

**HR001121S0024**  
**Computational Cultural Understanding (CCU)**  
**Frequently Asked Questions**

**As of June 2, 2021**

**Q103: I was wondering whether the page limit for Volume I of the proposal can be more than 50 pages when the proposal addresses the combination of TA1 and TA2.**

A103: Please adhere to the length requirements as specified in the BAA.

**Q102: Is a bibliography of works cited in the proposal text desired? If so, where should it be included (as a section in the main proposal, as an appendix, etc.), and is the page length of the bibliography limited?**

A102: This information is optional. If included, it should appear in the main proposal. Please adhere to the length requirements as specified in the BAA.

**Q101: We would like to know if the contract company will provide any technology in addition to human expertise?**

A101: TA1 and TA2 proposers are required to provide technology to meet all requirements of the TA(s) the proposal is addressing; please see the Deliverables section of the BAA.

**Q100: For TA3, does DARPA anticipate that any training data or other data provided to the successful proposer will be covered under DFARS 252.204-7000 or 252.204-7008? Does TA3 require the successful proposer to have completed its NIST 800-171 implementation?**

A100: All selectable proposals resulting in a procurement contract will have DFARS Clause 252.204-7000 in the contract and DFARS Provision 252.204-7008 will be applicable. In addition, DFARS 252.204-7019 (provision) will be applicable to all procurement contracts and DFARS 252.204-7020 will be in the resulting contract.

**Q99: How much time will be allotted between the release of data and evaluations to train models with the released data?**

A99: Development data will be made available in monthly increments with the completed development dataset provided approximately 4 weeks prior to evaluation. Figure 2 in the BAA describes the program evaluation schedule; evaluation conditions in Phase 2 involve surprise culture + language pairs where no data will be released prior to the evaluation of a specific culture + language pair.

**Q98: Will performers be allowed to use their full computational capabilities to train these models or will training also be subject to the TA2 limited form-factor and internet connectivity restrictions described on page 10?**

A98: Proposers may utilize any computational capabilities to train models.

**Q97: A58 provides a helpful table indicating the language of TA2 input and output to the operator. In addition to the operator, is it intended that a) the Native Speaker be a direct recipient of TA2 output, b) the Native speaker not be a direct recipient of TA2 output, or c) may proposers decide what is most effective?**

A97: Proposers may decide what is most effective so long as output to the operator is supported per the requirements of the BAA.

**Q96: The BAA refers to "shifts in norms...that are indicative of communicative failure or impending conflict." Please provide an example of a dialogue during which a norm shift occurs.**

A96: For instance, literature on style shifting (where the level of formality changes within a single conversation) contains a number of examples.

**Q95: The amended BAA states that "output mechanisms to provide remediation to the TA2 user should include both text and text-to-speech (TTS) as configurable options." Also FAQ A58 states that "In Operational Condition 2 ... the TA2 system output is in the local language". This appears to imply that remediation in Operational Condition 2 cases proceeds without any "side dialog" with the monolingual English operator. Is that an accurate assumption? Is the goal of TA2 to go ahead and "replace/rephrase" operator's output and present it directly in the local language to the participant? Or can we assume that there may be some flow of information from the TA2 system to the monolingual English operator that provides the latter with some understanding of the suggested remediation – and maybe request confirmation? The BAA asks for "dialog assistance service able to .... suggest culturally and socially-appropriate conversational actions for remediation".**

A95: Operational Condition 2 does not require any "side dialog" with the monolingual English operator. Systems may provide information to the operator explaining the remediation and/or requesting confirmation, however this is not a requirement of the BAA.

**Q94: FAQ A50 states that "[t]he TA2 application is silent until an integrated TA1 output signals the need for the remediation of an operator utterance." Is this just an example mode of operation or is this how we always expect the system to operate, even in Operational Condition 2? Should the system plan to interfere on its own before the violation actually happens (e.g. when there are indications in may happen), or give other guidance to the user, or should we assume that remediation only applies, as stated in A50, after problematic operator utterances have already occurred, been translated (for Operation Cond. 2) and presented to participant?**

A94: TA2 systems may propose methods to anticipate remediation. However, this is not a requirement of the BAA.

