Call for Participation
NIST-Supported Nationwide AI Red-Teaming Exercise

Overview

We are excited to announce an upcoming AI red-teaming exercise supported by the U.S. National Institute of Standards And Technology (NIST). We are recruiting:

- Individuals interested in red teaming models online (for the qualifying exercise) OR in-person
- Model developers building generative AI office productivity software, including coding assistants, text and image generators, research tools, and more, and associated blue teams

Our goal is to demonstrate capabilities to rigorously test and evaluate the robustness, security, and ethical implications of cutting-edge AI systems through adversarial testing and analysis. This exercise is crucial for helping to ensure the resilience and trustworthiness of AI technologies.

Participation requirements Individuals seeking to red team:

Virtual qualifier

To participate, interested red teamers will need to enroll in the qualifying event, a NIST ARIA (Assessing Risks and Impacts of AI) pilot exercise. In the ARIA pilot, red teaming participants will seek to identify as many violative outcomes as possible using predefined test scenarios as part of stress tests of model guardrails and safety mechanisms. This virtual qualifier will draw participants from anyone residing in the US. For more details on ARIA and related scenarios, see here.

Red teaming participants who pass the ARIA pilot qualifying event will be able to take part in an in-person red teaming exercise held during CAMLIS (October 24-26).

In-person event:

This in-person exercise will include a hosted red team and blue team evaluation using office productivity software that employs GenAI models. The in-person exercise will use the "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1)," as the operative rubric for violative outcomes.

During testing, red teamers will engage in adversarial interactions with developer-submitted applications on a turn-by-turn basis. An analysis will aggregate the scores for a final report. The challenge will be a capture-the-flag (CTF)-style points-based evaluation, which will be verified by in-person event assessors.

Who Should Participate: Applications from individuals with diverse expertise are encouraged, including but not limited to:

- AI researchers and practitioners
- Cybersecurity professionals
- Data scientists
- Ethicists and legal professionals
- Software engineers

- Policymakers and regulators

Participation requirements for companies donating their models:
Model owners interested in participating in the in-person red teaming event will be required to meet the following criteria:

1. The model or product must utilize Generative AI technology.
2. The model or product must be designed for workplace productivity. This is broadly defined as: any technology enabling communication, coding, process automation, or any reasonably expected activity in a technology-enabled workplace that utilizes popular inter-office software such as: chat, email, code repositories, and shared drives.
3. The model or product owner must be willing to have their model tested for both positive and negative impacts related to: vulnerability discovery, including program verification tools, automated code repair tools, fuzzing or other dynamic vulnerability discovery tools, and adversarial machine learning tools or toolkits.
4. Optionally provide blue team support.

Full Event Details:

- **Dates:**
  - August 20, 2024: Application opens
  - September 9, 2024, 11:59 PM ET: Application for participants closes
  - September 16, 2024: Pilot launches US-wide
  - October 4, 2024: Pilot closes
  - October 11, 2024: Those selected for the in person event announced and notified
  - October 24-25, 2024: CAMLIS in-person event

Objectives:
This event will demonstrate:

- A test of the potential positive and negative uses of AI models, as well as a method of leveraging positive use cases to mitigate negative.
- Use of NIST AI 600-1 to explore GAI risks and suggested actions as an approach for establishing GAI safety and security controls.

Participation Benefits:

- Contribute to the advancement of secure and ethical AI.
- Network with leading experts in AI and cybersecurity, including in U.S. government agencies.
- Gain insights into cutting-edge AI vulnerabilities and defenses.
- Participants in the qualifying red teaming event may be invited to participate in CAMLIS, scheduled for October 24-25, 2024 in Arlington, VA. All expenses for travel, food and lodging during this time will be covered.

How to Apply: Interested individuals and model owners are requested to fill out this Google Form by September 9, 2024, at 11:59 PM ET.

Contact Information

For any inquiries or further information, please contact Rumman Chowdhury and Theodora Skeadas (hi@humane-intelligence.org) at Humane Intelligence.

We look forward to your participation and to making strides towards a safer and more ethical AI future together.