

IARPA Request for Information on Novel Training Datasets and Environments to Advance Artificial Intelligence

The Intelligence Advanced Research Projects Activity (IARPA) is seeking information on novel training datasets and environments to advance artificial intelligence (AI). This request for information (RFI) is issued solely for information gathering purposes; the RFI does not constitute a formal solicitation for proposals. IARPA anticipates that responses to this RFI will be used to inform future funding opportunities for creating novel training resources for artificial intelligence algorithms. The following sections of this announcement contain details of the scope of technical efforts of interest, along with instructions for the submission of responses.

Background and Scope

Artificial intelligence, defined here as computer simulation of cognitive processes such as perception, recognition, reasoning, and control, have captured the public's imagination for over 60 years. However, artificial intelligence research has proceeded in fits and starts over much of that time, as the field repeats a boom/bust cycle characterized by promising bursts of progress followed by inflated expectations and finally disillusionment, leading to what has become known as an "AI winter" – a long period of diminished research and funding activity. Until recently, the conventional wisdom has been that new algorithms were the limiting factor in making steady progress towards artificial intelligence. However, recent advances in machine learning, a sub-field of artificial intelligence, have established that historical algorithms (e.g. backpropagation) in conjunction with high-performance computers can be used to achieve nearly human-level performance on diverse tasks such as image and speech recognition, language translation, and video game play. In each of these instances, and in many others, rapid progress was facilitated by the availability of massive amounts of training data well-suited to the problem under study. This realization raises the prospect that many additional artificial intelligence problems may be solvable in the near-term, without significant innovations in the underlying algorithms, if the right training resources become widely available.

Training resources for algorithms such as discriminative classifiers and generative models typically require a large collection of static labeled samples, each of which is represented as an (input, output) pair. For example, training data for an object recognition algorithm might consist of a set of images of objects or scenes, each of which is accompanied by a textual description of the object type(s) contained therein. In contrast, training resources for reinforcement learning algorithms typically provide a dynamic interactive environment for a simulated agent and supply evaluative feedback based on the agent's actions. Examples include simulators such as MuJoCo and the Arcade Learning Environment.

Through this RFI, IARPA is seeking input from the artificial intelligence research community on training resources that, if created, would be most likely to drive progress in new problem domains within the field of artificial intelligence (including machine learning). In particular, respondents are asked to answer one or more of the following questions:

1. Which problem domain(s) has the greatest potential to benefit from the availability of new training resources and why?
2. What new training resources are needed to achieve significant progress in this domain? How should these resources be structured? How do the proposed resources compare with currently available resources?
3. What kind of effort is needed to create and/or curate these training resources? What technical, logistical, and/or legal challenges would be associated with such an effort? How much would such an effort cost, and how long would it take? How much effort and money would be required to store, maintain, distribute, and/or utilize the proposed training resources?
4. Who would be the major stakeholders in the proposed training resources? How would these stakeholders use the proposed resources?
5. Annual challenges (e.g. ImageNet Large Scale Visual Recognition Challenge) employing a standard set of data for training and/or evaluation have helped to catalyze progress in many machine learning problem domains. Should a challenge be created in the proposed problem domain, and if so, how should it be designed, implemented, and judged?

Preparation Instructions to Respondents

IARPA appreciates responses from all capable and qualified sources from within and outside of the US. This announcement contains all of the information required to submit a response. No additional forms, kits, or other materials are needed.

All responses shall be formatted for printing on standard 8.5" x 11" paper with 1-inch margins. Text shall be no smaller than 12-point Times New Roman font. Responses shall include:

- A 1-page cover sheet that identifies the document as a response to IARPA RFI 16-03; lists all contributing authors, their respective organizations, and their email addresses; and provides a primary technical and administrative point of contact;
- A substantive, focused, one-paragraph executive summary;
- Responses to the questions enumerated above (limited to 10 pages);
- A list of citations; and
- Copies of any unpublished or otherwise inaccessible manuscripts referenced in the text.

Submission Instructions to Respondents

Responses to this RFI are due no later than 5:00pm Eastern Time on Friday, April 1, 2016. All submissions must be electronically submitted to dni-iarpa-rfi-16-03@iarpa.gov as a single PDF document. Inquiries to this RFI must be submitted to dni-iarpa-rfi-16-03@iarpa.gov. Do not send questions with proprietary content. No telephone inquiries will be accepted.

Disclaimers and Important Notes

This is an RFI issued solely for information and planning purposes and does not constitute a solicitation. Respondents are advised that IARPA is under no obligation to acknowledge receipt of the information received, or provide feedback to respondents with respect to any information submitted under this RFI. Responses to this notice are not offers and cannot be accepted by the Government to form a binding contract. Respondents are solely responsible for all expenses associated with responding to this RFI. IARPA will not provide reimbursement for costs incurred in responding to this RFI. It is the respondent's responsibility to ensure that the submitted material has been approved for public release by the information owner.

The Government does not intend to award a contract on the basis of this RFI or to otherwise pay for the information solicited, nor is the Government obligated to issue a solicitation based on responses received. Neither proprietary nor classified concepts or information should be included in the submission. Input on technical aspects of the responses may be solicited by IARPA from non-Government consultants/experts who are bound by appropriate non-disclosure requirements.