the Performer teams themselves. Evaluations will occur during M6 and M12 that will exercise Performer solutions across technical challenges.

Proposals must specify the processing dependencies needed to carry out the proposed research and what architecture and library characteristics are necessary for their approach(es) to be successful at meeting program objectives. Performers will have specific Deliverable Milestones at which all subcomponent and system algorithms and software will be delivered to IARPA and its designated T&E Team. The T&E Team will then conduct evaluations at the direction of the ARCADE PM and with the objective of characterizing the quality, functionality, and performance of the ARCADE Deliverables. In addition to quantitative measurements, T&E will be carried out to establish a thorough understanding of the progress, status, and limitations of the Performer's research. T&E results and feedback will be provided to Performers at regular intervals to keep them abreast of current independent performance measurements and to inform and improve their R&D approaches and methods. T&E results from all Performers will be shared with all teams to establish an understanding of the current state and progress of ARCADE research; T&E results will also be shared with United States Government external stakeholders, including contractors, for Government purposes.

Technical Exchange Meetings/Workshops/Site Visits/Travel Requirements

Performers are expected to assume responsibility for administration of their projects and to comply with contractual and program requirements for reporting, attendance at program workshops, and availability for site visits. The following paragraphs describe typical expectations for meetings and travel for IARPA programs as well as the contemplated frequency and locations of such meetings. In addition to ensuring that all necessary details of developed software, algorithm, and operational instructions are clear and complete, each Performer will be required to be available for questions and troubleshooting from the T&E Team during technical status meetings.

All Performer teams are expected to attend a collaborative workshop, to include Key Personnel from prime and subcontractor organizations. The workshop will focus on technical aspects of the program and on facilitating open technical exchanges, interaction, and sharing among the various program participants. Program participants will be expected to present the technical status and progress of their efforts to other participants and invited guests.

Site visits by the Government Team will generally take place at M3 and M8 during the life of the program. These visits will take place at the Performer's facility. Reports on technical progress, details of successes and issues, contributions to the program goals, and technology demonstrations will be expected at such site visits.

Anticipated Timeline

ARCADE is a 12-month Program comprised of two 6-month phases.

2. COSMIC – COMMERCIAL OBSERVATION FOR SPATIO-TEMPORAL MONITORING FOR INDICATIONS OF CHANGE (FY26) - ACTIVE

Program Overview

The aim of the Commercial Observation for Spatio-temporal Monitoring for Indications of Change (COSMIC) program is to formulate a methodology to leverage commercial remote sensing technologies and open source geolocated information to generate pseudo-persistent data (PPD). This will be done by combining and translating complex and novel data, such as non-nadir imagery and non-Red,Green, Blue (RGB) spectral bands, using Artificial Intelligence (AI) and computer vision-based methodologies into a layered, temporal geospatial model understandable by commercial agentic AI systems. COSMIC will enable the development of an agentic AI analytic system trained by commercial vendors and equipped to answer intelligence questions by bridging the gap between the Intelligence Community (IC) data characteristics and the simpler solutions offered by commercial vendors.

COSMIC will foster routine updating and modelling of baseline geospatial information in areas through multi-source, persistent surveillance. New sensors will fluidly be incorporated into the generation of layered geospatial models as constellations and sources change over time. In addition to improving the temporal resolution of layered geospatial models, COSMIC also aims to reduce resource demands by streamlining the derivation of actionable intelligence from these models. By integrating commercial remote sensing data with existing geospatial information, the program aims to create a harmonized and up-to-date representation of the physical world that can be used by agentic processes to answer questions relevant to the IC.

Objectives

COSMIC aims to create a methodology to rapidly integrate multiple sources of information into a unified framework to allow for incorporation into a layered GEOINT model or PPD. Additionally, this source should allow the prediction and visualization of resolution and bands of data not provided for PPD construction through predictive processes such as image diffusion or other generative AI approaches. The COSMIC capability will enable the rapid assessment of intelligence through an AI interrogatable multi-sensor, multi-int data source that combines satellite images and geospatial data from a variety of systems into an interrelated framework. This framework should then be compatible with various state of the art change detection algorithms, object detection, and agentic systems identified by the Government for the aid of intelligence analysts. For example, a change detection algorithm developed for a SAR system might be applied to imagery from a EO system, and the framework would need to be able to translate between the two domains.

Similar to how frontier model LLMs use grounding and retrieval augmented generation to provide up to date information, COSMIC solutions should rapidly be able to ingest new imagery,

data, and data types to update its models and augment the PPD as needed to answer critical questions.

Proposers shall address the following technical challenges (TCs) and objectives to meet the COSMIC goals to develop methodologies to rapidly integrate disparate remotely sensed imagery and geospatial intelligence (GEOINT) into a visualizable framework to create a PPD that is useable by agentic AI in their candidate approaches:

- Incorporation of multi-resolution imagery and GEOINT data;
- Creation of a PPD over varying time scales;
- Incorporation of electro-optical, multispectral, and radar imagery and imagery collected at angles other than NADIR without manual manipulation;
- Incorporation of imagery collected at different altitudes with varying atmospheric effects such as unmanned aerial vehicle (UAV) or satellite imagery;
- Incorporation of non-image-based GEOINT information with temporal elements (eg. weather events, traffic data, building permit location);
- Prediction of missing spectra, resolutions, and temporal images and two-dimensional information;
- Visualization of PPD elements including imagery, data availability, and confidence metric reflective of the likelihood of hallucinations;
- Compatibility with existing change detection algorithms;
- Compatibility with Agentic systems
- Deployability on stand-alone systems
- Ability to identify and discriminate ephemeral events compared with persistent change;
- Generation of imagery of various modalities based on the PPD.

