

QUESTIONS & ANSWERS - (BETTER) IARPA-BAA-18-05

Responses # 1 to 116 (Rounds 1 & 2)

#	Response
1	<p>Q: The parsing system UDPipe (http://lindat.mff.cuni.cz/services/udpipe/run.php) currently has parsers for 50 languages: can we assume the target languages will be among these? (Examples of languages which are *not* on UDPipe are Bambara (Mali), Wolof (Senegal), and Pashto and Dari (Afghanistan))</p> <p>A: The BETTER program focuses on high-resource languages. Performers may assume that the languages used in BETTER will be among the top 10 world languages</p>
2	<p>Q: For the "constrained" variant, are the various resources available from the Linguistic Data Consortium (including corpora) considered "other linguistic resources"? What about the various versions of Wikipedia? Or am I misunderstanding the sentence (first paragraph on page 9) "The constrained condition does not cover other linguistic resources such as dictionaries, lexicons, or ontological resources." : does the phrase "does not cover" mean that such resources can be *included* in the constrained condition or that they must be *excluded*?</p> <p>A: The constrained vs. unconstrained condition focuses solely on the use of parallel corpora for the purposes of training data-driven methods. Other linguistic resources such as lexicons or NLP tools, whether they include cross-lingual information or not, are not covered by the constrained condition.</p>
3	<p>Q: Since ICEWS is being used as the baseline, will performers have access (subject, of course, to appropriate agreements protecting intellectual property) to at least some subset of the government version of ICEWS (which includes the coded source texts, while the public version of ICEWS on Dataverse does not)?</p> <p>A: As stated in the BAA Section 1.B.2, ICEWS and the BBN ACCENT software are only provided as baselines for planning purposes. Program-specific baselines will be developed prior to program kickoff, with all relevant information provided to performers.</p>
4	<p>Q: In Section 5.A.2.a. it states that multiple awards are anticipated. Can the government provide information on how many awards are anticipated?</p> <p>A: No. That is unknown at this time.</p>
5	<p>Q: In Section 5.A.2. the government references "program budget constraints." Can the government provide clarification to industry what its program budget constraints are?</p> <p>A: The funding range is dependent on the number of performers selected, program balance, and the availability of funds; however in recent years, IARPA performer teams doing comparable research have received from roughly \$1 million to \$5 million per year. IARPA is not committed to being limited to this range.</p>
6	<p>Q: In Section 1.A.2 the BAA states that there are three major research thrust areas: information extraction (IE), information retrieval (IR), and human computation. Will the government consider making awards to proposals that focus on only one of these research thrust areas? Or, will awards only be considered for proposals that address all three research thrusts?</p> <p>A: As stated in the BAA Section 1.A.2, performer teams must address all three research areas.</p>

7a		<p>In Section 1.B.1 the BAA states that "Performers will be presented with annotated training data in English. Models will be assessed against a target-language test set for which no training annotations are provided." Will there be any training data provided on the target-language annotated in either English/other language?</p> <p>Q:</p> <p>A: Performers will receive training data annotations in English only for the information extraction tasks.</p>
7b		<p>Q: Can the performer use machine translation to convert target-language documents into English?</p> <p>A: IARPA welcomes novel contributions that will meet the requirements outlined in the BAA.</p>
8		<p>In Section 1.B.1 the BAA states: "In addition to the seed documents, performers will be provided with a corpus of documents annotated according to the program definition of relevance." Will there be any subset of corpus in target language?</p> <p>Q:</p> <p>A: The program will not provide training data in the target language. One of the goals of the program is to encourage/require performers to develop cross-language approaches that do NOT depend on target language training data for the specific analytic tasks.</p>
9		<p>In Section 1.B.1 the BAA states: "Using this English-annotated corpus, performers would be expected to extract from a target-language corpus." Will any English-annotated, target-language corpus be provided?</p> <p>Q:</p> <p>A: No. Performers will receive annotations only on the English-language corpus.</p>
10		<p>In Section 1.B.1 the BAA states: "Performer teams will be expected to run human-in-the-loop interactions using their own systems and recruited analysts or annotators." Can performer teams use human-in-the-loop interactions to train on the target language documents?</p> <p>Q:</p> <p>A: Performers will be allowed to use any data available to them to perform the human-in-the-loop interactions.</p>
11		<p>In Section 1.B.1 the BAA states: "Small, unannotated test-language development set". Can the performer teams annotate/train on the test-language development set? Can human-in-the-loop training be performed by performer on this development set?</p> <p>Q:</p> <p>A: Performers will not be allowed to train their systems on the test-language development set. The only interactions with target-language data allowed within the scope of the program are on the documents returned to the performers as part of the human-in-the-loop interactions.</p>
12		<p>Q: How do the goals of this BAA differ from those achieved by the BBN ACCENT (Accurate Events from Natural Text technology) software?</p> <p>A: BBN Accent is used to aid in the establishment of program baselines for the purposes of planning purposes. Additional baseline models will be developed for use within the program. IARPA welcomes novel contributions that satisfy the requirements outlined in the BAA.</p>
13		<p>Q: In 1.A.2, "... using English-language training data." Does this mean that annotations are in English but data is parallel?</p> <p>A: No. There is no guarantee that the provided annotations will be on a parallel corpus.</p>
14		<p>Q: Could you tell us what information is going to be included in annotation? What are the definitions of sentence-, paragraph-, and document-level annotations?</p>

