Artificial Intelligence Exploration (AIE) Opportunity
DARPA-PA-20-02-01
Recovery of Symbolic Mathematics from Code (ReMath)

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 recovering symbolic mathematical expressions instantiated in software's code. This AIE Opportunity is issued under the Program Announcement for AIE, DARPA-PA-20-02-01. 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, as 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

Our society increasingly depends on automation of physical processes, in industrial production, transportation, communication, and other advanced technological domains. This automation relies on embedded computers that implement mathematical algorithms to control physical processes. This reliance on automation creates an increasing need for rapidly understanding which standard mathematical algorithms and which variations thereof are implemented by embedded software, including in legacy industrial control systems and other cyber-physical systems (CPS).

Today, extracting symbolic representations of software's algorithmic parts, such as control laws for a physical process encoded in a CPS, requires fully manual analysis by highly specialized experts. There is no mechanized capability to translate and route relevant parts of the software to experts, such as control engineers, in an appropriate form for them to effectively analyze the mathematical expressions. In contrast, malware analysis has become considerably automated with aspects such as provenance and behavioral characterization gaining considerable traction in recent years.

Recent research demonstrates effective use of Artificial Intelligence (AI) translation techniques for tasks like mechanized symbolic differentiation and integration, in which going in one direction is relatively straightforward. In contrast, inverse problems like symbolic integration are more challenging and require expert ingenuity in repeated recognition and application of patterns.

Although programming mathematical equations is not as straightforward as symbolic differentiation, it is still a relatively straightforward skill with massive quantities of available examples and broadly taught practices. The inverse problem of recovering a mathematical description of software, even when the function of the software is known, has defied automation. This effort will seek to explore this problem space.

B. Objective and Technical Scope

The ReMath AI (Artificial Intelligence) Exploration program will discover whether a combination of recent advances in AI techniques, such as neural machine translation, sequence-to-sequence encoders, etc., can effectively recover mathematical structures implemented in software (both in source code and in compiled binary code) into their natural mathematical forms of symbolic expression. These techniques will improve the understanding of complex software and may enable future methods for analysis and testing of cyber physical systems.

ReMath proposals will identify and address challenges to application of AI methods, including (but not limited to) overcoming any potential scarcity of training data and developing novel effective representations that capture how mathematical symbolic units such as mathematical functions, equations, and formulas used by subject matter experts (SMEs) to describe and communicate about the domain concepts ("domain communication units") are instantiated in source and binary code. In particular, proposers should discuss effective ways of automatically augmenting existing training sources or corpora, or automatically generating new training corpora, and should discuss problem representations that take advantage of expert analyses' outcomes or captured domain SME work flows when understanding a system (if available).

Proposers should demonstrate the proposing team's experience and expertise in analysis of Cyber Physical Systems (CPS) and embedded systems, combined with relevant program and binary analysis and AI methods expertise. Strong proposals shall focus on scenarios from a selected CPS domain(s) where the proposer's team has SME-level expertise, which includes communicating the resulting findings to experts in other domains for collaborative analysis. Proposers should make effective use of SME effort and inputs, in ways transferable to other CPS domains. Proposers should plan to demonstrate efficacy of their approach(es) on at least two computing architectures representative of industry CPS computing platforms and illustrative of a variety of architectural platforms.

Static, dynamic, or hybrid program and binary analyses are all in scope, using any combination of symbolic, concrete, or emulation-based approaches, or orchestrated mixes of thereof, including hardware approaches, if required, for the proposed use cases. Proposers should demonstrate integration with leading open-source tools used by industry SMEs, improving the tools' outputs and SME's work flows.

C. Structure

ReMath AIE will consist of a 9-month Phase 1 followed by 9-month Phase 2. Phase 1 will conduct a feasibility study of recovering symbolic mathematical formulas, or classes of formulas, from CPS binary code implementing them, aided by source code, via a variety of AI methods. Phase 2 will explore recovery of symbolic mathematical formulas, or classes of formulas, from CPS binary code alone.

Proposals submitted to DARPA-PA-20-02-01 in response to this AIE Opportunity must be UNCLASSIFIED and have an 8-page limit. Proposals 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 9 months for the Phase 1 (base) effort and 9 months for the Phase 2 (option) effort. Combined, Phase 1 and Phase 2 efforts for this AIE Opportunity should not exceed 18 months. The Phase 1 (base) award value is limited to $500,000. The Phase 2 (option) award value is limited to $500,000. The total award value for the

award is limited to $1,000,000. This total award value includes government funding and performer cost share, if required.

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, Task Description Document (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 2 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 efforts 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 formulaic expressions in these primitives (aided by source-level representations, at 6 months), to extraction of SME domain communication units (at 9 months), with corresponding milestones for binary-level representations unaided by source code, in 12, 15, and 18 months respectively.

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

Month 1: Detailed experimental plan for the identified SME 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.

Month 3: Reliable identification of mathematical primitives aided by source-level representations (source code, pseudo-code, or code in a domain-specific language). The report for this milestone will include refinement of automatic generation of training data.

Month 6: Reliable extraction of formulaic expressions aided by source-level representations. The report for this milestone will include an initial demonstration of integration with modular program analysis tools, providing SME-aiding automation in the selected use case(s).

Month 9: Extraction of SME domain communication units aided by source-level representations. Reports of successful case studies with application of AI methods, with methodological summaries, algorithms, datasets, and finalized formalisms consolidated in a report. For unsuccessful case studies, a report of the encountered obstacles and summaries of the approaches attempted.

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

Month 12: Reliable identification of mathematical primitives from binary-level code, unaided by source-level representations. The report for this milestone must include details of automated training corpora augmentation and generation for binary code.

Month 15: Reliable extraction of formulaic expressions from binary-level code (compiled executable code or bytecode), unaided by source-level representations. Demonstrated integration with modular binary analysis tools with a detailed case study of SME-aiding automation on binary code in the selected use cases.

Month 18: Extraction of SME domain communication units from binary-level representations, final prototype and report. Reports of successful case studies with application of AI methods, with methodological summaries, algorithms, datasets, and finalized formalisms consolidated in a report. For unsuccessful case studies, a report of the encountered obstacles and summaries of the approaches attempted.

For planning and budgetary purposes, proposers should assume a program start date of November 17, 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 one-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 one-day meetings over the course of 18 months: two meetings in the Washington, D.C. area and one meeting in the San Diego, CA area. In-person meetings may be replaced with virtual meetings, should conditions dictate.
- 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.

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 calendar 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 calendar days from the posting date (August 20, 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 November 17, 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-20-02-01 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 ReMath@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)).

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 ReMath@darpa.mil. Questions regarding submission contents, format, deadlines, etc. should be emailed to ReMath@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.

C. Proposal Due Date and Time

Proposals in response to this notice are due no later than 12:00 PM on September 18, 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 the AIE Program Announcement 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

Sergey Bratus, Program Manager, DARPA/I2O, ReMath@darpa.mil

VIII. Frequently Asked Questions (FAQs)

All technical, contractual, and administrative questions regarding this notice must be emailed to ReMath@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 business 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-20-02-01), 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).