Program Scope and Limitations

Proposers are expected to address the following:

- Methods for relating data sources to each other such as:
    - Super-resolution;
    - Co-registration;
    - Resolution variance compensation methods;
    - Spectral harmonization;
    - Shared embedding spaces;
- Methods of predicting novel spectral bands not represented by data inputted into the PPD;
- Methods of predicting novel points in time at specified resolutions or image types not represented by data inputted into the PPD;

- Methods to incorporate and utilize non-image, high temporal frequency GEOINT data, such as weather or traffic information;
- Methods to improve the temporal resolution of layered GEOINT models
- Research & Development (R&D) approach: White Papers shall describe the technical approach to robustly meet program objectives;
- Software Development: White Papers shall describe the approach to software architecture, modularization, and integration. If the Proposer leverages backbone diffusion models, the ability to "switch out" and retrain using future model releases is highly desirable.

Data sources that are to be ingested into the PPD may include, but are not limited to:

- Multispectral imagery collected from airborne and satellite altitudes and varying view angles;
- Pan EO imagery collected from airborne and satellite altitudes and varying view angles;
- SAR imagery collected from airborne and satellite altitudes and varying view angles;
- Existing GEOINT models
- Traffic data;
- Weather data.

Additional data sources may be available to selected performers and data sources may be removed at the discretion of the Government. Proposed approaches should address how the above data sources may be incorporated.

The following areas of research and approaches are out of scope for the COSMIC program:

- Research that does not have strong theoretical and experimental foundations;
- Development of sensor systems or computation hardware;
- Development of sensor platforms, such as UAVs, vehicles, aerostats, towers, or camera systems;
- Research that utilizes proprietary data that cannot be delivered to the Government;
- Methods that require a human-in-the-loop as part of the integrated end-to-end system;
- Approaches that consist merely of integrating currently existing software;
- Approaches that require non-cooperative real time augmentation to provided data, e.g., opportunistic data from sensors integrated into handheld devices, computers, or automobiles;
- Approaches that create a fixed data cube or embedding space that cannot accept new data;
- Development of novel change detection algorithms.
- Development of novel Agentic AI systems

The use of additional sensors, imaging or sensing modalities, filters, calibration targets, and publicly available ancillary data to assemble training data sets or to support modeling and simulation research is in scope if it is relevant to the proposed research approach.

Delivered software will be evaluated by an independent Test and Evaluation (T&E) team on sequestered and shared evaluation datasets. Performers will build prototype algorithms, and subcomponent modules, and/or systems for end-to-end that will be run and evaluated by the T&E team. Testing protocols do not allow for expert operators, human-in-the-loop operation, or any operations not deemed "turnkey." However, systems or algorithms that have been trained using human-in-the-loop methods may be submitted, provided they run autonomously. The systems will be tested using established change detection algorithms.

Program Specific Terms/Definitions

Pseudo-Persistent Data Source: A pseudo-persistent data (PPD) source is comprised of imagery and geospatial intelligence from multiple sensors, collected at multiple resolutions, and at different times that may be interrogated to answer questions such as "Is their large scale construction occurring in this area" Through the incorporation of many different information sources that are harmonized, the pseudo-persistent data source allows for constant monitoring of an area of regard.

GEOINT: Geospatial intelligence (GEOINT) refers to the intelligence derived from the analysis and exploitation of imagery, signals, or signatures with geospatial information. It involves describing, assessing, and visually depicting physical features and geographically referenced activities on Earth. GEOINT can be collected from multiple sources such as satellites, aircraft, maps, and commercial databases.

Harmonization: Harmonization refers to the processes by which disparate data sources collected over the same geospatial location, but at different times and from different sensors are made compatible to be interrogated over a reference frame.

Agentic AI systems: Agentic AI systems are AI systems that operate with minimal human intervention to solve complex problems in near real time.

Human or Animal Subject Research applicable to this Program

N/A

Government Furnished Equipment/Information

The Government intends to provide example data sets for the purpose of data ingestion. The volume of data is not anticipated to be sufficient for training of algorithms.

Program Metrics

Delivered methodologies will be evaluated by the T&E team(s) on curated test data sets. Performers will build prototype systems that the T&E team will run and evaluate in a closed compute environment. The T&E team will calculate metric scores. Delivered methodologies will be tested against a range of image and GEOINT data emulative of the variety that are available for the employment of change detection algorithms in a variety of locations distributed around the world. Performers will receive information about the change detection algorithms that will be employed in testing.

Candidate evaluation metrics for COSMIC are designed to test the ability of the PPD to facilitate change detection, predict images from novel temporal points, predict images from novel spectra, super-resolution of images, and image coregistration. Metrics may be added, modified, or removed at the discretion of the Government. Proposers are encouraged to suggest metrics and evaluation methods in their white paper submission.

Table 1: Candidate metrics for COSMIC

Evaluation point	Description	Target Performance
Predicted image accuracy (for image products)	Structural similarity between synthesized image and hold out image	0.95
Relative Edge Response (for image products)	Standard image quality metric to reflect changes in smoothness of features in a super-resolved image	0.4
Change detection performance	Performance of change detection algorithm when applied to PPD versus a non-harmonized image collection	F1 improvement of 25% versus non-harmonized image collection. 25% reduction in the time to detect change
Sensor to Sensor Harmonization	Relative uncertainty across PPD measurements of the same location through time	±2.5%
Geolocation Error	Radius of circular error in which 90% of spatial misregistration exists (CE90)	10m
Answer fidelity	Accuracy of answers against benchmark intelligence questions	0.95

Program Waypoints, Milestones, Deliverables

COSMIC is a single phase 18-Month program with 6 evaluation milestones to evaluate Performers' progress in meeting the program metrics. The program will culminate in the delivery of documented software that can be easily deployed by Government analysts in a cloud-based environment. Table 2 shows a timeline for the program with Government-defined milestones and deliverables.

