# AI ATAC Prize Challenge Q&A

All questions due by August 30th, 2019 at 1700 EDT

## Non-Technical

**NT-Q1:** I*s the AI ATAC Challenge open to federal government employees?*

**NT-A1:** No, per the Terms and Conditions posted on the Challenge.gov website, Federal Government employees, PMW 130 support contractors and their employees, and Oak Ridge National Laboratory (ORNL) employees are not eligible to participate in this Challenge.

**NT-Q2:** *When is the last day to submit questions?*

**NT-A2:** All questions must be submitted by August 30th, 2019 at 1700 (5:00 PM) EDT.

**NT-Q3:** *The AI ATAC posting on the Challenge.gov website states, "The Navy's Information Assurance and Cybersecurity Program Office (PMW 130) seeks to automate the Security Operations Center (SOC) using artificial intelligence and machine learning (AI/ML) beginning with the endpoint." Does this imply there will be future AI ATAC Prize Challenges that cover other cybersecurity tools that utilize AI/ML?*

**NT-A3:** NAVWAR and PEO C4I may conduct additional AI ATAC Prize Challenges in the near future that cover other cybersecurity tools that utilize AI/ML, such as advanced network threat detection, behavioral analytics, and SOC automation. Participation in this challenge will not have any bearing on future challenges.

**NT-Q4:** *Will challenge winners be advertised to more than the Challenge.gov website?*

**NT-A4:** Winners of the challenge may be advertised beyond the Challenge.gov website, potentially including but not limited to: the Navy site and/or blog, DOD/government-focused media outlets, commercially focused media outlets, social media, etc.

**NT-Q5:** *Will the government be doing the testing or are we able to provide an individual(s) to conduct the testing of our tool?*

**NT-A5:** The government will conduct all testing and evaluation of the submitted tools.

**NT-Q6:** *Should it be assumed that if we'd like to participate in the testing, we will have the chance? In other words, should go ahead and kickoff a shipment of our equipment to Oak Ridge (since it is mentioned that the equipment will need to be there for the start of testing NLT 30 Sep)?*

**NT-A6:** All whitepapers and tools must be submitted before September 30, 2019. Participants will not be invited to the test event.

**NT-Q7:** *If it is a requirement to send an individual to conduct the testing, will it be acceptable to have multiple of the same appliances in the testing mix, if there are partners desiring to utilize our technology for this challenge?*

**NT-A7:** Individuals cannot be sent to conduct the testing or participate in the challenge; however, the participant should ensure a technical point of contact (POC) can be reached promptly during the challenge for questions or assistance with the technology. **If the technology submitted for evaluation CANNOT be installed and configured correctly according to the challenge guidelines and the technical POC cannot be reached and assist to enable evaluation of the technology, the participant's entry will be disqualified. The evaluators will make their best effort to get every technology correctly configured and running according to the documentation provided by the participant.**

**NT-Q8:** *Will you be sharing the results of the testing with the participants?*

**NT-A8:** The government hopes to share test results with the participants but it depends the number of submissions received and resources available. Specific results from one participant will not be shared with another.

**NT-Q9:** *What happens after an award for the first or second prize is made? Is the goal of this challenge to push this to deployment or small-scale deployment within the Navy?*

**NT-A9:** As indicated under the Prizes section of the AIATAC challenge on the Challenge.gov website, NAWARSYSCOM may award, pursuant to Title 10 U.S.C. § 2371b, a follow-on production contract or transaction to one or more participants who successfully demonstrated an effective AI/ML approach under this Challenge. This Challenge, however, does not in any way obligate NAVWARSYSCOM to procure any of the items within the scope of this Challenge from the winners.

# Technical

**T-Q1:** *Will the user interfaces of the technologies or situational awareness through a management console be a part of the Challenge evaluation?*

**T-A1:** No, the participating technologies must provide programmatically collectable classification results (i.e. local log file or syslog message) with no user interaction, though a user interface as part of the overall technology is still allowed (e.g. many commercial products provide both a management console, endpoint UI for the tool, and programmatically accessible logs via syslog, WMI, ELK stack, etc.).

**T-Q2:** *If a proposed solution is only compatible with Windows malware, or Linux malware, and not both, are we still eligible?*

**T-A2:** Yes, the solution is still eligible; however, said solution will only be evaluated against what the technology states compatibility with. To that end, since the competition is geared to general performance against a variety of samples, the score for the proposed solution may not be competitive with solutions that classify the broadest set of sample malware.

**T-Q3:** *Are certain file types more prevalent versus others in the testing corpus of benign and malicious files?*

**T-A3:** There is a balance of Windows versus Linux binaries (e.g. PE and ELF), document formats (e.g. PDF, DOCX, ODT, etc.), non-executable files (e.g. HTML, TXT, etc.), compressed formats (e.g. ZIP, 7z, RAR, etc.).

**T-Q4:** *Will malware samples be transformed to attempt to evade detection?*

**T-A4:** Provided solutions should have no expectations for what kinds of transformations, if any, will be applied to samples from the testing corpus. For example, malware samples may be packed, encrypted, compressed, or adversarially perturbed to attempt to test the robustness of the technologies being evaluated.

