

HR001119S0005
Machine Common Sense
Frequently Asked Questions

As of December 4, 2018

Q41: The level of effort spreadsheet provided with the BAA only has three phases. If we should consider the program to be four phases (one phase for each year), can the spreadsheet be modified?

A41: The level of effort spreadsheet provided with the BAA is only a template. Proposers should modify the spreadsheet as needed (e.g., insert an additional row for a fourth phase of the program, or modify the spreadsheet to reflect the corresponding government fiscal years).

Q40: Is a permanent resident of the U.S. (i.e., a Green card holder) eligible to submit a proposal?

A40: Yes, a permanent resident of the U.S. (i.e., Green card holder) is eligible to submit a proposal in response to the BAA.

Q39: Does the program cap or limit the indirect cost rates of foreign subcontractors?

A39: The program does not cap or limit the indirect cost rates of foreign subcontractors. However, per the BAA, proposers must clearly identify all indirect cost rates (including Fringe Benefits, Overhead, G&A, Facilities Cost of Money, etc.) and the basis for each.

Q38: For the cost volume of the proposal, the BAA indicates that two tables should be provided: one table with costs per fiscal year of the program and a second table with costs per phase of the program. However, the BAA does not mention "program phases", so how should the second table be structured?

A38: Proposers should consider each year as a phase, and provide a table with costs broken down by phase (four phases, one for each year of the program). DARPA anticipates a June 2019 start date for the MCS program that will run for a duration of 48 months. Proposers should also provide a second table with costs broken down by government fiscal year.

As of November 5, 2018

Q37: Are proposers restricted to using the cost estimate that is submitted with the abstract at proposal stage or will proposers be able to make adjustments if necessary when the complete proposal is submitted?

A37: Proposers are not restricted to using the cost estimate that is submitted with the abstract at proposal stage. Proposers can make adjustments for the complete proposal if necessary.

Q36: Can the PI submit [an abstract or proposal] to the electronic submission system or does submission need to be made by an authorized organizational representative?

A36: The decision regarding who is authorized to submit an abstract or a binding proposal is made internally by the proposing organization.

Q35: Although the MCS program can definitely benefit from the involvement of developmental psychology expertise (especially in the conceptualization of training and testing sets/problems/scenarios), can you confirm that TA2 is not a developmental psychology project, but is more a technology test and evaluation project?

A35: Yes, that is correct. The TA2 task needs developmental psychology expertise, but it is not primarily a developmental psychology research effort. But there is some significant developmental psychology work to be done.

As of October 30, 2018

Q34: Is it appropriate to propose additional downstream applications beyond the official test bed? How does common sense affect the ability to do machine translation?

A34: This would go beyond the expectations from a proposal (i.e., not only developing the commonsense service but also using it for machine translation) and does not need to be part of a proposal. Although if you believe you may have something that can do this in 2-3 years, DARPA is always interested in transitions out of these projects. This is a 6.1 program and is not DARPA's primary concern for this program, but it would be great if that did happen.

Q33: For TA3, we notice that for the AI2 commonsense natural language inference test, there are 70K instances released for training, 20k for development, 20K for blind evaluations. Does DARPA expect that the other four test sets in TA3 will be comparable in size?

A33: Yes, although there may be some variation. They will be relatively comparable in size, but the main constraint, as AI2 develops the tests is the cost of development. They are constructing the blind test sets out of their budget as a service to DARPA and the research community. So they may have fewer in certain instances due to cost constraints. But even if there are fewer, there will be ample training, development, and test to allow for strong training using modern methods, and allow for clear, statistically significant results on the blind evaluation tests.

Q32: Do all TA3 efforts need to address the AI2 tests?

A32: See A9. DARPA prefers the TA3 efforts not specialize in a niche in one particular question or the other. You can sequence the development. You might do really well on just one or two tests early on and then work on the other test sets later. However, DARPA desires proposals that take on all of the test sets.

Q31: The AI2 tests do not mention problem solving or decision making. Is there interest in having TA3's AI librarian take in a situation and recommend an action or a sequence of actions?

A31: All of the AI2 tests are multiple choice. You are given a language description of the question, or a language and image description, and you have to select the right answer from a multiple choice list. That might involve seeing a situation and then having to understand what might happen next. Or have to reason about something in the situation in order to select the answer. You will not be required to create a plan as an output of the system to be judged; but you will need some reasoning capability in order to answer the questions. Although there are a lot of approaches out there on just training the system to have the right reactions to everything versus having a rich reasoning system that takes in the situation and then reasons about what the answer should be.

Q30: Can we tell the system explicitly what problem to solve or does it have to figure it out by itself?

