Optical Character Recognition (OCR) Translation Assessment Event (AE) Q&A Telecon Transcript

1. **What metrics/data sets are you planning to use to evaluate OCR and translation performance? What is considered "bad" or "good"?**
   I think the OCR is going to be in part reliant on the language for the actual translation. I think what we're looking at is primarily the capture. What's good or bad is going to depend on what performance capabilities other devices have. So clearly a device that can capture a billboard at you know 500 yards away is going to be preferable to one that can only capture a billboard at 100 yards away, one that can translate a poster across the street in low light is going to be more preferred than one that has to have bright sunlight directly on it. That's really going to be the criteria that we're going to be evaluating on. How well all the devices can do and then we're going to try to pick the best one that meets our needs.

   > **Follow up question: Assuming that you capture it well, the OCR system and translation system are both still going to have errors. Do you have an idea of what level of accuracy is considered good or bad?**
   > Well, I think the big thing is, and you know I think everyone on the call can appreciate that the translation aspect of this is still very much an art form with lots of errors, so if we're getting clean capture and then that's getting converted to the text accurately, then whatever the translation that comes our based on our software will live with that we already know kind of where that is. As long as it can get a clean capture and an accurate translation to the text, then our software will be able to handle the rest. If the software comes close to the mark of the gist of what the image is saying, then that will be acceptable performance.

2. **Which PM would this ultimately reside at?**
   Currently we don't have a PM to transition to right now. We're working with one about network data management, so they may find a home for it once we get everything ironed out. Mr. Smith, our Acquisition Executive, said that in time those things will be understood and not to worry about the program of record to transition to right now.

3. **Will it be important for the OCR and accompanying AI/ML capabilities to be handled at the edge and not have 3rd party calls to a cloud service?**
   100%. We're very much wanting this to be able to operate without having to be connected. You know, for all kinds of scenarios where that that's desirable. Obviously if we're in an environment where being connected is good, then it should have that ability to do that through the cloud as well, but ideally what we're after is the full capability, without the cloud.

4. **Is the software envision to run as an .apk (Android Application).**
   I believe so. I'm hoping ultimately we'll get where we can run on all platforms, but I think primarily everyone's looking at the android form factor. Clarification: For the phone application of it, yes, but if it was running on a processor, like a tensorbook or some other little bit more robust processing in a vehicle, it may not necessarily be android. It could be running on the native hyper-enabled operator system, or possible even Linux or something else windows-based.

5. **Is the primary objective for real-time OCR doing scene-text resolution or is this needed for latent data (i.e. docs as well)**

Both. Ideally what we'd like to get after is you want to make copies of documents and translate. You want to be able to read newspapers, want to be able to read magazines, as well as street signs, billboards and just give our people a more full situational awareness and understanding. Also, if that thing is able to capture news and other media in a different country, that would be great also.

6. **What level of AI/ML integration do see being using within the OCR process, i.e. existing trained models to enhanced the results.**
   We're working on that right now. I'm not too sure that the level of the AI/ML integration is as important as if the OCR process can work with it. I think it's fairly standard stuff when you're talking about ultra and Kaldi and other sort of Google format language translation, so I don't think that that component is going to be a significant factor.

7. **Do you want multiple models for different hardware (The machine models that are being produced.)? Mobile has a lightweight OCR model and laptop was more accurate and robust model?**
   We are not married to a particular model or hardware. We are focusing on the capability first. If we have a good capability, then we can focus on the size and weight of the hardware.

   > **Follow up question: Regarding hardware, what about the machine models that are being produced? A lightweight OCR model can be created which can be uploaded to a phone that will be usable compared to a more heavyweight robust model which can be loaded onto a laptop.**
   > We might need two different form factors. If you can provide us with both forms (one that can be uploaded to a phone and one that can uploaded to a laptop), we might be interested in that.

8. **Is there a list of your top 3-5 languages you want for OCR? Literally can do OCR for basically all languages but it will be meh compared to focused development.**
   We are hoping for the OCR capability to be funneled through our software. After the language is converted to text, then our software can handle the rest.