**Q93: There are many sociocultural norms affecting communication that are nonlinguistic in nature, for instance, appropriate physical distance between conversation participants. Remediation of a violation of such a norm cannot be implemented as a manipulation of an operator utterance (but could, for instance, be affected by some other communication to the operator). Should we therefore assume that such norms and related remediations are out of scope for the program?**

A93: Detection of nonlinguistic sociocultural norms is in scope for TA1. Remediation of nonlinguistic norm violations is not the primary evaluation objective for TA2 but also is not explicitly excluded under the TA2 requirement to "suggest culturally and socially-appropriate conversational actions."

**As of May 25, 2021**

**Q92: Are you planning to update the package with an excel file? Or it is best to use the standard file?**

A92: As specified in the BAA, links to resources such as the DARPA Standard Cost Proposal Spreadsheet can be found here: <https://www.darpa.mil/work-with-us/contract-management>. The format of the Proposal Summary slide can be found in APPENDIX 1 – PROPOSAL SUMMARY SLIDE of the BAA. A Level of Effort Summary by Task is not required.

**Q91: What background data (without norm violations) is being collected?**

A91: TA3 data produced in support of TA1 development and evaluation are not specifically required to illustrate norm violations.

**Q90: Regarding the "emulator for a suitable mobile environment" described on page 7, does "mobile environment" mean a hardware environment such as a smartphone OS, a social environment such as a village town square, a game environment, or some other interpretation of the phrase "mobile environment?"**

A90: "Mobile environment" describes a hardware environment, such as a smartphone OS.

**Q89: The BAA mentioned that negotiation will be the main task for TA2, would TA1 also be limited to the negotiation task?**

A89: TA1 data is not limited to a negotiation task.

**Q88: Should we consider both distributive and integrative negotiations? (Note: In a distributive negotiation, every negotiator focuses on meeting his personal interests, regardless of the loss the others may have to face. In contrast, integrative negotiation focuses on mutual interests of all the parties and thus, comes up with constructive solutions that will be beneficial for all).**

A88: Proposals should consider all negotiations deemed to be relevant by the proposer.

**Q87: How many unique actors/participants are required to be in the 50,000 pieces of data recorded in each language/culture?**

A87: There is no specific requirement for the number of unique participants. However, the data must demonstrate a variety of sociocultural norms and support TA1 development for use in TA2 systems.

**Q86: What is the intended purpose of collecting 15,000+ hours of un-labeled video data (80% of 75%+ of 6 sets of 50,000 instances of 5-minute video clips)? Will later annotation of some of that video be necessary to evaluate TA1/TA2 performance?**

A86: Per the BAA, all TA3 data for TA1 development (including the unlabeled data) will demonstrate a variety of sociocultural norms as well as provide evidence of emotion. All data utilized in evaluation will be annotated.

**Q85: Will DARPA alone choose the cultures or will the data provider and/or eval team have input into those decisions?**

A85: DARPA will select culture + language pairs in consultation with the evaluation team. There will be a total of 6 culture + language pairs included in the program.

**Q84: In the end use case scenario, will video come from a button camera or equivalent? For the TA3 data, are both dialog participants required to be in frame or can it be filmed from one party's point of view?**

A84: TA2 frameworks will utilize input mechanisms that include the operator's perspective of the interlocutor as well as visual observations of the operator's behavior (e.g. facial expressions). TA3 data for TA2 development may include these TA2 perspectives and/or a third-person point of view; however, proposals should describe how the point of view will support TA2.

**Q83: Communication breakdown may happen for many reasons apart from norm violation, but still**

**cause similar emotional response and impact on the conversation. Is training data of such instances required to allow TA1/TA2 performers to distinguish between these scenarios?**

A83: Training data is not required. However, such instances could appear in TA2 evaluation scenarios.

**As of May 17, 2021**

**Q82: According to the RFP, it does not specify the start date.**

A82: The projected start date is January, 2022.

**Q81: I'm writing to inquire whether the government might consider the acquisition strategy to be limited to an 8(a) direct award, particularly for tasks 1 & 2.?**

A81: The acquisition strategy will not be revised to support a direct 8(a) award. All potential proposers are welcome to make teaming arrangements. Each proposal must address all requirements of the TA(s) included in the proposal.

