Responsible Artificial Intelligence Test & Evaluation (RAITE)

Transforming the Department of Defense Through **Assured** AI

Productizing Expertise to Empower RAI Novices

Getting responsible artificial intelligence (RAI) right is difficult and demands expertise. All AI-relevant skill sets, including ethics, are in high demand and short supply, especially regarding AI's intersection with test and evaluation (T&E). The Joint Artificial Intelligence Center (JAIC) and Chief Digital and Artificial Intelligence Officer (CDAO) are creating frameworks, guidance, and tools to empower working-level personnel across the Department of Defense (DoD) to generate defensible drafts of RAI evaluation plans. The goal is to free up experts to review and fine-tune plans across broader program portfolios.

Affirming T&E, A Necessary Piece of the RAI Puzzle

Testing provides a feedback mechanism for system improvement and builds public and warfighter confidence in our systems, and RAI should be treated just like performance, reliability, and safety requirements. DoD cannot take for granted that any AI program—including non-combat systems— has successfully implemented RAI. For many RAI principles, it will be the role of T&E to build part of the case that legal, moral, and ethical considerations are infused into the system design and operations.

Intertwining T&E and RAI Across Product Lifecycle

T&E and RAI should be integrated into a program early, often, and always. Early concept and design choices are critical to effectively testing the system and achieving RAI. Programs must embrace T&E and RAI from inception; as development proceeds, these two streams must be integrated in tight feedback loops to ensure effective RAI implementation. However, DoD must also extend RAI and T&E activities after fielding. The technological complexity makes testing all operating conditions impossible; furthermore, many AI systems, along with their operating environments and use cases, will continue to update and evolve. By continuing to monitor performance and RAI, we can build appropriate warfighter and public trust in our AI capabilities.

RAITE AT THE CDAO

"The RAITE team aims to provide justified confidence that our AI-enabled systems provide stakeholders with systems that meet requirements and support missions through ethical action. We have a variety of stakeholders that need different assurance levels, arguments, and evidence, and that's what the AI assurance framework allows us to do."

— Dr. Jane Pinelis, CDAO Chief of AI Assurance

Translating Principles into Actionable Guidance

The five DoD AI Ethical Principles are a necessary north star, but, alone, they are not enough to implement or ensure RAI implementation. The CDAO is operationalizing these guiding principles so that programs have a framework for generating an RAI test strategy for their use cases. However, AI stakeholders must understand that RAI integration is not a guarantee and that there will not be magic "RAI" numbers to set requirements against. Programs will have to integrate multiple methodologies and sources of evidence to build an assurance case establishing how programs have reduced or mitigated RAI risks.

Developing, Testing & Evaluating RAI in Context

T&E lacking operationally relevant context will fail to ensure that fielded tools achieve RAI. Mission success depends on technology that must interact with warfighters and other systems in complex environments, while constrained by processes and regulation. AI systems will be especially sensitive to operational context and will force T&E to expand what it considers. Furthermore, although the principles are sharply delineated into five discrete categories and other T&E elements are siloed, in operation, these principles and parameters are interdependent. RAI cannot be separated from system effectiveness, safety, cybersecurity, or other elements, so stakeholders must interweave these critical contexts into RAI implementation and evaluation.

11 April 2022

Dr. Jane Pinelis | yevgeniya.k.pinelis.civ@mail.mil

U.S. Department of Defense

ai.mil