Trusted AI Challenge Series

Air Force Research Laboratory (AFRL), State University of New York (SUNY), IBM, NYSTEC, National Security Innovation Network (NSIN)

Presented by Innovare Advancement Center

Request for White Papers
Deadline: June 4, 2021, 5:00pm (EST)

I.   Overview

Innovare Advancement Center is a globally connected, world-class facility acting as a lightning rod for top scientific, engineering, and entrepreneurial talent to leverage highly specialized resources and accelerate both expertise and innovation in critical research areas, including artificial intelligence/machine learning, cybersecurity, and quantum information science. As part of Innovare Advancement Center's outreach, it is announcing the Trusted Artificial Intelligence (TAI) Challenge Series. Interested participants will have an opportunity to submit a two page white paper after the competition is announced virtually on April 29, 2021.

The TAI Challenge Series will cover one of four distinct topic areas:
Topic #1 -   Verification of Autonomous Systems
Topic #2 -   Human-Artificial Intelligence Performance Optimization: Trust and Joint Action for Digital Data Analysis
Topic #3 -   Dynamic Bi-Directional Trust in Human-AI Collaborative Systems
Topic #4 -   Trustworthy AI Certification

Each topic represents critical areas for AFRL and its partners, and the goal of this competition is to help advance the mission to build a magnetic ecosystem in which the world's leading scientific and entrepreneurial talent tackle the greatest challenges to national security and economic competitiveness for the TAI realm. Please see Section IV for topic details and eligibility for academic, small business, and international R&D communities.

II.   Background

This TAI Challenge Series event follows Event 1 of the series "Building the Vision," held Oct 14, 2020 that covered a set of thought-provoking talks and included an interactive panel offering industry, research, and government perspectives, and insights into the critical path requirements for building reliable, robust AI and autonomous systems that can be widely adopted. While current machine learning and AI technologies are focusing on many issues for static data and systems, dynamic systems such as autonomous vehicles, drones, and unmanned aerial vehicles are increasingly being deployed in both civilian and military contexts. Of special interest to this forum are formal methods, protocols, and standard certifications for testing, validation, and certification of trustworthy systems along with the supporting infrastructure and tools. Further, the next generation of technologies will involve evolutionary computing that focuses on the system's ability to learn, prioritize and discount knowledge as it evolves through interaction with people, the environment, and other systems. Through these challenge problems, we hope to uncover novel solutions that move the community closer to addressing these needs, in the context of today's concrete problems.
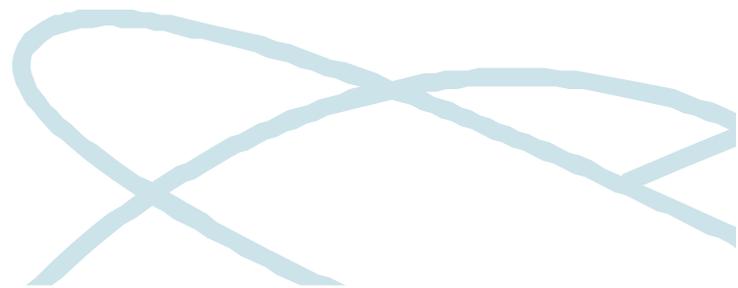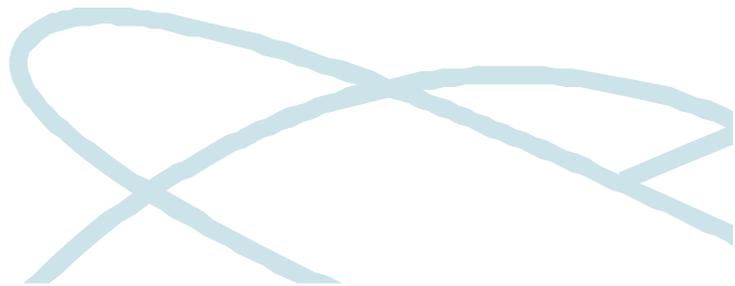
A recording of Event 1: Building the Vision" can be found using the following links:

III.   Competition Details

Funding: Approximately $500,000 will be available to fund up to one (1) year grants to successful proposers, of approximately $50,000 - $100,000 per grant.

IV. Topic Descriptions and Request for White Paper Submission Details:

Topic #3: Dynamic Bi-Directional Trust in Human-Artificial Intelligence (AI) Collaborative Systems

Sponsor: National Security Innovation Network and NYSTEC

Eligibility: US-based Academic Institutions & Small Businesses

Targeting: $75,000 per effort, $10,000 at time of award, $15,000 after justification of effort and $50,000 at completion. Period of Performance will be from June 12, 2021 to September 1, 2021. Expecting to fund 1 proposed effort.