Table 2: Program Waypoint, Milestone and Deliverables Testing Timeline

Event	Months after Kickoff	Deliverables
Program Wide Milestone Kickoff	1	Read-ahead package due from Performers to the Government 7 days before meeting. If required by the PM, updates after the meeting are due 3 days after the meeting date.
Technical Review Meetings	Monthly	Read-ahead package due from Performer to the Government 1 day before meeting. If required by the PM, updates after the meeting are due 3 days after the meeting date.
Site Visits and Critical Design Review	4, 7, 10,13,16	Site visits
Evaluation Milestones (Software and Data Deliverables)	3, 6, 9,12, 15, 18	Performers will submit their software, development datasets and documentation to the T&E partner on the first business day of the month. Source code, build scripts and documentation for facilitating reproduction of system operation by an independent T&E Team. All deliverables must be clearly labelled to identify IP assertions, which may be UNLIMITED or Government Purpose Rights.
Financial and Technical Reports	Monthly	Monthly financial and technical reports are due by the 15th day of the following month in the IARPA provided template.
Collaborative Workshop	12	A Collaborative Workshop held in the DC area
Draft Final Report	17	A draft of the final technical report outlining technical progress, methodology, challenges, successes, and recommendations for future efforts
End of Program, Final Report	18	Period of Performance Ends. Performers shall provide a final technical report outlining their technical progress, methodology, challenges, successes, and recommendations for future efforts.

Software Deliverable Formatting

Performers will be required to provide algorithm and software deliverables in a manner that conforms to a standardized industrial method or methods that will be provided at program Kickoff. To facilitate planning, Proposers may assume that the standardized configuration will require the use of software containerization technology (e.g., Docker and a REST API). This means that the entirety of a Performer's system, including pre- and post-processing, must be included within the delivered software container. For models that require training, the expectation is for the initial model training to occur on Performer systems, with the ability for the T&E Team to re-train and test the model with the same and/or other data.

Each team is required to include, among their key personnel, a Lead System Integrator (LSI) who shall be responsible for preparing software Deliverable subcomponents, modules, and systems, performing quality control of Deliverable, and integrating key components into the primary COSMIC system(s). The LSI will also oversee communication and coordination across a Performer's research teams including subcontractors, if applicable, to ensure research products are functional, integrated and following software coding best practices (e.g., inline comments, documentation). Additional team members and roles are dependent on the proposed research, as such, there is no predetermined or required skill mix.

Place of Performance

Performance will be conducted at the Performers' recommended sites with Government's concurrence.

Test & Evaluation

T&E will be performed by a team or teams selected by the IARPA COSMIC Program Manager and may be modified throughout the program.

T&E will be conducted by an independent team of Government and contractor staff carrying out evaluation and analyses of Performer Deliverables using program test datasets and protocols. Evaluations will occur during months 3, 6, 9, 12, 15, and 18 that will exercise Performer solutions across technical challenges.

Performers are encouraged to develop methodologies for a self-contained / self-hosted environment for use by Government transition partners on their own network/systems. White papers must specify the processing dependencies needed to carry out the proposed research and what architecture and library characteristics are necessary for their approach(es) to be successful at meeting program objectives. Performers will have specific Deliverable Milestones at which all subcomponent and system algorithms and software will be delivered to IARPA and its designated T&E Team. The T&E Team will then conduct evaluations at the direction of the COSMIC PM and with the objective of characterizing the quality, functionality, and performance of the COSMIC Deliverables. In addition to quantitative measurements, T&E will be carried out to

establish a thorough understanding of the progress, status, and limitations of the Performer's research. T&E results and feedback will be provided to Performers at regular intervals to keep them abreast of current independent performance measurements and to inform and improve their R&D approaches and methods. T&E results from all Performers will be shared with all teams to establish an understanding of the current state and progress of COSMIC research; T&E results will also be shared with United States Government external stakeholders, including contractors, for Government purposes.

Technical Exchange Meetings/Workshops/Site Visits/Travel Requirements

Performers are expected to assume responsibility for administration of their projects and to comply with contractual and program requirements for reporting, attendance at program workshops, and availability for site visits. The following paragraphs describe typical expectations for meetings and travel for IARPA programs as well as the contemplated frequency and locations of such meetings. In addition to ensuring that all necessary details of developed software, algorithm, and operational instructions are clear and complete, each Performer will be required to be available for questions and troubleshooting from the T&E Team in weekly or bi-weekly status meetings.

All Performer teams are expected to attend a collaborative workshop, to include Key Personnel from prime and subcontractor organizations. The workshop will focus on technical aspects of the program and on facilitating open technical exchanges, interaction, and sharing among the various program participants. Program participants will be expected to present the technical status and progress of their efforts to other participants and invited guests.

Site visits and critical reviews by the Government Team will generally take place at months 4, 7, 10, 13, and 16 during the life of the program. These visits will take place at the Performer's facility or as a virtual meeting at the discretion of the COSMIC PM. Reports on technical progress, details of successes and issues, contributions to the program goals, and technology demonstrations will be expected during these visits.

Anticipated Period of Performance

COSMIC is an 18-Month Program.

3. DECIPHER - (FY26) - ACTIVE

Program Overview

Analysts often struggle to understand specialized language (e.g., slang, jargon, technical terminology) in various situations where dictionary definitions are not readily available, such as:

- Non-English research articles with novel technical concepts and terms
- Communications using slang
- Undefined acronyms used within limited contexts
- Web forum posts using coded language to mask illicit activities

In these and similar contexts, analysts dedicate significant effort to triaging content pertinent to their domain of interest and understanding the meaning of unknown or novel specialized terms used within that domain. Adding to this challenge, language can change quickly, and adversaries can rapidly adapt their language to evade detection.

The Decipher program's goal is to create capabilities to detect and define specialized language in diverse, multilingual text collections. In Decipher, "specialized language" refers to single or multiword expressions that are difficult to disambiguate (e.g., acronyms, terminology with multiple meanings), are used to deliberately obfuscate interpretation (e.g., coded language), or for which accurate translations or explanations are not readily available (e.g., emergent jargon or slang) to typical non-expert users. Decipher will flag and extract candidate specialized terms in a document or corpus and then will provide candidate plain language translation and annotation of specialized terms, including jargon, slang, and acronyms. The program will also develop technology to annotate text with relevant social and contextual factors, capturing nuanced meaning, as well as to detect concept drift by monitoring use of terminology as it evolves over time.

