Broad Agency Announcement

Artificial Intelligence Quantified

INFORMATION INNOVATION OFFICE

HR001124S0029

June 04, 2024

This publication constitutes a Broad Agency Announcement (BAA) as contemplated in Federal Acquisition Regulation (FAR) 6.102(d)(2) and 35.016 and 2 CFR § 200.203. Any resultant award negotiations will follow all pertinent law and regulation, and any negotiations and/or awards for procurement contracts will use procedures under FAR 15.4, Contract Pricing, as specified in the BAA.

OVERVIEW INFORMATION:

- **Federal Agency Name** – Defense Advanced Research Projects Agency (DARPA), Information Innovation Office
- **Funding Opportunity Title** – Artificial Intelligence Quantified (AIQ)
- **Announcement Type –** Initial Announcement
- **Funding Opportunity Number –** HR001124S0029
- **Assistance Listing Number:** Not applicable
- **Dates/Time - All Times are Eastern Time Zone (ET)**
  - Posting Date: June 4, 2024
  - Proposers Day: June 14, 2024
  - Proposal Abstract Due Date: June 25, 2024 at 12:00 PM
  - Question Submittal Closed: August 2, 2024 at 4:00 PM
  - Proposal Due Date: August 13, 2024 at 12:00 PM
- **Anticipated individual awards** - Multiple awards are anticipated.
- **Types of instruments that may be awarded** – Procurement Contract, Cooperative Agreement, or Other Transactions
- **NAICS Code**: 541715
- **Agency contact**
  - Points of Contact

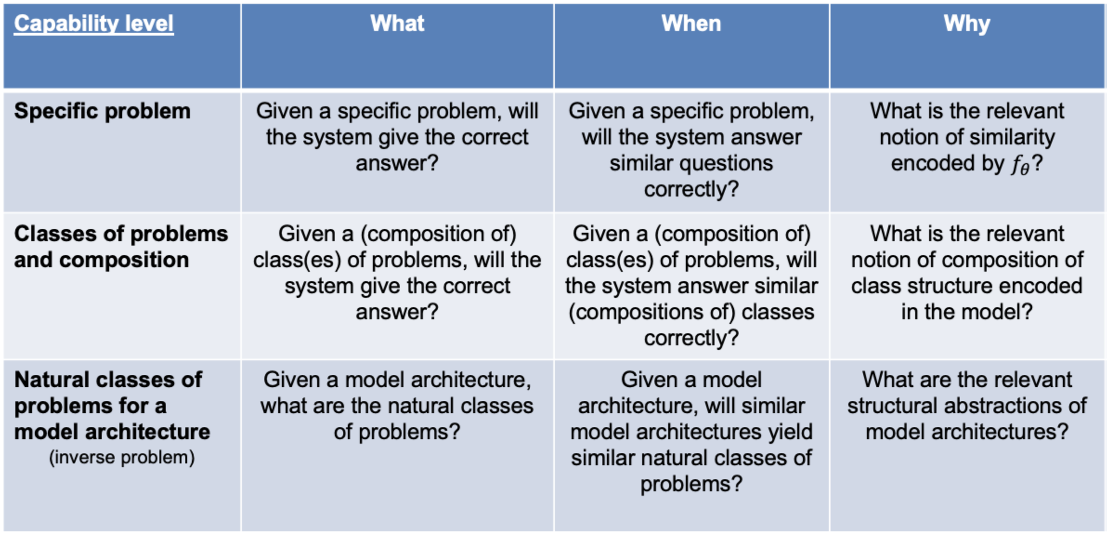
    The BAA Coordinator for this effort may be reached at:
    AIQ@darpa.mil
    DARPA/ I2O
    ATTN: HR001124S0029
    675 North Randolph Street
    Arlington, VA 22203-2114

Section I: Funding Opportunity Description

The Defense Advanced Research Projects Agency (DARPA) is soliciting innovative proposals in the technical areas of assessing and understanding the capabilities of artificial intelligence (AI) to enable mathematical guarantees on performance of generative AI. Proposed research should investigate innovative approaches that enable revolutionary advances in science, devices, or systems. Specifically excluded is research that primarily results in evolutionary improvements to the existing state of practice.

