

Deep Learning BAA FAQ

Q69. We aim to differentiate our proposed work from that of other groups working in the field. We are considering adding a member to our team from an area that is not mentioned in the BAA, but that I believe could contribute significantly to the Deep Learning paradigm. Will this disqualify our proposal?

A69. If you would like to expand the scope of your response to the Deep Learning BAA beyond what is asked for, you should make a case for it in the proposal. It is up to you to assemble the team you believe will do an excellent job.

Q68. In both the BAA and at the industry day, it was stated that the output of a phase 1 deep learner was to be used by a standard linear classifier (e.g., an SVM) for the purpose of evaluating the deep learner. Linear classifiers are based on feature space representations. So, I only see two possibilities here for deep learners: 1) they are feature set selectors/optimizers serving as the preprocessor for a linear classifier, or 2) more sophisticated deep learners (e.g., semantic networks) that must export some sort of feature set that only partially represents what has been learned.

A possible resolution of this dilemma appears in Figure 3 on page 13 of the RFP, which includes the phrase "or cluster evaluator". But this alternative appears nowhere else in the document or in the industry day presentations. Can you please clarify this?

A68. The idea of evaluation in phase 1 is to use a weak classifier like a simple linear classifier to compare to a more sophisticated machine learning method like a Support Vector Machine with a task-specific kernel. The SVM performs linear classification on a feature space created by the kernel, so we are asking whether a deep learner can do better than a task-specific kernel since both give their results to a linear classifier.

If you have another idea to evaluate the learnability of representations created on unlabeled data by a deep learner (like a cluster evaluator), you are welcome to make a case for it in your proposal and compare it to the simple linear classifier. If your proposal is selected for funding, you can argue for an addition to or a change in evaluation criteria in phase 1.

Q67. Can I arrange to have you see a prototype of our system?

A67. No. We are interested in seeing proposals that address the Deep Learning BAA, not in technology demonstrations.

Q66. Would proposing a shorter Phase 1 disqualify our team?

A66. Not unless you are proposing to be the evaluation team. If you are proposing to be a development team, your shorter schedule would not control the other performers' schedules.

Q65. Should we buy or lease equipment? Is it possible to use equipment that DARPA already has?

A65. The equipment you plan to use and its costs should be included in your proposal. DARPA will not directly provide any equipment to performers for this program.

Q64. Shouldn't throughput statistics be included in evaluations? What is the expected response time for a successful machine learner?

A64. There is no guidance in the BAA on response time or throughput. You are welcome to make a case to include these as additional criteria in your proposal.

Q63. To what extent is teaming with industry desirable? For the development effort, do you prefer relatively small-scale single-institution academic teams, or involvement or leadership of larger industrial partners for system integration and eventual transitioning of the developed software to defense applications?

A63. You should plan for a level of effort, team composition, and team size that allows you to address the issues in the BAA as best as you can. There is no reason to prefer industry or academia in the composition of a team.

Q62. Will run-time complexity, algorithm performance, or parallelization be considered either as a proposal evaluation criterion or as a program evaluation metric?

A62. Run-time issues are not evaluated as a metric but will be important for scalability and so may be considered in the evaluation of proposals. This applies to dataset processing enabled by fast hardware as well. However, Deep Learning is about algorithms and architecture, not hardware.

Q61. Does all software have to run on the same hardware platform (e.g., Linux Beowulf cluster with a gcc compiler) or would there be interest in exploring GPU or special hardware platforms for higher implementation efficiency?

A61. Software does not have to run on the same platform. We are interested in software that is suitable for parallel processing, less interested in software that can run only on a specialized platform. Deep Learning is about algorithms and architectures, not hardware.

Q60. Would there be interest in reinforcement learning mechanisms embedded with the hierarchical deep learning to supplement the sparsely labeled data with additional feedback obtained in the form of discrete rewards?

A60. If you think additional criteria (such as compatibility with reinforcement learning) or additional test methods should be considered in evaluation, you should make a case for them in your proposal.

Q59. Can you give us an idea of the size of the teams that you are looking for?

What is your expectation regarding the number of awards, and the size of awards?

Do you envision that all awards will be similar in size?

What, if any, budget has been set for this program?

A58. The answer to these and similar questions is that funding will depend on the specifics and merits of the proposals. Propose a plan and budget that you believe can meet the requirements of the BAA technical area you are proposing to, and justify the size of your team. Selections and funding will depend on the proposals.

Q58. Are “rough order of magnitude” cost estimates acceptable for Phase 2/Phase 3 options?

A58. The proposal needs to be fully priced so that we can exercise the option in a timely manner.

Q57. Are there any maximum funding limits, given that this is a 6.1 program?

A57. No.

Q56. Must the technical section’s top-level structure be by phase, or can the phases be described at a sub-section level?