	A:	Event annotation will be provided at the document level, meaning that for each event in the document of a type of interest, there will be one template-style annotation that will not be anchored in the text.
		For all levels of annotation (IE task 1, IE task 2, and IE+IR task), the template will include the event type, the Agent(s), and the Patient(s). Each Agent and Patient will be represented by a list of strings that name that entity within the document. These will be either proper names or nominals. Some event types will take or permit multiple entities in the agent and/or patient slot.
		For some of the fine- (IE task 2) and finer-grained (IE+IR task) events there will be an additional slot available for Related Event(s). Related Events slots will be filled by pointers to other event templates from the same document.
		For the IE+IR task events, analogous templates will be provided, but there will be additional event types not present at the <u>fine-grained</u> level, some of which might be sub-types of <u>fine-grained</u> events, and many of which will be additional event types that further capture a set of related phenomena. There may be some additional fields added at this level, too.
		Event <i>mentions</i> will not have to be identified by systems. However, a list of “anchor” or “trigger” words/phrases will likely be included in the reference data template provided to performers. The inclusion of these is to aid in system development and training. These would not be required when scoring system output.
		Performer systems will be expected to return templates, with the following important differences:
		· Triggers will not be required
		Only a single string fill will be required for each Agent or Patient, with full credit being awarded for the most specific name for the entity available in the document, and partial credit for less specific names or nominals.
15	Q:	Is the human in-the-loop bilingual?
	A:	The human in the loop is allowed to be bilingual.
16	Q:	Will there be any limitations on the types of human interactions that are allowed, beyond the count of the number of documents?
	A:	IARPA welcomes novel contributions that meet the requirements outlined in the BAA.
17	Q:	The human-in-the-loop stage involves collecting human annotations on test documents to be included in the “enriched documents” that will be used for feedback. Do these annotations need to resemble the annotations provided on the original English-language training documents? Will the government specify a format for these annotations or will performers be allowed to define their own formats?
	A:	As noted in BAA Section 1.B.1, page 11, all human-in-the-loop interactions will be performed by the performer teams. Given this, there are no limits on format for performer-developed annotations. The description of the human-in-the-loop is left largely unconstrained in order to encourage a wide range of novel contributions.
18	Q:	The BAA states, 'Performer algorithms must... nominate the top n test set examples for use in their human-in-the-loop interactions.' Will performers be allowed to choose any n test set examples for feedback (i.e., the n documents used for feedback are not the n documents that the system considered most relevant)?
	A:	Yes. Any nomination technique is within scope subject to the constraints outlined in the BAA, e.g., number of documents.

19	Q: Will performers have access to the full held-out test set, in parallel with the T&E team, during the human-in-the-loop experiments?
	A: No. As stated in BAA Section 1.B.1, page 12, performers will only have access to the n documents chosen by the developed algorithms.
20	The BAA states that performers “will not be allowed to acquire native language speakers to support data annotation or acquisition.” Will performer teams be allowed to use native speakers for other purposes, such as system evaluation? Will the humans involved in the human-in-the-loop experiments be required to read/understand the target language?
	Performers are not allowed to recruit native-language support in order to facilitate large-scale data annotation within a target language. A: Performers are allowed to recruit native-language support for tasks such as the human-in-the-loop experiments.
21	The BAA refers to 'the ability to apply new training data in an online manner' and 'requires performer algorithms to be trainable using small batches of data in a streaming manner.' We request further clarification of these points. Table 2 'Program parameters' gives examples of system training, for example, ~1000 document annotations + 50 human-in-the-loop interactions. Will performers' systems receive a single, batch update of these 50 documents (and then be evaluated), or might they receive and be evaluated on multiple, smaller updates using subsets of these 50 documents (at the extreme, one could imagine doing updates one-at-a-time, so 50 different updates and evaluations of the system, with the 50 documents presented one-by-one in arbitrary order -- is this what is meant by 'streaming'?).
	A: Yes, the performers' system may be presented with different sized batches of training data during system test and evaluation. It could be that training time constraints will limit the number of different runs attempted.
22	Q: During the evaluation, will the T&E team perform multiple IE+IR tasks? If so, will the top-n documents be per task or across multiple tasks?
	A: The T&E team will perform a single IE+IR task.
23	The BAA says that the T&E team will provide “the raw, unannotated test set examples” to the performers for use in the human-in-the-loop experiments. Will performers' systems be allowed to apply any meta-data to these documents, such as to indicate the reason for the selection of each document or to specify the type of human feedback to request?
	A: Performers are allowed to apply arbitrary metadata to the documents in order to facilitate their own human-in-the-loop experiments. These metadata fields will not be used during the program T&E evaluation, however.
24	The BAA states that the “constrained condition does not cover other linguistic resources such as dictionaries, lexicons or ontological resources.” Would you please clarify this statement. Does that mean that dictionaries, etc. are prohibited under the constrained case?
	A: The constrained vs. unconstrained condition focuses solely on the use of parallel corpora for the purposes of training data-driven methods. Other linguistic resources such as lexicons or NLP tools, whether they include cross-lingual information or not, are not covered by the constrained condition.