**Q80: Does DARPA expect that the video documents will be filmed and produced at any OCONUS locations? Would it be acceptable to film and produce the video documents at CONUS studios using culturally appropriate actors?**

A80: Specific filming/production locations are not a requirement of the BAA. TA3 proposals should address the realism of data supporting TA2 development. See Amendment 01, page 12.

**Q79: What is the deliverable timeline for TA3?**

A79: TA3 data will be delivered incrementally to the program per the schedule defined in the BAA. Subsets of data should be delivered to the government evaluation team (and the program as appropriate) on a monthly basis. See Amendment 01, page 11.

**Q78: Are there required formats for audio and video content produced under TA 3?**

A78: There is no required format. However, proposals should address the recording quality. See Amendment 01, page 12.

**Q77: Does DARPA expect to receive a bi-lingual corpus under TA 3 that can be easily ingested in TA 2 and 3 or is monolingual what is expected?**

A77: Data created in support of TA1 development and evaluation are not expected to be bilingual. Per the ceiling condition described in the BAA, a portion of the data created in support of TA2 development will be bilingual.

**Q76: What is the name and contact for the BAA's Coordinator?**

A76: Please reference Part I of the BAA.

**Q75: Is the cover sheet included in the page limit for a full proposal submission?**

A75: Yes.

**Q74: Is there an abstract due for CCU prior to the submission deadline for proposals?**

A74: No.

**Q73: The government is going to suggest/require a specific machine translation engine. Will the government also provide a recommended/required speech recognition engine?**

A73: The specific machine translation system(s) and processing components utilized in the program will be decided by the government evaluation team in collaboration with performers. Components for (1) input processing, (2) Machine Translation, and/or (3) visualization that may be provided by performers are designed to be shared resources and should be made available to the program as a whole. See Amendment 01, page 14.

**Q72: The BAA notes that we need to fill out certain forms to ensure we are compliant with National Defense Authorization Act (NDAA) for FY 2019, Section 1286. In addition to the forms requested, are there any constraints on foreign nationals working on this program? In particular, we are talking to universities with grad students from China and Iran. Will those students be allowed to work on this program?**

A72: Please refer to the BAA. The BAA states that non-US performers are fine subject to applicable restrictions.

**Q71: The BAA states that CUI might be added to the program. Because of this, will all subcontractors (including universities and independent contractors providing subject matter expertise) be required to have safeguarding measures in place to protect CUI? Or will only the prime be required to have CUI safeguarding if we have a plan for keeping CUI data from any subs or consultants that are not able to handle it?**

A71: At this time, it is not anticipated that TA1/2/3 performers will be required to handle CUI. Any CUI materials that might be included in the program are anticipated to be kept within the Government evaluation team.

**Q70: Is it expected for TA1 teams to “team up” with a specific TA2 team in order to get their algorithms integrated, or are all algorithms from any TA1 team expected to be provided in a form that any TA2 team could easily integrate and leverage for their system development?**

A70: All algorithms from any TA1 team are expected to be provided in a form such that any TA2 team could integrate. TA1 and TA2 teams are expected to collaborate for integration and development. All TA1 systems should be interoperable with all TA2 systems.

**Q69: Could you please tell me if this is a new requirement, or if there is a current or previous contractor who has performed similar services? If available, can you please send along the incumbent contract or task order number?**

A69: CCU is a new DARPA I2O program.

**Q68: Will TA2 need to produce more than just text output during interaction -- i.e. will it need to produce speech (e.g. TTS)?**

A68: For flexibility of use, both text and speech (TTS) should be provided as configurable options for output. Proposers may leverage open source or other existing TTS components provided that these components may be provided to the program as a whole. See Amendment 01, pages 10-11.

**Q67: Should the task of communicative change detection also focus on explainability? I.e. also understand why the miscommunication is happening (ASR errors, MT errors, operators expressions or body language etc.) or just detect the event?**

A67: This is not a requirement of the program.

**Q66: Has this initiative been funded?**

A66: The CCU program has an authorized budget.

**Q65: Is there a past performance requirement?**

A65: No.

**Q64: Is pricing a CPFF or CPAF? Is there a cap on fee?**

A64: If applicable and appropriate a CPFF contract type can be awarded. See FAR Part 15 for the Government's fee limitations.

**Q63: Compared to the LORELEI program's level of effort in the past, is CCU envisioned as: Much larger, Larger, About the same level of effort, Smaller, Much smaller?**

A63: Proposals should describe a level of effort commensurate with the tasks proposed.

