**Transcript**

March 21, 2025, 12:51PM

 **Amanda Green** 0:03
Good morning, everyone.
We are about two minutes until the top of the hour.
We are waiting for our principal stakeholder to pop into the meeting, so as soon as he arrives, we will get started.
Just a reminder to please keep yourself muted.
Cameras are turned off. This session will be recorded.
We will post the recording and transcript later on our website.
Looks like Mr. Mollenkopf is just popping into the meeting now, so as soon as he hops in, we will greet him and get started.
Good morning, Mr. Mollenkopf.

 **Al Mollenkopf** 15:33
Hey, Amanda. Good morning.
I'm working through a little bit of a technical issue on my end and I think that we're gonna be good enough to go.
Can you hear me OK over.

 **Cox, Jacob** 15:58
I can hear you fine, Mr. Mollenkopf.

 **Al Mollenkopf** 16:01
OK. All right, good.
Just trying to make sure everybody can hear.
I've got my teams, is acting really glitchy, so I apologize for that. If something happens to my teams, I'll jump out and jump right back in as quickly as I can.
I got the IT guys here trying to help me figure this one out.
I appreciate everybody taking time out to.
Talk about Ghost crew and kind of the big idea around where we are.
I have a long list of questions that I'm going to read through, but before I start reading through the questions, I'm gonna read some notes that I made to try and help you all understand what the objectives are and just make some things a little bit clear because I want everybody to kind of understand where we're headed.
I'm going to start right here with #1. The big idea behind Ghost crew is to accelerate Red team operators experience accumulation using AI assisted decision making and simulated attack operations built into the tool chain.
The outcome is we're going to have human machine teams able to perform red team activities at above expert human levels.

 **Amanda Green** 17:25
It looks like we are having some technical difficulties getting audio on our end.
I have our IT specialist in here. Again, apologizing for starting late.

