Artificial Intelligence Exploration (AIE) Opportunity
DARPA-PA-20-02-02
Reduction of Entropy for Probabilistic Organization (REPO)

I. Opportunity Description

The Defense Advanced Research Projects Agency (DARPA) is issuing an Artificial Intelligence Exploration (AIE) Opportunity inviting submissions of innovative basic or applied research concepts in the technical domain of Artificial Intelligence Applications. This AIE Opportunity is issued under the Program Announcement for AIE, DARPA-PA-20-02. All proposals in response to the Reduction of Entropy for Probabilistic Organization (REPO), as described herein, will be submitted to DARPA-PA-20-02-02. All awards will be made in the form of an Other Transaction (OT) for prototype project. The total award value for the combined Phase 1 base and Phase 2 option is limited to $1,000,000. This total award value includes Government funding and performer cost share, if required or if proposed.

To view the original DARPA Program Announcement for AIE visit beta.SAM.gov under solicitation number DARPA-PA-20-02: https://beta.sam.gov/opp/667875ba2f464ccfa38688ea1a718fe7/view

A. Introduction

The goal of AIE REPO is to provide novel tools to analysts to enable the processing of large, unstructured and unorganized documentation datasets that contain electrical computer-aided design (ECAD) schematic files, mechanical CAD (MCAD) drawings, and flowchart/block diagram data. Successful REPO research tools to be developed by performers will:

- Conduct initial flow chart analysis and link block diagram data to text documentation in a single cohesive structure;
- Automate analysis of ECAD and MCAD to identify and export notable features in graphs and human-readable summary format for analysts; and
- Generate a chord diagram showing relationships between dataset entities and permit an analyst to consider relative magnitudes in comparison to alternative arcs to understand previously undiscoverable connections.

The effect of successful REPO research will reduce the technical and program risk of obsolescence management and re-engineering of legacy designs.

B. Objective/Scope

Design and development programs in aerospace and defense often feature high complexity and long design times. This translates into long sustainment times, often measuring decades, to pay off the incurred development costs. Examples include designs such as passenger aircraft, geosynchronous satellites, and jet turbines. When the inevitable technical refresh effort for such designs begins, the design intent for the equipment in question can often be problematic. REPO will provide the following tools to the researcher who has to analyze large, unstructured, and organized documentation datasets:

- Initial triage of the dataset will be performed, linking files by subject matter and topic.
- Flowchart and block diagram data will be linked to text documentation in a single cohesive structure.

- Machine analysis of ECAD schematic files and drawings to identify design intent and purpose, prepared in a summary format for analysts.
- Machine analysis of MCAD schematic files and drawings to identify design intent and purpose, prepared in a summary format for analysts.

It is desired that REPO deliverables would be able to perform an analysis similar to the following:

- Identify key performance microcircuits, such as field programmable gate arrays (FPGAs), central processing units (CPUs), and digital signal processors (DSPs), from part numbers found through machine analysis of PDF, JPEG/PNG, and common circuit design entry software, including Altium Designer, Cadence OrCad, Cadence Allegro, and Mentor Graphics DxDesigner.
- Parse makefiles, project build scripts, total command language (TCL) tool extensions, and integrated development environment (IDE) text files to include Xilinx ISE, Xilinx Vivado, Intel Quartus, Eclipse, ARM Keil, and IAR to identify which ASCII-format source files and binary blobs belong to the particular design project in question. File header information can be used to inform this analysis, and ASCII-format source files can be intended for languages including VHDL, Verilog, SystemVerilog, C/C++, Python, Java, Perl, Delphi, Lua.
- Use identified microcircuit information and parsed source code information to infer signal flow of the schematic by identifying key landmarks including principal system I/O, to include RF, PCIe, Ethernet, Serial, VME, VPX and analog interfaces.
- Associate identified schematic microcircuits with parsed source code to create an integrated model of the system, with analysis output formatted in a standardized intermediary text file that can be parsed programmatically.

REPO proposals will include:

- An ability to perform ECAD schematic identification to determine principal interfaces and hardware capabilities.
- An ECAD analyzer that can be easily augmented with additional schematic file formats matching alternative ECAD tool suites.
- A parser tool suite that can build on-the-fly test projects containing ASCII and binary source identified during analysis, to match IDE project files found, and can be easily expanded to include additional compilers, programming languages, and IDE projects/file formats.
- A text processing tool capable of identifying key schematic, test specification, statement of work documents, drawing title blocks, drawing/schematic titles, document numbers, and associating that information with source code and IDE projects.
- Version control system interfacing with industry standard tools, such as Git and Subversion, permitting introspection of repositories for identifying and parsing files of interest and producing a timeline of changes files of interest.
- Test data archives that will be used to train REPO applications and delivered to DARPA as part of the REPO effort.
- Continuous documentation generation built into all human-readable work products using Sphinx or a similar documentation generation tool.