25	<p>Will the use of publically available language translation tools (e.g., Google Translate) be permitted under either one or both of the data regimes?</p>
	<p>IARPA welcomes novel contributions that meet the requirements outlined in the BAA. Performers should confirm that any approaches satisfy the intellectual property requirements outlined in BAA Section 4.B.1.d.</p>
26	<p>Will performers need to deliver two separate systems (constrained and unconstrained) for evaluation? Will both be involved in the human-in-the-loop experiments? Will performers need separate sets of human evaluators for the two systems?</p>
	<p>Performers will need to deliver two separate systems, one trained under the constrained data regime and one trained under the unconstrained regime, for each evaluation. Only the unconstrained system will be used in the human-in-the-loop experiments.</p>
27	<p>Will the document set describing the program definition of relevance be in English or the target language? Will the IR task be conducted against an English-language or target-language test set?</p>
	<p>The documents used to define relevance "by example" will be English language documents. The IR evaluation task will be conducted against a corpus of target-language documents.</p>
28	<p>What constitutes a "query" for the IE+IR task? (e.g., a single document with a single document-level annotation, a set of documents with a common set of document-level annotations, a set of documents with multiple and varied document-level annotations, any of the above supplemented with a targeted query phrase or Boolean query structure)</p>
	<p>Performer systems will be expected to operate under a "query by example" setup wherein a potentially wide range of example documents will serve as the seed for the query.</p>
29	<p>In the IE+IR task, the BAA states that "document-level information will be used to extract such events from the target language corpus". Would you please clarify what is meant by "document level annotations", perhaps with an example? Will these document level annotations correspond to the ontology that is implied by the annotations provided in the English language training set?</p>
	<p>Please refer to Question 14 for an explanation of the annotation scheme for the program.</p>
30	<p>The list of potential data sources includes a "Small, unannotated test-language development set". While the list is described as "non-exhaustive", this entry seems to imply that performers might not receive ground truth in the target language for evaluating the IE tasks. Is that the case?</p>
	<p>Performers will receive training data annotations in English only for the information extraction tasks.</p>
31	<p>The BAA indicates that performer systems may have to identify grammatical mood and veridicality. If so, will the government be providing associated annotated data for training? Using what format?</p>
	<p>The identification of grammatical mood will be required in order to accurately perform the information extraction tasks. Explicit identification of mood is not a requirement of the program, however. In other words, aspects such as mood and veridicality will be implicit in the provided annotations.</p>

32	<p>It will be challenging to adequately cover the extensive set of items to be included in “Section 2: Summary of the Proposal” in 5 pages. Would the government consider expanding the number of pages allocated to the summary or moving some of the details to other sections? (Note: Q: several of the elements included in Section F of the Summary are not requested in the body of the proposal.)</p> <p>A: Section 2: Summary of the Proposal will be changed to 12 pages.</p>
33	<p>In the constrained scenario, are performers only allowed to use the parallel corpora? If so, does that mean that models, such as word vectors or language models, pre-trained on material other than the provided corpora would not be allowed? Q:</p> <p>The constrained vs. unconstrained condition focuses solely on the use of parallel corpora for the purposes of training data-driven methods. Other linguistic resources such as lexicons or NLP tools, whether they include cross-lingual information or not, are not covered by the constrained condition. If models are trained on monolingual corpora, e.g., monolingual English word embeddings, they are not covered under the constrained condition. If models require parallel corpora, e.g., bilingual word embeddings trained on an aligned corpus, they <u>are</u> covered under the constrained condition. A:</p>
34	<p>Q: Will news articles in the corpora also have metadata associated with them , such as the author and name of the publication?</p> <p>A: There is no guarantee of any additional metadata in the training data corpus.</p>
35	<p>In the human-in-the-loop (HIL) portion of the evaluation, what language will supplemental documents returned to us use, English or foreign? BETTER wants all annotation to be in English, which would suggest English, but retrieval and extraction will be performed against foreign, Q: which would suggest foreign.</p> <p>A: Systems may choose either English (training) or foreign language (test) documents for HITL annotation subject to the constraints outlined in the BAA such as number of document annotations allowed.</p>
36	<p>Is the HIL aspect of the eval intended to explore personalization? If so, will a query (or a series of queries) by a hypothetical end user be available to establish context? Q:</p> <p>The human-in-the-loop experiments are designed to demonstrate the manner in which IE or IR algorithms can be refined in order to provide more performant solutions to each task. A:</p>
37	<p>In processing the HIL documents, are we constrained by the same annotation strictures (e.g., no span-level annotations) as elsewhere? Or are we free to explore user interactions not sanctioned by the non-HIL aspects of the evaluation? Q:</p> <p>As noted in BAA Section 1.B.1, page 11, all human-in-the-loop interactions will be performed by the performer teams. Given this, there are no limits on format for performer-developed annotations. The description of the human-in-the-loop is left largely unconstrained in order to encourage a wide range of novel contributions. A:</p>
38	<p>On a more general note, can you give us any context about the kinds of queries that users will issue? Is there a specific query language you have in mind? Will they use BOW lexical? Boolean key phrase? Extraction factoid template? Some combination of the above? Q:</p>