**Q62: Is there any information you can provide about the size of the budget that is anticipated per team?**

A62: No, proposed budgets should align with the proposed solution.

**Q61: Is there any information about the size of teams that you expect?**

A61: The Government will not place limits on the size of the proposed team.

**Q60: Will TA1 be evaluated separately from TA2? Will TA2 evaluation depend on results from TA1?**

A60: Yes, to both questions.

**Q59: What is available to performers for surprise cultures? For instance, will documents be provided for the surprise culture and how much before the evaluation?**

A59: No. The essence of a surprise culture is that the materials will not be provided ahead of the evaluation.

**Q58: In Table 1 Testing Conditions, can we interpret Column 3 as follows: Operation Condition 1, "TA2 System as integrated with TA1 modules/interpretations/markup of raw or translated input" Operation Condition 2, "TA2 system without TA1 interpretations, and only externally-provided MT input" Could you clarify the difference between Operational Condition 1 and Operational Condition 2 in Table 1? Is it responsive to propose, for Operational Condition 1 of the TA2 evaluation, that the TA2 system apply machine translation to its output (i.e. from English into Participant 2's language)?**

A58: In Operational Condition 1 (which involves an operator with a limited but existing ability to communicate in a local language), the inputs to and the outputs from the TA2 system are all in the local language. In Operational Condition 2 (which involves an operator with no ability to communicate in the local language), the operator's input to the TA2 system is machine translation from English into the local language and the TA2 system output is in the local language. This is summarized in the following table:

	TA2 Input from Operator	TA2 Output to Operator	TA2 Input from Native Speaker
Operational Condition 1	Local Language	Local Language	Local Language
Operational Condition 2	Machine-translated English into local language	Local Language	Local Language

**Q57: In a fielded setting, only contactless biometrics would be feasible for the emotion recognition component. For program development purposes, is it in scope to use neuro-physiological sensors involving contact as part of developing emotion recognition capabilities?**

A57: Yes. Proposers should outline any sensors required for their proposed approach. However, such sensors may only be utilized in research and will not be available during evaluation.

**Q56: Regarding evaluation, is the program metric about deciding when to remediate or also about the content of the remediation?**

A56: Yes, the program metric evaluates both detection of the need for remediation and the content of the remediation. Please refer to the Program Metrics section of the BAA.

**Q55: Will the evaluation assume that the foreign speaker has no knowledge of English? Knowledge of English and American culture might change the ceiling metrics.**

A55: No. That is not the assumption.

**Q54: In TA1, for the 25% of media that will be composed of audio and text, is the text a speech transcript of the audio?**

A54: No, this will consist of text documents.

**Q53: How will video, speech, and language labels be provided?**

A53: The data format will be determined by the TA3 performer in collaboration with the evaluation team.

**Q52: If a culture uses more than one language, should we assume that only one of the languages will be used for that culture in data production and evaluation? If not, how will this be handled with respect to data and evaluation?**

A52: Yes, only one language will be used for that culture for evaluation purposes.

**Q51: TA3 is tasked with producing "50,000 documents per language" yet the program schedule denotes data releases by "culture" and the BAA uses the term "cultures/languages." Is it anticipated that each culture will generally employ a distinct language, or might some of the distinct cultures evaluated in the program share a language? In the latter case, would we expect 50,000 documents to be produced per culture, rather than language? Further clarification: Let culture 1 use language A let culture 2 use language A Will TA3 produce 1) 50,000 docs of A, to be split among culture 1 and culture 2 2) 50,000 of A for culture 1 and 50,000 of A for culture 2 3) something else? Is any given instance of the data (e.g. a video) expected to be mono-cultural? Mono-lingual?**

A51: 50,000 documents are anticipated for each language + culture pair. Each instance of the data will be associated with a language + culture pair.

**Q50: Can you walk through a sample interaction and what you would expect the system to do and how the user would interact with it and the person they are speaking to?**

A50: An operator uses the CCU system to engage in conversation with a non-English-speaking interlocutor. Depending on the operator's ability to speak the local language, translation may be turned on or off. Either the operator or the interlocutor initiates the interaction. The TA2 application assesses the ongoing conversation via visual, auditory, and language information from input mechanisms. The TA2 application is silent until an integrated TA1 output signals the need for the remediation of an operator utterance. The TA2 application then supplies via an output

mechanism an alternative, socioculturally-appropriate utterance so that the operator may rephrase. The interaction then proceeds. Please see also Q58.