C. Structure

The REPO program contains a single technical area and a proposal may contain no more than eight (8) pages including the cover sheet and all figures, tables, and charts. The page limit does not include a submission letter, the Task Description Document (TDD), or bibliography references (optional). All pages shall be formatted for printing on 8-1/2 by 11 inch paper with 1-inch margins and fixed-width font size not smaller than 12 point. Font sizes of 8 or 10 point may be used for figures, tables, and charts. Document files must be in .pdf, .odx, .doc, .docx, .xls, or .xlsx formats. Submissions must be written in English or VHDL. All pages should be numbered.

REPO will consist of a 12-month Phase 1, followed by 6-month Phase 2. Phase 1 will conduct a feasibility study of automated analysis of ECAD schematics, flowcharts, diagrams, and natural language documents, via a variety of AI methods. Phase 2 will explore combining information from Phase 1 with parsed source code files to create as complete a description of the system's design intent as possible.

Proposals submitted to DARPA-PA-20-02 in response to the technical area of this AIE Opportunity must be UNCLASSIFIED and must address two independent and sequential project phases (a Phase 1 Feasibility Study (base) and a Phase 2 Proof of Concept (option)). The periods of performance for these phases are 12 months for the Phase 1 base effort and 6 months for the Phase 2 option effort. Combined Phase 1 base and Phase 2 option efforts for this AIE Opportunity should not exceed 18 months. The Phase 1 (base) award value is limited to $600,000 and the Phase 2 (option) award value is limited to $400,000 to include performer cost share, if required or if proposed. The total award value for the combined Phase 1 and Phase 2 is limited to $1,000,000. This total award value includes Government funding and performer cost share, if required or if proposed.

D. Schedule/Milestones

Proposers must address the following Research Project Objectives, metrics, and deliverables, along with fixed payable milestones in their proposals. The task structure must be consistent across the proposed schedule, TDD, and the Volume 2 - Price Volume. Proposers must complete the "Schedule of Milestones and Payments" excel Attachment provided with this AIE Opportunity as part of submitting a complete proposal and fulfilling the requirements under Volume 2 - Price Volume. If selected for award negotiation, the fixed payable milestones provided will be directly incorporated into Attachment 3 of the OT agreement ("Schedule of Milestones and Payments"). Proposers are encouraged to use the TDD template provided with the Program Announcement DARPA-PA-20-02, which will be Attachment 1 of the OT agreement.

The effort will show successful progression with the following significant milestones from reliable identification of mathematical primitives aided by source-level representations at 3 months, to reliable extraction of key schematic identifiers (at 6 months), to source and binary file triage (aided by ECAD identifiers) at 9 months, to parsing and re-creation of IDE and design tool projects (informed by ECAD documents/source files analyzed previously) at 12 months, with corresponding milestones for informing for parsing version control systems and natural language documents in 15 and 18 months respectively.

Phase 1 fixed milestones for this program must include the following:

Month 1: Milestone Report to include detailed experimental plan for the identified use cases. The plan

must identify domain use case(s) and representations to be used, outline AI approaches to explore, and their training sources, and detail approaches to automated augmentation and generation of training corpora. Deliverable for this milestone will exclusively consist of a written report defining the experimental plan.

Month 3: Milestone Report to include initial look at input data. Performers should be able to carry out a rudimentary pass against unstructured datasets and provide some extraction of data items from schematics and other visual circuit representations.

Month 6: Milestone Report to include key schematic identifier extraction. Performers should be able to identify the "primary" integrated circuit devices on visual circuit representations that would drive ancillary data set items. An example of this is identifying the use of a CPU in a schematic for a circuit card, and determining that a VHDL source file is likely not related to that CPU for the board design in question.

Month 9: Milestone Report to include reliable fusion of source code with visual circuit representations. Performers should be able to fuse the developed methods of parsing visual circuit representations with relevant text source code and object code that has been identified from data sets. The ability to identify key integrated circuit devices in the visual documentation and correlate that to source code in the data set needs to be present at this point.

Month 12: Milestone Report to include dataset project/IDE recreation. Performers should be able to parse project configuration files, associate the information extracted with source files that belong to that project, and be able to automatically re-create the project with sources while identifying any gaps or missing items in the project structure. An example of this is parsing the *.xpr extension project file for Xilinx Vivado, identifying the HDL, checkpoint, netlist, and constraint files that belong to this project, and determining what is missing in the dataset that would prevent successful synthesis from taking place. Similar exercise would be performed for an Altium schematic/PCB design project and its *.PrjPcb project file, or another equivalent tool.