 **Al Mollenkopf** 17:46
Hey, Amanda, I'm here, and I think we're doing OK.

 **Sandefer, Jesse [USA]** 17:52
I'm hearing Mr. Mollenkopf just fine.

 **Cindi Brothers** 17:58
I hear him as well.

 **Al Mollenkopf** 18:01
Y'all let me know if you need a pause or stop, and I think somebody's recording the session, so we should be able to transcribe some notes out as needed.
So, keeping the focus on what Ghost crew is trying to do here, I want to try to abstract some of this out to bring it up to the top before I start answering specific questions.
So we are or the focus is on assisting humans by rendering and interpreting complex data such as tool results and recommending courses of action and by understanding the most likely ramifications of a selected course of action.
So capturing live event data during operations so it can be centralized and used for other to inform other operations is really important.
Regarding the course of action outcomes, where tool simulation is expected, we talked about this during the industry day, but the big idea behind the tool simulation is it could be a range of capabilities. For example, you could use a virtual you know.
Set of infrastructure like VM Ware to go off and execute the simulations.
Or you could do semantic simulation where the results are derived from analysis of numerous previous results, or even hybrid where you use both of those. Those functions and formats.
So.
OK.
So those are those are a couple of key points there.
There was a lot of questions around the tool simulation.
So, you know, I don't want folks to get choked out on that.
It would be great if we could run off and virtually simulate every single throwing of a tool at A at a particular target to see what, given the situation that might be too costly in terms of trying to get things done.
The semantic simulation aspect where results are derived from analysis of previous results of similar situations might be a better way to go.
It just depends on what the what your particular implementation.
What would come up with?
Regarding the form factor, we've had a lot of questions about form factor.
So I just wanted to kind of break it down. You know, the big idea is, you know, we're probably looking at.
A range of cost here of about 500 K and the expectation of that would be to have two ghost crew prototype clients built as desktop PCs.
And.
The expectation there is they would be delivered with whatever ghost crew software stack.
That that needs to run.
On those client systems.
And then there's also an expectation of a centralized lessons learned repository or a centralized function.
System that that would take the lesson.
Learned data from those two client systems.
Reformat it into, perhaps using the reinforcement learning or some type of system where you're using rag or something to aggregate the lessons learned from the OPS periodically, like a monthly on a monthly basis.
And then the server component would be the aggregation point to redistribute to the Ghost crew clients. The prototype client systems.
And that server doesn't have to be super sophisticated in order to do that.
But that's kind of what we're looking at here for the prototype delivery is to client systems in the form factor of a desktop computer and then a small server computer that performs.
Forms the aggregation functions.
An appropriate timeline. I also wanted to touch on that 'cause. There was a ton of questions about that.
So we understand that AI is a bit of a moving target in this range and so the appropriate timeline that we're thinking and it's going to be largely dependent on the concepts and methodologies that y'all present. But the big idea is to have something built within 90.
To 180 days of contract award, which we expect to be in in the May time frame.
And remember, most of all, Ghost crew is not an AI auto poning system. Ghost crew is about.
Interpreting.
Event data and enabling humans to select courses of action that helps them to gain experience but also makes them, you know, select better at over time and helps their functions.
To be more effective as a pentester, red teamer, whether they're doing the thread emulation mission or whether they're doing.
A common pin test operation.
So with that as kind of the precept, I'm going to go through some of these questions and then I'm not sure I have answers to all of the questions.
So we may have to have another question and answer session to make sure that we're answering as many of the key questions as possible.
So first question is the objective of ghost crew to provide AI driven operational red team capabilities for general IT and cyber IT systems?
Yes, the focus of the prototype is commodity. IT systems, router, switches, windows and Linux computers, that sort of thing.
But again, it's not an auto pone system, it's.
Assisting humans and making courses of action decisions that helps them to avoid mistakes while they accumulate more experience on a given pin test or threat emulation operation.
And when I say gain more experience, if you're a pen tester and you, you execute a pen test on a network, you're only going to get the experience from using which tools you selected.
But using Ghost crew, you'll gain even more experience because you'll be able to simulate courses of action that could be virtually executed offline or or using semantic, you know, analysis and then.
Then the operator will learn oh, if I execute course of action, Bi will have to go make cleanup logs over at this location. I'll have to take these other follow on actions.
That's a really complex thing for me to do, so I don't want to actually take that particular course of action just yet.
So gaining experience helping the human understand which decisions to make is is really the focus.
Second question, our proposals focused exclusively on AI driven Red teaming of AI assets, machine learning models. OK, so know the focus is not red teaming of AI assets, it's assisting humans in red teaming a general IT network for the prototype.
So think common router switches, Windows computers, Linux computers, maybe a server.
To our numerous servers, it just depends.
But we're not gonna the focus of the prototype is not like an OT network or some specialized network that's gonna require serial.
Supergranular access to things that that we're just trying to get this prototype off the ground so that we can start iterating on it and make it multiple versions for multiple things, multiple, multiple different use cases.