**Q49: Have any of the languages/cultures have been decided? Are we allowed to know what they are? Have you decided on the specific types of emotions (anger, sadness, frustration, etc)?**

A49: The first language + culture pair will be Chinese (Mandarin). Additional language + culture pairs will be announced to performers incrementally during the program. See Amendment 1, pages 10 and 11. The final inventory of emotion categories utilized in the program will be determined in consultation with performers and the evaluation team. Proposers should refer to existing literature for examples of specific types of emotions.

**Q48: For TA1 task 1, please characterize the granularity of input (e.g. sentence/paragraph/document/individual/population) at which it is anticipated gold annotated data will be labeled. For TA1 task 1, please characterize the nature of the envisioned label space of gold annotation.**

A48: The granularity of input will be dependent on what is proposed by TA3.

**Q47: Will the training/test data capture actual interactions (with attestable emotions) or just actors potentially feigning emotions?**

A47: The test data utilized will be as realistic as possible within applicable constraints. Training data characteristics will be dependent on the TA3 performer's approach. TA3 proposals should address the generality of the data provided for TA1 development and evaluation and the realism of the data provided for TA2 development. See Amendment 01.

**Q46: In terms of changes in norms or emotional states, will that be defined across timesteps/segments, with respect to the participants in a conversation?**

A46: Yes, changes will be defined across timesteps/segments with respect to the participants.

**Q45: Will the program establish an ontology of sociocultural norms, emotional states, and the associated changes, shared across multiple modalities? Will the ontology be partially complete (e.g., only about emotions, or only cover some labels/characters), in order to promote "discovery" of new norms, etc? How many sociocultural norms is each data expected to represent? Is there a predefined list of sociocultural norms to be provided? How many sociocultural norms are expected to be discovered in each culture?**

A45: TA3 will provide a set of norms for development and corpus evaluation, but this will not constitute an exhaustive ontology. Per the BAA, TA3 data should demonstrate a variety of sociocultural norms; however, the number of norms required for annotation is not predefined.

**Q44: Will language identification (including regional variant and/or creole) be considered within scope, or will it be assumed that participant languages are known a priori?**

A44: Yes, it is possible that language identification could provide benefit and is in scope. However, at a basic level participant languages will be known a priori.

**Q43: Human interactants can occupy multiple roles that impact appropriate behaviors (e.g., a married couple who also work together professionally); has DARPA considered this in emphasizing the use of unlabeled training data?**

A43: Yes, DARPA has considered that participants could occupy multiple roles depending on context.

**Q42: Are there any guidelines on the sourcing of training data, given that existing video/text/etc. data will have been generated by someone for a specific reason, and thus cannot be viewed as representative (e.g., security camera footage more likely in settings subject to security risks)?**

A42: Characteristics of development and corpus evaluation data will be dependent on the TA3 performer's approach. TA3 proposals should describe the generality of the data proposed.

**Q41: Are TA3 datasets expected to feature negotiation interactions, as those types of interactions are the focus of the demonstration/evaluations?**

A41: Negotiation interactions are the focus of evaluation. TA3 will consult with the evaluation team in determining scenarios for development data in support of TA2. TA3 data is anticipated to represent a range of scenarios; please see Amendment 01, page 12.

**Q40: Will TA3 data include labeled examples of all the emotion and socio-cultural norm labels that TA1 is expected to recognize?**

A40: Yes, TA3 will have for evaluation a subset of the labels perceptible by the TA1 systems.

**Q39: If TA3 provides a corpus of labeled and unlabeled socio-cultural interaction data for the target culture, it is likely that the labels applied to the target culture will not span the range of possible socio-cultural labels for all cultures. (For example, the target culture data may show instances of highly hierarchical behavior, but no instances of highly egalitarian behavior.) Will TA3 provide a similar dataset for at least one reference class culture, such as the US, in order to provide examples for contrast?**

A39: No. However, CCU will provide data corresponding to multiple cultures over the course of the program.

