OFFICIAL – COMMERCIAL IN CONFIDENCE

Document Details: Clarification Q&A in response to the call for proposals
Challenge: AI / Novel Technology Red Agent Penetration Testing
Deadline for questions: Wednesday 30th October 2024

#	Question	Answer
1	What level of maturity is the Authority aiming for with regards to the Red Agents being evaluated? - Within Workstream 2, the nature of the evaluation environment could range from a more abstract OpenAI Gym type environment for low maturity Red Agent models that can only currently really function in simulated/abstract environments, through to more realistic fully virtualised/emulated environment with 'real' nodes and services for higher maturity Red Agent models than can work in 'real' environments. This would have a significant impact on the resources, time and effort that would be required to establish the evaluation environment.	We are open to testing a range of tools with varying levels of maturity, however a good outcome from the project would be to evaluate in an environment that closely represents the real world, e.g. VMs on servers or in the cloud etc. Think creatively to save costs by simplifying the environments needed or where existing environments can be re-used. The purpose of this work is to assess products in a consistent way.
2	As part of Workstream 2, is it expected that the Solution Provider may be required to modify or enhance any of the Red Agents being evaluated? - e.g. to add the capability to execute specific penetration testing actions such that all Agents have the same suite of actions/tools available, and thus the evaluation can truly focus on their ability to understand and decide which action to use and when, and not just about what actions they currently have 'baked in')	We envisage that the Red Agent tool supplier(s) may need to tailor their solution to be able to use the specific actions and tools needed to get through the evaluation. We would not expect the Solution Provider for WS2 (i.e. the provider of the test environment) to modify or enhance the Red Agent tools directly. The level of tailoring and effort required will be of interest to the overall evaluation of a solution. There is a scenario where the agent to reports back while it's in motion for the human agent to then reconfigure and redeploy if new information highlights any particular paths of interest.

3	Would you be looking for the evaluation environment to reflect a specific type of environment (potentially reflective of a specific use case / scenario), and evaluate the Red Agents identified from Workstream 1 'as-is', or would you be looking for the design, architecture and make-up of the evaluation environment to reflect the capabilities of the identified Red Agents? - Essentially we are trying to understand if the nature of the Evaluation Environment will be an easily defined 'constant' and form a 'line in the sand' for the selection and evaluation of the Red Agents, of if the nature of the environment will be a to be determine variable based on the capabilities/features of the to be identified Red Agents.	The environment should be static, and the red agent should adapt to it. We would look to work with the solution provider to deploy a baselined/constant environment, within which there will be a series of 'easy', 'medium' and 'hard' challenges for the Red Agent.
4	With regard to Workstream 1, are there currently defined and captured use cases / scenarios for evaluating the Red Agents against, or would capturing and understanding them be part of the delivery of Workstream 1? - Whether 'Red Agent A' is better than 'Red Agent B' will typically depend on what you want to do with them, and thus the use cases / scenarios.	These would be captured during the delivery of Workstream 1.
5	Given the nature of this challenge, with the two workstreams, one identifying Red Agents for evaluation following a literature review and 'paper evaluation', and the other undertaking a practical evaluation of a set of to be identified and agreed Red Agents, within a to be constructed environment, is the Authority accepting that the delivery will be an Agile delivery, working towards a prioritised backlog and managing a trade-space? - e.g. precisely how many Red Agents might be physically evaluated will likely depend on the nature/maturity of each, and therefore the corresponding level of effort associated for establishing the necessary environment and undertaking each respective evaluation.	Agreed.

6	What would the call-off involve?	The call off allowance is to cover reasonable third-party charges for selected Red Agent tools to be prepared, deployed and tested in the Workstream 2 test environment. The call off may also be used to fund collaborative development of the most promising Red Agent tools during the project - if applicable. The Authority would make the call-off fund available to the Workstream 2 provider, so that the provider can subsequently pass call-off costs to the Red Agent suppliers through their own contracting mechanism.
7	Do bidders need to consider the creation of the "Real-World" equivalent environment as part of the proposal, or is this provided by HMGCC?	Bidders need to consider the creation (or re-use) of a 'real world' equivalent environment as part of the proposal.
8	Will details of the SoC (SIEM tools used) be shared with those looking to carry out WS2? For example if using ELK with specific IDS / IPS this can be an a consideration for the test bed	We are not looking to replicate a specific environment or SOC so there are no constraints in this regard - just a suitable environment to evaluate the real-world applicability and capability of emerging red agent tools.
9	When identifying red agent tools, are there a set of tools which are in, or out of scope? For example, are tools which perform intelligent reconnaissance more interesting than tools which can exploit vulnerabilities?	We are interested in the full range of functions.
10	Would a gaming laptop or intel NUC be considered as a commercially available laptop/device	Yes.
11	Can you clarify the £60k max per single organisation. Is this per workstream or for the whole project?	Up to £60k per workstream. For example, a single supplier who was successful for both workstreams could be awarded up to £120k (plus call-off).
12	Are there specific threats that are out of scope for project? I.e. are red agents with potential for social engineering within scope for WS1?	No - we are interested in the full range of functions.
13	Can the red agent utilise existing penetration testing tools such as Nmap etc?	Yes.