Next question, is there an expected TRL level?
That we are proposing or is the expectation for this to be a novel technology?
Umm.
So really the expectation is to build from scratch.
When either you leveraging something you already have.
Built from scratch after award is probably the acceptable direction. Will go in and like I mentioned before, upon award we're thinking 98 to 180 days to get the prototype built.
We are interested in getting as much into the prototype as possible so that we can start iterating to quickly add features and capability 'cause. We know that the there's a lot of moving targets in the space that we're talking about today.
Next question, can you share any details about appropriate timeline and cost expectation?
So I've hit on this a couple of times.
An appropriate timeline is to have is to have the prototype systems built within 92180 days of award.
Next question, an appropriate timeline depends on the maturity of the proposal.
Yes, the exact timelines would be a part of the contract and we understand getting from concept to prototype will take time.
We are interested in getting as much into the prototype as possible within 92180 days of contract award with options to extend the contract to add additional features and capabilities. We expect to make this a longer term.
Effort. If it's provided, we can kind of prove it out.
Technology wise, with there's a lot of interest in this capability from Cybercom and others.
Next question, can you provide details on the available historic data sources, database log types, document stores relevant to this course of action?
Please specify the data formats rough total size, EG OK so I can out at this time tell you that we're going to have any historic Red team data that's going to be available to you because a lot of our data is CUI and I don't really think that.
It contains granular enough data.
To provide you with what I think you're looking for there.
So I will double check that, but I don't.
I would not expect us to provide you with historic operational data from the Red team operations and Pinterest that we've done up till now.
Next question, can you provide a range of the number of users you anticipate?
Will concurrently use this application.
So again, this goes back to the expected if it's a 500K project that we contract.
Out 500K were we're anticipating to have two Red team.
Or I'm sorry. 2 ghost crew.
Colliant systems that would mean two red team folks.
And then the server component that that would be you know, providing those centralized data aggregation and reinforcement learning services to the two clients.
What types of data?
Next question, what types of data is available to provide contextualized recommendations?
What format we covered that already?
We're not going to be able to provide lessons learned at the non cui level for pen tests at this time.
But I will double check that to make sure we don't have any unclassified PIN test data that we could share.
Next question, are there operational security restrictions that would prevent collection of session data from live exercises or for lab?
Again, this is a.
This is not an exercise tool per se.
It may be used on exercises, but the bottom line is.
It's a it's anticipated that the system will be designed to record everything that it needs in order to.
Either through rag or some type of reinforcement learning, share the lessons learned to other operators for future operations.
So no, there are no risk, no operational security restrictions that would prevent collection of session data from live use.
Next question, should the course of action recommendations be personalized to individual users or tailored to specific role and rank and categories?
Not for the prototype.
Future intent is to have a version of Ghost crew for multiple roles.
But first we have to make the prototype work and gain momentum there.
So when I say multiple roles, we may have a thread emulation role where.
The purpose is to have a much higher tradecraft at OPSEC threshold.
We're we have to be more careful.
And you may have a pen test role where the pen test, there's a lot more acceptable risk.
So, so there may even be other rules where we have a ghost crew client designed specifically for OT environments or for other uses offensively.
Next question is the currently available lessons learned in training data correlated to these ranks and roles?
Yeah. So there are some lessons learned data, but it's not going to be sufficiently granular to be useful and its cui data.
So again, I'll check on that to see what's available to be shared at the non cui level.
And the next question is local cloud deployment of the applications permissible or is on node or on premises hardware only permission permissible?
So the objective is for the initial prototype to desktop form factors. Is the required platform for the two clients and then you know some type of small server to do the data aggregation lessons learned.
And so that it can be redistributed.
Out to the clients as necessary, so you know if I was building this and we did actually build an internal prototype, there's some high speed HP desktops that have dual.
GPUs and you could probably get away with using something like that that has a couple of high speed.
GPUs in it and you know TB.
Of disk storage for the clients and then maybe one for the server as well, and those were around 1010 K each. I believe the last time we ballpark that so.
That that's what we're looking at there.
Next question, are there any specific vulnerability databases or threat intelligence feeds that are currently available and the penetration testing target environment? Or do they need to be integrated into the application?
So Army Cyber Command does have some commercial threat intelligence that would be made available.
I think that not for the prototype we're going to be able to do that.
We probably won't be able to provide that to each one of you.
For the prototype build.
Next question, are there preferred specific vulnerability databases or threat intelligence sources?
Again, that's gonna be up to you to kind of sort out what, what, what feeds your concepts and methodologies best initially for the prototype and then, you know, there could be some government teaming when we get to the contract writing part where we'll say, yeah, we'll throw in.