**Q38: It sounds like TA3 will decide on the set of emotion labels and the set of socio-cultural norm labels that TA1 needs to recognize. How will this categorization scheme be selected? Will DARPA specify the cross-cultural dimensions framework to use, or will TA3 make the decision, or will the performers as a group be expected to develop a consensus ontology?**

A38: TA3 will decide, in consultation with the government evaluation team, on the set of labels that are available for TA1 evaluation. The scheme will be developed in coordination with program participants. There is not so much a defined ontology as a set of evaluation data that will be used to evaluate the outputs of TA1.

**Q37: Approximately how many emotion labels and how many socio-cultural norm labels does the program expect to use? Will the effect of label set size on recognition error rate be accounted for in setting program performance metrics?**

A37: DARPA's interpretation of metrics will account for task complexity. There is no pre-defined number of labels.

**Q36: Are TA1/TA2 performers prohibited, allowed, or expected to supplement the TA3 training data with their own labeled and unlabeled data for the target culture(s)/language(s)? Beyond data made available by the program, which types of data are we permitted to collect ourselves for training purposes?**

A36: Allowed. Any type that can be shared with other program participants, subject to privacy and other constraints specified by DARPA.

**Q35: How much and what type of background information will TA2 dialogue assistants have prior to the interaction? E.g., participant roles, histories, preferences, past interactions?**

A35: Minimal to none.

**Q34: Would it be in scope for TA1/TA2 to generate small amounts of sociocultural data for testing previously unseen/unlabeled languages to mitigate risks in advance of the introduction of surprise cultures during eval?**

A34: Yes, proposers may generate data provided that this data is made available to the program as a whole.

**Q33: Since culture 1 is planned to be introduced in month 0, is it an already extant corpus?**

A33: This data is expected to be available by the beginning of the program.

**Q32: What kind of data/modality is expected for TA1 Task 1 (norm discovery)? Is the “unlabeled discourse” data in the form of text, audio, ASR/system output, or something else? Will the data be provided by the program, or will performers be able to obtain their own data, or both?**

A32: Data will cover all modes. Data will be provided by TA3, however this does not preclude performers from obtaining their own data, as long as the independently-obtained data are available to the program as a whole.

**Q31: Dialogue assistance task: The task is to observe conversations for evidence of misunderstandings. Not clear if these are human-assisted, or device assisted interactions. Is there a speech-to-speech system in the loop? If so, will the whole system be provided, or only the MT component? It is possible that the performance of current MT or speech-to-speech systems in these kinds of interactions may be problematic – depending on the language?**

A31: Please refer to the BAA. The ceiling condition is human-assisted and TA2 systems (with incorporated TA1 components) are evaluated in device-assisted interactions. Proposals should outline approaches for operating with inputs from speech-to-speech machine translation.

**Q30: Is there an option for the teams to bring in their own MT engines/capabilities in order to augment or remediate available MT system?**

A30: Yes. Teams are not prohibited from supplying MT capabilities. However, to allow comparison between performer systems, these capabilities must be made available to the program as a whole. Additionally, the BAA emphasizes creating and leveraging Open Source capabilities where feasible. Components for (1) input processing, (2) Machine Translation, and (3) visualization are designed to be shared resources and should be made available to the program as a whole; please see Amendment 01, page 14.

**Q29: The idea that off-the-shelf MT can be plugged into the pipeline and just work, presupposes that MT effectiveness, performance and errors are not impacted by cultural context, understanding and situational interaction. This is generally not the case. How do performers deal with problems caused by the mistranslation of inputs? Redrafting the inputs and running it through a “black box” MT system and hoping this fixes the output often is not viable, because the process is not deterministic. Would part of this process involve post-processing translations from the “black box” to dress errors and cultural appropriateness? This sounds like translation by other means.**

A29: TA2 systems should have methods for accommodating errors from input/pre-processing.

**Q28: Are innovations in video and/or speech processing expected to be part of the technical activities, or is the expectation that off-the-shelf techniques will be used? Are all performers expected to include members with expertise in video analysis and/or speech processing?**

A28: Development of novel processing technologies is not a primary goal of the CCU program. If novel processing approaches are proposed in the course of achieving the objectives of a TA, proposals should describe how this innovation directly supports the program objective. Teams should include members with the expertise required to execute the work proposed.

**Q27: How about situations such as non-understanding handled? Detecting and reacting to situations such as the operator couldn't get a culture-specific joke. Would that be in the scope of this project?**

A27: Systems should be able to handle a variety of communication failure types and causes.