A30: For TA1 and TA2, early on in the program, it will be shown a video and then give an expectation signal. We want this to be as general as possible but we envision that early in the program you might give it an explicit problem to solve somehow. It could be a very important learning capability to turn it loose in a simulated world and have it learn through curiosity, experiences, etc. on its own. It is encouraged that it has its own ability to figure out what to do, but we expect a lot of the test problems, especially in the early years, might have a more explicit starting point.

Q29: How general does it have to be? Is it possible to focus on one of objects, agents, or places?

A29: DARPA believes it is important to address all of these together in order to capture a general common sense capability. It is not necessary to perform equally on all of these over the course of the program – it is fine to sequence what you want to do – but it is important that you really try to look at these general underlying capabilities that happen with all three of these.

Q28: Does it have to learn everything from scratch or can it build on basic knowledge?

A28: DARPA is open to either approach and anticipates proposals with a range of solutions. Children are born with a lot of machinery – some psychologists describe it as mammal commonsense – it has some predefined notion of how to deal with objects and space. DARPA is open to proposals that build something in, but all of it should not be built in. This is an important part of the experience learning and problem solving tasks. The agent needs a learning capability so it can learn from new experiences. On the other hand, DARPA expects that some proposals will propose to learn from scratch (i.e., they will not build in any kind of artificial representation) and essentially simulate evolution. Expose it to millions of experiences until it learns its own representation of intuitive physics. DARPA expects a range of possible approaches to be proposed.

Q27: Are you looking for psychological realism in the performance? E.g., in language learning, children make classic errors in verb tense, they over generalize until they are able to be more specific, etc. Are you looking for errors?

A27: This won't be a prominent part of the test. But there are areas, for example, where you see a scene, and you don't remember the details but you remember the gist of the situation. This is an important other side of generalization. There are some cases where the errors are an important companion to the capability we want to capture. We won't intentionally seek out creating these subtle errors to see if your developed capability matches that. We do want you to make sure you match important capabilities like generalization, and so the test may try to capture that, but not intentionally seek out the errors.

Q26: Is it necessary to mimic human cognition?

A26: To some degree, yes. It is important that you have a thorough understanding of what is happening for human child cognition and you need to try to mimic important characteristics of that. You do not have to mimic the exact development sequence. You do not have to mimic all the parts of the brain. It is at some level of abstraction, it is part of the problem, trying to find out what is it that is important about human cognition, especially in TA1 that you want to mimic. You would want to mimic it in terms of how the test is created (e.g., look at the performance of children at 18 months, on all of the tasks, understanding objects, agents, and places. Then devise tests to see if your system performs the same way a child does. That is the level of mimic that DARPA desires.

Q25: Will images and audio be a part of the training/test data? Audio track on video?

A25: For TA1 and TA2, there will certainly be vision and you may have some audio later on. This will be required for some of the tasks, e.g., the milestone capability examples such as the one for when a child learns to name objects, or name the function of an object. This would be used when the simulated agent has to interact with other agents. Don't foresee a lot of natural language interaction in TA1, but certainly having some audio would be possible, although this would be in the later part of the program. For sounds in the environment, audio will be a part of it.

Q24: How much computer vision will be needed for TA1 and TA3? Is TA1 expected to use computer vision to perceive the simulations or will TA1 have programmatic access to the simulation?

A24: TA1 is expected to do perception and vision of the environment, build the representations, understand the world, etc. from vision. For TA3, there will be image-based questions. At least one of the AI2 tests is all image-based and we expect there to be image-based questions in some of the other tests as well. The exact percentage of that is unknown at this time. But as we add questions later in the program, it will be important to have image-based questions as well as natural language questions for TA3.

Q23: Are you open to developing a new 3D simulated world or extending current ones such as Virtual Home to be more realistic?

A23: DARPA is open to select, modify, or create from scratch the test environment. There are many simulations that already exist, and you may want to choose one to start with and then modify or extend it. DARPA is open to any option that is effective.

Q22: Is an approach that gradually builds up cognitive capabilities like a child does within scope?

A22: Yes, an approach that gradually builds up cognitive capabilities like a child does is within scope. You need to sequence the development. DARPA is not necessarily concerned that your approach completely mimic the exact developmental sequence for a child. There are a lot of interesting reasons why the brain is primed to learn language at some point; and a lot of interesting characteristics of cognitive development in young children. You don't need to duplicate that, but you are expected to have some trajectory (e.g., you develop some initial capabilities, and refine them and extend them). For the TA1/TA2 track, the goal is to be able to mimic the cognitive functions of an 18-month-old by the end of the program. This is how the tests will be designed and you should develop your own trajectory on how to get there.