	<p>Strictly speaking there are no “users” performing queries. During IR evaluation, systems will be presented with a set of documents, each of which can be examined by the system and for which a relevance judgment is required to be emitted. During HITL, recruited analysts or annotators will receive a small set of documents that were selected by the performer system only, purely automatically, based on all the documents presented to the performer system.</p> <p>A:</p>
39	<p>The BAA states “The Excel document, in the format provided in Appendix B, shall include intact formulas and shall not be hard numbered.”</p> <p>Q: Will IARPA provide the required Excel sheet when a Cost Volume is requested, or is the Offeror to create the form in Excel? (4.B.2.b)</p> <p>A: The Offeror is to create the form in Excel ensuring that it provides the information in 4.B.2.b.</p>
40	<p>The BAA states “A. Completed Cost/Price Template - Offerors shall submit a cost element breakdown for the base period, each option period and the total program summary in the format provided in Appendix B3.” What is the “3” referencing? (4.B.2.b.A)</p> <p>Q:</p> <p>A: The footnote number was a typo and has been removed and the BAA has been updated accordingly.</p>
41	<p>Regarding the “annotated training data and parallel corpora” – can we assume the parallel corpora will be translations (or vice versa) of the documents that are in the annotated training set? (As opposed to documents only loosely similar in content to the training documents, or a corpora of English and identical/translated foreign documents but be a separate set of documents from the annotated training set.) (1.A.3)</p> <p>Q:</p> <p>A: The English-language training data will be not be applicable to a target-language corpus.</p>
42	<p>Will we be able to utilize publically available foreign language parsers/syntactic analyzers for the identified foreign language in the “constrained” (1.A.3) scenario (or will you be providing this type of tool for the constrained scenario?) Or will this only be allowed for the unconstrained scenario? (1.A.3)</p> <p>Q:</p> <p>A: The constrained vs. unconstrained condition focuses solely on the use of parallel corpora for the purposes of training data-driven methods. Other linguistic resources such as lexicons or NLP tools, whether they include cross-lingual information or not, are not covered by the constrained condition.</p>
43	<p>For the span+role identification of the IE tasks,</p> <p>a. Is the role just the string as it appears in the sentence (as opposed to mapping to some normalized format or ontology)? For example, will the annotations in the training set just refer to the text span, or use annotations such as “MATERIAL CONFLICT (source: MILITARY FORCE, target: CIVILIANS)” for the text “the soldiers attacked the villagers”</p> <p>b. In the case of “the military forces attacked the citizens”, you mention “military forces” is the correct span, and just “forces” is incorrect. Would “the military forces” also be considered correct?</p> <p>Q: (1.B.1)</p>

		<p>a) The role fillers will be strings as they appear in the document; however, performer systems are expected to resolve pronouns and nominals to a name mention occurring in the document wherever possible. There may be multiple sentences mentioning the event in the document, and the string fill for the event template may be taken from any of these mentions, OR from elsewhere in the document where the relevant entity is mentioned. Pronouns will not be accepted. A nominal mention will only be accepted if there is no name mention of that entity anywhere in the document. . (Systems should not attempt to perform normalization of entity mentions to CAMEO categories, as they used in the question.)</p>
	A:	<p>b) In this case "the military forces" would also be considered correct. In general, we will be permitting a modicum of language-specific morphological variation (such as the presence or absence of articles/determiners, small variations in word-forms such as singular-vs-plural-vs-possessive, etc.), ignoring variations in whitespace and perhaps other punctuation, and modest differences using a language-independent string edit distance measure.</p>
44	Q:	<p>Will we be required to generate document level annotations in the third task? E.g. will the IE F1 score for task 3 include document level annotations or just sentence/paragraph level annotations? (1.B.1)</p>
	A:	<p>IE scores at all 3 levels will be based on annotations generated at the document level. We are scoring a single template for each event occurring in the document, even if the event is mentioned multiple times. Performer system submitted event templates will not need to be anchored in the text at either the sentence or paragraph level. Submission of multiple event templates for separate mentions of the same event within the same document will count as a precision error.</p>
45	Q:	<p>In the third task, can we assume that the document level annotations use the same structure as the paragraph/sentence level annotations? a. E.g. the same event categories (even though I imagine they may be more specific than any annotation in that document, e.g. each sentence reflects at most a "material conflict" on its own but the document taken as a whole is a "military attack") b. We assume there will be roles specified in the document level annotations? (and therefore be required to determine that the role fillers in the sentence/paragraph extractions semantically match the document level annotations of the seed document for the IR task?)</p>
	A:	<p>Yes.</p>
46	Q:	<p>Are the documents used in the "human in the loop" process all English documents? (1.B.1)</p>
	A:	<p>Systems may choose either English (training) or foreign language (test) documents for HITL annotation subject to the constraints outlined in the BAA such as number of document annotations allowed.</p>
47	Q:	<p>Are the "annotations and enrichments" restricted to the annotation structure used in the Gov't provided annotation set? Do these annotations include document level annotations as well as individual spans? Are there limits to the extent of the annotations provided by the human team during this process? (1.B.1)</p>
	A:	<p>As noted in BAA Section 1.B.1, page 11, all human-in-the-loop interactions will be performed by the performer teams. Given this, there are no limits on format for performer-developed annotations. The description of the human-in-the-loop is left largely unconstrained in order to encourage a wide range of novel contributions.</p>

