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 #4: Trustworthy AI Certification

Sponsors: IBM and SUNY

Eligibility: Faculty members from all SUNY state-operated colleges are eligible to participate.

Targeting: $100,000 per effort for 1 year. Expecting to fund one proposed effort. Project may be invited to submit a renewal proposal based on satisfactory progress at the end of Year 1.

Objective: This topic seeks to create novel methodologies and tools for eliciting and fusing accuracy, fairness, robustness, and explainability metrics and beliefs to certify trustworthiness of AI systems.

Description: AI systems that are worthy of trust in mission-critical and high-stakes domains have high aptitude not only in basic accuracy, but also in distributional robustness, adversarial robustness, algorithmic fairness, and explainability. In recent years, there has been significant progress in developing metrics for these considerations beyond accuracy and in toolkits for computing them. However, the metrics for the different pillars of trust are operationalized in different ways depending on the precise characteristics of the application along with its regulations and policies, and also depending on the normative stances and politics of the consumers of the metrics. It is also not clear how the different elements of trust interact with each other and with accuracy: are they tradeoffs or non-tradeoffs that can be achieved simultaneously? Thus, certification of AI systems for trustworthiness is not as simple as only computing fairness, robustness, and explainability metrics.

Guidance: Using expertise from human factors engineering, machine learning, software engineering, game theory, cognitive psychology, policy, law, and other disciplines, we challenge researchers to develop novel methods for:

- Eliciting feasible trust-related policies from committees of multiple stakeholders, including people from traditionally marginalized groups,
- Operationalizing these trust policies as means for certifying AI models, including by fusing beliefs towards different elements of trust in summaries that can be regulated, and
- Creating end-to-end tools for developers and third parties to govern AI models.

Summary: Certifying AI systems for trustworthiness is not straightforward because trustworthiness involves several dimensions whose interrelationships are not clear and that have different politics. Develop components of a system that measures trust metrics and elicits acceptable ranges for these trust metrics from stakeholders.

White Paper Submission: Submit a 2-page white paper of proposed research project in the format shown in Attachment A. Applicants should restrict the white paper proposal to non-proprietary content. 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. Submissions are limited to one proposal per individual investigator. An individual may submit only one proposal as either a PI or co- PI under Topic 4. SUNY faculty members submitting proposals for this topic,

are however, eligible to submit one additional proposal under Topics 1, 2 or 3 subject to eligibility criteria.

2. Email Topic 4 submissions to SUNY-IBM-AI-Alliance@suny.edu.

3. For questions please email to SUNY-IBM-AI-Alliance@suny.edu.

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 Process and Criteria:

1. White papers will be evaluated by members of the joint SUNY-IBM selection committee using the following evaluation and selection criteria:
   A. Relevance to the challenge topic, technical merits and innovative aspects of the proposed research, and
   B. Relationship of the proposed research and development to the goals of the SUNY-IBM Collaborative AI Research Alliance.

2. White papers that meet the evaluation and selection criteria above will be invited to present at a joint SUNY-IBM workshop. The project selected for the award will be required to submit a detailed Statement of Work jointly developed by SUNY and IBM researchers.

3. 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.

Award Information and Administration:
The selected project will be awarded $100K for a one year period. The project will be subject to annual review, at which time the Principal Investigators may be invited to submit a proposal for renewal based on satisfactory progress in the current year.

Terms and Conditions:
Lead investigators are required to provide a progress report six months from the award date and at the end of the project. A final report detailing the team's accomplishments, artifacts produced including invention disclosures and publications and lessons learned will be required within 30 days of the end of the project.