Advanced AI/AA Analytics

Department of Defense
Dept of the Army -- Futures Command GENERAL

INFORMATION

Document Type:	Grants Notice
Funding Opportunity Number:	W911NF-17-S-0003 SPECIAL NOTICE FOR ADVANCED AI/AA ANALYTICS
Funding Opportunity Title:	Advanced AI/AA Analytics
Opportunity Category:	Other
Opportunity Category Explanation:	SPECIAL NOTICE - Awards Made from Topic Area 7.a.KCI-AA-1 & 9.b of W911NF-17-S-0003-09
Funding Instrument Type:	Cooperative Agreement
Category of Funding Activity:	Science and Technology and other Research and Development
Category Explanation:	
Expected Number of Awards:	1 (However, the Government reserves the right to make multiple awards.)
CFDA Number(s):	12.431 -- Basic Scientific Research
Cost Sharing or Matching Requirement:	No
Version:	Synopsis 1
Posted Date:	10 MARCH 2021
Original Closing Date for Applications:	12 APRIL 2021
Estimated Total Program Funding:	$40,000,000.00 - 45,000,000.00
Award Ceiling:	$40,000,000.00 - 45,000,000.00
Award Floor:	$0

ELIGIBILITY

Eligible Applicants: See W911NF-17-0003 for eligibility criteria

ADDITIONAL INFORMATION

Agency Name:
Dept of the Army -- Futures Command
Description:
METRIC DEVELOPMENT FOR HUMAN AI/AA ENSEMBLES

The U.S. Army Contracting Command – Aberdeen Proving Ground, Research Triangle Park Division, on behalf of the CCDC Data & Analysis Center is soliciting proposals under:

"7. ANALYSIS AND ASSESSMENT (AA) CAMPAIGN a. KCI-AA-1: Methodology for A&A of Complex Systems and Technologies across Multiple Domains" &

"9. AMRY AI TASK FORCE RESEARCH INTERESTS b. Artificial Intelligence and Machine Learning (AI/ML)"

of the ARL Core Broad Agency Announcement (BAA) for Basic and Applied Scientific Research, W911NF-17-S-0003-09.

DEVCOM DAC is soliciting proposals under this Special Notice of the BAA for the performance of applied research focused on developing a suitable analytical and simulation framework for artificial intelligence and assistive automation (AI/AA) systems to ensure these systems can be assessed and evaluated properly by the Army. The framework should include live, virtual, and constructive simulations to model the AI/AA in relevant Army operational situations. The framework should enable the DEVCOM DAC to perform: direct measurement, trade off analysis, performance and effectiveness analysis, development of data for higher-level force-on-force models, and sustainability analysis. These analyses will enable Senior Army decision makers to better integrate AI/AA systems into the Army structure. As part of this framework, metrics and their underlying variables should be developed, defined and delivered. These experimental, calculated, or modeled metrics must have high explanatory value in appropriately capturing the most important determinants of the performance and effectiveness of the AI/AA technology-human operator ensemble.

The Army, like world industry, has many automated or partially automated systems under consideration that will employ artificial intelligence (AI). Some of the planned assistive automation (AA) will only be as complex as an alarm clock, but longer term commercial and Army goals will push against current AI state-of-the-art. For example, future Automatic Target Recognition (ATR) will employ advanced machine learning (ML) techniques. Future Cyber-security applications must be able leverage self-learned mitigation techniques to incoming attacks faster than a human can respond. Development, validation, and application of these sophisticated new technologies require the Army to figure out how to specify, analyze, and evaluate AI/AA that may react in unforeseen and unintended ways to new stimuli. Even more challenging will be self-driving technologies, ground-based and airspace auto-navigation capabilities, and air-defense systems required to react so quickly to approaching obstacles or incoming threats that they must be virtually autonomous.