Objectives

The Decipher capability will enable analysts to quickly interpret small text corpora of interest (e.g., a text message exchange containing limited messages). Additionally, it will enable the Intelligence Community (IC) to scour large, diverse collections, identify those which contain communications pertinent to IC missions, and furnish likely definitions in an easy-to-understand format.

Solutions must perform the following two automated tasks:

Task Area 1 (TA1): Detection: Find specialized language in a large corpus

For a given text collection, detect *specialized language*, which in the Decipher program refers to single or multiword expressions that are difficult to disambiguate (e.g., acronyms, terminology with multiple meanings), are used to deliberately obfuscate interpretation (e.g., coded language), or for which accurate translations or explanations are not readily available (e.g., emergent jargon or slang) to typical non-expert users.

A key challenge of this TA is defining a threshold between specialized and non-specialized language, as this may differ depending on the end user's knowledge of a domain as well as other factors. The Proposer is encouraged to develop creative approaches and draw on techniques from the research literature on lexical complexity prediction to address this challenge. Solutions should work on collections of multilingual content, especially high- and medium-resource languages.

Task Area 2 (TA2): Definition: Generate a glossary of specialized terms using limited context

Given a set of specialized language within a collection, generate a corresponding set of English-language definitions and plain language explanations or glosses for these terms with confidence scores.

Generated definitions should follow lexicographic best practices. Solutions should work on multilingual content and small collections of content containing limited attestations of a term. The TA2 capability should also accurately define common terms, though the focus of this TA is developing a tool that can generate definitions for specialized terms that are rare, unfamiliar, or previously unattested.

Additionally, Performers will develop a simple, user-friendly interface that presents the functionality of the technology for both TA1 and TA2, including specialized terminology definitions in glossary form and as annotations to a document.

Program Scope and Limitations

Proposed approaches shall explicitly address the following:

- Software Development: White papers shall describe the approach to software architecture, modularization, and integration. If the Proposer leverages backbone language models, the ability to "switch out" and retrain using future language model releases is highly desirable.
- Anticipated dataset development plans or preexisting datasets the Proposer plans to use in developing their system.
- Methods to address how the proposed solution will maximally leverage context to generate accurate definitions in situations where only a small number of attestations occur in a collection.

The following areas of research are out of scope for the Decipher program:

- Research that does not have strong theoretical and experimental foundations.
- Approaches that are unable to detect and generate definitions for novel future terms (e.g., slang terms not attested until years after the completion of the program).
- Development of hardware.
- Approaches to develop new methods in cryptology (encoding) and cryptanalysis (decoding).
- Research that utilizes proprietary data that restricts the Government's ability to transfer the technology or data.
- Approaches that consist merely of integrating currently existing software.
- Approaches that only work on English-language materials.

Program Specific Terms/Definitions

Specialized Language: Single or multiword expressions that are difficult to disambiguate (e.g. acronyms, polysemic terminology), are used to deliberately obfuscate interpretation (e.g. coded language), or for which accurate translations or explanations are not readily available (e.g. emergent jargon or slang) to typical non-expert users. The Government, in consultation with the Test and Evaluation (T&E) Team may modify this definition to facilitate evaluation or optimize alignment with mission use cases.

Human or Animal Subject Research applicable to this Program

If human subject research is proposed, it must follow all federal HSR guidelines and requirements, as outlined in the general ETA Program Announcement.

Government Furnished Equipment/Information

While the Government may provide limited sample test data and information about the format of the test set, Performers are expected to develop their own dev-test datasets.

Program Metrics

Performers will build and deliver prototype software systems that the independent T&E Team will run and evaluate in a closed compute environment on sequestered evaluation datasets. Testing protocols do not allow for expert operators, human-in-the-loop operation, or any operations not deemed "turnkey." The T&E team will calculate metric scores. The test set will contain novel specialized language developed to evaluate Performer system performance in a variety of conditions. This collection may contain, for example, a set of text messages containing a conversation about smuggling weapons. Performers will not receive information about the contents of the collections or their topic domains.

The Government encourages proposers to recommend their own metrics to measure success. Good metrics for the purpose of this effort maximize interpretability, allow easy implementation and replicability, allow calculation of confidence bounds, furnish results that are comparable over time, and are actionable (i.e., inform changes in technical approach).

Detection

This metric focuses on the ability to accurately identify specialized language terms. The T&E Team will establish testing protocols and will communicate requirements and standards to Performer teams in advance of test events. For example, within the sequestered T&E environment, the Performer system may be provided at test time with a prose description of the domain of interest, the channel of communication, or other metadata. The Government tentatively plans to use information extraction metrics (e.g., precision, accuracy, and F1) to evaluate how well the Performer systems identify specialized language terms relevant within a given domain of interest. Performance in retrieving documents from a target domain of interest may also be measured relative to a baseline, which may be a general-purpose, state-of-the-art information retrieval capability.

Definition Generation

Metrics for this task focus on measuring how well the Performer systems can generate definitions for novel specialized terms. System-generated definitions for specialized terms in the test set will be evaluated against hidden ground truth definitions. Systems will be evaluated on a combination of automated and human-derived metrics, which will be announced at program kickoff. The Government is especially interested in definition generation capabilities that work on multilingual collections and accurately generate likely definitions despite limited attestations and context.

As measure of baseline performance, the T&E team may use state-of-the-art large language models prompted to generate definitions. Additionally, evaluation may consider how well systems perform on the definition generation task with progressively smaller amounts of available context.

Program Waypoints, Milestones, Deliverables

Decipher is a single phase 12-Month program with four evaluation milestones to evaluate Performers' progress to meet the program metrics. The program will culminate in the delivery of documented software that can be easily deployed by Government analysts using a standard laptop with a Graphical Processing Unit (GPU). Table 1 shows a timeline for the program with Government-defined milestones and deliverables.