9. **Item 14, do you want (Text to Speak) TTS capabilities in the native language or does the captured content need to be translated and then TTS in English?**
   Yes, the main output that we are looking at right now is English.

10. **Do you just want a software solution or also a hardware one?**
    We are open to both.

    > **Follow up question: If industry has a software solution, but does not have a hardware component, could we possibly have help in finding another company with a hardware solution to partner with?**
    > Initially we are focusing on using the software solution and we will use the software in a mobile vehicle. We do not need a customized processing capability; we will probably use the best edge processing commercially available solution.

11. **"USSOCOM S&T JATF would like to enhance the capabilities of their language translation devices." So you are only looking for OCR?**
    We are primarily looking for the OCR.

12. **Items 4 and 8 on the Requirements Request document seem to be somewhat contradictory. Can you clarify what these two statements about form factor mean?**
    We would like for the item to be as small as possible. It can be in the form of a phone, but we are also open to suggestions at this time.

13. **Do the submissions need to provide language translation capabilities functionality or just character recognition?**

    The submission needs to provide language translation capabilities functionality as well as character recognition. We are looking for the ability to read images with the language on it then have it go through the software that processes the language that is being used for translation. We would like for it to read signs and billboards then translate. We are looking for the ability to visually capture the writing, put it in the software once it is converted to text, then be translated.

14. **If you want this software placed in your existing hardware, what specs are we looking at in that hardware? Can't give reliable speed estimate on unknowns.**

    We have a tensorbook and S21 phone. The phone is the interface for the translation capture for the voice to voice. The computation will go on the laptop of the tensorbook. That is the current form factor that we are working with. We might take bits and pieces and add it to a smaller form factor, because we do not need all of the existing parts for language translation. This is the current equipment that we are using. A counterpoint to that is that we are open to other form factors. For example, if there is a camera type device that we need to plug into the phone or the laptop to get the OCR capability that we are after, that is also okay. We are open to hearing what is practical at this time.

    Follow up question: Is that the lambda labs tensorbook laptop?
    Yes, it is.

15. **Is this being used for on-the-fly OCR and/or collection and analysis?**

    Yes, we would like for on-the-fly OCR. In an ideal scenario, the Warfighter can put the item on their dashboard, like a go-pro and it would read street signs and translate to "one way" or "do not enter" in real time or as close to it as possible. We are not looking into collection and analysis as much as we are looking into helping users navigate and communicate.

16. **Any idea of speed in terms of mobility in the vehicle?**

    I would not expect anything to be translating as they're passing signs at 60 miles an hour on Interstate, for example. However, if it can look out just like a human could and see something a few 100 meters down the road and give a translation like "here's the exit for the museum", then I think that would be good. We realize that this is new tech in that's growing, so if we figure out that the fastest speed that we can be moving to get a good capture is 30 miles an hour for example, then that becomes an operational parameter that we incorporate into. The bigger thing is going to be for you guys to come back and tell us what some of those limitations are.

17. **How much is cost a factor?**

    We're going to come up with an area to look at and evaluate proposals that will come in, but right now we don't have it. Currently this is still a major balance set of fact finding for us.

18. **What keeps the group from using something like google lens with translation that is readily available on android and can be used offline/on airplane mode?**

    We haven't found Google to be super even when connected, and certainly haven't found it to be very good at all when disconnected and that's just with language translation put into text. Google has been relying heavily on the cloud and that's one of the things we're trying to get away from.

19. **Will you provide sample data for testing the accuracy of the model/software?**

That's something that we can look into. We have standardization for folks but we're going to have models that we use to test the internally. I'm assuming you're asking for something that you can use to test before you submit, I think that's something that we can work on providing.

20. **Would there be interest in specific dialect models versus broader language models? i.e Iraqi, Yemeni, instead of only MSA Arabic**
Yes, we want all of it with integration of dialects.

21. **When processing documents, are you looking for an ability to detect specific keywords within a document and apply different behavior based on those keywords?**
At this point in time, our main focus is the ability to read the documents and translate accurately. At the end of the day, I don't know how deep we need the OCR capability to be able to do things like detect specific keywords. If it can just capture and help us translate, I think that would still be the number one objective.