A56. There is no guidance in the BAA, but the Scope section, page 9, gives top-level descriptions. The basic requirement is: make clear what you are doing in each phase.

Q55. Can we include a glossary as a separate section of our proposal?

A55. If you want the source selection committee to read it, please respect the page count.

Q54. Is it expected that this 6.1 program will be 0% fee?

A54. Nothing is specified as far as the fee range.

Q53. Can a researcher participating at no cost be a co-PI?

A53. Put the description of the level of effort and percentage of time in the proposal and cost it, even if it is zero.

Q52. Can performers collaborate?

A52. There is a competitive process in each phase that may result in a downselect, so there would seem to be limited motivation for development teams to collaborate. We certainly expect that the evaluation team will not collaborate with development teams in a way that gives advantage.

Q51. Can a team be an evaluator in, say Phase 1, but propose to be a developer in Phase 2?

A51. No, teams will be selected to be either the evaluator or a developer, and proposals can only bid to one area or the other.

Q50. Must the one-page summary graphic be in quad or pentachart format?

A50. Please refer to the language of the BAA. The content specified there is more important than the format, other than the graphic being limited to one page.

Q49. Do references count in the page count for the proposal?

A49. They do not count in the page count, but please do not consider them to be “required reading” for the source selection committee.

Q48. Where are the interim and final evaluations to be held?

A48. We don’t know at the moment. A reasonable assumption is that either all evaluations will be conducted at the evaluation team’s site, or your evaluations will be conducted at your site.

Q47. In table 2, page 48 of the BAA, there are example tasks for each data modality. Could you give a rank order/priority for DARPA of these tasks for each data modality?

A47. The program does not have such a ranking.

Q46. What if a learned representation is not in a form that can be directly applied to linear classifiers?

A46. If you don’t want a weak linear classifier, then propose an alternative and make a case for it.

Q45. For Phase 1 evaluation, is an interface to the linear classifier required?

A45. The evaluation team is likely to require this.

Q44. For evaluations, can developers use data in addition to government provided training data for unsupervised training?

A44. For purposes of evaluation, you can only use the provided datasets.

For development of algorithms and architectures, you can use whatever you wish.

Q43. What are the licensing rights expectations for offerors that have existing intellectual property?

A43. Please see the BAA for information concerning proprietary information, software, and data.

Q42. Will you be trying to transition this program to one of the armed services?

A42. Not within this program. This is a basic research program.

Q41. Can separate sectors of the same company be developer and evaluator if there is an organizational conflict of interest mitigation plan in effect?

A41. No.

Q40. The industry day presentation mentioned the possibility of non-continuation of phases. Are there any "hard stops" to prevent continuation?

A40. Yes, for example if no development team passes the go/no-go's the program would end. There could be other reasons that we cannot foresee.

Q39. Will developers get a test harness or "evaluation shallow learner" with which to do internal evaluations and benchmarking?

A39. Not necessarily. This will depend on the selected evaluation proposal.

Q38. Can you submit a "one-sided" proposal, e.g. video-only or text-only, with the idea that DARPA will match you with someone else.

A38. No, that is why we encourage and provide opportunities for teaming, but it is up to you to find a team. See the program's teaming web page at

<http://csc-ballston.com/baa/DLteaming.htm>

Q37. Will developers receive some labeled data with which to perform self-evaluations and development?

A37. In Phase 1 you will not receive any labels; in latter phases you will.

Q36. Do “text” and “language” mean the same thing in this BAA?

A36. Language could include text or speech.

Q35. Could you give several examples of tasks for text/language?

A35. Document recognition/classification, text mining, and summarization. There are examples at the end of the BAA with datasets.

Q34. Universities will typically submit multiple proposals from different teams. Can a team be the evaluation team while other teams within the same university are development teams?

A34. You can propose to both areas, but you will not be selected for both areas, per the BAA.

Q33. Will proposals that include commercial technologies that have paid commercial licensing, but free research licensing be graded negatively or considered non-compliant?

A33. Proposals will be judged on their merits, relative to the parameters of the BAA. See the BAA for information on proprietary information, software, and data.

Q32. Will the evaluation team choose the preprocessing procedure? Will they actually perform the preprocessing? Will this be the same for all three phases?

A32. There will probably be some differences between phases. We certainly expect that the evaluation team will work with the development team to ensure that evaluations go smoothly and are not affected by extraneous factors. Development teams will choose their own preprocessing, within the boundaries of the BAA.

Q31. Could dual developer/evaluation bids be submitted and the proposer choose which area to perform under if both bids are chosen.

A31. If both areas are bid and determined to be selectable, DARPA will select what it considers to be the stronger bid.

Q30. While no organization will be selected for both evaluation and development work, can a *person* share time between the evaluation team and a development team?

A30. No, there would be a clear conflict of interest.