48	<p>Can you give any more detail on what you mean when you refer to document-level annotation? Is it simply a label attached to a document or are you envisioning something with more structure? What is your reason for transitioning from sentence/paragraph-level annotation to document-level annotation? Understanding the intent behind the transition to document-level annotation will allow us to better address the government's goals in our technical approach.</p> <p>Q:</p>
	<p>A: Please refer to Question 14 for an explanation of the annotation scheme for the program.</p>
49	<p>Can the "recruited analysts or annotators" be system developers? If yes, this opens up more options for the type of human interaction we can explore.</p> <p>Q:</p>
A:	<p>No. Teams will need to recruit analysts/annotators who are not part of the development team. Teams will have to submit in advance the types of annotation they plan on doing during the HITL portion of the evaluation, and this is subject to approval by the government.</p>
50	<p>The Data and Evaluation section mentions that performers will submit to T&E source code in addition to executables. Will T&E be expecting to build and/or evaluate the source code itself?</p> <p>Q:</p>
A:	<p>As stated in BAA Section 1.C.1, performers are expected to submit buildable Docker containers, which implies that all software dependencies are expected to be included within the container.</p>
51	<p>In the BAA, you write "The constrained condition does not cover other linguistic resources such as dictionaries, lexicons, or ontological resources". Does this mean 1.) dictionaries, lexicons or ontological Resources ARE allowed in the constrained regime, or 2.) they are not allowed.</p> <p>Q:</p>
A:	<p>The constrained vs. unconstrained condition focuses solely on the use of parallel corpora for the purposes of training data-driven methods. Other linguistic resources such as lexicons or NLP tools, whether they include cross-lingual information or not, are not covered by the constrained condition.</p>
52	<p>In Information Extraction, are the "slots" for an event (e.g. date/time) known in advance and fixed for a particular evaluation, or are they received in the form of a query from T&E?</p> <p>Q:</p>
A:	<p>The relevant slots for the information extraction tasks will be fixed at the beginning of each program phase.</p>
53	<p>Q: For Human Computation, is an IRB and a plan to access subject-matter experts mandatory?</p>
A:	<p>Performers should consult with the relevant IRB to determine whether local IRB approval and a human subjects research plan is needed.</p>
54	<p>Q: Is the limitation on Human Computation (Number of accessible documents) shared by the IE and IR processes, or is it independent?</p>
A:	<p>The program constraints are shared between the IE and IR tasks.</p>
55	<p>According to the BAA, analyst interaction will result in "annotated or enriched documents" which will be provided to T&E. What are the limitations on the annotation or enrichment of these documents?</p> <p>Q:</p>
A:	<p>As noted in BAA Section 1.B.1, page 11, all human-in-the-loop interactions will be performed by the performer teams. Given this, there are no limits on format for performer-developed annotations. The description of the human-in-the-loop is left largely unconstrained in order to encourage a wide range of novel contributions.</p>

56	Q: What form does the annotated training data for IR take?
	A: Performers will receive documents that provide annotation of relevancy "by example." Please see Questions 27 and 28 for additional clarification.
57	Q: Will the annotated documents provided in Task 3 also contain the sentence-level annotations from Tasks 1 and 2?
	A: Please refer to Question 14 for an explanation of the annotation scheme for the program.
58	Q: Are the documents in Task 3 different from those in Task 2?
	A: For planning purposes, offerors should assume that the documents will be different.
59	Q: Regarding human in the loop, can a human annotator read the document and tell if it is relevant?
	A: Yes.
60	Q: You stated that performers will not be allowed to acquire native language speakers to support data annotation. Can native language speakers be used for running human-in-the-loop interactions?
	A: Yes.
61	<p>"In the constrained condition, performers will be required to use parallel corpora, either annotated or unannotated, provided or identified by the Government team. The constrained condition does not cover other linguistic resources such as dictionaries, lexicons, or ontological resources. In the unconstrained scenario, performers are allowed to use linguistic resources harvested from other sources subject to some conditions."</p> <p>It is in particular the scope of the negation in the second sentence that is leading to debates. My understanding is that we are using linguistic resources that are partially embedded in our engineered solutions in both scenarios.</p> <p>The constrained condition does not restrict this, or does it? My understanding is that it only restricts our training of data driven methods to the corpora that are identified by the Government team, that is:</p> <ul style="list-style-type: none"> - we are allowed in the constrained condition to have dictionaries and other linguistic tools integrated in our technologies, various NLP tools, including language independent ones - the training of probabilistic and Deep Learning models is restricted to the corpora that you will provide <p>In the unconstrained condition we are allowed to use in addition other corpora (given some general constraints).</p> <p>Q: Is this the correct interpretation?</p> <p>A: Yes. The constrained vs. unconstrained condition focuses solely on the use of parallel corpora for the purposes of training data-driven methods. Other linguistic resources such as lexicons or NLP tools, whether they include cross-lingual information or not, are not covered by the constrained condition. In this context, "covered" means subject to the restrictions of the constrained condition, while "not covered" means not subject to the restrictions of the constrained condition. Again, the restrictions are focused solely on the use of parallel corpora to train new, corpus-based resources.</p>

62	Q:	How are these two data regimes applied, tested, evaluated, and when during the 3 phases of the project?
	A:	Performers will need to deliver two separate systems, one trained under the constrained data regime and one trained under the unconstrained regime, for each evaluation.
63	Q:	Will a *single* domain be used across the three phases of the program: that is, can the information derived from the earlier, larger sample phases be used in the later, small-sample phases? While the BAA explicitly excludes *language* expertise, is it permissible to use consultants with *domain* expertise?
	A:	There will be multiple domains across the life of the program. Performers should not assume that the domains are transferable. There are no explicit prohibitions against utilizing domain experts.
64	Q:	There seems to be a contradiction between line 13 on page 8 which refers to the relevance of "such as the previously mentioned political events" and line 11 on page 11 which states "Explicit event coreference, either within- or across-documents, is not an area covered under the BETTER program." Please clarify.
	A:	Performers will be required to extract events, such as political events, from text. Performers will not be required to perform event coreference resolution. Please see Q14 for more information about the data annotation scheme.
65	Q:	Can the humans-in-the-loop use resources such as Google Translate to see whether the information they are providing is improving what they are looking for?
	A:	IARPA welcomes novel contributions that meet the requirements outlined in the BAA. Performers should confirm that any approaches address the intellectual property requirements outlined in BAA Section 4.B.1.d.
66	Q:	Page 11, line 17 states The tasks described above are correct-or-not evaluations Is this constraint going to be applied to all of the systems, or will it be possible to use probabilistic classifications, particularly on the IR task, where this approach is fairly standard and plenty of evaluation metrics using probabilities exist? What about situations where the text itself is ambiguous, even for a IE task, for example in technical terms (e.g. does "LDA" refer to "linear discriminant analysis" or "latent Dirichlet allocation"?) and locations (e.g. geonames.org returns 13,750 records for "San Jose", and 638 in Colombia alone). Also in some cases, analysts/humans-in-the-loop find it useful to know whether cases are central or on the edge of the existing classifications: this is particularly helpful if (as in this case) only a limited set of new cases can be applied, since central cases provide relatively little new information.
	A:	The correct-or-not evaluations apply to information extraction tasks, to include both event classification tasks as well as argument ID and classification. The scoring system will take into account some linguistic differences such as morphology. Please see Q43 for more information on IE scoring.
67	Q:	How big will the test set be for IR? (Will algorithmic efficiency be a significant factor?)
	A:	The size of the test set will be determined by program kickoff. Algorithm efficiency will not be an explicit factor on which performers are scored.