Army analysis and experimentation for previous Army systems have been conducted within a well-established framework that includes use of accepted methods for assessing expected performance and effectiveness. AFC Combat Capabilities Develop Command's Data and Analysis Center (DAC) has the Army role to make credible performance and effectiveness estimates for all classes of equipment including guns, bullets, missiles, communications devices, networks, protective equipment, tracked and wheeled vehicles, and manned/unmanned air vehicles. These estimates --- both empirically derived and from modeling and simulation (M&S) --- are developed for both benign laboratory-type environments and combat environments that are contested in multiple domains. To assure that the estimates are credible, DAC analyses endeavor to consider all relevant factors at an appropriate level of engineering detail including vulnerability to realistic threat attacks in various domains, hardware and software reliability, relevant human factors, maintenance and other sustainment challenges.

There is not yet a well-established framework, analytic metrics, measurement techniques, analytical tools, analytical and simulation environments, and data collection architectures for assessing AI/AA in either benign or contested environments. This gap is serious now, and will become even more critical as commercial and military system continue to operate in more dynamic environments. AI/AA systems will require analysis for many different uses: developer driven trade-offs; evaluator assessments of effectiveness, suitability, and survivability; support of senior decision makers; and sustainability analysis for life-cycle costing. In addition, the Army has additional criteria to include inputs to the Army's force-level models that will assess the overall contribution of the AI/AA to Army operations.

The gap cannot be credibly filled by sole focus on the new hardware. Any adequate analytical framework and supporting tools must include human interactions with the AI/AA systems. Such interactions run the gamut from traditional human factors engineering issues such as ease of use and avoidance of excessive automation "modes" through more complex issues such as the transparency of AI/AA reasoning, new or excessive human attentional demands, unplanned or unintended uses of AI/AA with potential negative consequences, and trust and complacency issues.

One of the goals of this research program is to enable integrated coherent analysis of the behavior of AI/AA-human ensemble across multiple domains in tactically realistic environments. Adequate analysis will require detailed consideration of the performance of the AI/AA technology and its supporting algorithms, the human component of the Soldier as user of the AI/AA technology, and the details of the interface between the technology and the using Soldier(s). The specific focus of the initiative is the development and articulation of an analysis and assessment framework or frameworks; experimentation, test and evaluation observational methodologies; M&S or other technical tools; and metrics for evaluating the performance of the AI/AA-human ensemble.

As first steps towards developing a suitable analytical framework for AI/AA systems, DEVCOM DAC is seeking eligible entities to develop, define, and deliver new analytical metrics as well as the underlying variables that will drive those metrics. These experimentally observed, measured, calculated, or modeled metrics must have high explanatory value in appropriately capturing the most important determinants of the performance and effectiveness of the AI/AA-human ensemble. Moreover, it is critically important that the framework enforce consistency across the relevant metrics and underlying variables whether the values come from early experiments, developmental tests, operational tests, or M&S.

Metrics will be of particular interest for cases where values are greatly different for the benign and contested environments. Eligible applicants will define the vignette and scenario details of the contested environments, but as a minimum they must include impact on the operational performance of the AI/AA-human ensemble when confronted with contested use of the electromagnetic spectrum, contested network availability, enemy attempts to defeat friendly sensors, electromagnetic pulse, and heavy enemy fire.

The general framework contemplated is shown in the attached PowerPoint chart. To the maximum extent possible, the Army needs to identify variables and metrics that are susceptible to being captured both by direct measurement or observation from test and experiments and by calculation and analysis from M&S. Reliance on only formal Test and Evaluation (T&E) or only M&S would not provide Army leadership with the robust analysis they will need to make multi-billion dollar decisions about AI/AA technologies. The metrics must be meaningful for both benign and contested environments, and they must be able to satisfy all or at least most of the analytical requirements of Evaluators, Developers, the broader analytical community, and senior decision makers.

Four exemplar AI/AA-human ensembles spanning the conceptual space between very simple and very sophisticated AA are given below:

1. A decision support system that filters incoming message traffic and provides the manager/commander with a common operating picture based on that traffic.

2. An automated target recognition system (military system or commercial cyber-security product) that draws on an arbitrary network of sensors and determines the highest value targets and how they should be serviced.

3. An optionally manned semi- or fully autonomous vehicle with several different classes of armament.

4. An autonomous system that reacts outside of human-in-the-loop timescales and then notifies the controller about what has been done and queries the responsible Soldier/individual when there is time.