14	For WS2 what is the expected level of innovation - For example. We assume there are existing cyber test beds that could be utilised but have been considered unsuitable for task? There was also mention of the test bed working offline and further mention of cloud deployment. Are we aiming for a private cloud instance?	We are not anticipating a high level of innovation for WS2 environment provision, more that the test environment is appropriate for the type of technology we are evaluating. Innovation will be more of a measure for Red Agent tools that are deployed during the project. A private cloud instance would be considered as an option for the WS2 test environment.
15	For WS2 / a test bed is there then an expectation of Human in the loop - To facilitate social engineering?	We don't anticipate this as a primary focus at this stage, however we would not preclude this.
16	Can you please explain criteria for the assessment framework that is needed for workstream1?	The basis of the WS1 assessment could amongst other considerations evaluate the current TRL of each identified Red Agent tool, credibility, potential TRL within 12 months, and ability for it to execute a real-world scenario. Specific criteria would be collaboratively developed by combining Authority insights and supplier expertise during the project.
17	If a consortium of partners applied does the 60k get split up?	Each workstream is up to a maximum of £60k funding.
18	For the application itself - Does the 6 page limit apply to say a separate attached document or include the mandatory fields in application portals (such as UKRI)?	Wording submitted in the mandatory fields within the portals will contribute to the overall page limit.
19	In a situation whereby we would want to apply to deliver both workstreams, should we submit separate applications for both workstreams, or a combined proposal addressing how we would deliver both workstreams?	Individual proposals for each workstream or a combined application would be accepted, however if combined please make it clear which elements of the proposal relate to which workstream so that these can be considered on their own merit.
20	Do you have suppliers for the red agent already or are you open to applicants building one during this challenge?	Our aim is for the Workstream 1 (landscape mapping) provider to identify and assess the capabilities of Red Agent tools, so that these may be potentially considered for further practical testing and development in Workstream 2. No funds would be provided for the development of red agent tools prior to their assessment in Workstream 1.

21	Does the royalty fee license just apply to the project IP developed during this project and not any background IP brought by the supplier?	Pursuant to Clause 10.3 of the Co-Creation Terms and Conditions (V5) for this Challenge, for the duration of the Contract the Supplier (Solution Provider) shall grant Cranfield a non-exclusive, royalty free licence to use both its Background IP and Project IP if Cranfield has a genuine need to use it in order to perform its duties under the Contract and in meeting the needs of the Funding Party, but for no other purpose. In relation to licences upon completion (see Clause 10.4), this does not solely apply to the Project IP – a non-exclusive, world-wide, non-transferable licence is also to be granted to the Funding Party for any of the Supplier's Background IP used in the provision of the Services and which is wholly and necessarily required for sole use of the Deliverables. Please contact Co-Creation@cranfield.ac.uk for any further clarification required here.
22	Does the Authority in this case refer to HMGCC?	Yes.
23	Will we be required to provide a testing/assessment environment or will they provide one for us?	The successful WS2 supplier will provide the test environment. Red Agent tool providers that are considering bidding for WS2 (provision of the test environment) will need to demonstrate appropriate governance/controls to ensure impartiality and fairness in the evaluation process. For clarity, Red Agent tool suppliers who are not directly bidding for WS1 or WS2 at this stage can express an interest to the Authority (cocreation@hmgcc.gov.uk) for their tool to be considered for evaluation as part of the WS1 (landscape mapping) and WS2 (practical testing) workstream.

24	Is there a focus or priority on a particular vulnerability class, such as SQLi or requirement to test against many?	No - there will be a need to decide what vulnerability classes to use in the environment so that the Red Agents can be prepared appropriately. It is anticipated that the current maturity of red agents is that they may have limited techniques. The purpose of the environment is to test the red agent, not play guess the vulnerability.
25	Are there any hardware limits given the LLM should be run locally, which requires significant compute?	There are no hardware limits for the WS2 infrastructure environment, however we would not expect any specialist/bespoke compute.
26	Are there any other technology requirements or constraints - I.e. is it expected to be delivered in a containerized environment?	Not at this stage. The successful WS2 provider for the test environment will inform potential Red Agent tool suppliers of any technical constraints prior to the tools being tested.
27	The project scope reads: "HMGCC will provide the supplier of Workstream 2 with additional reasonable call-off costs of up-to £65k (exc. VAT) during the project to support third-party charges for selected Red Agent tools in the test environment." My understanding of the above is that the call-off costs can only cover the licenses/costs of the Red Agent tools and no other licenses for software that will be installed on the test environment, e.g. Windows licenses, or any additional licensed software that could arise upon discussions with the Co-Creation team. Could you please confirm if the above is valid?	Correct. The WS2 environment should include core environmental components expected in a real-world environment (e.g. windows OS licenses). Any specialist tools, configuration or license costs required by a specific Red Agent tool to then operate in that environment could be covered by the call-off.