AI has achieved near human level performance in domains including text generation, planning, and game playing[i], raising the prospect of widespread integration with human partners in the military and society. As highlighted in the Pentagon's 2023 Data, Analytics and Artificial Intelligence Adoption Strategy, AI is essential to Department of Defense (DoD) missions and will enhance its competitive edge by addressing keying operational problems identified in the 2022 National Defense Strategy, filling validated gaps to enhance the warfighting capabilities of the Joint Force, and strengthening the enterprise foundation required to sustain enduring advantages[ii,iii]. However, integration of AI technologies in the military requires ensuring safe and responsible operation of autonomous and semi-autonomous technologies to minimize the probability and consequences of failure.[iv] Methods for guaranteeing *what* the capabilities (and limitations) of AI systems are, as well as *when* and *why* those capabilities do or do not manifest, do not currently exist.

Artificial Intelligence Quantified (AIQ) will develop technology to assess and understand the capabilities of AI to enable guaranteed performance. The program will test the hypothesis that mathematical methods, combined with advances in measurement and modeling, will allow guaranteed quantification of AI capabilities. Specifically, the program will address three interrelated capability levels: 1) specific problem level, 2) classes of problem level, and 3) natural class level, aiming to address the quantification and assessment challenges at each level (see Fig. 1).



Capability level	What	When	Why
Specific problem	Given a specific problem, will the system give the correct answer?	Given a specific problem, will the system answer similar questions correctly?	What is the relevant notion of similarity encoded by f_θ?
Classes of problems and composition	Given a (composition of) class(es) of problems, will the system give the correct answer?	Given a (composition of) class(es) of problems, will the system answer similar (compositions of) classes correctly?	What is the relevant notion of composition of class structure encoded in the model?
Natural classes of problems for a model architecture (inverse problem)	Given a model architecture, what are the natural classes of problems?	Given a model architecture, will similar model architectures yield similar natural classes of problems?	What are the relevant structural abstractions of model architectures?

Figure 1: Three capability levels addressed in AIQ. f_θ is the AI model represented as a parameterized function.

The state-of-the-art (SOTA) methods for assessment are ad hoc, deal with the simplest of capabilities, and are not properly grounded in a rigorous theory, which has led to negative consequences. The abilities observed and thought to be unpredictable properties that suddenly emerge with scale have recently been shown to be a consequence of uninformed evaluation based on improper scoring metrics.[v] Moreover, the lack of theory impedes safe operation. For example, recent research shows that simple prompts can be systematically constructed to extract training data, up to thousands of characters at a time, a significant issue for defense and intelligence applications. Evidence suggests that leakage of training data does not improve with scale and performance; rather, susceptibility to training data leaks worsens with scale.[vi]

AIQ Technical Approach:

AIQ brings together two Technical Areas (TAs) and a government team to test the program hypothesis. The goal of TA1 is to provide rigorous foundations for understanding and guaranteeing capabilities across levels; teams proposing for TA1 are expected to be led by individuals with deep technical expertise, such as pure or applied mathematics, theoretical computer science, or statistics, or other relevant expertise and demonstrate relevance to AI. The goal of TA2 is to develop methods for evaluating AI models, integrating and evaluating TA1 results at scale using appropriate research datasets; teams proposing for TA2 are expected to comprise computational, cognitive, and/or behavioral scientists with expertise in AI evaluation.

A single proposal may address only one of the technical areas; no combined proposals for TA1 or TA2 will be accepted. An organization may submit individual proposals for each TA; however, no prime or subcontractor proposer will be selected as a performer on both TAs. All proposals for must address all the requirements for the proposed TA as described in this BAA.

TA1: <u>TA1 proposal teams should directly address at least one of the three capability levels above</u>. They will be charged with deriving mathematical results related to generalization, empirically validated, that demonstrate an understanding of factors that affect generalization. Strong TA1 proposals will explain which level(s) will be their focus, what mathematical approaches will be used, and why these approaches are promising, both mathematically and practically. Proposals may follow any approach that is likely to be successful. In particular, the proposal should argue, presenting any preliminary results and evidence, how their approach is likely to meet the metrics and milestones of the program. TA1 proposers should develop software to empirically demonstrate their results and demonstrate their potential to scale and perform well in program evaluations (see TA2 section below) using their own datasets.

<u>Specific Problem Level</u>: At the level of specific problems, the problem is one of point testing, as in current evaluation methods. Proposers can adopt any mathematical approach but must argue their relevance for the program in terms of quantitation efficacy, robustness, its potential for scaling, and applicability in real problems. Recent research points to the possibilities for promising methods. For example, mathematical tools, including Lipschitz continuity,[vii] bounding curvature of information manifolds,[viii,ix] empirical scaling laws[x] , and information bounds,[xi] enable guarantees on generalization ability and inability by constraining the complexity of the function, or geometry, relating cross-entropy loss to capabilities, providing proofs relating

mutual information of input and output, respectively. This list is not exhaustive; proposals may follow any approach, and performers should argue persuasively for why their approach is promising.