**Q26: How active a role can the TA2 system play in the assistant role. For a real interpreter, they will be directly interacting with both participant1 and participant2, translating each to the other, but can also engage in side conversations if necessary, e.g. to manage the flow of information or if a word is not understood. Is TA2 assumed to be just a resource for Participant 1, or also translating participant1 to participant2 (using COTS MT) and vice versa, or in any case able to communicate with participant2, e.g. to ask for clarification or to smooth things over, or influence how something gets translated in the current cultural and situational context?**

A26: Within the requirements of the BAA, proposers should include the level of assistance they believe to be necessary.

**Q25: Does DARPA intend to provide trained versions of SotA translation algorithms and sentiment analysis, or will TA1/TA2 performers be responsible for selecting their own SotA tooling?**

A25: DARPA does not intend to provide trained algorithms.

**Q24: Which TA2 input mechanisms / sensors are of interest? E.g., 3D cameras, ultra high resolution cameras, infrared cameras? Might cameras and microphones be mounted to a fixed positions (instead of worn)? Can we use any sensors beyond video and audio?**

A24: Proposers should define input mechanisms or sensors of benefit to the proposal. Input mechanisms are assumed to consist of primarily wearable components and proposals should address the realism of the sensors proposed for a field setting. See Amendment 01, page 10.

**Q23: There were discussions regarding compliance/development in concert with the underlying frameworks (e.g., APIs, etc). Who will be responsible for the frameworks? Collaborative among TAs? Led by Evaluation Team?**

A23: Each TA is responsible for their own underlying development framework(s). Aspects of the overall integration will be addressed by the evaluation team in collaboration with performers.

**Q22: Integration with machine translation tools was discussed. What tools are currently being considered? Who will be providing them or making that determination? TA1? TA2? TA3? Eval? Do you foresee Machine Translation to be required by TA1 along with other SOTA language processing?**

A22: Machine translation components will be chosen by the Evaluation Team in consultation with performers. Machine translation may be optionally utilized in TA1 approaches. Proposers are not prohibited from supplying MT capabilities provided that these are made available to the program as a whole. Speech-to-speech machine translation component(s) must run within the integrated TA2 prototype. Proposers should budget appropriately for any proposed use of commercial utilities. See Amendment 01, page 14.

**Q21: The BAA notes the visualization framework(s) is/are to be determined by the program. Is the visualization framework(s) an existing system(s)? or is it possible for TA1 or TA2 to propose a visualization framework? or is the visualization framework(s) part of the govt evaluation team's work?**

A21: The visualization framework is only relevant to TA1. The development of novel visualization frameworks is not a goal of CCU. Proposals must outline the ability to provide any proposed visualization resources to the program as a whole; CCU prefers Open Source solutions. See Amendment 01, page 14.

**Q20: Will the government provide recommended/required body language/facial expression detection?**

A20: No. Performers are expected to propose detection components for use in the program.

**Q19: Should a team include an ASR expert or can teams rely on SOTA publicly available tools? Is it recommended to have a psychologist or social scientist on the team?**

A19: Team composition is up to the proposer. Teams may rely on SOTA publicly available tools.

**Q18: Should we assume that TA1 focuses on sociocultural analysis, and that SOTA capability that provides raw extractions/analyses from text, audio, video should exist prior to the program?**

A18: Yes. If proposers feel the need for additional processing capability development, it should be included in the proposal. Performers are responsible for supplying their own NLP tools, other than Machine Translation, which DARPA will specify.

**Q17: By using “negative emotions” from TA2 in TA1, does this indicate a feedback loop between TA2 and TA1?**

A17: No. Negative emotions are being annotated by TA1 Task 2, so the feedback loop is between TA1 Task 2 and TA1 Task 1.

**Q16: Are TA1 APIs to both discover norms on culturally relevant data and recognize the application of a norm in the evaluation context?**

A16: Yes, because they have to be able to do that in the corpus evaluation module.

**Q15: Modern conceptions of culture view it as a set of tools that are strategically as well as reflexively deployed by participants; is this consistent with DARPA’s intent?**

A15: DARPA’s intent is that proposers define their motivations for their proposed representations of social and cultural behaviors.