This commercial threat feed and we'll provide that in addition to what you're providing to do the training.
And that sort of thing.
Next question.
Is there an existing contractor currently providing a similar solution or support?
And I'm going to say the answer to that is not that I'm aware of. I don't.
I don't know of anyone that's providing an AI assisted pen testing platform that that helps our red teamers and thread emulators with you know making decisions on the fly and providing them course of action and.
You know, recommendations.
Next question, estimated budget range. Can you provide a rough budget range or expected cost structure for this effort?
The estimated budget is anywhere from.
One to $500,000, that's the. That's the limit of our prototype.
So for 500 K that that goes back into the form factor and the objectives.
There is direct, you know, I'm. I'm telling you, the we intend to continue evolving the prototypes.
There's a lot of interest in this capability, but the objective is to get those three systems.
With some.
With the prototype capability at.
Somewhere in the one to 500 K range is, the is the cost.
I'm sorry not the cost, but that's the price range that we're looking for the prototype.
Next question, does the platform need to be deployable in an air gapped environment?
Currently, the plan is to deploy the system as hardware in the loop on pcte.
That's one deployment usage pattern and the second usage pattern is for use on internal networks that are not Internet accessible.
So we could not count on having a connection back to the cloud GPUs.
Because that that runs into Ato issues and things that that we don't want to mess with the prototype.
So yes, all three systems would be, you know, able to work.
And the heart as a hardware in the loop for pcte and for use on internal networks that that are not Internet accessible.
Yeah, I'm sorry.
I keep getting feedback every once in awhile on the on my headset of some kind, so I apologize for these breaks.
Next question is the phase four event. In May, an actual demonstration of a prototype or a virtual presentation of the proposed solution.
The beginning of the release reads as arfp for development.
But some of the assessment criteria cannot be graded until there's an actual product present.
Yes, the assessment criteria for the prototype.
Will depend on the prototype being built so.
Exact timelines are gonna be part of the contract and we understand getting from concept to prototype will take a little bit of time. So we are interested in getting as much into the prototype as possible within 90 to 180 days of contract award of 90 to.
180 days of contract award with options to extend the contract to add additional features and capabilities and capabilities.
We're gonna be flexible in the timelines because we know that there's a lot of moving targets in the AI space and so.
We it's gonna be difficult to get the compute that you need and the form factor that's necessary and that that could be a constraint that we're trying to figure out solutions for. But right now, we're gonna have to go with the form factor that we described earlier.
Next question, what is the TRL level for this prototype? OK.
So ideally the TR level would be like A5 or A6.
The function the technology will function as intended in the real world and can perform most of the intended tasks.
Indicating a significant step towards a final product. However, we understand the domain of IA and AI and assisted capabilities is a rapidly changing domain, so we will rely heavily on the evaluation of concepts and methods.
Ologies when performing the bound select assessment.
The opportunities.
Next question.
The opportunity appears to be targeted at building mathematical model to assess the probability of pen test success and proposing courses of action for doing so.
Is there a need desire to also schedule the resources for this activity as well? So I'm not sure how to answer that question. I hope I've answered that in the.
Form, form and function.
Format.
Discussions earlier so.
Next question, if we join a teaming list, should we submit an individual White paper with our specific capabilities or respond as a team with any partner we find suitable?
So I would recommend that you do both.
But during the assessment, we will consider pairing industry partners where feasible.
So we have done that in the past and yes, we will absolutely do that.
So recommend you do both.
If if that's feasible for you, next question for the prototype.
Would the government please confirm if llm's need to be locally hosted on premises or if they can be outsourced to open AI, anthropic, etc. Again, the local desktop form factor for the initial prototype is where we're headed.
And you know, if there's additional questions on that form factor, the desired output there, we can have a follow up question and answer session and talk about that 'cause. I know that that's a significant limiting factor for all of you.
You.
Next question.
And I didn't write an answer on this one, so let me see output.
Granularity. What level of detail is expected in the AI generated course of action? Recommendations, for example, should the output include specific tool commands and parameters or just high level tactical suggestions?
I would recommend being as granular as possible and explaining why a course of action is selected. So if I run inmap with this.
Particular.
Flag set.
That happens to blue screen. A box then that should be, you know, thrown in there as a warning, yes.
Recommend trying this. It looks like it's a Linux box so.
But yeah.
The answer is be as granular as you can.
Next question, simulated outcome Fidelity.
How detailed should the simulated attack outcomes be?
Are you expecting probabilistic outcomes?
Deterministic predictions based on specific inputs. So the answer to this question is, you know, we'd like them to be as good as possible, and referring back to the course of action outcomes where tool simulation, you know, we talked about tool simulation earlier. This can be a range of.
Capabilities. For example, you could have virtual tool use in a virtual environment that's mocked up, you know VM Ware or some type of.
Similar type or you could use semantic simulation where.