Phase 2 fixed milestones for this program must include the following:

Month 15: Milestone Report to include version control system integration and rudimentary interface identification. Performers should be able to inform the analysis done thus far with information obtained from machine analysis of commits and other records preserved in tools such as Git/Subversion. Additionally, performers should be able to obtain some design intent information from the schematics analyzed, specifically with regards to primary I/O interfaces employed. The automated recognition of common I/O standards, such as RJ45 connectors for Ethernet use, PCIe connectors/gold fingers, or USB connectors for USB use, would be achieved at this point.

Month 18: Milestone Report to include natural language document integration. Performers should be able to inform the analysis done thus far with information obtained from non-structured visual representations such as flowcharts and block diagrams, as well as text documents such as reports. Additionally, performers should provide a file-based and graphical representation of relationships obtained between dataset items. An example of this would be the automated generation of a link between schematic files, an FPGA on the schematic, VHDL text files for the FPGA, constraint file for the FPGA, and the test specification for the board design, along with recognition of key interfaces that would define the overall board product with FPGA as being an Ethernet appliance.

For planning and budgetary purposes, proposers should assume a program start date of **December 4**, **2020**. Schedules will be synchronized across performers, as required, and monitored/revised as necessary throughout the program.

All proposals must include the following meetings and travel in the proposed schedule and costs:

- To foster collaboration between teams and disseminate program developments, a two-day Principal Investigator (PI) meeting will be held approximately every six months, with locations split between the East and West Coasts of the United States. For budgeting purposes, plan for three two-day meetings over the course of 18 months: two meetings in the Washington, D.C. area and one meeting in the San Francisco, CA area. If necessary, meetings may be held virtually.
- Regular teleconference meetings will be scheduled with the Government team for progress reporting as well as problem identification and mitigation. Proposers should also anticipate at least one site visit per phase by the DARPA Program Manager during which they will have the opportunity to demonstrate progress towards agreed-upon milestones.

E. Deliverables

Every milestone, unless otherwise noted, will deliver the following:

- Remote access to the Git repository storing all work products developed under REPO, along with the specific tagged release of the Git repository corresponding to the milestone being delivered.
- Remote access to the datasets used for training and testing REPO products.
- Remote access to the training and inference hardware used for training and testing REPO products.
- All source code required to make, compile, build, configure, train, and run REPO source code on all training and inference platforms and targets employed (on/off premises), with no limitations.
- All generated output from the REPO products, including any intermediate files produced during the training or inference run. Results should include a file representation that specifies the contents of the data set, and can be parsed with automated tools in a convenient manner.
- Properly annotated and generated output from the continuous documentation tools, with additional annotations to make a cohesive human-readable document on PDF format, accompanying the milestone report.

Negotiated deliverables specific to the objectives of the individual efforts. These may include registered reports, experimental protocols, publications, intermediate and final versions of software libraries, code, and APIs, including documentation and user manuals, and/or a comprehensive assemblage of design documents, models, modeling data and results, and model validation data.

II. Award Information

Selected proposals that are successfully negotiated will result in award of an OT for prototype project. See Section 3 of AIE Program Announcement DARPA-PA-20-02 for information on awards that may result from proposals submitted in response to this notice.

Proposers must review the model OT for Prototype agreement provided as an attachment to the AIE Program Announcement DARPA-PA-20-02 prior to submitting a proposal. DARPA has provided the

model OT in order to expedite the negotiation and award process and ensure DARPA achieves the goal of AIE which is to enable DARPA to initiate a new investment in less than 90 days from idea inception. The model OT is representative of the terms and conditions that DARPA intends to award for all AIE Awards. The task description document, schedule of milestones and payments, and data rights assertions requested under Volumes 1, 2, and 3 will be included as attachments to the OT agreement upon negotiation and award.

Proposers may suggest edits to the model OT for consideration by DARPA and provide a copy of the model OT with track changes as part of their proposal package. Suggested edits may not be accepted by DARPA. The Government reserves the right to remove a proposal from award consideration should the parties fail to reach agreement on OT award terms and conditions. If edits to the model OT are not provided as part of the proposal package, DARPA assumes that the proposer has reviewed and accepted the award terms and conditions to which they may have to adhere and the sample OT agreement provided as an attachment, indicating agreement (in principle) with the listed terms and conditions applicable to the specific award instrument.

In order to ensure that DARPA achieves the AIE goal of award within 90 days from the posting date **(September 8, 2020)** of this announcement, DARPA reserves the right to cease negotiations when an award is not executed by both parties (DARPA and the selected organization) on or before **December 4, 2020.**

III. Eligibility

See Section 4 of the AIE Program Announcement DARPA-PA-20-02 for information on who may be eligible to respond to this notice.

IV. AIE Opportunity Responses

A. Proposal Content and Format

All proposals submitted in response to this notice must comply with the content and format instructions in Section 5 of DARPA-PA-20-02. All proposals must use the templates provided as Attachments to the Program Announcement and the "Schedule of Milestones and Payments" Excel Attachment provided with this AIE Opportunity and the and follow the instructions therein.