Objective: This task seeks novel ideas to address dynamically establishing trust in human-AI collaborative systems.

Description: Consider an aircraft with a (human) pilot and an AI assistant (agent). The pilot can rely on the AI for assistance, especially in stressful situations where the pilot may be overwhelmed by the flight conditions (bad weather, poor visibility, wake turbulence, cockpit alarms and warnings). Moreover, the AI can rest control of the aircraft away from the pilot if it believes a dangerous situation (such as a crash) is imminent or if it thinks the pilot is behaving irresponsibly, unreliably, or even dangerously (intent to cause harm). Similarly, the pilot can disable the AI if the tables are turned and they think the AI is untrustworthy.

In such aircraft, it is crucial that bidirectional trust be established and maintained throughout the mission or flight. What are the mechanisms for doing this, both pre-flight and in-flight? Answering this question will involve expertise in a number of disciplines, including control theory, avionics, human factors, AI, human psychology, stress management, and more.

The problem can be made more complex by adding a (human) co-pilot into the mix. Then the two humans must establish bidirectional trust in each other and with the AI, and vice versa. Additionally and regrettably, there are several documented cases of suicidal / homicidal commercial aviation pilots whose actions have led to loss of life and aircraft. The onboard AI can, pre-flight, perform a psychological assessment of the pilot/co-pilot, and choose not to hand over control of the aircraft to the humans should the assessment show that at least one of them is not psychologically flight (trust-)worthy. Similarly, incidents involving sleep-deprived pilots and those under the influence of drugs or alcohol have occurred. The AI can similarly perform pre-flight assessments of this nature.

Guidance: Prospective performers should assemble a multi-disciplinary team with expertise in areas such as control, AI, avionics, psychology, and human factors. A detailed written report on how research in this problem area might progress over a 2-5 year period is advisable. This challenge problem is not limited to establishing bidirectional trust in the cockpit. A working demo, which includes a properly trained Deep Neural Network, on some facet of the problem would be desirable but not necessary.

Summary: This task seeks novel ideas on how to address the problem of establishing trust in human-AI collaborative systems. Techniques for establishing trust in the AI-component(s) are likely to involve testing, formal verification, demonstrations, and also human factors. Techniques for enabling the AI to establish trust

in the human(s) may revolve around psychological and performance testing, and techniques for rapid assessment of a person's state of mind, alertness level, and performance qualifications.

White Paper Submission: Submit a 2-page white paper of proposed research project in the format described below. The deadline for 2-page white papers is **5:00pm (EST), June 4, 2021**. Proposals should be submitted in PDF form via e-mail as instructed below.

1. Proposers are limited to one submission. Multiple submissions or a single proposal addressing multiple problem areas will not be accepted or further evaluated. Proposers are eligible to submit additional proposals under Topics 1, 2 or 4 subject to eligibility criteria identified therein.

2. Email submissions and questions should be addressed to 2021TAIChallenge3@nystec.com.

3. For questions please email to 2021TAIChallenge3@nystec.com.

4. All white papers should be 11-point Times or Arial font, single spaced and be <u>a maximum of two (2) pages (not including references)</u>.

5. Applicants must follow the white paper template at Attachment A.

6. White papers must clearly address the challenge problem identified in each submission.

Evaluation and Award Process: The Government (AFRL and NSIN), in partnership with NYSTEC will employ a two-step process to select proposals for grant funding:

1. White papers will be reviewed by the selection committee using the following evaluation and selection criteria:

   A. The technical merits and innovative aspects of the proposed research and development; and,
   B. Relationship of the proposed research and development to the Office of the Undersecretary of Defense Research & Engineering technology verticals of interest, missions, outstanding problem sets and extant transition pathways (follow-on funding opportunities).

2. Submitters will be notified on **June 18, 2021** if they have been selected for award. One proposal will be selected, and the awardee will submit a full-length white paper (~30 pages), format to be provided, due on September 1st, 2021. The awardee will receive $10,000 once selected, $15,000 after supporting documentation of effort is supplied per the Sponsorship Agreement and the remaining $50,000 of funding once the final white paper has been inspected and accepted by NSIN and NYSTEC.

Awardees will be invited to present their challenge solution as part of the third and final TAI event in the series, i.e. the "Trusted AI at Scale" event. Trusted AI at Scale is scheduled to take place virtually and will feature top researchers, as well as leaders from governmental, academic, and industrial organizations, from Jul 27-28, 2021.

With the direct assistance from NYSTEC, the Awardee will develop and create marketing collateral to be used for future marketing by NYSTEC, NSIN and Awardee to include a video about the experience, digital

marketing materials, written experience, and quotes.