Table 1: Program Waypoint, Milestone and Deliverables Testing Timeline

Program Wide Milestone Kickoff	**1**	**Read-ahead package due from Performers to the Government 7 days before meeting. If required by the PM, updates after the meeting are due 15 days after the meeting date.**
Technical Review Meetings	Every 2 Weeks	Read-ahead package due from Performer to the Government 1 day before meeting. If required by the PM, updates after the meeting are due 7 days after the meeting date.
Site Visits and Waypoint Review	3,7	Critical System Design Reviews at Performer Sites
Evaluation Milestones (Software and Data Deliverables)	2,5,8,11	Performers will submit their software, development datasets and documentation to the T&E partner on the first business day of the month. Source code, build scripts and documentation for facilitating reproduction of system operation by an independent Test and Evaluation Team. Performers shall also provide documentation and training to the Government for proper usage. All deliverables must be clearly labelled to identify IP assertions, which may be UNLIMITED or Government Purpose Rights.
Independent T&E	3,6,9,12	Upon receipt of the deliverables, T&E will be conducted. Performers may expect test results and analysis by the final business day of the month.
Financial and Technical Reports	Monthly	Monthly financial and technical reports are due by the 10th day of the following month in the IARPA provided template.

Delivery of deployable software with documentation	11	Performers will submit their software and documentation to the T&E partner on the first business day of the month. To enable transition of the capability, software should run on a standard laptop with a GPU. Performers shall also provide documentation and training to the Government for proper usage. The software packaging must clearly identify IP assertions, which may be UNLIMITED or Government Purpose Rights.
Collaborative Workshop	12	A Collaborative Workshop held in the DC area at the end of Program
End of Program, Final Report	12	Period of Performance Ends. Performers shall provide a final technical report outlining their technical progress, methodology, challenges, successes, and recommendations for future efforts.

Software Deliverable Formatting

Performers shall be required to provide algorithms, software deliverables, design documentation, and operating instructions to execute the automated software in a manner that conforms to a standardized industrial method or methods that will be provided at program Kickoff. To facilitate planning, Performers may assume that the standardized configuration will require the use of software containerization technology (e.g., Docker and a REST API). This means that the entirety of a Performer's system, including pre- and post-processing, must be included within the delivered software container. For models that require training, the expectation is for the initial model training to occur on Performer systems, with the ability for the T&E Team to re-train and test the model with the same and/or other data. Performers shall also provide source code to the Government and T&E team to enable modification of Performer containers to suit deployment constraints. Proposer teams that do not include the requisite expertise to conduct such software development shall include costs in their white paper/proposal to obtain software development support.

Each Proposer team is required to include among their key personnel a Lead System Integrator (LSI) who shall be responsible for preparing software deliverable subcomponents, modules, and systems, performing quality control of Deliverables, and integrating key components into the primary Decipher system(s). The LSI will also oversee communication and coordination across Performer's research teams including subcontractors, if applicable, to ensure research products are functional and follow software coding best practices (e.g., inline comments, documentation).

Place of Performance

Performance will be conducted at the Performers' recommended sites with Government's concurrence.

Test & Evaluation

T&E will be performed by a team or teams selected by the IARPA DECIPHER Program Manager and may be modified throughout the program.

T&E will be conducted by an independent team of Government and contractor staff carrying out evaluation and analyses of Performer Deliverables using program test datasets and protocols. Evaluations will occur during months 3, 6, 9, and 12 that will exercise Performer solutions across technical challenges.

Proposers are encouraged to develop methodologies for a self-contained, self-hosted environment for use by Government transition partners on their own network/systems. White papers must specify the processing dependencies needed to carry out the proposed research and what architecture and library characteristics are necessary for their approach(es) to be successful at meeting program objectives. Performers will have specific Deliverable Milestones at which all subcomponent and system algorithms and software will be delivered to IARPA and its designated T&E Team. The T&E Team will then conduct evaluations at the direction of the Decipher PM and with the objective of characterizing the quality, functionality, and performance of the Decipher Deliverables. In addition to quantitative measurements, T&E will be carried out to establish a thorough understanding of the progress, status, and limitations of the Performer's research. T&E results and feedback will be provided to Performers at regular intervals to keep Performers abreast of current independent performance measurements and to inform and improve their R&D approaches and methods. T&E results from all Performers will be shared with all teams to establish an understanding of the current state and progress of Decipher research; T&E results will also be shared with United States Government stakeholders, including contractors, for Government purposes.

Technical Exchange Meetings/Workshops/Site Visits/Travel Requirements

Performers are expected to assume responsibility for administration of their projects and to comply with contractual and program requirements for reporting, attendance at program workshops, and availability for site visits. The following paragraphs describe typical expectations for meetings and travel for IARPA programs as well as the contemplated frequency and locations of such meetings. In addition to ensuring that all necessary details of developed software, algorithm, and operational instructions are clear and complete, each Performer will be required to be available for questions and troubleshooting from the T&E Team in weekly or bi-weekly status meetings.

All Performer teams are expected to attend a collaborative workshop, to include Key Personnel from prime and subcontractor organizations. The workshop will focus on technical aspects of the program and on facilitating open technical exchanges, interaction, and sharing among the various program participants. Program participants will be expected to present the technical status and progress of their efforts to other participants and invited guests.

Critical System Design Reviews will be conducted by the Government Team during months 3 and 7 during the life of the program. These visits will take place at the Performer's facility. Reports on technical progress, details of successes and issues, contributions to the program goals, and technology demonstrations will be expected during these visits.

Anticipated Period of Performance

Decipher is a 12-Month Program.

4. LocUS – LOCATION USING SOUND (FY26) - ACTIVE

Program Overview

The LocUS program aims to create technology that automatically and accurately geolocates multimedia content by maximally leveraging audio and visual information. LocUS will improve the geolocation capabilities of the Intelligence Community (IC) considerably beyond imagery-only methods and thereby increase the volume of content that can be accurately geolocated. This program will increase analysts' ability to rapidly and accurately determine where a video, image, or audio clip was collected in the absence of accurate metadata indicating location. Applications for national security include human trafficking interdiction, hostage recovery, localization of malign actors using social media or confiscated devices, and other intelligence and law enforcement use cases.