Q29. Are there any constraints with regards to working with a startup company?

A29. The startup must conform to all government-mandated rules, e.g. have an approved accounting system and cost accounting.

Q28. What is the anticipated funding profile for each phase?

A28. We do not know at this time. This will depend on the specifics and the quality of the proposals.

Q27. Does “identical core learner” mean some principle with adjustments for the data format, or does it mean something stricter?

A27. Basically, it means minimizing the amount of adjustment based on the dataset, if you believe that adjustment is required.

Q26. What are the expected sizes of the unlabeled datasets? How will the size of the test datasets be determined?

A26. This is up to the selected evaluation team proposal. We do not see an upper limit to the size of the datasets. One of the objectives of the program is to use large amounts of unlabeled data and small amounts of labeled data.

Q25. Will this become a classified program?

A25. We do not anticipate that this program will become classified.

Q24. Where can I find the Deep Learning program manager’s briefing from the Deep Learning Industry Day?

A24. The presentation is available on the Deep Learning website.
(<http://www.schafertmd.com/conference/deeplearning2009/downloads/Deep%20Learning%20Industry%20Day%20Brief.pdf>)

Q23. Will the evaluation team be given linear learning comparison tools for Phase 1?

A23. No, the evaluation team will be expected to provide a weak linear classifier in Phase 1.

Q22. Will evaluators be given audio, video and text to create datasets from?

A22. No. This is a key responsibility of the evaluation team.

Q21. What can you tell us about the input data, for example, will the data be segmented into things that are similar, e.g. 30 videos of horses, or 100 videos of 30 horses and 70 other animals?

A21. This depends on the selected evaluation team. We are very interested in the transfer and sharing of representations.

Q20. In Volume 2, Cost Proposal, Part E of the BAA states: DARPA currently anticipates using 6.1 funding for this program. If the offeror can represent that its proposed research is beyond 6.1, will DARPA consider using other than 6.1 funding for the program?

A20. Typically program funding is not re-programmed, and this program is projected to be a 6.1 program.

Q19. Are there any requirements for open source?

A19. No. Please refer to the BAA for conditions related to proprietary information, software, and data. Obviously, the outputs of the program must be usable by the government, and the government would like to control its technology costs.

To quote the BAA:

“To the greatest extent feasible, offerors should not include background proprietary software and data as the basis of their proposed approach; offerors wishing to use proprietary software and data as the basis of their proposed approach should also recommend an amelioration strategy (replacing proprietary components with open source alternatives, research versions, etc.) to address, among other things, possible public releases. Offerors expecting to utilize, but not to deliver, open source tools or other materials in implementing their approach must ensure that the government does not incur any legal obligation due to such utilization.”

Q18. Will proposals based on approaches other than Deep Belief Nets be considered for funding under the Deep Learning program?

A18. Proposers are encouraged to propose any approach that generates hierarchical representations with at least three layers of compositionality, per the BAA. Deep hierarchies of HMMs, SVMs, Bayes Nets, Perceptrons, and novel approaches are encouraged if the proposer can plausibly demonstrate they can meet the requirements of the BAA.

The BAA specifically excludes "proposals based on multiple shallow learners that do not reflect the central architectural idea of a hierarchy of representations using a common set

of algorithms.” It does not exclude any type of learner that *does* reflect this central architectural idea.

Q17. The BAA states that the Phase 1 goal is to develop a learner that can ingest unlabeled data and output representations that, when run through a simple linear classifier, “will produce output classification results comparable in quality to a shallow state-of-the-art classifier system, such as an SVM using a task-specific kernel.” Isn’t this a high hurdle for Phase 1?

A17. Yes, it is. Learning from unlabeled data is meant to be one of the hard problems in this program. We want to establish in Phase 1 how good the performers’ architectures are at learning from unlabeled data.

Please note that the BAA says “comparable in quality.” We plan to control for the advantage that the shallow classifier will have in using labeled data when doing our evaluation. We encourage proposals to include additional experiments involving small amounts of labeled data in Phase 1, as well as additional metrics for evaluating performance.

Note that the BAA states that the end-goal of the program is a machine that is “best at learning given a few labeled examples and a huge number of unlabeled examples.” The Phase 1 unlabeled data challenge is a leap towards that goal.

An equally important criterion for evaluating teams will be their results against other teams in the program, as stated in the BAA.

Q16. What kind of background/world knowledge can the teams be assumed to have access to?

A16. No special, proprietary, or classified knowledge is needed. Expertise in machine learning and knowledge of the state-of-the-art as derived from public sources is assumed.

Q15. Are the deep learners expected to learn all the concepts in a bottom up manner, or would it be acceptable to automatically prime the learner with world and domain specific concepts available from knowledge resources like WordNet, CYC, or the KNEXT systems?