Q21: Will the natural language elements of the program be in English? Or is there a requirement to operate across languages?

A21: All of the work in the program will only be performed in English, and this will be difficult enough. The common sense concept base/repository might end up to be useful for language translation but this is not part of the program and DARPA does not intend to make it part the program.

Q20: Do you anticipate supporting human subjects research under the program? Or should TA1 and TA2 leverage prior published results?

A20: TA1 and TA2 should rely heavily on prior published results. It is an option for TA1, or possibly TA2, to conduct additional experiments. If this is the case, then HSR is fine, but such research must comply with the federal regulations for human subjects protection. Further, research involving human subjects that is conducted or supported by the DoD must comply with 32 CFR 219, Protection of Human Subjects (and DoD Instruction 3216.02, Protection of Human Subjects and

Adherence to Ethical Standards in DoD-Supported Research. For all proposed research that will involve human subjects in the first year or phase of the project, the institution must provide evidence of or a plan for review by an Institutional Review Board (IRB) as part of their proposal. The time required to complete the IRB review/approval process varies depending on the complexity of the research and the level of risk involved with the study. The IRB approval process can last between one and three months, followed by a DoD review that could last between three and six months. Ample time should be allotted to complete the approval process. DoD/DARPA funding cannot be used toward human subjects research until ALL approvals are granted.

Q19: Do the TA1 performers need to agree to a specific set of experiments or will TA2 be supporting possibly three independent evaluations?

A19: TA2 will not need to support independent evaluations. DARPA wants the set of experiments to be as common as possible. TA2 will define a set of tests and provide one simulation environment. All of the TA1 teams will use this same environment and take the same tests. When developing your capability over time, you can determine which test you want to emphasize first, and you may do well on some, and not on others. DARPA desires a common simulation environment, a common set of tests (i.e., not have these tailored to each TA1 team).

Q18: Does TA1 design experiments for TA2 to conduct?

A18: No, TA1 does not design experiments for TA2 to conduct. TA1 has the option of having team members that design and conduct developmental psychology experiments to refine their theory. TA1 may have a computational model that makes predictions on how the model would respond in some situation, and this could be compared to how children respond to the situation so you can see how that matches human performance. TA1 has the option, but it is not required, to do developmental psychology experiments to refine their models/work. TA2 would also involve developmental psychologists to design the test problems for TA1 to take. TA2 does not need to conduct developmental psychology experiments, although TA2 may need to conduct some experiment to refine their test problems. TA1 or TA2 has the option to include developmental psychology experiments but DARPA does not expect interplay between these two.

Q17: Is video out of scope for TA3?

A17: There are no video tasks in the AI2 tests. At this time, AI2 only has image-based questions; there are no video-based questions. There will be image questions and you will have to be able to do vision on images in order to take in some of the questions and provide answers, but you don't have to perceive video for the TA3 AI2 tests. Video and simulations in the 3D environment is important for TA1 and TA2, but this is not required for TA3.

Q16: Is there interest in proposals that involve learning and experiments with real robots in the physical world?

A16: Under TA1, we are interested in proposals for machine common sense systems that can successfully perform on the TA2 evaluation tests. It is up to TA1 proposers, how they would like to develop and train their machine common sense agent. Training could be done with a physical robot or with a simulated robot in a virtual environment.

Q15: Is an embodied perception/cognition approach within scope?

A15: Yes, an embodied perception/cognition approach is within scope. For TA1, this may be an important part of the problem. You have to create a simulated agent that exists in this simulated world, and it's using perception of the visual scene, and the action it takes is the input/output.

There won't be a symbolic description of the problem to this agent – it will have to be embedded in perception.

Q14: Q14: Will the 3D simulation have an API allowing the execution of actions within the environment to learn by active exploration, e.g., pushing objects?

A14: Yes, the 3D simulation will have an API for allowing for the execution of actions within the environment. However, we do not know at this time what that API will be, or how rich the actions will be. There will be a list of actions you can take, such as move, walk, open, etc. At this time we do not know if the environment will have very rich, fine motor interactions (this may be possible later in the program). You will be expected to take actions, and you can learn by taking actions in the environment. It is unclear how rich the vocabulary of those actions will be. The input will be vision, e.g., a pixel image of the simulation environment where you can move your head/camera around to see different parts of the scene (i.e., you would see the raw pixel images of what is happening in the environment).

Q13: Is it a negative to have a small team with a small budget but a good idea?

A13: It is not a negative to have a small team with a small budget that has a good idea. DARPA's preference is to have teams take on the full problem in the two areas. DARPA does not preclude small teams, but recommends sending in an abstract to see how it fits within the program.