Classes of problems and composition: At the level of classes of problems and their compositions, the challenge is to understand collections of problems and explore a compositional approach to quantification. Again, a few recent mathematical papers, though studied in limited contexts, point to the possibilities in terms of inferring latent skills[xii] and assessing (in)consistency of local dimensionality across the class.[xiii] Additional approaches for classes of problems include scaling laws and/or combinatorics,[xiv] composition,[xv] topology, and curvature of manifolds.[xvi] Again, this list is not exhaustive; proposals may use any approach.

Natural classes of problems: At the level of natural classes of problems, the challenge is to find invariant properties of model architectures to answer the inverse problem: which architectural choices give rise to suitable classes for a given application? Recent work, for example, highlights Optimal Transport and related methods for understanding and analyzing transformer architectures.[xvii,xviii] Importantly, these approaches enable understanding model behavior *before* training, and hence choices of architectures that suit particular classes of problems. Other methods for understanding invariant properties of architectures and inverse problems include methods from algebraic geometry,[xix] perturbation theory, geometry, and topology.[xx] Again, this list is not exhaustive; proposals may use any approach.

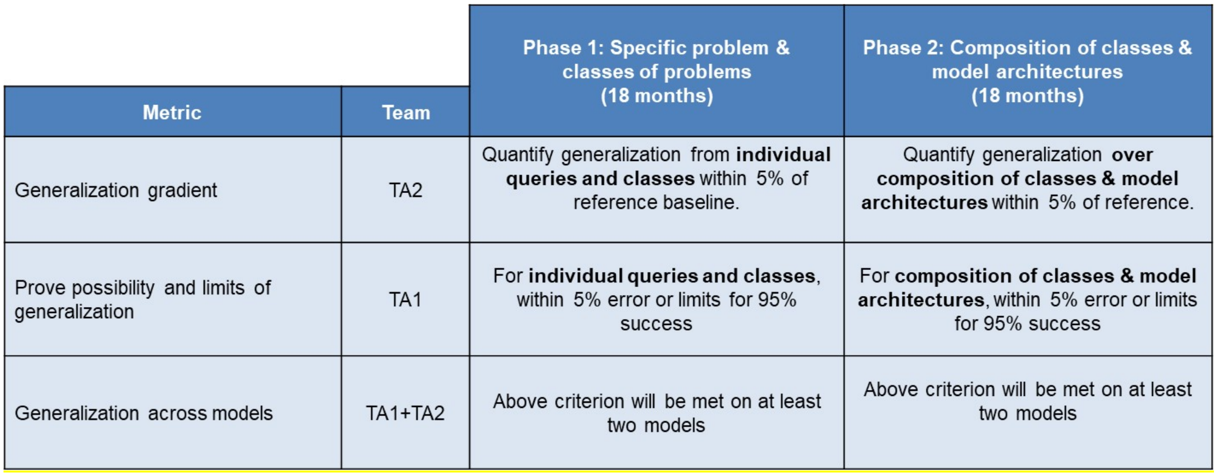
All TA1 proposers should include a plan for collaborating closely with TA2 to facilitate replication, integration, and scaling in their proposals. The plan should include delivery, schedule of results, software implementations, the tests to be conducted and datasets to be used along with conformance to software interfaces defined by TA2. The proposal should also identify any unique requirements of the developed methods (such as sampling) for TA2 integration, and the key assumptions underlying predicted accuracy at scale (typical of large-scale valuation problems).

TA2: TA2 will focus on empirical verification and integration of TA1 mathematical results and software into evaluation suites at scale. Specifically, TA2 will empirically document generalization behavior and verify practical utility and scalability of theoretical results and provide empirical investigations into model performance across questions and capability levels. Strong TA2 proposals will demonstrate the capability to evaluate at scale, and describe the plans to engage the broader research, policy, and industry communities, as well as plan for collaborating closely with TA1 teams. The government evaluation team will work with TA2 on developing tests of TA1 technologies and documentation of best practices for assessing evaluation methods.