The results are derived from analysis of numerous previous you know, results of a tool being thrown in a similar situation, or some hybrid of both. If you don't have sufficient semantic simulation background and you're not comfortable with that, then you're going to have to throw the tool in.
A offline virtual network and.
Something to that to that nature.
Next question visualization requirements.
What specific visualization formats or standards are preferred for the AI assisted visualization of courses of action in operational states?
So I think we gave you like our prototype.
The UI from the internally you know developed prototype.
Which was it was OK.
It wasn't that great, honestly.
But visualization is focused on assembling the output of tools.
So that the courses of action can be calculated effectively, and so that the operator understands the current logical and physical topology of the targeted network.
So I'll just leave it at that.
Next question risks and assessment metrics.
What specific metrics or scales should the risk assessment feature?
Use to quantify stealth impact, OK.
Let me read that one again.
Risk assessment metrics.
What specific metrics or scales should the risk assessment feature use to quantify stealth, impact and risk for each course of action recommendation?
So it's gonna be up to your particular concept and methodology that you know your implementation, but I recommend keeping them as simple as possible with limited range like one through 4 or one through 5 rather than one to 20 for course of action selection.
A lot of that course of action selection was likely going to depend on the probability of tool success and the impact to post use cleanup and which factors.
Are going to be factored into the risk complexity quotient.
Um are the most complex course of action is not necessarily the best course of action in terms of risk.
Especially where stealth is concerned so.
Next question.
Lesson learned format.
What format or structure is expected?
For the centralized lessons learned repository, should it include tagged metadata, natural language summaries?
So I'm going to say on this you should implement this however you like, ensuring that lessons learned and related data can be shared between Ghost crew client systems using some type of client server architecture.
And that client server architecture may not always be fully connected.
So for example, if I have one of the desktops.
Clients connected to as hardware in the loop on pcte and then the other client is off being used.
To do a PN test of, you know, an AMC network.
Over a specialized network connection, the server may not be able to reach both of those clients to pull their data and to and to push updated populated data.
So you may have to have some type of CD-ROM or DVD, you know package tar G zip package that can be delivered in offline capability.
Unfortunately, that's the state of where we are.
I can't guarantee that there will always be connected and available on the network.
Rick State updates what is the expected latency or refresh rate for quick granular system and operational state updates to be considered effective?
I think this again depends on the operating concept.
So it's going to depend on.
You know the course of action selection, how you've implemented the course of action.
You know decision tree or whatever your methodology is there.
So I'm not gonna.
I'm not gonna be too prescriptive there.
Regarding how often should the clients be updated from the server, I'm thinking like monthly ish is probably a good thing to be able to collect up all of the data from the clients.
Let's say there's eventually 100 of these clients out there and we would like to get to a point where they could all connect back to a server over a network and then be able to.
Pull the data, the event data and then either do some type of rag update to it.
In a specific way that that enables them to.
To have updated lessons learned.
Or if it's reinforcement learning, be able to do that reinforcement learning and then push out the necessary models in order to do that.
That's going to be largely up to you.
Next question thread Intel integration scope.
Which specific threat? Intelligence feeds or formats? Sticks. Taxi. Mitre attack mapping? Should the platform prioritize for integration?
This is going to be up to you, and there's another question later on that talks about which tools.
Should be supported by this system and I'm going to say that that should be up to you if the recommendation there is for you to probably you know, pick a tool.
That and like cobalt strike or metasloid or, you know, Metasploit plus nmap and burp suite it. Whatever your tool set is just kind of standardize on that and focus that that's that'll be OK because long term it will be a tool agnostic type of a platform that will.
It will be able to do learning on all tool types.
Next question.
AI model preferences are you prioritizing a specific type of AI model for the decision making simulation, or is this left to the vendor discretion?
So yeah, totally left to our industry partners, although using government friendly licensed models is certainly something that would be a good consideration if you using a model that like forbids the army to use it or the DoD, then that.
That's probably not gonna workout very well for us, so.
Next question.
Does the AI need to operate entirely in real time live during operations, or is there a tolerance period for offline processing?
So anticipated only needs to be able to operate during, you know, real time operations.
You know it has to be able to recommend courses of action during the time.
Time that the operations is occurring and there could be like a prep a preparatory session before the operation where you could go through courses of action to accelerate the operator's ability to more quickly go through the. That might be something there in a pre OP type.
Of methodology.
Next question Explainability requirements, to what extent is explainability required for AI?
For AI recommendations, for instance, should the AI provide human readable justifications for each course of action or simulated outcome?
So explainability in AI we understand this is a moving target and we some risk for the prototype, but some measure of explainability is definitely gonna be desired in the objective system.
Next question.
Context awareness depth.
How deeply should the AI tailor recommendations to the operator experience level?
Are there predefined experience tiers? Novice, intermediate. Expert.
So.
They're there.