Information not explicitly requested in DARPA-PA-20-02, its Attachments, or this notice may not be evaluated.

B. Proposal Submission Instructions

Responses to DARPA-PA-20-02-02 shall be submitted through electronic upload to DARPA's BAA Portal (https://baa.darpa.mil).

DARPA will acknowledge receipt of complete submissions via email and assign identifying numbers that should be used in all further correspondence regarding those submissions. If no confirmation is received within two business days, please contact REPO@darpa.mil to verify receipt.

When planning a response to this AIE Opportunity, proposers should take into account the submission time zone and that some parts of the submission process may take from one business day to one month

to complete (e.g., registering for a Data Universal Numbering System (DUNS) number or Tax Identification Number (TIN)).

C. Electronic Upload

First time users of the DARPA BAA Portal must complete a two-step account creation process. The first step consists of registering for an extranet account by going to the URL listed above and selecting the "Account Request" link. Upon completion of the online form, proposers will receive two separate emails; one will contain a user name and the second will provide a temporary password. Once both emails have been received, the second step requires proposers to go back to the submission website and log in using that user name and password. After accessing the extranet, proposers may then create a user account for the DARPA submission website by selecting the "Register your Organization" link at the top of the page. Once the user account is created, proposers will be able to see a list of solicitations open for submissions, view submission instructions, and upload/finalize their proposal.

Proposers who already have an account on the DARPA BAA Portal may simply log in at https://baa.darpa.mil, select this solicitation from the list of open DARPA solicitations and proceed with their proposal submission. Note: proposers who have created a DARPA submission website account to submit to another DARPA Technical Office's solicitations do not need to create a new account to submit to this solicitation.

All full proposals submitted electronically through the DARPA submission website must meet the following requirements: (1) uploaded as a zip file (.zip or .zipx extension); (2) only contain the document(s) requested herein; (3) only contain unclassified information; and (4) must not exceed 100 MB in size. Only one zip file will be accepted per full proposal. Full proposals not uploaded as zip files will be rejected by DARPA. Technical support for the DARPA submission website is available during regular business hours, Monday – Friday, 9:00 a.m. – 5:00 p.m. Requests for technical support must be emailed to BAAT_Support@darpa.mil with a copy to REPO@darpa.mil. Questions regarding submission contents, format, deadlines, etc. should be emailed to REPO@darpa.mil. Questions/requests for support sent to any other email address may result in delayed/no response.

Since proposers may encounter heavy traffic on the web server, DARPA discourages waiting until the day proposals are due to request an account and/or upload the submission. Note: Proposers submitting a proposal via the DARPA submission site MUST (1) click the "Finalize" button in order for the submission to upload AND (2) do so with sufficient time for the upload to complete prior to the deadline. Failure to do so will result in a late submission.

D. Proposal Due Date and Time

Proposals in response to this notice are due no later than 4:00 PM ET on October 7, 2020. Full proposal packages as described in Section 5 of DARPA-PA-20-02 must be submitted per the instructions outlined in this AIE Opportunity *and received by DARPA* no later than the above time and date. Proposals received after this time and date may not be reviewed.

Proposers are warned that the proposal deadline outlined herein is in Eastern Time and will be strictly enforced. When planning a response to this notice, proposers should take into account that some parts of the submission process may take from one business day to one month to complete.

V. Proposal Evaluation and Selection

Proposals will be evaluated and selected in accordance with Section 6 of DARPA-PA-20-02. Proposers will be notified of the results of this process as described in Section 7.1 of DARPA-PA-20-02.

VI. Administrative and National Policy Requirements

Section 7.2 of DARPA-PA-20-02 provides information on Administrative and National Policy Requirements that may be applicable for proposal submission as well as performance under an award.

VII. Point of Contact Information

Walter Weiss, Program Manager, DARPA/I2O, REPO@darpa.mil

VIII. Frequently Asked Questions (FAQs)

All technical, contractual, and administrative questions regarding this notice must be emailed to REPO@darpa.mil. Emails sent directly to the Program Manager or any other address may result in delayed or no response.

All questions must be in English and must include name, email address, and the telephone number of a point of contact. DARPA will attempt to answer questions publically in a timely manner; however, questions submitted within 7 days of the proposal due date listed herein may not be answered.

DARPA will post an FAQ list under the AIE Opportunity on the DARPA/I2O Opportunities page at http://www.darpa.mil/work-with-us/opportunities. The list will be updated on an ongoing basis until one week prior to the proposal due date. In addition to the FAQ specific to this notice (DARPA-PA-20-02-02), proposers should also review the Program Announcement for AIE General FAQ list on the DARPA/DSO Opportunities page under the Program Announcement for AIE (DARPA-PA-20-02).