The TA2 framework proposed should be general and applicable to generative AI approaches and be suitable for evaluating AIQ methods relative to SOTA which have been applied to text generation,[xxi] and integrated into a range of other tasks.[xxii,xxiii] TA2 proposers should propose ways to use the baselines that are already available, and develop new ones for those capabilities where SOTA baselines are not available. The baselining methods, datasets used, and hardware (servers) used should be described, explaining the need and relevance of each. Evaluation will be on open-source models such as Llama plus one of the following to demonstrate generality: phi-2, ViT, Llava, Kosmos-2, Stable Diffusion. These open-source models are chosen to span domains (language, vision, multimodal), sources (Facebook, Microsoft, Google), and designs (transformer, diffusion). The evaluation will "close" open-source models via finetuning.

Both TA1 and TA2 proposals should address experimental design; the former, for their internal evaluation and the latter, for program level evaluation. Across both TA1 and TA2 the goals are, given an evaluation result, to accurately predict performance on non-evaluated scenarios, specifying the training and testing datasets.

Experimental design is expected to vary by capability level. For the specific problems (point evaluations) case, the evaluation should consider methods such as leave-k-out cross-validation. Wherever relevant, the problems should be drawn from existing evaluation methods to facilitate comparison and demonstrate improvements in quantification. The TA1 methods should predict the test values based on the training data, and the program specified metric is that predictions should be within 5% of the test value. For classes of problems, the evaluation will be based on classes of capabilities necessary for successful interpretation and implementation, which is broken down into (language) ability, types or problems, and domains. The performance is quantified in terms of correct prediction of responses (and brevity). Whereas the previous level focused on generalizing to the neighborhood around a point, this level is focused on regions of space defined by coherent capabilities, which includes many possible specific problems. For natural classes (the inverse problem), the evaluation will be of the following form. Given a class of problems, choose the model that will perform best. Classes will be based those listed under classes of problems and models will be the same as the baselines listed above. For example, would Llama or Phi-2 perform better on logical reasoning?



Metric	Team	Phase 1: Specific problem & classes of problems (18 months)	Phase 2: Composition of classes & model architectures (18 months)
Generalization gradient	TA2	Quantify generalization from **individual queries and classes** within 5% of reference baseline.	Quantify generalization **over composition of classes & model architectures** within 5% of reference.
Prove possibility and limits of generalization	TA1	For **individual queries and classes**, within 5% error or limits for 95% success	For **composition of classes & model architectures**, within 5% error or limits for 95% success
Generalization across models	TA1+TA2	Above criterion will be met on at least two models	Above criterion will be met on at least two models

Figure 2: Metrics across TAs and phases.

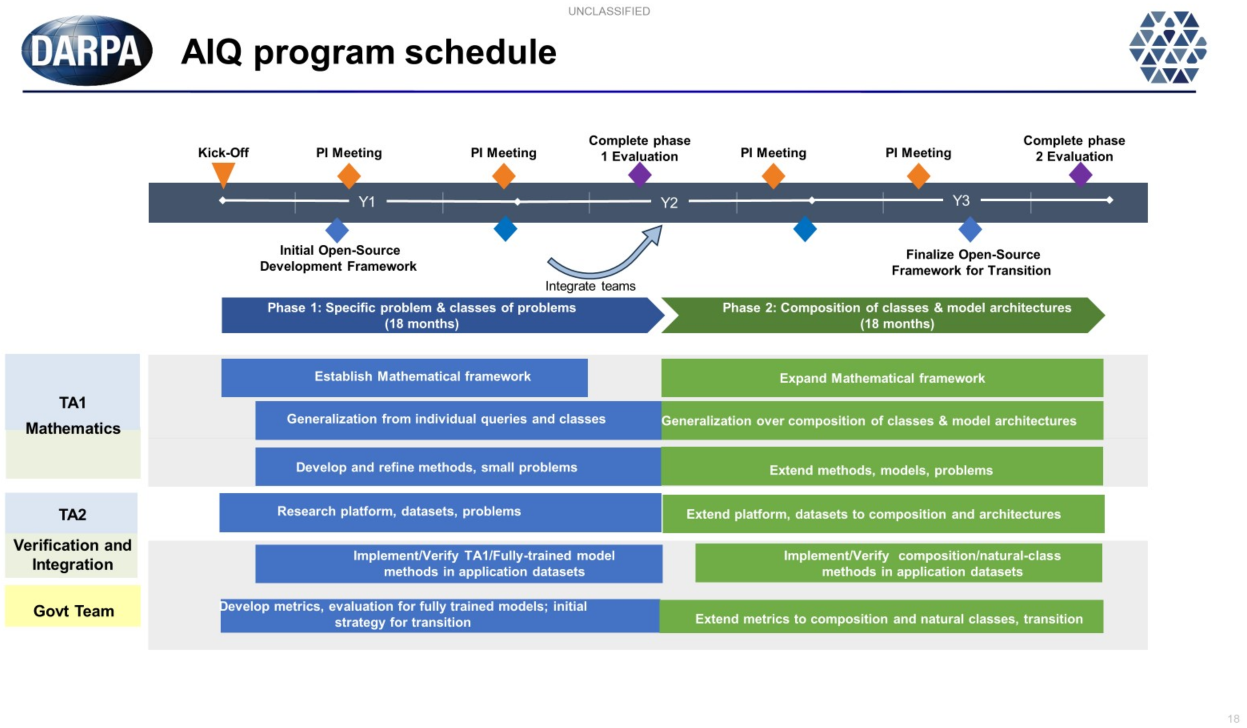
Program Phases and Metrics:

The program is divided into two phases of 18 months each. Phase 1 focuses on specific problems and classes of problems. Phase 2 focuses on compositions of classes and architectures. See Fig. 2 for metrics and Fig. 3 for the schedule.

Performance will be assessed across three metrics. For TA1, metrics emphasize mathematical precision regarding the possibility and/or limits of generalization from evaluations. For TA2, metrics emphasize empirical precision regarding generalization. Both TAs are expected to have results for more than one model.

Each phase will end with evaluations and demonstrations on the selected problems, which will provide an opportunity to evaluate the progress made against the program objectives. Proposers should submit, as a part of their Technical Volume, a detailed schedule of logically sequenced tasks and subtasks that in sum constitute a constructive plan for achieving the proposed technical objectives while appropriately managing risk. Schedules will be synchronized across performers, as required, and monitored and reviewed throughout the AIQ program's period of performance. For budgeting purposes, use January 15, 2025, as a start date for both TAs.

The Government will specify the locations for Principal Investigator (PI) meetings during program performance. There will be kick-off meeting and two PI meetings in Phase 1 held approximately six (6) months and twelve (12) months after the kick-off meeting. The evaluation meeting/workshop is held at the end of Phase 1. In Phase 2, there will be two PI meetings and a final evaluation meeting. PI meeting locations are likely to be spread across performer locations, and the proposers should plan to host at least one three day PI meeting with 40 participants over the duration of the program. The goals of the PI meetings will be to present new research findings and accomplishments, review plans for the next period, discuss implementation milestones, and resolve any programmatic, budgeting, or logistics issues. In addition to these program-wide events, the Government team will conduct site visits and will hold monthly teleconference meetings with each PI to enhance communications with the Government team. For travel planning and costing, assume seven (7) trips during the two phases per the program schedule shown (Fig. 3), alternating between Washington, DC and San Diego, CA, with each trip requiring 3-days and 2-nights.



UNCLASSIFIED

AIQ program schedule

Kick-Off

PI Meeting

PI Meeting

Complete phase 1 Evaluation

PI Meeting

PI Meeting

Complete phase 2 Evaluation

Y1

Y2

Y3

Initial Open-Source Development Framework

Integrate teams

Finalize Open-Source Framework for Transition

Phase 1: Specific problem & classes of problems (18 months)

Phase 2: Composition of classes & model architectures (18 months)

TA1 Mathematics

Establish Mathematical framework

Expand Mathematical framework

Generalization from individual queries and classes

Generalization over composition of classes & model architectures

Develop and refine methods, small problems

Extend methods, models, problems

TA2 Verification and Integration

Research platform, datasets, problems

Extend platform, datasets to composition and architectures

Implement/Verify TA1/Fully-trained model methods in application datasets

Implement/Verify composition/natural-class methods in application datasets

Govt Team

Develop metrics, evaluation for fully trained models; initial strategy for transition

Extend metrics to composition and natural classes, transition

18

Figure 3: Program Schedule

There are multiple points of essential collaboration among TAs, and the Government expects all performers to collaborate effectively. Proposers should read the descriptions of all TAs and the Program Assessments/Schedule section to ensure a full understanding of the program context, structure, and anticipated relationships required among performers. To facilitate the open exchange of information, all program performers will have Associate Contractor Agreement (ACA) language included in their award.