22. **Is it ok to have a setting/selector for language?**
If it can do the main job, that's what we're after but I think it's a good option.

23. **How long would the bridge be in place for if a transition partner is not ID'd this FY?**
If you're not ready this year, then we'll see what happens next year

24. **Is the software responsible for the storage of translated text and allowing search capabilities on top of it (From item #10)?**
I'm assuming you're asking would the Osseo OCR software be responsible for the storage of that? I think the answer is no. Whatever we do, we're going to have our own storage capability on our hardware so that we can use it. Now the searching capability if that can be provided with the OCR then that just makes it that much more attractive. However, we store our data, we're going to have to have it searchable in some way, and that will be, I think, primarily the onus on us to do that part.

25. **If a sign says One Way after 3 PM and the OCR reads it as One Way after 8 0M, that would not be acceptable performance? Example of performance error.**
I would say that particular example would probably not be an acceptable level of performance. If you guys tell us, hey, that's just as good as it's going to get this year, and for the next five years, then we're going to have to evaluate if that works for us.

26. **What level of AI/ML integration do you see being used within the OCR process, i.e. existing trained models to enhanced the results?**
We are going after the very best that we can possibly humanly get for AIML and language translations, so as far as I'm concerned, the most important part is that the OCR can capture and convert that to something that our AIML can translate.

> **Follow-Up Question: When you look at sensor base, meaning the devices that can process down at the source of information what in your opinion, do you want that OCR as close to for source of data? Do you want that OCR as close to the actual data type and manipulating the data?**
> I think we answer that when we say we'd like it to be able to do this without being connected to the cloud. If you look at the cloud reach back as a luxury, but primarily the heavy lifting has to be done right there at the source.

> **Follow-Up Question: You mentioned analysis, can that data later be moved to improve the translation capability of say, a software application?**
> We're going to save data so we can help our AIML learn better when it's in a good environment for doing that later.

27. **Would the language be known in advance, or would we also need to detect the language? Also, would in involve handwritten text or only printed?**
I think, for most people when they go travelling, they know where they're going, so yeah, we will know general languages in advance. But if the ability to detect the language is an option that can be provided, I think we would very much welcome that as well. Regarding the second question, I mean imagine traveling around in a foreign country and looking at the billboards all over in some big city it's going to come in all types of fonts, so I think its ability to recognize characters, whatever their form and shape and convert those to text that can be translated, I think, is going to be key.

> **Follow up: Are you looking for multiple translation at the same time?**
> All the research indicates that multiple languages being translated simultaneously on the same platform is probably going to be the future in the next couple of years. At this time, I think what we're focusing on is very simple, with our language device being able to actually talk with another human being, without the internet and have it be as accurate as possible, and so the same thing goes with the OCR devices, if we can just get it to where it can, if we know we're going to China, you know, then we just wanted to be able to do Chinese. For now, let's stay focused on one language at a time.

28. **Would you like for the software to automatically separate and sort documents within a file that contains multiple documents?**
I think that is a great option to have in there, so yes please.

29. **Can the government confirm if the application should provide accuracy at 99%+ or higher on handwritten and digital text?**
Yes, that is the standard.

30. **What software would our capability feed downstream for further analysis?**
I'm not sure it's in the scope of what we're after here. If we just focus on our ability to actually translate what we're seeing, then that would be fine. Everything else that we want to do after that you know is secondary concern.

> **Follow up: Does the file format matter AKA would it be acceptable for the output to be in a json file format, or would you like xml, for example?**
> The more common or universally understood the form factor is, the better. Traditional forms would be best, but if there were some real powerful reason to go with something very specific because it did it just so much better I think we'd be totally open to that.

31. **What does "Variable optics are acceptable" mean?**
We realized that you know the camera that comes on a phone might not necessarily be good enough for some of the things that we want to get after especially when we're talking about things at a distance and things at low light, so if it happens to come with a kit like you can interchange these lenses depending on the need that's completely acceptable

32. **What are the communications protocols for this capability? Should it connect via Bluetooth or SIM? Or something else? What are the restrictions?**
We're very open to things like Bluetooth and Wi-Fi but also, it would be nice if there weren't any type of signal being emitted that could be monitored, observed, or interrupted, so if it can be actually a device or cable that plugs into a device in that I think would be preferred. But it's good to have those options for environments where that's not a concern.