Objectives

The LocUS program aims to develop technology that can automatically and accurately geolocate an indoor or outdoor image, video, or audio clip taken anywhere in the world.

Task Area 1 (TA1): Multimodal Geolocation

Develop capability to automatically and accurately geolocate multimedia content.

The Test and Evaluation (T&E) team(s) will collect and curate test data for LocUS test events.

Program Scope and Limitations

The LocUS program develops capabilities for indoor and outdoor terrestrial geolocation of previously unseen images, video, or audio clips. The LocUS program is collection-sensor agnostic.

The following areas of research and approaches are out of scope for the LocUS program:

- Methods that utilize biometric identification methods (e.g. facial recognition, voice recognition)
- Methods that work exclusively for single modalities (e.g. audio only), or only for types of data within single modalities (e.g. human speech data only)
- Solutions that work only outdoors, or only for specific regions, terrains, or architectures
- Development of hardware solutions, or solutions that require custom hardware
- Methods that require a human-in-the-loop as part of the integrated end-to-end system

Geolocation Explainability is required for the LocUS program. Geolocation Explainability refers to the ability of an automated system to provide plain language descriptions of features relevant to geolocation of the query. These may be features utilized by the system to arrive at, or dispense with, a candidate location or locations. The intention of Geolocation Explainability is to extract

potentially meaningful features from image, video, or audio that would (1) assist analysts to accurately geolocate the query, and (2) increase analyst justified confidence in automated system outputs.

Performers will build prototype systems that an independent T&E team or teams will evaluate on sequestered evaluation datasets. Testing protocols do not allow for expert operators, human-in-the-loop operation, or any operations not deemed "turnkey." Systems or algorithms that have been trained using human-in-the-loop methods may be submitted, provided they run autonomously.

Program Specific Terms/Definitions

Geolocation Explainability refers to the ability of an automated system to provide plain language descriptions of features relevant to geolocation of the query.

A **query** is the input image, video, or audio clip to be geolocated. The prototype system should be able to geolocate the camera (or, if requested, an object in the scene that will be specified with the query).

Reference Data is any type of information that could aid in geolocation and may be supplied by the T&E partner in advance of Test Events. T&E may distribute image, video, or audio materials with certain qualities or characteristics that will be representative of queries of interest in Test Events. For example, if a Test Event were to include indoor imagery from hotel rooms, the T&E partner may potentially supply reference data to aid geolocation of that query.

Test Events are T&E-conducted evaluations that will exercise performer solutions across technical challenges to be determined by the United States Government and T&E partner. Before each test event, performers will deliver prototype software to be evaluated by T&E against program metrics in a variety of challenge conditions. There are five test events planned for the LocUS program.

Human or Animal Subject Research applicable to this Program

N/A

Government Furnished Equipment/Information

With the exception of potential Reference Data (for example, imagery or video for the Performers to optionally use to ensure that Performer systems can process expected file types), the Government does not plan to furnish data or equipment to LocUS Performers. Minimal reference data may be provided to ensure interoperability, which is not expected to be enough data for training purposes. The LocUS T&E partner will supply queries for testing and evaluation, and may supply additional queries to Performers for development.

Program Metrics

Candidate LocUS metrics include the following. Targets may be refined, and additional metrics may be added. Proposers are encouraged to include potential metrics as part of the white paper submission. Good metrics for the purpose of this effort maximize interpretability, allow easy implementation and replicability, allow calculation of confidence bounds, furnish results that are comparable over time, and are actionable (i.e., inform changes in technical approach).

Table 1: Notional LocUS Program Metrics

Task	Metric	Target
Geolocation Accuracy	Circular Error (CE) 90	< 250 meters
Geolocation Time to Prediction*	Time	< 5 minutes

*For a search area of Continent (2,500 km).

250m Circular Error (CE) 90 refers to a confidence value (90%) that points or features fall within the horizontal radius (250m) of the true location.

Geolocation Time to Prediction refers to the time it takes for the prototype system to provide an answer to a geolocation query.

The LocUS program will evaluate program metrics over a variety of search areas, ranging from neighborhoods to continents, to better characterize solution performance.

Program Waypoints, Milestones, Deliverables

The LocUS program is planned as a 15-Month Program. Table 3 provides the notional LocUS Program Schedule for Task Area 1 (TA1):

Table 2: Notional LocUS Program Schedule (Task Area 1)

Activities	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kickoff	X														
Critical System Design Reviews at Performer Sites				X						X					
Software Delivery		X			X			X			X			X	
Test Event			X			X			X			X			X
PI Meeting Workshop							X								
Demo Day Workshop													X		
Monthly Status Reports	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Final Report															X

During the 15-Month Program, there will be a total of five test events, or one test event every ~3 months. Prior to each test event, Performers will deliver Performer systems as containerized software (e.g. Docker container) compatible with LocUS program API standards, to be developed in the Pre-program Phase. At test events, T&E will test and evaluate Performer systems against numerous queries from T&E-collected, sequestered datasets. Performer systems are required to geolocate multimodal content. While image-only and audio-only performance will be evaluated at each test event to better characterize the contributions that individual modalities provide to overall geolocation capabilities, the goal of the program is improved geolocation using all available multimodal information. Proposed solutions for the LocUS program shall be a single complete system.

Images, videos, and audio clips will be highly varied in content, location, and in the quality of the query (e.g. length, blur, resolution, audio bit rate, bit depth, and sample rate). The queries will be provided in a file format that is widely available. Proposers should assume that the metadata will be removed from all queries.

Following each Test Event challenge, performance analysis results and challenge data may be provided, if appropriate, to Performers for review and methodology improvement.

Performers may incorporate the use of other existing data sources, provided these sets can be made available to the Government for evaluation purposes.