Open-source methodology and Software:

Intellectual property rights asserted by proposers are highly encouraged to align with open-source regimes, fostering a collaborative and transparent environment. The program places a strong emphasis on creating and leveraging open-source development, along with advocating the use of liberal open-source licensing (e.g., Apache, MIT). This strategy includes the establishment of open-source repositories (e.g., GitHub), which are accessible for review by the government team, other performers, and the wider research community. Such an approach is pivotal in promoting a culture of open innovation and shared knowledge. By facilitating this openness, the program aims to spur rapid innovation and continuous improvement. Openness and transparency are achieved by providing a robust foundation for future users and/or developers of the program’s technologies and deliverables. Moreover, this open-source methodology ensures that the advancements and learnings are not siloed but rather contribute to the collective intelligence of the field, leading to more significant and impactful technological progress.

Section II: Evaluation Criteria

- Proposals will be evaluated using the following criteria listed in ***descending order of importance***: Overall Scientific and Technical Merit; Potential Contribution and Relevance to the DARPA Mission; and Cost Realism.

    - **Overall Scientific and Technical Merit**: The proposed technical approach is innovative, feasible, achievable, and complete. The proposed technical team has the expertise and experience to accomplish the proposed tasks. Task descriptions and associated technical elements provided are complete and in a logical sequence with all proposed deliverables clearly defined such that a final outcome that achieves the goal can be expected as a result of award. The proposal identifies major technical risks and planned mitigation efforts are clearly defined and feasible.

    - **Potential Contribution and Relevance to the DARPA Mission**:
The potential contributions of the proposed effort bolster the national security technology base and support DARPA's mission to make pivotal early technology investments that create or prevent technological surprise. The proposer clearly demonstrates its capability to transition the technology to the research, industrial, and/or operational military communities in such a way as to enhance U.S. defense. In addition, the evaluation will take into consideration the extent to which the proposed intellectual property (IP) rights structure will potentially impact the Government's ability to transition the technology.

    - **Cost Realism**: The proposed costs are realistic for the technical and management approach and accurately reflect the technical goals and objectives of the solicitation. The proposed costs are consistent with the proposer's Statement of Work and reflect a sufficient understanding of the costs and level of effort needed to successfully accomplish the proposed technical approach. The costs for the prime proposer and proposed subawardees are substantiated by the details provided in the proposal (e.g., the type and number of labor hours proposed per task, the types and quantities of materials, equipment and fabrication costs, travel and any other applicable costs and the basis for the estimates). It is expected that the effort will leverage all available relevant prior research to obtain the maximum benefit from the available funding. For efforts with a likelihood of commercial application, appropriate direct cost sharing may be a positive factor in the evaluation. DARPA recognizes that undue emphasis on cost may motivate proposers to offer low-risk ideas with minimum uncertainty and to staff the effort with junior personnel to be in a more competitive posture. DARPA discourages such cost strategies and is interested in the most disruptive, transformational approaches.

- Unless otherwise specified in this announcement, for additional information on how DARPA reviews and evaluates proposals through the Scientific Review Process, please visit: Proposer Instructions and General Terms and Conditions

Section III: Submission Information

- This announcement allows for multiple award instrument types to be awarded to include Procurement Contracts, Cooperative Agreements, or Other Transactions. Some award instrument types have specific cost-sharing requirements. The following websites are incorporated by reference and contain additional information regarding overall proposer instructions, general terms and conditions, and each specific award instrument type.
  - **Proposer Instructions and General Terms and Conditions**: Proposer Instructions and General Terms and Conditions
  - **Procurement Contracts**: Proposer Instructions: Procurement Contracts
  - **Assistance (Cooperative Agreements)**: Proposer Instructions: Grants/Cooperative Agreements
  - **Other Transaction Agreements**: Proposer Instructions: Other Transactions

- This announcement contains a required abstract phase. Proposers are required to submit a two page abstract. Proposers will receive a written response to their abstract either encouraging or discouraging a full proposal submission. This written response will include a rationale for the decision. Proposers may only submit a full proposal if their abstract received an encourage response.Abstracts are due June 25, 2024, at 12:00 p.m., as stated in the Overview section. Additional instructions for abstract submission are contained within **Attachments A and B**.