68	<p>We assume the documents provided back to the team during human-in-the-loop feedback will be in the foreign language, but that the humans enriching them will be English speakers. Is that correct?</p> <p>Q: humans enriching them will be English speakers. Is that correct?</p> <p>A: The recruited human analysts are allowed to be bilingual.</p>
69	<p>Can we use general English linguistic resources, even in the constrained condition? For instance, dependency treebanks or the IE training data from the ACE program? (We are talking about general research community resources, not proprietary resources.) What about publicly-available external knowledge bases, e.g. DBPedia?</p> <p>Q: available external knowledge bases, e.g. DBPedia?</p> <p>A: Yes. Please see Q61 for more information.</p>
70	<p>Can we use general linguistic resources for the foreign language? For instance, if the Phase 1 language is French, can we use a French universal dependency treebank for French, if it's available to everyone? What about foreign languages OTHER than the target language? (e.g. if the language is French, can we use a Spanish treebank?)</p> <p>Q: if the language is French, can we use a Spanish treebank?)</p> <p>A: Yes. Please see Q61 for more information.</p>
71	<p>The BAA states that "Each performer team will be afforded a limited number of human interactions in order to refine system judgments." The BAA does not appear to comment further on how the interactions will be "limited". Is there anything more that can be said here about those limitations? We are trying to understand the complexity of what will be allowed during this phase, which would be different if human participation is limited to, say, ten minutes total, compared to eight hours.</p> <p>Q: participation is limited to, say, ten minutes total, compared to eight hours.</p> <p>A: IARPA welcomes novel contributions that meet the requirements outlined in the BAA. The limitations mentioned in the BAA represent constraints on the number, not type, of interactions.</p>
72	<p>The documents returned during the human-in-the-loop phase are described as "seed" documents. Does that mean we are allowed to expand to documents outside those returned from the test set (e.g. those harvested from some other corpus) and return them as part of the process?</p> <p>Q: process?</p> <p>A: The seed documents are intended to describe, by example, the types of documents that are relevant to a query. IARPA welcomes novel contributions that meet the requirements outlined in the BAA.</p>
73	<p>One obvious way to improve an IE or IR system will be simply to directly annotate additional training data in English. Is this intended to be disallowed in the constrained condition and allowed in unconstrained? Or something else?</p> <p>Q: disallowed in the constrained condition and allowed in unconstrained? Or something else?</p> <p>A: The unconstrained condition does not affect such interactions. Performers will be allowed a limited number of interactions with program datasets as outlined in BAA Section 1.B.2. As outlined in BAA Section 1.A.3, the acquisition of additional datasets is not within scope.</p>
74	<p>Header address says ATTN: Jason Matheny but salutation says Dear David Markowitz. Is this correct?</p> <p>Q: Header address says ATTN: Jason Matheny but salutation says Dear David Markowitz. Is this correct?</p> <p>A: No. The salutation should say John Beielser. The BAA will be updated accordingly.</p>
75	<p>The BAA says that '...performers may be asked to identify some types of grammatical mood and/or veridicality, such as the negation of a potential event'. How is that different from sentiment which is out of scope?</p> <p>Q: potential event'. How is that different from sentiment which is out of scope?</p> <p>A: The BAA statement addresses grammatical moods, with BETTER focused on moods related to the assertions of facts. Sentiment captures feeling, such as positive or negative, toward a specific topic or entity.</p>
76	<p>Q: Are GPUs going to be provided as mentioned in Proposer's Day?</p>

	A:	The T&E team will have GPUs to run developed performer algorithms. IARPA will not provide performers with GPUs for the training/development of algorithms.
77	Q:	What will be the order of magnitude of the document collection from which the IR component will retrieve? Thousands? Millions? Bigger?
	A:	The exact size of the IR corpus will be determined at program kickoff. For planning purposes, though, the corpus will likely have an upper ceiling in the tens of millions of documents.
78	Q:	If an event type for the fine-grained and template filling tasks is not present in the annotated data, will we still be responsible for extracting it from English and other language documents? For instance if the fine-grain IE annotates "bombing" but not "hostage-taking", or "attack" but not "disinformation" will we be responsible for all four or just the two annotated ones occurring in different contexts with different actors and victims across all the documents?
	A:	BETTER is not focused on explicit lexical event triggers; all event types will be present in the annotated data. Please see Q14 regarding the data annotation scheme.
79	Q:	Can we use COTS MT? Our own MT developed prior to this project? Improvements or new MT for this project? Bi-lingual dictionaries? What are the constraints on software and resources for cross-language aspects of BETTER?
	A:	Please see Q61 for more information on the constrained vs. unconstrained regime. In general, IARPA welcomes novel contributions that meet the requirements outlined in the BAA. Performers should confirm that any approaches address the intellectual property requirements outlined in BAA Section 4.B.1.d.
80	Q:	Will the events (or event types) under each course-level category (or category combination) be predefined? Or, will they be entirely open (e.g., every verb could be an event type)? If the events will be predefined, what is the expected number of them? Tens, hundreds or thousands? Will labeled training data be provided by the BETTER program for the Fine-grained IE task?
	A:	The event types will be predefined and present in the annotated data. Performers can expect event types that number in the tens. Some form of labeled, English-language training data will be provided for all IE tasks.
81	Q:	The BAA states that "In the constrained condition, performers will be required to use language resources, e.g., annotated training data and parallel corpora, provided or identified by the Government team." Could you elaborate on these resources? Specifically, will this include: i. Monolingual annotated resources such as the Penn Treebank, OntoNotes, or FrameNet in English. ii. Analysis tools trained on the above (such as syntactic parsers or semantic role labelers). iii. Same as the above, but in the target language.
	A:	Please see Q61 for more information on the constrained vs. unconstrained regimes. In general, the resources listed will be allowable under the constrained regime.
82	Q:	Will the human-in-the-loop speak only English, or will that person also speak and be capable of interacting in the target language?