DAC intent is not that the investigation be limited to these four exemplars, but rather that the investigation must span a broad range of dual-use applications so that new AI/AA applications

will properly “fit” with the developed framework and metrics. DAC’s framework or framework must span that broad range.

An important question for early investigation is whether the broad range of potential AI/AA applications can be captured by a single framework in a single scenario or vignette and a single set of metrics with the associated driving variables. If there will be a need for multiple frameworks with multiple scenarios or vignettes and multiple sets of metrics for different classes of technologies, a taxonomy and an ontology (or perhaps alternative taxonomies and ontologies) should be proposed that can consistently capture the full application space. The taxonomy should classify the factors/components of the overall analysis framework (e.g., AI/AA technologies, underlying variables, scenarios, metrics, measures, environments), while the ontology should specify the types of entities, the relevant properties of the entities, and their relationships. The factors specified should exist (either in the test, experiment or M&S) for the output ---measured or modeled --- to be credible and relevant. Articulation of this taxonomy and ontology is absolutely essential work to enable the Army to make apples-to-apples comparisons of competing AI/AA technologies and should be addressed in the proposal.

Specifying relevant frameworks, metrics, critical variables, and vignettes/scenarios for the AI/AA technologies will be very challenging, as analysis of new technology always is. For example, the quality of the AI/AA-human interface is critical, but it is not clear how we should capture the difference between a good interface and an opaque one in an M&S effort designed to capture operational benefit of the AI/AA. As a second example, there may be emergent behavior or output from the AI/AA that conflicts with designer intent and operational need. Properly capturing the consequences of such flaws is a necessary component of an adequate analysis.

Ensuring that the framework(s) developed properly illuminates the AI/AA-human ensemble will be even more challenging. Human factors aspects of AI/AA involve many subtleties that are not present, at least to a high degree, for Soldier interactions with non AI/AA hardware. Many of these subtleties will require experiments or tests as they appear to be beyond current M&S state-of-the-art.

- Determining what constitutes a "reliability failure" and potential maintenance burden of an AI-enabled system where there are important hardware, algorithmic, and human factors components.
- Identifying potential “emergent behaviors” of the AI/AA-human ensemble that could have detrimental impacts to performance, effectiveness, vulnerability, safety, and reliability.
- Assuring that we can assess the contribution of the AI/AA to mission success both “objectively” and in the subjective view of the using humans.
- Evaluating the simplicity and intuitiveness and/or the clumsiness and difficulty of the interface between the human and the AI/AA.

- Measuring and possibly modeling the transparency to the using human of such matters as exactly what the AI/AA is doing at a given time, what "mode" the AI/AA is in, and when the AI/AA is approaching or beyond its designed competence level.
- If observations show that tending to the AI/AA takes reasonable or unreasonable amounts of human time and effort both during slack times and during peak operations, impacts must be captured in any M&S.
- If observations show that humans trust the AI/AA where they shouldn't or are unduly complacent with it, then the impacts of that too must be captured in any M&S.
- If observations show that humans trust the AI/AA so little that they choose not to use it at all in stressful contested situations, then those impacts must be captured in any M&S.
- If observations show that humans attempt to use the AI/AA for unintended purposes or purposes out of the design envelope, then those impacts must be captured in any M&S.
- The expectation is that an eventual awardee will receive unclassified data in support of this research; however, there is also the potential for the awardee to receive classified data for use in performance.

Past work with AI/AA technologies reveals widely varying technological maturity. When exact specifications, measured performance data, validated M&S output, breadboard hardware, technology components, or actual Army systems are available they can be fully exploited in development of the analytical framework and associated metrics. Other use cases are more notional, but still replicate as closely as possible the researchers' understanding of both technical state-of-the-art and Army operational intent for the proposed technology. A spectrum of use cases of variable maturity, each with appropriate metrics, will be employed for the FY22 and beyond work.

During FY22 and beyond the framework and metrics developed in FY21 will be extended so as to yield more illuminating insight into the most critical variables that drive the effectiveness of human-AI/AA interfaces. Early work has shown that even when AI/AA technology performs perfectly according to specification when tested in a laboratory environment, poorly designed or sub-optimum human interfaces can render the technology useless or even detrimental for mission accomplishment.