- Full proposals are due August 13, 2024, at 12:00 p.m., as stated in the Overview section. Only proposers who receive a response encouraging a full proposal shall be eligible to submit a full proposal. **Attachments C, D, E, and F** contain specific instructions and templates and constitute a full proposal submission. Please visit Proposer Instructions and General Terms and Conditions for specific information regarding submission methods through the Broad Agency Announcement Tool (BAAT).

- **BAA Attachments**:
  - **(required) Attachment A**: Abstract Summary Slide Template
  - **(required) Attachment B**: Abstract Instructions and Template
  - **(required) Attachment C**: Proposal Summary Slide Template
  - **(required) Attachment D**: Proposal Instructions and Volume I Template (Technical and Management)
  - **(required) Attachment E**: Proposal Instructions and Volume II Template (Cost)
  - **(required) Attachment F**: MS ExcelTM DARPA Standard Cost Proposal Spreadsheet
  - **(informational) Attachment G**: Associate Contractor Agreement (ACA)

Section IV: Special Considerations

- This announcement, stated attachments, and websites incorporated by reference constitute the entire solicitation. In the event of a discrepancy between the announcement, attachments, or websites, the announcement shall take precedence.

- All responsible sources capable of satisfying the Government's needs, including both U.S. and non-U.S. sources, may submit a proposal that shall be considered by DARPA. Historically Black Colleges and Universities, Small Businesses, Small Disadvantaged Businesses, and Minority Institutions are encouraged to submit proposals and join others in submitting proposals; however, no portion of this announcement will be set aside for these organizations' participation due to the impracticality of reserving discrete or severable areas of this research for exclusive competition among these entities. Non-U.S. organizations and/or individuals may participate to the extent that such participants comply with any necessary nondisclosure agreements, security regulations, export control laws, and other governing statutes applicable under the circumstances.

- As of the time of publication of this solicitation, no proposal will be accepted that is classified. All proposal submissions are expected to be unclassified. Program work is expected to be unclassified.

- This program is subject to Attachment G: Associate Contractor Agreement.

- Federally Funded Research and Development Centers (FFRDCs), University Affiliated Research Centers, and Government entities interested in participating in the AIQ program or proposing to this BAA should first contact the Agency Point of Contact (POC) listed in the Overview section prior to the Abstract due date to discuss eligibility. Complete information regarding eligibility can be found at Proposer Instructions and General Terms and Conditions.

- As of the date of publication of this solicitation, the Government expects that program goals as described herein may be met by proposers intending to perform fundamental research and does not anticipate applying publication restrictions of any kind to individual awards for fundamental research that may result from this solicitation. Notwithstanding this statement of expectation, the Government is not prohibited from considering and selecting research proposals that, while perhaps not qualifying as fundamental research under the foregoing definition,[xxiv] still meet the solicitation criteria for submissions. If proposals are selected for award that offer other than a fundamental research solution, the Government will either work with the proposer to modify the proposed statement of work to bring the research back into line with fundamental research or else the proposer will agree to restrictions to receive an award. For additional information on fundamental research, please visit Proposer Instructions and General Terms and Conditions.

  Proposers should indicate in their proposal whether they believe the scope of the research included in their proposal is fundamental or not. While proposers should clearly explain the intended results of their research, the Government shall have sole discretion to determine whether the proposed research shall be considered fundamental and to select the award instrument type. Appropriate language will be included in resultant awards for non-fundamental research to prescribe publication requirements and other restrictions, as appropriate. This language can be found at Proposer Instructions and General Terms and Conditions.

For certain research projects, it may be possible that although the research to be performed by a potential awardee is non-fundamental research, its proposed subawardee's effort may be fundamental research. It is also possible that the research performed by a potential awardee is fundamental research while its proposed subawardee's effort may be non-fundamental research. In all cases, it is the potential awardee's responsibility to explain in its proposal which proposed efforts are fundamental research and why the proposed efforts should be considered fundamental research.

- DARPA's Fundamental Research Risk-Based Security Review Process (FERBS) (formerly CFIP) is an adaptive risk management security program designed to help protect the critical technology and performer intellectual property associated with DARPA's research projects by identifying the possible vectors of undue foreign influence. The DARPA team will create risk assessments of all proposed Senior/Key Personnel selected for negotiation of a fundamental research grant or cooperative agreement award. The DARPA risk assessment process will be conducted separately from the DARPA scientific review process and adjudicated prior to final award. For additional information on this process, please visit Proposer Instructions: Grants/Cooperative Agreements.

