CDAO

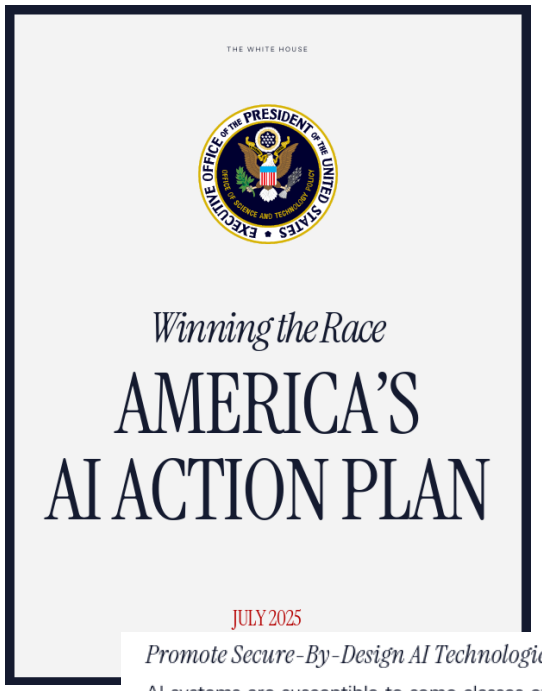
Chief Digital & Artificial Intelligence Office

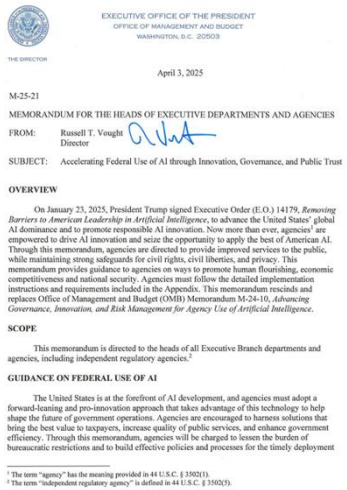
Operationalizing AI Assurance

Dr. Matthew K. Johnson
Chief of Responsible AI
US Department of Defense

ResponsibleAI ToolKit

Overview of the Policy Landscape for Trustworthy AI





Sec. 3. Unbiased AI Principles. It is the policy of the United States to promote the innovation and use of trustworthy AI. To advance that policy, agency heads shall, consistent with applicable law and in consideration of guidance issued pursuant to section 4 of this order, procure only those LLMs developed in accordance with the following two principles (Unbiased AI Principles):

Promote Secure-By-Design AI Technologies and Applications

AI systems are susceptible to some classes of adversarial inputs (e.g., data poisoning and privacy attacks), which puts their performance at risk. The U.S. government has a responsibility to ensure the AI systems it relies on—particularly for national security applications—are protected against spurious or malicious inputs. While much work has been done to advance the field of AI Assurance, promoting resilient and secure AI development and deployment should be a core activity of the U.S. government.

Recommended Policy Actions

- Led by DOD in collaboration with NIST at DOC and ODNI, continue to refine DOD's Responsible AI and Generative AI Frameworks, Roadmaps, and Toolkits.

CDAO

Three Conditions & Three Approaches for AI Assurance

THREE CONDITIONS

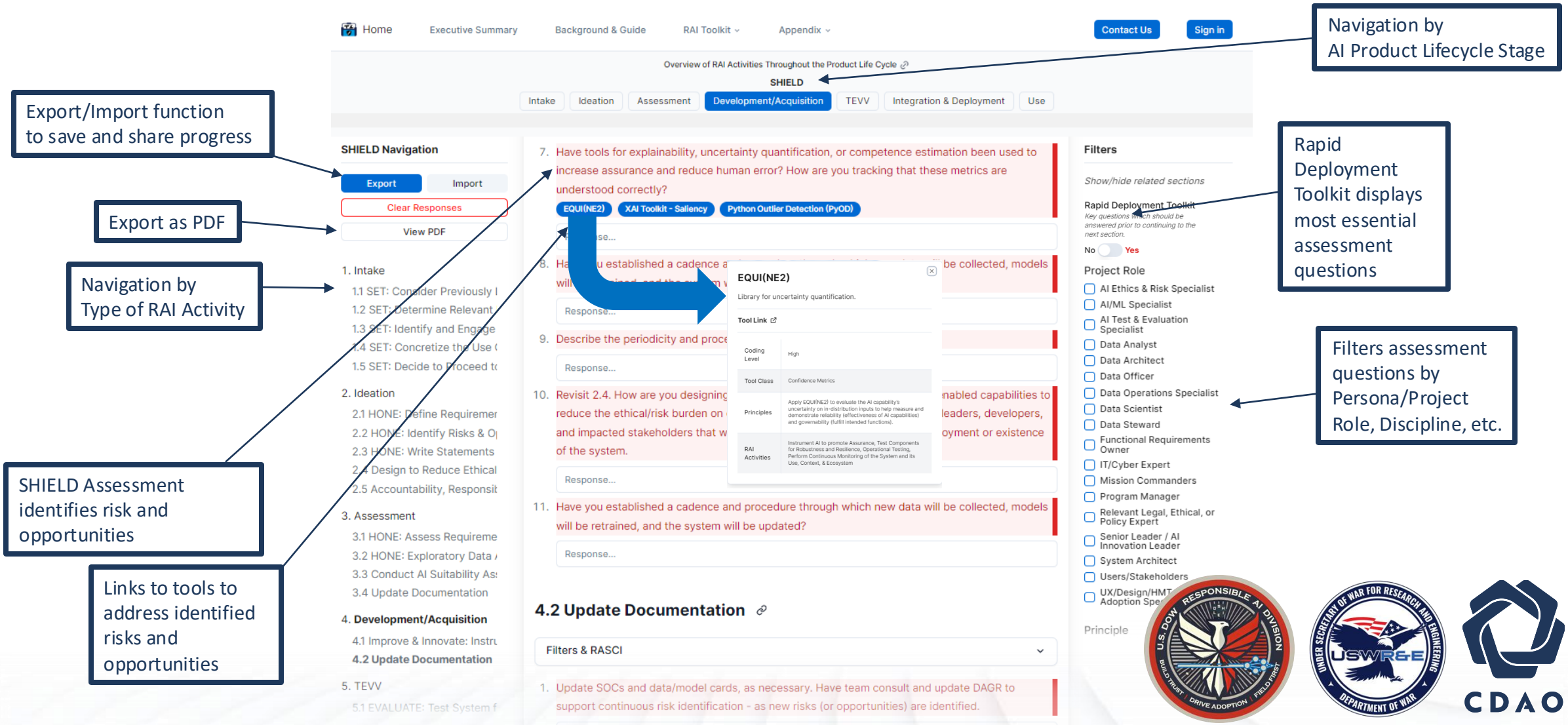
- Rapid pace of technological advancement
- AI Assurance Workforce Gaps
- Nascency of Alignment Tradecraft & Testing