Moreover, it was observed during FY21 that there were occasions when the human-AI/AA ensemble failed to accomplish the intended mission, and the observers were unable to ascertain the specific cause for the failure. Faulty equipment, poorly trained humans, and inadequate algorithms can sometimes be difficult to distinguish. Some algorithms may just be wrong for the intended application. Some may be largely correct, but have unintended and unforeseen consequences. Some may be technically accurate, but fail in implementation through either badly designed interfaces or poorly trained humans. New work is needed to address these challenges through experiments, by sophisticated M&S, and in all relevant Live,

Virtual, and Constructive, assessment environments. Credible Army analysis requires that both the human and the AI/AA component of the ensembles be characterized and understood at a much higher level of granularity than has been traditionally done.

It is the intent of this Special Notice to solicit the most creative, innovative, and flexible approaches to the ultimate goal of generating and exploiting research to solve pressing research gaps and issues impacting both the military and commercial sectors in regard to developing a suitable analytical and simulation framework for artificial intelligence and assistive automation (AI/AA) systems to ensure these systems can be assessed and evaluated properly by the Army. This Special Notice seeks proposals which will result in the award of a single cooperative agreement, although the Government reserves the right to award more than one cooperative agreement under this notice or to make no award based on the proposals received.

While the BAA and Topics under this notice remain open for proposal submission until 31 March 2022, in order to be eligible for award under this Special Notice with the funding available for FY 2021, proposals must be received in accordance with the due dates and instructions provided in this announcement. It is anticipated that a cooperative agreement under the authority of 31 USC 6305 may be awarded for the competitive selection of basic and applied research proposals pursuant to 10 USC 2358. All proposals will be evaluated in accordance the evaluation criteria under the BAA. DEVCOM DAC may ask an applicant to re-scope or modify their proposal in order to achieve maximum benefit for the program within the fixed available funding. The total amount for award is anticipated to be approximately $8,000,000.00 per year for up to five years, subject to available funds. A Proposal must include a detailed plan and budget for the first year (FY21 effort) and anticipated work and budget for the four option years (FY22 effort and beyond).

The award under this notice will consist of an initial 12-month Base award that will be executed, with the potential to exercise option periods for up to four additional years, with each option period consisting of a 12-month period, at the discretion of the Government. The period of the performance and funding amount of potential option periods will be based on the goals of the program, merits of the proposed optional research and available funding. DEVCOM DAC reserves the right to negotiate with an Applicant to re-scope their proposal or optional proposal technical focus, period of performance, and associated costs in order to maximize the available program funding, balance of research topics across the program, and overall impact to the program.

A Recipient of a cooperative agreement will work collaboratively with scientists from DEVCOM DAC to further the Advanced AI/AA Analytics. DEVCOM DAC will participate in the research and use its strong in-house technical expertise to jointly plan and execute the research program with a Recipient. In order to be eligible for a cooperative agreement under this program, applicants must submit proposals in accordance with the submission instructions under BAA W911NF-17-S-0003, Section II.D no later than 12 April 2021. A Proposal submitted after this closing date will not be considered by the Government. Criteria for eligible Applicants are given

in BAA, Section II.C.1. Proposals will be evaluated using the criteria listed in the BAA, Section II.E. Under this Special Notice, no formal Q&A will take place. If an applicant has a question regarding technical clarification related to this Special Notice, contact Dr. Thomas Stadterman at thomas.j.stadterman.civ@mail.mil. For general assistance, contact Joshua Houck at joshua.d.houck.civ@mail.mil. Upon submission, an Applicant is encouraged to provide notification to both Dr. Thomas Stadterman and Joshua Houck by email. This Special Notice expires 12 April 2021 at 11:59 PM (EDT).

Grantor Contact Information: If you have difficulty accessing the full announcement electronically, please contact:

CHRIS JUSTICE
Branch Chief, Branch B
ACC-APG RTP Division
800 Park Office Drive
Suite #4229
Research Triangle Park, NC 27709
(919) 549-4287