33. **Which languages should be focused on for translation?**

I think if someone wants to pin this to a language to use for the demonstration, I think, just look at all the big guys, you know, Russia, China, Arabic, Hindi - languages that a lot of people speak and from which we can work on derivative languages later so that would be my recommendation.

34. **Should the translations be displayed on the glasses/binoculars?**

I think that's pretty cool; it wasn't a thought that I had I was envisioning people either looking at a phone or a laptop, but I guess the form factor could be glasses or binoculars. And what I also am imagining here is that you know we might have a camera set up on on a dash or person might have just their phone in their hand that they're trying to use as the capture device, but I think you know glasses and binoculars as options are certainly open. Any device that can do the job, is welcome.

35. **What are the languages of interest?**

Russian, Chinese, Arabic. But we're open to anything.

36. **The information given states optics. Is this request for both hardware and software?**

I think the answer is yes and no, but I think it's going to have to be both because when i'm talking about optics, I imagine just because of other you know familiarity with cameras and whatnot your phone can capture some things, but sometimes you need a real camera or a big zoom lens or something like that or something that you can actually change the aperture to capture in low light images, so I imagine it's going to have to be a combination of both hardware and software, in order for it to work efficiently, especially if it's going to be disconnected from the cloud.

37. **Can you elaborate on the existing "language translation devices" or other "ongoing efforts" in order to provide a clearer picture of how integration could work?**

That that's an interesting question brad, but I think that the main thing Is we're just looking for the OCR capability to be able to work with whatever form factor or hardware or software, we have for language translation. We are looking at a couple of different entities to provide the language translation device and software and I think Brad's company is one of the them. But I think the big thing is as long as the OCR can be integrated, then it can be of service to us.

38. **Is a solution that combines third party commercial Software Development Kits (SDKs) with our own capabilities acceptable?**

I don't think we're overly bothered with whomever produces it or if it's a combination, if it does, what we need, I think that's the happy place.

39. **To get a better sense of a possible scenario, you could have a situation where someone is in a vehicle driving in a city with billboards written in a specific foreign language, just say Russian, and ideally you'd want to be able to capture that information with a device, download it to a laptop and immediately provide translation of that to get a better sense of what specifically is stated on the Billboard. Is that an accurate use case?**

More or less yes, but the form factor may vary. It might just be we have it all packed in a phone or a phone sized device. It might be that it's going from the phone to the laptop. The main thing is that it can capture as much as possible in the environment and convert it in as close to real time as possible.

40. **As far as the translation capability in the OCR device it was mentioned in the requirements, for example a tourist utilizing binoculars. Is there a set focus on an environment where the translation capabilities will be used?**

There might be places where we'd want to use binoculars and that's not considered polite or good etiquette, so they might use a different form factor. Whatever the OCR form factor ends up being, whether it's a telescopic lens or button camera, we realize that there are going to be parameters and constraints on that such as low visibility and how far of a distance it can go, so I wouldn't worry too much about that.

41. **A lot of translation services rely on cloud services to provide that analysis and data back or host a database that would be facilitating. A host of analyzed data, if it's capable to create a virtualization capability on the OCR device and host the entire database, as well as an AI system that would normally be hosted in the cloud, is that something that you guys would be interested in?**
    We'd very much like to look at something like that.

42. **Based on the current form factor that USSOCOM, JTS and DTRA use are they utilizing the Samsung S21 with 16 gigabytes of physical RAM? And is it possible to test the OCR software applications on those devices that they currently use? We're looking at just strictly an agnostic software solution that can be used on any of those devices: Is this something we can test prior to giving a demonstration or when we go past TRL 7?**
    Yes is the first answer, we are using the best S 20 ones we have as well as the best tensor books that I think are being made right now. But that's just one form factor. Other form factors that we're using are S 21 and note 10, so we're still looking at all those. I think the answer is that you would probably have to acquire your own device to work with. But if we take a look at your technology, we'll certainly test it on those devices and give you feedback.