The following table presents the waypoint, milestone, and deliverable schedule for the LocUS program (Task Area 1):

Table 3: LocUS Waypoint, Milestone, and Deliverable Schedule (Task Area 1)

Month	Event	Description	Comment	Deliverable
All	Waypoint	Monthly Status Report (MSR), Technical and Financial	Due on 15th of each month	MSR in provided template
1	Waypoint	Kickoff	DC metro area and/or virtual	Overview presentation provided 1 week in advance of meeting
2	Deliverable	Containerized solution delivery for Test Event 1	Interim	Container, Report Describing Container, and Source Code
4	Waypoint	Critical System Design Review	At performer site	Presentation

5	Deliverable	Containerized solution delivery for Test Event 2	Interim	Container, Report Describing Container, and Source Code
7	Waypoint	PI Meeting Workshop	DC metro area and/or virtual	Presentation, Demonstration
8	Deliverable	Containerized solution delivery for Test Event 3	Interim	Container, Report Describing Container, and Source Code
10	Waypoint	Critical System Design Review	At performer site	Presentation
11	Deliverable	Containerized solution delivery for Test Event 4	Interim	Container, Report Describing Container, and Source Code
13	Waypoint	Demonstration Day Workshop	DC metro area and/or virtual	Presentation, Demonstration
14	Deliverable	Containerized solution delivery for Test Event 5	Final	Container, Report Describing Container, and Source Code

Software Deliverable Formatting

Performers will deliver containerized software (e.g. Docker container) compatible with LocUS program API standards, to be developed in Preprogram Phase. Source code for capabilities must be developed and delivered in a manner such that T&E can install, build and execute the software, including the ability to retrain models. Any external libraries or other dependencies must be included in deliverables. Proposers are encouraged to convey any specific software framework requirements or constraints as part of the white paper submission.

Place of Performance

Performance will be conducted at the Performers' recommended sites with Government's concurrence.

Test & Evaluation

T&E will be performed by a team or teams selected by the IARPA LocUS Program Manager and may be modified throughout the program.

The LocUS T&E team(s) will compile robust sets of diverse and relevant multimedia corpora to support research goals. The T&E team(s) will conduct independent test and evaluation for the

program with curated data sets and will be responsible for integrating containerized solutions. Performer systems will be evaluated on an interim basis to track progress and ensure compliance with system requirements. T&E results from all Performers will be shared with all teams to establish an understanding of the current state and progress of LocUS research; T&E results will also be shared with United States Government stakeholders, including contractors, for Government purposes.

Technical Exchange Meetings/Workshops/Site Visits/Travel Requirements

- One kick-off meeting in Washington DC Metro Area and/or virtual (Month 1)
- Two in-person Workshops in the Washington DC Metro Area and/or virtual (Months 7 and 13)
- Critical System Design Reviews (virtual and/or in-person where LocUS PM travels to Performer Site) every 6 months (Months 4 and 10)

Meetings, Workshops, Site Visits, and Travel Requirements are subject to revision.

Anticipated Timeline

Table 4: LocUS Anticipated Timeline

Task Area (TA)	PoP
TA1	15 months (est.)

LocUS is planned as a 15-Month Program.

5. MOVES – MOVEMENT OBSERVATION AND VIDEO-BASED EVALUATION SYSTEM (FY26) - ACTIVE

Program Overview

The intelligence community (IC) would benefit from the ability to diagnose, from open-source data, medical conditions in individuals and populations. This ability would enable, e.g., monitoring of the spread of health threats (disease, chemical accidents or attacks, environmental toxins, etc.) and other medical intelligence analyses. The MOVES program aims to demonstrate the feasibility of this long-term goal by addressing a subset of the challenges associated with generalized diagnosis from video.

The MOVES program will focus on diagnoses of neurological conditions that typically affect people's movement and demonstrate that it is feasible to diagnose such conditions from challenging input video that may:

- Be taken from non-ideal angles
- Include subjects wearing a wide variety of clothing styles
- Show partially obstructed subjects

Fig. 1 shows examples of frames from the types of videos that may be used. Clothing may include attire for a wide variety of weather conditions and from a wide variety of cultures. Activities may be slow or rapid, the number of people in the video can range from small to large, and the videos may be of arbitrary length.

Figure 1: Examples of frames from the types of videos that MOVES algorithms will evaluate.

The program requirements are selected to ensure a path to rapid adoption and trust in the algorithms' performance by medical professionals. Key challenges include the nature of the videos as described above, especially finding movements that are diagnostic of conditions, differentiating between similar symptoms of different disorders, and providing enough information to the medical professionals about how the diagnosis was made so that they can verify the diagnosis and build confidence in the algorithm performance.

Objectives

The MOVES program will develop algorithms that can diagnose a small set of neurological movement disorders from video of subjects in non-clinical and less-than-ideal settings. A key difference between the types of videos MOVES will analyze and existing research into algorithmic detection of movement disorders is that existing research tends to focus on diagnosis in a clinical setting, where the subject is asked to perform motions that are diagnostic of movement disorders.

In the MOVES context, subjects will be performing everyday activities and may consciously or unconsciously minimize or conceal any movement symptoms. Furthermore, subjects in MOVES test videos may be wearing a wide variety of clothing styles, some of which may conceal movement more than others.

MOVES will develop 2 prototype algorithms that perform two different functions:

1. **Extract skeletal movement and body shape from underneath clothing**

   Algorithm 1 will take a video as input and allow the user to select a subject of interest. The algorithm will output both an animation of the skeletal movements of the subject's body (i.e., a skeletal stick figure) with sufficient time resolution to capture typical tremor or involuntary movement symptoms, including hands and fingers, as well as a simulated unclothed full-body animation. The output animations will be matched to the movements of the source video to allow medical professionals to perform diagnoses directly from the original video with the aid of the skeletal and unclothed body animations. Keypoint locations for skeletal motion will also be output to a text file.

2. **Rank order video clips by diagnostic significance and perform diagnosis**

   Algorithm 2 will take a video as input and allow the user to select a subject of interest. The algorithm will divide the input video into roughly 15 second video segments, rank-ordered by a diagnosis score. The diagnosis scores should be a monotonic function of the likelihood of the presence of that condition in the subject of interest in that segment. The end user will be able to set the score threshold for detecting a condition.

Performers must select three conditions from the list below, which their algorithms will be able to diagnose. As the program progresses, if availability of datasets changes, IARPA may direct teams to change conditions they have initially selected.