	A:	The human analysts are allowed to be bilingual.
83	Q:	In the context of the IE+IR task, the BAA mentions “seed documents” and “document level annotations”. How many seed documents will be provided? (Figure 1 illustrates a single seed document.) Will each seed document represent a separate IR query? How many document level annotations will be included in each seed document?
	A:	Each seed document will represent a single query. The number of seed documents and annotations therein is currently to be determined.
84	Q:	In the description of the IE+R task, the BAA indicates that the document level annotations will represent the finest granularity of event classification and states that “Performers are expected to use these document-level annotations to extract relevant semantic information and perform an information retrieval task.” It seems counter-intuitive that the finest level of classification is applied to coarsest unit (the whole document). We would like to clarify: The fine classification could vary from one event sentence/paragraph to another; in that case would a document be given multiple document level annotations?
	A:	Please see Q14 for an explanation of the data annotation scheme.
85	Q:	The BAA states that the T&E team will provide performers “with a corpus of documents annotated according to the program definition of relevance.” Will these documents serve as examples of relevance ranking for a specific IR query? Will performers receive a set of relevance-annotated documents for each seed document (or set of seed documents) that constitute a IR query?
	A:	Performers will receive relevance-annotated documents for a given analytic task. Since there will be a single IR query for each IR task these documents will serve as the complete set of annotated relevance examples. The seed documents are separate from the annotated relevancy corpus.
86	Q:	Will the performer systems be expected to handle personalization for more than one analyst? Will we receive more than one English Language training set covering the domain for each phase?
	A:	No, performers may assume one analyst. Performers will receive one English-language training set.
87	Q:	Will the T&E team provide the performers with the seed documents along with the Top-N Test Set documents during the human-in-the-loop evaluation?
	A:	Yes.
88	Q:	The BAA describes the potential need for performers to develop “interfaces to facilitate human interactions.” Will there be constraints on the types of interfaces that performers are allowed to construct and/or on the types of feedback that can be collected from the human subjects?
	A:	There are no such constraints.
89	Q:	The BAA states that the “constrained condition does not cover other linguistic resources such as dictionaries, lexicons or ontological resources.” Are linguistic functions, such as tokenizers, sentence chunkers and part-of-speech taggers for English and/or the target language, prohibited in the constrained regime? Would it be possible to explain the specific rationale for the constrained and unconstrained regimes to facilitate performer understanding?

	A:	Please see Q61 for more information on the constrained vs. unconstrained regimes.
90	Q:	The BAA states that “data annotations are provided at the sentence- or paragraph-level”. Does this mean that event nuggets/triggers will not be tagged? Does this mean that the English language training data will not tag the entities of interest (e.g., actors, locations, dates,)? At what level will author characteristics such as mood and veridicality be annotated?
	A:	Please see Q14 for more information on the annotation scheme.
91	Q:	Will the English-Target Language parallel corpora be aligned with the domain of interest, (i.e., matching topics found in the English language training data)?
	A:	Performers should not assume that any parallel corpora are domain aligned.
92	Q:	Are transfer learning methods, or similar machine learning procedures, allowed in the constrained data regime?
	A:	In general, transfer learning methods trained on monolingual corpora are allowed. Please see Q61 for more information on the constrained vs. unconstrained regime.
93	Q:	Will performers be allowed to collect telemetry and other usage information, via the submitted tool, from analysts, annotators, and evaluators as input to the system during testing and evaluation?
	A:	The human-in-the-loop occur on performer systems, so teams are free to collect any information desired during those interactions. Performers should not anticipate receiving such information from the T&E team.
94	Q:	Are commonly available, non-annotated data sources permitted in the constrained condition? Specifically pre-trained word embeddings?
	A:	Yes. Please see Q61 for more information.
95	Q:	May performers use existing off-the-shelf technologies such as dependency parsers in the constrained condition?
	A:	Yes. Please see Q61 for more information.
96	Q:	Can we assume availability of dependency parsers in the target language under the constrained condition?
	A:	Performers may use dependency parsers in the constrained condition.
97	Q:	Could you define "proprietary" in section 1.A.3? If performers publicly release data, does that make it non-proprietary?
	A:	Yes, if data is widely, publicly available it would be considered non-proprietary.
98	Q:	May performers self-fund data collection for use in the unconstrained condition if the data is made publicly available (i.e. is non-proprietary)?
	A:	Self-funded data collection during the course of the program is not within scope.
99	Q:	Will the foreign-language corpus domain be the same as the English training data domain?
	A:	For testing purposes the foreign-language corpus will be in the same domain as the training data.
100	Q:	Does the model need to translate the target language into English?
	A:	Translation is not a program requirement.
101	Q:	What are the requirements for information retrieval?