THREE APPROACHES

- Crowdsource
- Automate
- Leverage Interdisciplinarity

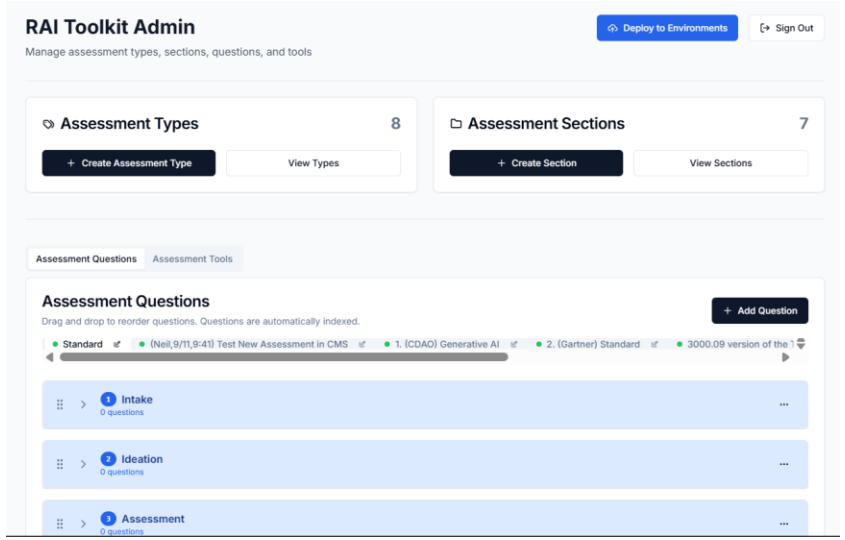
CDAO

Operationalizing AI Assurance with the RAI Toolkit



Additional RAI Toolkit Features

- Mapping to other requirements, frameworks, & standards (MIL-STD-882E, NIST AI RMF, IEEE 7000, etc.)
- Additional use-case focused versions of the Toolkit
- Automation Enhancements
- Dashboard Enhancements
- Content Management System



DoD Responsible AI (RAI)

Build Trust, Drive Adoption, Field First

RESPONSIBLE AI | ToolKit

ai.mil/Initiatives/Responsible-AI/

Operationalizing AI Assurance
(*Build Trust*)

- RAI Toolkit + Assurance Portal
- AI Assurance Tools & Best Practices
- Partner Validation & Technical Support
- Frontier AI Red Teaming & Assurance Capability Development

Outward Focus:

- Risk-informed AI investment decision-making, delivering reliable, mission-ready AI for warfighters

Drive AI Assurance Policy & Strategy
(*Drive Adoption*)

- Identify & Mitigate Policy Barriers
- AI Assurance Policies
- AI Acquisition Resources

Inward Focus:

- Strategy and Policy
- DoD Leadership, Authorizing Officials, & Risk Owners

Interoperability with Partners
(*Field First*)

- Chair of White House's CAIO AI Assurance Working Group
- International Partnerships
- Industry Partnerships

Lateral Focus:

- Interagency
- Industry
- Allies & Partners

RAI Contacts

Chief of RAI
Dr. Matthew Johnson
matthew.k.johnson94.civ@mail.mil

Contact the RAI Team:
osd.pentagon.cdao.list.rai@mail.mil

Your Questions Answered

- **Why is RAI important for the DoD?**
  - RAI ensures that AI-enabled capabilities (1) Do what they're supposed to do, (2) Don't do what they're not supposed to do, and (3) Warfighters trust them. Baking in RAI ensures the system will be reliable and avoid downstream issues. Ensuring warfighter trust catalyzes adoption.
- **What do the RAI Toolkit & AI Assurance Portal do?**
  - The RAI Toolkit provides technical tools and processes for ensuring the performance of RAI capabilities and documenting an assurance case. The Portal enables sharing of tools and artifacts, to enable reusability.

U.S. DOW RESPONSIBLE AI DIVISION — BUILD TRUST | DRIVE ADOPTION | FIELD FIRST

UNDER SECRETARY OF WAR FOR RESEARCH AND ENGINEERING — USWR&E — DEPARTMENT OF WAR

CDAO