Q12: Can a TA3 team propose to use the TA2 environment for learning, and still use the AI2 benchmarks for evaluation?

A12: Yes, a TA3 team can propose to use the TA2 environment for learning, and still use the AI2 benchmarks for evaluations. However, the proposal would have to say that you propose to use that environment for your training and not pay for it, but we don't know exactly what that environment is at this time. The proposal would have to describe what you expect in that environment so DARPA can evaluate your plan.

Q11: Do we need to have clarity on the exact collaborators/partners we plan to work with for the abstract?

A11: DARPA prefers that abstracts have clarity on collaborators/partners proposers plan to work with. An important part of your proposed solution is the organizations and key personnel that will perform the work on the effort. It is not necessary to have the cost proposal and subcontract agreements, etc. completed by the abstract phase. It is also possible to change it, e.g., if you do not have this information available and send in an abstract, then receive feedback on the abstract, and decide to rearrange your team. It is possible to send in an abstract with only one team member and then add a team member later. However, DARPA prefers to have as complete a story as possible in the abstract in order for DARPA to provide you with the most accurate feedback so you don't waste time/money preparing a full proposal that is not of interest to the program.

Q10: Is TA1 more important than TA3, or are they equally important?

A10: There are very different qualities DARPA is looking for in these two TAs – one TA is not more important than the other TA – and there is not more funding allocated for one versus the other.

Q9: Do we have to propose a full solution, or can we propose to solve part of the problem?

A9: DARPA's preference is to have teams propose full solutions to both of the problems. So either propose to do all three systems (objects, agents, and places) and all of the test problems for child cognition; or propose to take all five of the tests from AI2 for broad common sense. DARPA may

have interest in a small, low cost, niche proposal on some small piece of the problem, but award of a large number of this type of proposal is unlikely. DARPA is really looking for (and it is an important characteristic of the problem) a general solution to address all of these things (e.g., there are a lot of interactions between objects, agents, and places in the childhood cognition piece; there is a lot of general knowledge behind answering the broad common sense questions). DARPA is not as interested in small niche proposals (e.g., a proposal for physical common sense for a robot, and nothing else); but it is possible, if you have a great idea, that there may be interest. Submit an abstract describing the partial solution for feedback, with the understanding that the preference is to have full proposals for the two problem areas.

Q8: Could a TA3 team use the TA2 training environment to train their agent to have physical common sense that they would use to answer TA3 questions?

A8: Yes, a TA3 team could potentially use the TA2 environment to train their agent to have physical common sense that they would use to answer TA3 questions. However, you would need to describe that well in the proposal. There is obviously a lot of synergy you could get from doing that. DARPA is receptive to this, but it is not an explicit milestone or deliverable in the program plan.

Q7: Would TAs share results in the later years of the program? Would you want to integrate the TAs in the later years of the program?

A7: There certainly is a lot of synergy that could happen; however the program has not been organized for that programmatically in any way, with the exception of joint PI meetings to share results.

Q6: Does each proposal have to be limited to a single TA? What if a proposal for TA1 contains ideas that could help a TA3 goal?

A6: Yes, each proposal must be limited to a single TA. Each proposal will be evaluated based on the TA under which it was submitted.

Q5: Can the same institution propose to TA1 and TA2?

A5: Yes, the same institution may propose to both TA1 and TA2. However, you must provide a mitigation plan, clearly describing how you will separate the two groups (e.g., how you will ensure that the testing team will not share/leak any information to the team under test) and avoid organizational conflict of interest.

Q4: Can you propose to both TA1 and TA3?

A4: Yes, you can propose to both TA1 and TA3. However, you will need to submit separate abstracts and separate proposals for each TA. It is acceptable to have shared common text in the abstracts/proposals, but they must be separate, one for each TA.

Q3: What is the total dollar amount available for the four year program?

A3: The total dollar amount available for the entire four year program is \$70M.

Q2: Will DARPA be executing the contract itself or will an agent be used? If an agent will be used, what agent?

A2: DARPA plans to use SPWAWAR as the agent for the program.

Q1: Will the DARPA PM be restricted from answering clarification/scope questions about the BAA technical areas?

A1: The DARPA PM can respond to questions in accordance with the DARPA Proposer Communication Plan found at

<http://www.darpa.mil/attachments/DARPAVendorCommunicationPlanMarch2014a.pdf>. If you have any questions, send an email to mcs@darpa.mil. If the question is not specific to your approach, it will be added to the public FAQ. Questions that are specific to a solution may be reworded as a generic question and posted on the FAQ. The MCS BAA allows for the submission of abstracts, which guarantees direct feedback from the PM.