	As BAA Section 1.B.1, page 10 outlines, performers will be required to rank a list of documents according to a program-defined notion of relevancy. Queries will be performed by example. A:
102	Q: During the unconstrained data regime, performers are able to use government provided resources. What type of resources will be available? A: Please see BAA section 1.B.1, page 12 "Data and Evaluation" for an example of government furnished data.
103	Q: Can elaborate what is in and what is out in the constrained setting. I interpret "does not cover" to mean that non-corpus linguistic resources are permitted in the constrained setting. Is this correct? If so, what are the precise features that would cause a resource to be disallowed? Is it simply any resource generated through corpus analysis or training against text? What about something like Freebase, a resource derived from corpora through a variety of means? How about models trained from corpora commonly in use by the NLP community, such as NER models, parsing models, or models for ACE information extraction? A: Yes, non-corpus linguistic resources are permitted in the constrained setting. The constrained vs. unconstrained condition focuses solely on the use of parallel corpora for the purposes of training data-driven methods. Other linguistic resources such as lexicons or NLP tools, whether they include cross-lingual information or not, are not covered by the constrained condition.
104	Q: Table 1 in the BAA indicates there will be four languages in the program, but Table 6 mentions only three languages. Is there a typo in one of the tables? A: BETTER will make use of four languages. Table 6 in the BAA will be corrected.
105	Q: In the constrained setting, can performers assume the availability of named entity (person, organization, GPE) recognizers (NER), Part-of-Speech (POS) taggers and parsers in English? Are English Semantic Role Labeling (SRL) tools allowed in the constrained setting? How about the availability of NER, POS tagger, parser, and SRL for the foreign language in the constrained setting? A: All of these pre-trained tools are allowable in the constrained setting. Please see Q61 for more information.
106	Q: For the constrained setting, can performers use unlabeled monolingual corpora, for example, to train word embeddings? A: Yes. The constrained setting explicitly addressed cross-lingual resources.
107	Q: Is the annotated training data for IR tasks in English or the target language? A: The annotated training data for IR will include target-language examples.
108	Q: BAA states that the performers can nominate a fixed number of documents for human in the loop. Are there any constraints on innovative ways to perform annotation on these documents? Is there a limit on the amount of time an annotator can spend on a document? Are there constraints on what form of annotation can the annotators provide? A: As per BAA Section 1.B.2, performers are allowed to perform a certain number of human-in-the-loop revisions. Within that constraint, however, offerors are allowed to propose novel approaches that satisfy the requirements listed in the BAA. There is no fixed time limit on annotator interactions.

109	<p>For human in the loop, BAA “requires algorithms to be trainable using small batches of data in a streaming manner?” Can performers perform multiple iterations of human in the loop, which could enable more efficient ways to use the limited amount of human effort?</p> <p>Q:</p>
	<p>As per BAA Section 1.B.2, performers are allowed to perform a certain number of human-in-the-loop revisions. Within that constraint, however, offerors are allowed to propose novel approaches that satisfy the requirements listed in the BAA.</p> <p>A:</p>
110	<p>For human in the loop, is the “held-out test set” in English or the foreign language? If this is in the foreign languages, does this imply that the annotators will be foreign language speakers?</p> <p>Q:</p>
	<p>The IR data will include foreign-language data. The human-in-the-loop is allowed to be bilingual.</p> <p>A:</p>
111	<p>The BAA stated that performers need to provide the correct number of events per annotated section of text. In the following paragraph “The 2013 terrorist attack in Boston killed 3 people. This attack happened on Boylston Street.” Should a system output one or two events (one in Boston, the other on Boylston Street)? When there are multiple event mentions of the same type in the sentence or paragraph, should the system output all of them?</p> <p>Q:</p>
	<p>Please see Q14 for more information on the data annotation scheme.</p> <p>A:</p>
112	<p>In a paragraph, there might be multiple mentions of the same location (e.g., “US” and “United States”) that are the location of an event. Does a performer system need to output both as argument spans or just one of them?</p> <p>Q:</p>
	<p>Please see Q14 for more information on the data annotation scheme.</p> <p>A:</p>
113	<p>Does the program-provided IE annotation for first and second way points contain span and role annotation for event arguments, in addition to event types? Do the seed documents for the third way point (IE+IR) contain span and role annotation for event arguments?</p> <p>Q:</p>
	<p>The answer to both questions is yes.</p> <p>A:</p>
114	<p>How many training documents or events should the performers expect for the first, second and third waypoints, respectively? How many unique event types are there for each way point? Will these numbers vary by phase?</p> <p>Q:</p>
	<p>Performers should anticipate on the order of a thousand annotated examples per event type. The exact number of event types is currently unknown, but performers should anticipate less than a hundred per phase. As stated in BAA Section 1.A.2 “the amount of training data provided to performer teams may be variable over the life of the program.”</p> <p>A:</p>
115	<p>Table 6 (timeline) doesn’t list the time when the performers will receive the IR annotation dataset in each phase. Is the date going to be decided later?</p> <p>Q:</p>
	<p>Performers will received annotated IR data at the same point as the second IE task, e.g., IE Task 1.2. Table 6 will be updated to reflect this information.</p> <p>A:</p>
116	<p>BAA mentioned that the IR Relevancy will be defined at kickoff. Are performers expected to propose definitions for relevancy?</p> <p>Q:</p>
	<p>No.</p> <p>A:</p>