**Q14: How should proposers think about within-culture versus cross-culture differences given that cultures are not uniform over virtually any size or time scale?**

A14: Proposers should outline accommodation of within and cross-cultural differences as motivated by theoretical and empirical evidence.

**Q13: Will multi-person interaction (e.g., one participant gesturing to a non-speaking colleague) be considered part of the scenarios explored?**

A13: Possibly. There might be observation of multiple participants considered as part of the analysis.

**Q12: Are performers required to address all of the areas for expressing norms within societies, languages, and group affinities or can they focus on just one (or more) of these areas?**

A12: Proposals may emphasize specific relevant areas, but each proposal must cover the whole of a TA.

**Q11: Are performers required to address emotional, social, and cultural norms or are they able to focus on just one of these areas?**

A11: Proposals must address all aspects of the TA, including emotion recognition and discovery of sociocultural norms.

**Q10: Are research outputs intended to be operated in the field with limited computational resources?**

A10: Yes, research outputs are intended to be operated in the field with limited computational resources. Please consult the BAA for a description of Phase 1 and Phase 2 requirements.

**Q9: Is TA1 responsible for developing visualizations or will another party develop visualizations with outputs created by TA1?**

A9: The program does not include efforts specifically on visualization development. Open source visualization frameworks will be utilized with the minimum configuration required to achieve the

program objectives. Proposers may describe specific visualization frameworks for use with TA1 outputs; however, such frameworks must be available to the program as a whole.

**Q8: How and when will the CCU API's be determined? They seem to be needed by month 4.**

A8: APIs will be determined during month 1, in discussion with performers.

**Q7: Is the TA2 Phase 2 requirement for deployment on a "smartphone form-factor" intended to specify a literal smartphone, or a device with a similar computational footprint and size (e.g. Raspberry Pi or Intel NUC)? Using an actual smartphone may rule out use of external microphones and cameras that were found beneficial on the Phase 1 laptop configuration.**

A7: The Phase 2 requirement describes a literal smartphone, possibly with augmentation from external equipment (e.g. a microphone). Additionally, Phase 2 will include evaluation runs where TA2 units must operate without reliance on internet connectivity. See Amendment 1, page 10.

**Q6: Are both semantic and non-semantic indicators of emotion considered in scope?**

A6: Yes, all relevant indicators of emotion are considered in scope.

**Q5: Does the program consider an assessment of a culture's value along general cultural factors/dimensions (e.g. degree of collectivism versus individualism, degree of long-term orientation) to be in scope for the type of social norms that TA1 will be recognizing?**

A5: Yes, all relevant cultural factors are in scope.

**Q4: Does the program consider highly specific behavioral social norms (e.g. make eye contact with the speaker, bow when introduced to a person of higher status, do not interrupt) to be in scope for the type of social norms that TA1 will be recognizing?**

A4: Yes, any norms that affect communication are in scope.

**Q3: Which state-of-the-art mono-cultural emotion recognition engines represent the baseline performance for emotion recognition that DARPA has in mind?**

A3: DARPA will select one or more baseline systems from the existing literature prior to evaluation.

**Q2: Could you please define what is meant with "culture"? Is it to be understood as "language"? Can you provide some insight to the alignment between language and culture from the program's perspective? Is the expectation that all 6 cultures will be linguistically distinct?**

A2: Culture is not understood to be equivalent to "language". Language and culture are distinct but may overlap. There may be multiple cultures using the same language or there may be a one-to-one correspondence between a language and a culture. We expect that proposers will define the relevant aspects of culture. However, the BAA includes examples of cultural facets that may influence discourse behaviors, including the performance of individual and group identity.

**Q1: Are solutions welcome for individual modes, or is a multimodal solution required, encompassing all of video, speech, and language? Would a TA1 team be required to address all data modalities (video, audio, text), or can a TA1 focus on one type of data? Please describe DARPA's expectations with regard to performer's performance of multi-media analysis. Specifically, is each TA1/TA2 performer expected to implement technologies to process audio, video and image data?**

A1: All solutions must be applicable to all modes of data, including text, speech, image, and video. TA1 proposals must address all data modalities. Processing technologies utilized are anticipated to be off-the-shelf, although proposers are not prohibited from including additional technologies provided that these improve upon state-of-the-art open source components and can be made available to the program as a whole.