Really shouldn't.
I mean the AI should kind of auto scale to the OP.
And I'm not sure that having a an experienced tier is going to be necessary except for the different roles.
And so what I would say if you're worried about a find experienced here, I would focus on novice.
And.
You know we can iterate on that as we add features and capability to follow on capabilities.
Simulation engine is the AI expected to include a full simulation engine digital twin of the target environment? Or can it rely on precomputed heuristics and historical data?
Yeah, I talked about the simulation stuff earlier, so it's kind of up to you to implement that.
How you want you can have a hybrid approach.
You could do it semantically through learning.
You could do it through having offline range.
So there's a couple of different approaches there.
Next question, training data sources, what type of data will be provided to train the AI historical Red team logs synthetic attack data, so I don't think we're going to have any non cui data that can be shared for the prototype.
Once we get.
To the contract there, there could be something that might be able to be shared.
I'm going to look into that and circle back on that.
I just don't think that we have a lot of data that's of the granular.
Nature that you're gonna want.
To do that next question, initial training expectations is the AI expected to be pre trained with a baseline capability before deployment or should it Start learning from scratch?
During initial operations, I think the capability is going to be expected to work as outlined kind of at delivery at least to a limited you know degree it should.
It's going to be expected to learn during operations.
But it should be delivered to function.
You know the course of action recommendation.
They may not be as effective as they would after two years of learning on the job.
Obviously we get that.
So yeah, there you go.
Reinforcement learning, scope for reinforcement learning component. What reward functions or success criteria should guide the Ai's learning process?
Yeah, so reinforcement learning scope is going to be up to all of you.
How you decide to implement that?
There's a couple of different ways we.
We tried a couple of ways.
Reinforcement learning is it requires a little bit of time and that's why we were thinking, you know, we're going to leave it up to each particular partner to figure out how they want to capture the log events from the clients, how they want to factor that into less.
Learned. And then how they want to redistribute that back out to the clients?
There's a number of ways that could be ranging from super sophisticated that would be really good, but maybe overkill.
All the way back down to, you know, less than optimal.
So we're going to leave that to you all to come up with something innovative.
Next question, target environment, will the AI be trained on standardized it specific OS versions, network topologies?
Or should it be adaptable to unpredictable target systems?
For the for the prototype, we're looking at a general IT environment to be the target environment.
So think common switches, routers, windows and Linux servers and clients computers you know.
Nothing. Nothing sophisticated, no cars, no OT devices, no elevators.
We're not gonna throw any curveballs.
We're just trying to make this thing go and crawl, walk, run.
Lessons learned. Retraining. How frequently should the AI retrain or update its models based on centralized lessons learned?
And new threat Intel.
So the reinforcement learning we are, I think I mentioned this earlier, but we're expecting that to be approximately monthly and it could be more, it could be longer depending on how hard that's going to be to do given the hardware capabilities of the centralized server and.
The clients.
Let's see.
So next question.
Tool integration. Are there specific penetration testing tools?
Metasploit and Map, Burp suite etc.
That platform must integrate or is it tool agnostic compatibility preferred so this you know depends on your concept of employment for the prototype. But if I was recommender here I would say select a set of tools and focus on those tools like.
I don't cobalt strike or you're a metasploit shop or metasloit or cobalt strike.
Plus, nmap and a burp suite. Or you know, just have a fixed set. Focus your AI on that fixed set of tools. And then because if the methodology works, it will scale to work with other tools over time.
And that's what we're trying to get proven here is, is this this, can this can work?
Stealth constraints.
Next question, how should the platform balance stealth versus performance and its connectivity and resource usage on target systems?
Are there specific threats thresholds for acceptable network or system impact?
OK.
So I don't see any specific connectivity thresholds for this type of system prototype.
You know the systems are they're going to be networked into pcte as hardware in the loop, and then the other use cases is going to be connected to a network that has at least a 10 megabit. You know network link and much larger most likely.
It'll probably in the Gigabit range, but.
But if you plan for Pcte hardware in the loop and the 10 megabit network access.
As a minimum, I think you'll be in good shape for the pcte hardware in the loop. If this goes as planned.
And that's a that's an if.
But if this goes as planned, we may take a couple of ghost crews and just take them down to Florida and, you know, install them locally.
At where?
The hardware is so that they can be available so that folks can just remote to them and use them as needed.
And that will certainly reduce any network choke points there.
So that's the that's the trajectory.
I have a quick drink of coffee here.
Hold on one second.
Error recovery scenarios.
What are the most critical failure scenarios?
The platform should recover from EG network disconnection, tool crashes, AI prediction errors.
I think that the threshold for the prototype is going to be pretty robust, so I wouldn't focus.
I wouldn't get choked out too much over if the system crashes, it just needs to be able to be restarted and brought back.
Back to the point from which we can conduct the assessment.
I think that the error concern there would be if it crashes too often during the assessment at the you know that that would be the that would be the challenge.