- Parkinson's Disease
- Essential Tremor
- Dystonia
- Stroke
- Mild Traumatic Brain Injury
- Multiple Sclerosis

Program Technical Deliverables and Performance Metrics

Technical Deliverable 1 will be the algorithm that extracts skeletal movement and body shape from underneath clothing. Output videos will have the same pixel resolution and frame rate as the input video.

The performance of Technical Deliverable 1 will be based on two metrics, Percentage of Correct Keypoints as a fraction of head height (PCKh) and frequency response. PCKh@a > b% means that across all skeletal keypoints and frames in the video, keypoints will be within a head heights of the actual location more than b% of the time. Frequency response specifies that the algorithm must be capable of reproducing tremors at sufficient frequency for the diagnosis of neurological conditions.

Deliverable 1 target performance metrics: PCKh@0.5 > 95%; Frequency response > 12 Hz

Technical Deliverable 2 will be the algorithm that rank orders video clips by diagnostic significance and performs diagnosis. It may take as input the output videos from Technical Deliverable 1.

Deliverable 2 target performance metric: Diagnostic accuracy of 80% on disorders the performer team has selected. For this metric, score threshold will be chosen after the test is run to maximize accuracy. Cases in the test set will be roughly balanced in number of instances of subjects with and without known movement disorders.

Performers may suggest other metrics or evaluation approaches that they feel highlight the value and performance of their approach.

Performers should expect that during independent Test and Evaluation (T&E), the outputs from one team's Technical Deliverable 1 software, including the simulated stick figure, simulated unclothed full body, and keypoint text file, may be used as inputs for another team's Technical Deliverable 2, if that is an appropriate evaluation. This will be conducted by the T&E team; data or outputs from one Performer team will not be shared with other Performer teams.

Other Technical Deliverables: Any datasets compiled and any source code created during the execution of this effort.

Program Waypoints, Milestones, Deliverables

MOVES is a single-phase 14-month program. Table 1 shows a timeline for the program with Government defined milestones and deliverables.

Table 1: Program Waypoint, Milestone and Deliverables Testing Timeline

Event	**Program Month**	Deliverables
Program Kickoff	1	Performer and T&E teams will be in attendance. Read-ahead package due from Performers to the Government 7 days before meeting.
Technical Review Meetings (Virtual)	Monthly, or as requested by the PM	Read-ahead package due from Performer to the Government 1 day before meeting.
Financial and Technical Reports	Monthly	Monthly financial and technical reports are due by the 15th day of the following month.
Site Visits	4, 13	PM will visit performer sites for a technical review.
PI Technical Exchange Meeting	5, 10	Performer and T&E teams will be in attendance. Read-ahead package due from Performers to the Government 7 days before meeting.
Technical Deliverable 1, and related datasets and source code	7, 12	Performers will submit their software and documentation to the T&E partner on the first business day of the month. Datasets collected/compiled and source code created in support of this technical deliverable will also be due.
Technical Deliverable 2	12	Performers will submit their software and documentation to the T&E partner on the first business day of the month.
Final Report, datasets, and source code	14	Performers shall provide a final technical report outlining their technical progress, methodology, challenges, successes, and recommendations for future efforts. Performers shall provide datasets and source code created under this program to the Government.

Software Deliverable Formatting

Performers will be required to provide algorithms in a Docker container. The entirety of a Performer's system for each Technical Deliverable, including pre- and post-processing, must be included within the delivered software container. Performers shall also provide documentation required to install and run the software, a change log, and the source code to the Government and T&E team.

Test & Evaluation

T&E will be performed by a team or teams selected by the IARPA MOVES Program Manager and may be modified throughout the program.

Independent T&E will be conducted by an independent team of Government and contractor staff carrying out evaluation and analyses of Performer Deliverables using program test datasets and protocols. Evaluations will occur during months 7 and 13. In addition to Independent T&E, IARPA will regularly gauge interim progress of Performer research and development on the basis of Performer self-reported technical updates.

The T&E team will also collect a dataset for training and testing purposes. The training dataset will be distributed to all Performer teams. We expect that the training dataset will be collected from roughly 100 individual volunteers, with movement disorders, wearing inertial measurement unit motion capture devices and filmed from multiple directions. There will be a complementary set of roughly 100 individual volunteers without known movement disorders captured the same way.

T&E results from all Performers will be shared with all teams to establish an understanding of the current state and progress of MOVES research; T&E results will also be shared with United States Government stakeholders, including contractors, for Government purposes.

Technical Exchange Meetings/Workshops/Site Visits/Travel Requirements

All Performer teams are expected to attend the program Kickoff and Technical Exchange Meetings in person. These meetings will facilitate open technical exchanges, interaction, and sharing among the program participants. Program participants will be expected to present the technical status and progress of their efforts to other participants and invited guests.

Site visits by the Government Team will take place during months 4 and 13. These visits will occur at the Performer's facility. Reports on technical progress, details of successes and issues, contributions to the program goals, and technology demonstrations will be expected during these site visits.

Anticipated Period of Performance

MOVES is a 14-Month single-phase program.

Human or Animal Subject Research applicable to this Program

IARPA anticipates that the bulk of the training data used to train the algorithms under development will be open-source data. Additionally, the T&E team will collect a dataset using human volunteer subjects for distribution to all Performer teams. This dataset will consist of high-fidelity motion-capture data of volunteer subjects performing a variety of normal-life activities, with each individual wearing a variety of clothing styles ranging from minimal and/or form-fitting to bulky and/or loose and flowing. IARPA does not anticipate that Performers will also collect human subjects data, though Proposers are not precluded from doing so. If human subjects research is proposed, it must follow all federal HSR guidelines and requirements, as outlined in the general ETA Program Announcement.

Place of Performance

Performance will be conducted at the Performers' recommended sites with Government's concurrence.

Government Furnished Equipment/Information

The Government will not provide equipment or software to the Performer. The Government will distribute data collected by the T&E team to the teams.