- The APEX Accelerators program, formerly known as the Procurement Technical Assistance Program (PTAP), focuses on building a strong, sustainable, and resilient U.S. supply chains by assisting a wide range of businesses that pursue and perform under contracts with the DoD, other federal agencies, state and local governments and with government prime contractors. See https://www.apexaccelerators.us/ for more information.
  APEX Accelerators helps businesses:

  o Complete registration with a wide range of databases necessary for them to participate in the government marketplace (e.g., SAM).
  o Identify which agencies and offices may need their products or services and how connect with buying agencies and offices.
  o Determine whether they are ready for government opportunities and how to position themselves to succeed.
  o Navigate solicitations and potential funding opportunities.
  o Receive notifications of government contract opportunities on a regular basis.
  o Network with buying officers, prime contractors, and other businesses.
  o Resolve performance issues and prepare for audit, only if the service is needed, after receiving an award.

- DARPAConnect offers free resources to potential performers to help them navigate DARPA, including "Understanding DARPA Award Vehicles and Solicitations," "Making the Most of

Proposers Days," and "Tips for DARPA Proposal Success." Join DARPAConnect at www.DARPAConnect.us to leverage on-demand learning and networking resources.

- DARPA has streamlined our Broad Agency Announcements and is interested in your feedback on this new format. Please send any comments to DARPAsolicitations@darpa.mil

- Project Spectrum is a nonprofit effort funded by the DoD Office of Small Business Programs to help educate the Defense Industrial Base (DIB) on compliance. Project Spectrum is vendor-neutral and available to assist businesses with their cybersecurity and compliance needs. Their mission is to improve cybersecurity readiness, resilience, and compliance for small/medium-sized businesses and the federal manufacturing supply chain. Project Spectrum events and programs will enhance awareness of cybersecurity threats within the manufacturing, research and development, as well as knowledge-based services sectors of the industrial base. Project Spectrum will leverage strategic partnerships within and outside of the DoD to accelerate the overall cybersecurity compliance of the DIB.

  www.Projectspectrum.io is a web portal that will provide resources such as individualized dashboards, a marketplace, and Pilot Program to help accelerate cybersecurity compliance.

[i] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354-359.

[ii] https://www.gao.gov/blog/how-artificial-intelligence-transforming-national-security

[iii] Accordingly, the DoD has issued directives (https://www.defense.gov/News/News-Stories/Article/Article/3578219/dod-releases-ai-adoption-strategy/) and made structural (https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf) changes to accelerate acquisition and integration of AI technologies.

[iv] https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf

[v] Schaeffer, R., Miranda, B., & Koyejo, S. (2023). Are emergent abilities of Large Language Models a mirage?. *arXiv preprint arXiv:2304.15004*.

[vi] Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., ... & Lee, K. (2023). Scalable extraction of training data from (production) language models. *arXiv preprint arXiv:2311.17035*.

[vii] Priebe, Carey (2023). Personal communication.

[xii] Arora, S., & Goyal, A. (2023). A theory for emergence of complex skills in language models. *arXiv preprint arXiv:2307.15936*.

[xiii] Stolz, B. J., Tanner, J., Harrington, H. A., & Nanda, V. (2020). Geometric anomaly detection in data. *Proceedings of the national academy of sciences*, *117*(33), 19664-19669.

[xvii] Sander, M. E., Ablin, P., Blondel, M., & Peyré, G. (2022, May). Sinkformers: Transformers with doubly stochastic attention. In *International Conference on Artificial Intelligence and Statistics* (pp. 3515-3530). PMLR.

[xviii] Geshkovski, B., Letrouit, C., Polyanskiy, Y., & Rigollet, P. (2023). A mathematical perspective on Transformers. *arXiv preprint arXiv:2312.10794*.

[xxi] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

[xxii] Du, Y., Watkins, O., Wang, Z., Colas, C., Darrell, T., Abbeel, P., ... & Andreas, J. (2023). Guiding pretraining in reinforcement learning with large language models. *arXiv preprint arXiv:2302.06692*.

[xxiii] Carta, T., Romac, C., Wolf, T., Lamprier, S., Sigaud, O., & Oudeyer, P. Y. (2023). Grounding large language models in interactive environments with online reinforcement learning. *arXiv preprint arXiv:2302.02662*.

[xxiv] 'Fundamental research' means basic and applied research in science and engineering, the results of which ordinarily are published and shared broadly within the scientific community, as distinguished from proprietary research and from industrial development, design, production, and product utilization, the results of which ordinarily are restricted for proprietary or national security reasons