Next question, operator customization, can operators override or customize AI recommendations and simulated outcomes?
And if so, how should these modifications feedback into the system?
Yes. So operators should be able to override and customize AI recommendations.
Because the AI is not going to execute the functions.
Remember, the AI is only doing recommendations.
So the AI should have a function that says here's 10 courses of action, or your, you know, select your own course of action and then the human operator can say, Yep, I'm gonna. I'm gonna do my own thing.
I'm popping into the command line and here's what I did.
That should be flagged so that the system logs that the operator took a custom recommendation, and so that can be captured. All the related you know associated events could be captured with that and brought back to the server and say hey, in this situation the human.
Did not select any of these ten courses of action.
This other course of action was better, and then however you want to, you know, RAG or CAG or whatever.
Reinforcement learning, whatever's methodology, you're going to use.
For redistributing those lessons learned is kind of up to you.
Multi operator support is the platform expected to support multiple operators simultaneously on a single operation.
If so, how should it handle conflicting decisions or resource allocation?
OK, for the for the prototype, it's just going to be one seat per desktop.
And that operator is going to be operating alone, not in tandem or in as a team.
There we have talked about having multiple ongoing operations in the cloud and a lot of other you know things that we would like to do along the lines of this.
But we're we don't want to get choked out.
On proving the methodology by getting choked out with, you know RMF and getting caught up in at OS and that sort of thing.
So for the prototype, no.
One person per seat.
And going back to the form factor discussion, you know if if the form factor, if your if your solution is able to provide more than two clients that would be great too.
You know, I see this as something that we're going to be able to, you know, discuss during the contract if you are down selected to.
So maybe it's two clients.
Maybe it's five clients?
Maybe you're able to.
Use a laptop form factor.
Ctor which I I'm suspect of, but you there are paths where you can have a very sophisticated like Mac has a really interesting set of GPU capabilities that are very robust and you know can extend computers.
So we're not going to constrain you to what the form factor is, but yeah, the form factor we're looking for is at least two clients and one server.
For the contract.
Assessment.
Next question, prototype maturity.
We talked about that.
You know the contract will be in May. So the down select will down select to at least one obviously and then maybe multiple partners asking partners to work together on something and that'll be the B2B contract that the civic you know lays out for.
Whoever the down selected folks are.
So that.
That answers that question.
Next question, is there a subset of listed capabilities that can be deprioritized or mocked up?
Eg partial threat.
Yes, for your for your consideration.
You know, if you have a concept for employment, you should absolutely mock things up.
We don't expect you by the end of May to have a full prototype built to demonstrate.
We expect you to have like, here's, here's the methodology.
And if you have some capability that you can demonstrate, that would be ideal.
But if you carefully lay out, here's the methodology.
Here's the concept.
Here's what we think we can have within 90 to 120 days of.
Of, you know, contract time.
Next question.
Will the Army provide a simulated or live test network for the May demo?
I mean we will have an environment that you can connect to.
If you come to the facility for the demo.
But if you are down selected, you should be able to, you know, use your own architecture to demonstrate over teams.
You know, you know, it's basically a demo of how you are conceptually going to build the capability and so.
There won't be like a requirement for us to test your capability in our environment for May if that's if that's what you're worried about on that question.
If we if we do contract, you know, with you to build the.
Three, you know capabilities and the contract says within 120 days you provide this prototype set and we take that prototype set into that.
When we would take it into our prototyping environment and test it.
So yes, that's the intent.
For the proof of concept, is sufficient to simulate outcomes using precomputed data or basic heuristics, or must it be dynamic simulation against a virtual range?
Now absolutely you can.
You can simulate what you need to simulate and you can use semantic.
We talked about this several different times, but yeah, there's a lot of wiggle room on. You know where we're going to go with the prototype.
Will Red team operators be actively testing the prototype during the demo?
No, the demo is just you're explaining to us you know, so how this works is we're going to read your RFPs, your responses via white papers, and we're going to read everybody's white paper that says, here's how. Here's how we would do it. Here's the.
Technologies we would apply and at the end of this length of time, we would be able to provide you with this capability.
And that's what would get down selected to and contracted for.
So that's really where we're headed. And then let's say it takes you 60 days or 90 days or 120 days to actually build that.
That's we would put that in the contract and then at the end of 120 days or whatever that length of time is that's in the contract, we would take the prototypes at that point, put them in our Lab. And then do the assessment testing to make sure that they meet.
The requirements and if they do, we would probably continue.
To extend that contract.
To add more features and continue the development of Ghost crew.
So that's the general plan. You know, a lot of things can happen, but that's the general plan.
And we are getting close to 10:00 here.
Hold on one second.
Let me see if I have time to continue into these questions.
OK, let me let me just answer a couple more of these questions.
I know where the time is getting short.
I really appreciate you all taking time out to discuss this and chat. Seems like it's turned off for this meeting so.
Hoping everybody can still hear me.

