INNOVARE
ADVANCEMENT CENTER

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:

https://youtu.be/Uuk0k59I7Y4
https://www.innovare.org/events/trusted-ai-challenge-series

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 #1: Verification of Autonomous Systems

Sponsor: AFRL and AFOSR

Eligibility: US and International Academic Institutions may apply

Target: Up to $100,000 per effort for 1 year. Expecting to fund 1-3 proposed efforts.

Objective: The verification of autonomous systems has largely relied on formal and heuristic testing approaches to verify simple behaviors such as safety properties (e.g. the avoidance of bad states). This topic seeks novel approaches for the verification of autonomous systems that learn and interact in their environments subject to more complex behavioral properties.

Description: Formal approaches to verification have enabled a sense of reliability and resiliency in systems whose dynamics are well-understood and whose complexity is reasonably bounded. However, the advent of autonomous agents powered by deep neural networks challenge these assumptions. Indeed, the dynamics of deep neural networks, while understood, are not interpretable in a form that is amenable to traditional verification algorithms. Some progress has been made for the formal verification of deep neural networks. For example, the use of linear programming and branch-and-bound techniques to mathematically model and verify properties of restricted neural networks of modest size at inference time has shown promise. However, the current literature focuses on verifying basic properties such as safety (i.e. avoid a given set of outputs), reachability (i.e. ensure a given output is observed at some point), and robustness (e.g. bound the proportion of misclassifications). While these properties provide a good starting point, the richness of properties available in traditional verification that come in the form of various logics such as Linear Temporal Logic (LTL) and the Mu-calculus is largely missing. A simple example specification that exists outside of the aforementioned properties is "ensure proposition A holds until B is reached" (e.g. avoid hostile areas of the map until ammo is replenished). We seek solutions in this direction for the verification of learning agents subject to a richer language of specifications. These solutions may be applied at design and/or test time of the underlying learning model, during the learning and/or inference phases of the learning algorithms, and can be within the context of various learning areas, such as data analytics (e.g. classification and regression) and decision-making (e.g. reinforcement learning and planning).Unlike formal approaches to verification, some heuristic approaches to verification focus on creating behavioral and edge tests to evaluate the output of a system given specific inputs. Recently, these concepts have been adapted for static machine learning tasks in Natural Language Processing. Unfortunately, unlike a static machine learning task, an autonomous agent's environment is ever changing, and it is infeasible to design tests for every feasible input. Although behavioral testing will still be an important aspect of the verification pipeline, it must overcome unique challenges. Particularly concerning is that, 1) undesirable behavior may only be exhibited after a large number or complicated sequence of events, and that, 2) undesirable behavior may appear stochastically as autonomous systems are not deterministic. The need for this is intuitive and well-documented as evidenced by the DARPA Assured Autonomy program, the Guarding AI Against Robustness Deception (GARD) program, and the AFOSR Agile Test and Evaluation portfolio of investments.

Guidance: Prospective performers should develop or adapt a formal verification or heuristic testing approach to verify autonomous systems subject to complex behavioral properties. Complex, in this sense, entails going beyond reachability and simple robustness properties and may entail the specification of behaviors to be derived from existing logics, such a Linear Temporal Logic (LTL) or Computation Tree Logic (CTL), among others. Potential solutions include the use of linear programming by representing the autonomous agent as a simple neural network that can be reasoned over as a simplex, the use of whitebox and differential testing by leveraging neuron coverage as an analogue to traditional code coverage metrics in software testing, or other approaches. The performers must clearly document any assumptions made on the autonomous system model and its learning and inferencing dynamics, to include datasets or simulation environments used and evaluation metrics proposed or adopted from the literature. There are no restrictions on the application domain and this may include classification, regression, natural language processing, reinforcement learning, and planning.

Summary: We seek novel approaches to assist users in creating baseline and edge tests for autonomous agents that address these challenges. Of primary interest are approaches that enable users to define and search for undesirable behaviors, and that provide statistical guarantees on performance.

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 **5:00 pm (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 2, 3 or 4 subject to eligibility criteria identified therein.

2. Email Topic 1 submissions to afrl.ri.taichallenge@us.af.mil.

3. For questions please email to afrl.ri.taichallenge@us.af.mil.

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

5. Applicants submitting white papers 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 will employ a two-step process to select proposals for grant funding:

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

   A. Primary Evaluation Criteria
      i. The technical merits and innovative aspects of the proposed research and development; and,
      ii. Relationship of the proposed research and development to United States Department of Defense missions.

   B. Other Evaluation Criteria

   i. The applicant’s capabilities integral to achieving U.S. Air Force objectives. This includes principal investigator’s, team leader's, or key personnel’s qualifications, related experience, facilities, or techniques or a combination of these factors integral to achieving U.S. Air Force objectives, and the potential risk of this effort to the U.S. Air Force.

2. Submitters will be notified on **June 18, 2021** if they have been selected for award. White papers selected for award will be invited to submit a formal proposal within 30 days to an AFOSR Broad Agency Announcement (BAA). Proposals selected under the BAA will be awarded a one year grant with awards up to $100,000 each.

3. Proposers from US and international universities are eligible to submit to this challenge competition. Proposers who have not previously received a grant from the USAF are also strongly encouraged to apply.

4. 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.