A15. No extra knowledge sources outside of those provided by the evaluation team is needed. DARPA welcomes demonstrations of how external knowledge can be used in a Deep Learning system. If additional criteria or metrics associated with such knowledge are compelling, a case can be made in the proposal to include additional evaluation methods and criteria.

Q14. What types of preprocessing will be allowed? What, if any, limitations will be put on data preprocessing? Is it acceptable if a team deploys preprocessors capable of producing more complex (higher level) features, such as boundaries in images or concepts and relationships between them produced from text?

- A14. Pre-processing should not include extra knowledge beyond that in the datasets. In phase 1, only unlabeled data is used and the representation codes are sent for evaluation. In addition, the same core deep learner (at least 3 levels) must be used for all datasets in phase 2 and 3 so all pre-processing for all datasets must be outside this core learner.
- Q13. Will the Phase I evaluation system include using each team's preprocessing techniques in conjunction with the representations its deep learner produces, or will it be required to operate directly on low-level data (such as words and pixels)?
- A13. The phase 1 evaluation system will operate only on the codes generated by the deep learner and not specialized pre-processors. You are still welcome to propose additional methods and criteria to be used for evaluation.
- Q12. In the evaluation of the proposals, how important will scalability to very large real-world data collections, such as real-world images and free text, be?
- A12. The applications envisioned require operation on large real-world data so scalability is important.
- Q11. What are the envisioned evaluation steps likely to be associated with a task that involves text data? Would these be low level recognition of parts of speech or named entity recognition, or would they involve higher level understanding, like concept extraction and relations between concepts?
- A11. To some degree this depends on the selected evaluation team's datasets. Some example datasets can be found at the end of the BAA. Envisioned horizon applications include situation understanding using anomaly detection.
- Q10. What complexity/difficulty should the deep learners be capable of solving in spatio-temporal tasks? Will these tasks be limited to standard clustering/classification problems with spatial or temporal aspects (video, maps at different scales), or will they involve the prediction of spatiotemporal events or system states ("it is possible that the truck will be arriving at its destination in 1 hour")?
- A10. To some degree this depends on the selected evaluation team's datasets. Some example datasets can be found at the end of the BAA. The envisioned horizon applications include identification of salient events in surveillance videos.
- Q9. Is fee/profit considered a direct or indirect under 6.1 funding when a candidate is a performer?
- A9. It depends upon contract vehicle type however, based upon the potential for a CPFF contract; fee/profit is not included.

Q8. We are proposing over \$650K worth of effort possibly as a procurement type contract which usually mandates it be TINA compliant but it sounds from the BAA that there will be competition due to the BAA wording indicating DARPA will downselect for phases II and III. Even if over the \$650K threshold, as long as there's competition, we don't usually have to comply with TINA. For this BAA do we need to comply with TINA since it's indicated there will be competition? Pg 34/39 (printed page numbers at bottom right) under the heading of "Volume 2 - Cost Proposal", it is a bit confusing as this seems to indicate that despite competition, a procurement type contract over \$650K will still need to be TINA compliant.

A8. A Certificate of Current Cost or Pricing Data will be required.

Q7. May I have a private meeting with DARPA's Deep Learning program manager to discuss my BAA proposal, or my team's ideas regarding the BAA?

A7. No, however, you are welcome to submit questions regarding the BAA by email to the Deep Learning BAA's mailbox (DARPA-BAA-09-40@darpa.mil). Answers will appear in this FAQ.

Q6. Could you please look at my paper and tell me if you are interested in my work as part of the Deep Learning program?

A6. If, after reading the Deep Learning BAA, you believe that your work is pertinent to the Deep Learning program you should submit a proposal that responds to the specifics of the BAA, or join a team that is submitting a proposal. A paper is not a proposal, and we cannot read and comment on papers. A proposal must follow the format specified in the BAA and address the BAA's goal of creating a universal, deep-learning machine that can learn from unlabeled data in multiple modalities and perform multiple tasks.

Q5. How can I go about adding my expertise to a team?

A5. One way of finding team partners is to use the Deep Learning teaming page at <http://csc-ballston.com/baa/DLteaming.htm>.

Q4. What start date should be assumed for the development of the financials in the proposal?

A4. For the purposes of the proposal, you may assume a start date of October 2009, realizing that this date is notional.

Q3. What, if any, guidance can be provided as to the scope of the proposals in FTE equivalents?

A3. Everything you need to propose is contained in the Deep Learning BAA. The work and staffing requirements should be based on your plan for meeting the goals of the Deep Learning program.

Q2. Are evaluators also subject to the 6.1 funding restrictions?

A2. Yes.

Q1. The URL for the teaming page in the BAA (<http://csc-ballston.dmeid.org/baa/DLteaming.htm>) does not work.

A1. The correct URL is <http://csc-ballston.com/baa/DLteaming.htm>