 **CFIC Smartboard** 1:08:05
Yes, Sir. We can still hear you.

 **Al Mollenkopf** 1:08:09
OK, OK, awesome.
I'm trying to skim through the rest of these questions.
I think I've answered a lot of these questions now.
Let's see.
Let's see.
What level of stealth is required for the proofs of concepts own footprint to avoid detection in a demo environment?
I'm not sure what the so the answer that I wrote on that is, you know it's up to, it's up to you to implement the tradecraft that you think would be expected in a general IT network.
We understand Edr will catch a lot of things, so we're probably not going to be super aggressive with.
You know, with that sort of thing, 'cause we're not really so much focused on is the tradecraft 100% exact.
We're trying to focus on.
Can we build a system that the AI will assist a human in selecting the right courses of action, helping them understand the dimension of what the current situation is, and then select the right course of action based on the in game that experience from multiple different courses of?
Action through simulation.
So it's kind of about building the person we expect, you know, honing in on the tradecraft that will come with time with the AI.
But we don't expect.
The AI to be a tradecraft expert right out of the gate.
I guess that's the answer.
That's what I was trying to say in my answer.
Next question, should the prototype be prepared to handle dynamic or unknown target environments during the demo?
No. So really focus your demo or your concept, your methodology.
If you're down selected to make sure you focus on a general generic it network, not like an OT network or some, you know, serial.
You know different type of network we want to avoid that if if possible.
Will the Army provide specific test networks or scenarios for validating the prototype performance during the May 2025 demonstration?
So if you're down selected to and you get selected to do a demonstration, it's really.
Your it's your opportunity to show what you would build.
You don't have to have your prototype completely built by May.
We don't.
We don't expect that.
We expect there to be a contract executed in May with the down selected individual or multiple teams that would then you know up to 180 days although.
90 days is ideal.
The time.
What we would.
Hammer out the time frame in the contract but yeah, so and at the end of that period, that's when we would take the prototypes and test them in our environment.
Next question, are there specific security standards that the POC must meet for data storage and transmission during the demo?
So that's going to be up to you.
We're not putting any.
We're not saying that the host have to be staged or anything like that.
I I think that as much stigging as you can do is probably ideal.
So it's not the wild Wild West and we're not.
We're not going to.
You know, we're not going to be able to implement something that's just.
Not has zero security at all, but we're not going to make it. You know you're not going to have to apply NIST 800 one 71 features to it right out of the gate.
Next question.
I answered that I answered that.
Are there plans to extend Ghost crew to larger, more complex networks in the future that the POC should account for conceptually?
There are great plans for ghost crew.
Ghost Crew has a huge upside to be focused not just on threat emulation and red teaming, but also on other offensive you know.
Concepts.
I'll just leave it at that.
So there may be multiple different versions of Ghost crew.
And each one learns on its own tool set and its own experiences.
And these might be different environments that they work in. If we can prove out that that the methodology works.
OK, that is.
That is pretty much it.
So what I'm gonna do is there's a few questions left here, but it looks like I've covered.
I've covered most of them.
And I'll tell you what, if I haven't answered your question, I absolutely want to.
So please do me a favor and reach out to the CFIC team and if you have questions that I didn't answer, we may have to do another session.
Just let me know and I will circle back at a later date, but I hope that helps you all understand kind of where we're going and what we're trying to do.
And answers as much of our questions as we can and in earnest.
And.
I hope everybody has a great rest of your Friday and a wonderful weekend.
And with that, Miss Green, I'm gonna turn it back over to you over.

 **CFIC Smartboard** 1:13:54
Thank you, Mr. Mollenkopf.
Apologies for our technical difficulties at the beginning of the session.
We seem to have gotten those worked out and we appreciate you hopping on this call.
Just a reminder to everyone who is planning to submit that deadline is April 9th. To get those whitepapers submitted.
So I know we still have a couple of weeks. Also, if anyone is getting into Vulcan
and they have any challenges they can reach out to me directly and I will be happy to point them in the right direction to make sure they are in Vulcan so they can submit. Anything else before we disconnect, sir?

 **Al Mollenkopf** 1:14:41
Nope. I just thank everybody for showing up.
We think there's a great capability here if we can prove it.
The concepts works. I think there's a great upside for everyone, so thank you all. Talk to you again soon out here.

stopped transcription