**T-Q5:** *What if we have an AI/ML classification technique that works on Java heap execution? I realize that this challenge includes other memory manage systems. But - if the technique works well for Java applications it could be applied to other systems as well. Just wondering how inclusive the solution needs to be - mostly as time is short.*

**T-A5:** The Challenge scoring framework was designed to reward the broadest coverage of malware file types, and therefore endpoint technologies that are more generally applicable will score higher. Technologies that are more focused, such as a classifier for execution within the

Java heap, are unlikely to score well with respect to the Challenge given the diversity of file types in the test corpus.
However, we still encourage the submission of these more focused technologies. The broader intent of this Challenge for NAVWAR is to find new and emerging technologies in AI/ML for endpoint defense that have high potential for operational impact. More specific analytic technologies may attract acquisition interest from NAVWAR should they add value for specific malware families or exploitation approaches where the more general technologies are weak.

**T-Q6:** *How many endpoints/hosts will be utilized during the testing?*

**T-A6:** 1 ephemeral (started and then shutdown after test) endpoint per sample tested, per endpoint technology, per compatible operating system.

**T-Q7:** *Will there be any additional security software running on the hosts that will need to be whitelisted (DISA HBSS, Windows Defender, etc.)?*

**T-A7:** No, all existing OS-bundled AV will be disabled, along with any firewalls, etc. No other AV software will be installed (e.g. HBSS). Additionally, no endpoint detect and response (EDR) or other policy-based software **not included within the evaluated technology** will be deployed.

**T-Q8:** *Is a hardware appliance, virtual appliance, or Cloud technology desired for this testing?*

**T-A8:** Cloud technologies are not eligible if no on-premise hardware or virtual appliance exists that can fulfill the cloud's role for the endpoint tool. In other words, if the endpoint agent requires a management appliance (or even does analysis in an appliance that the endpoint sends files to), then that appliance must be provided as a virtual machine image or as a hardware appliance with appropriate documentation for installation and configuration as needed. The hardware appliance or virtual appliance **must** be included in the submission of the endpoint agent, if an external appliance exists. Unfiltered Local network connectivity between the appliance and the endpoint agents will be allowed.

**T-Q9:** *What are plans for connectivity during the testing? Internet connectivity? Local connectivity?*

**T-A9:** Local network connectivity between the agent and the rest of the testbed (which may include running appliances for the agent being evaluated) will be allowed with no filters on ports, traffic, etc. **However,** no Internet connection will be present for this challenge, but for the dynamic analysis portion (if static analysis does not detect the file sample), then an Internet simulator tool such as https://www.inetsim.org/ will be used.

**T-Q10:** *If shipped directly to ORNL, will the participant be able to update to the latest code base and content from the cloud?*

**T-A10:** We will not allow the agents to update before or during testing; the latest versions of the technology models should be included at the time of submission.

**T-Q11:** *I have a submission question about the tool. The requirements say:*

*"Software for endpoint agents and/or management appliances, or hardware for management appliances must be shipped by trackable, non-postal delivery (FedEx, UPS, DHL, etc.) and received no later than 30 September 2019..."*

*My software tool is cloud based. Do I still need to ship it? Or, rather, would giving you account access suffice?*

**T-A11:** The official challenge guidelines, under Judging Criteria->Scope state that "Technologies must operate on-premises, whether solely at the endpoint or in coordination with a local network-level appliance or VM that emulates a cloud analytic capability." In order to best emulate the US Navy operational environment, which often lacks Internet connectivity, only technologies that can be deployed locally by the evaluators will be accepted. If you can run your cloud-based solution either on the endpoint along with the agent OR as a container, set of software, or deployment via a VM or on a hardware appliance, then such software or hardware can be shipped to ORNL and will then be included in the challenge.

**T-Q12:** *Will teams be able to run the Test Evaluation Process and/or Scoring prior to final submission? In other words, is there a remote server that team solutions can be run on or will identical testing software to that used for scoring be made available to teams?*

**T-A12:** The test evaluation process and scoring will be done exactly as specified in the official Challenge guidelines at Challenge.gov. To that end, it is possible to replicate the test evaluation process before or after submitting your technology and whitepaper with your own dataset. The dataset used for the official challenge scoring will not be released to challenge participants due to the nature of the malware in the dataset.

**T-Q13:** *Will the dataset used for the official challenge scoring and final rankings be made public and/or shared with the challenge participants?*

**T-A13:** No, the dataset will not be shared with the challenge participants; however, participants can expect the dataset to be a roughly balanced dataset of both malicious and benign executable files in the PE and ELF formats, along with document or language-specific files like JARs, MS Office documents, Open Office documents, PDFs, compressed or encoded formats (e.g. ZIP, 7z, RAR, TAR), and static files (e.g. HTML, TXT, JS, etc.). Some of the malicious samples are purpose-built to contain malicious functionality and have not been seen "in the wild". Because of this, we cannot release them as they would inevitably end up with online services such as VirusTotal and other vendors containing signatures to these samples.

**T-Q14:** *How can I build a dataset that roughly mirrors the challenge dataset?*

**T-A14:** A dataset closely modeling the challenge dataset could be constructed by amassing malware from sites such as VirusShare, VirusTotal, or similar repositories. To test the generalization of a proposed tool, ideally some malware samples would be purpose-built to contain malicious and working functionality. For benign files, DigitalCorpora has a large set of benign files collected over many years. To supplement the benign files with benign executables, one could collect installed executables from existing Windows or Linux hosts.