CDAO
Chief Digital and Artificial Intelligence Office

Responsible Artificial Intelligence (RAI)

Transforming the Department of Defense Through AI

Establishment of Ethical Principles

The U.S. Department of Defense (DoD) is deeply committed to the ethical development and use of Artificial Intelligence (AI) in combat and noncombat applications. To that end, the DoD officially adopted AI Ethical Principles in February 2020 (making it the first military to do so) and tasked the Joint Artificial Intelligence Center (now the CDAO) with leading coordination of oversight and implementation of the Principles across the Department.

The Principles are intended to assist the Department in addressing the novel ethical ambiguities and risks associated with the design, development, procurement, deployment, and use of AI and to contribute to the efficiency, effectiveness, and legitimacy of the Department's AI capabilities. They build off of foundational engineering practices, as well as the Department's existing legal, ethical, and policy frameworks and require a holistic, integrated, and disciplined approach that focuses, not only on the technology, but also on organizational operating structures and culture. This holistic approach to the responsible development and use of AI is what the DoD calls Responsible AI (RAI), and it is what provides our Service members, the American public, and our partners and allies the confidence that DoD AI-enabled systems will be safe and reliable and will adhere to ethical standards.

Why is RAI Important for the DOD?

- Provides a source of soft power in an age of great power competition and sets us apart from authoritarian state actors.
- Positions the United States to establish global norms for the responsible design, development, deployment, and use of AI in defense, reflecting our democratic values and strengthening our national interests.
- Strengthens international partnerships by assuring allies of our shared values in the application of AI for defense.
- Earns the trust of the American public, industry, and the broader AI community to sustain our technological edge.

DOD AI ETHICAL PRINCIPLES

Responsible: DoD personnel will exercise appropriate levels of judgement for the development, deployment, and use of AI capabilities.

Equitable: The Department will take deliberate steps to minimize unintended bias in AI capabilities.

Traceable: The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedures and documentation.

Reliable: The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.

Governable: The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.

"...ultimately, AI systems only work when they are based in trust. We have a principled approach to AI that anchors everything that this Department does. We call this Responsible AI, and that's the only kind of AI that we do. Responsible AI is the place where cutting-edge tech meets timeless values."

—General Lloyd J. Austen, U.S. Secretary of Defense

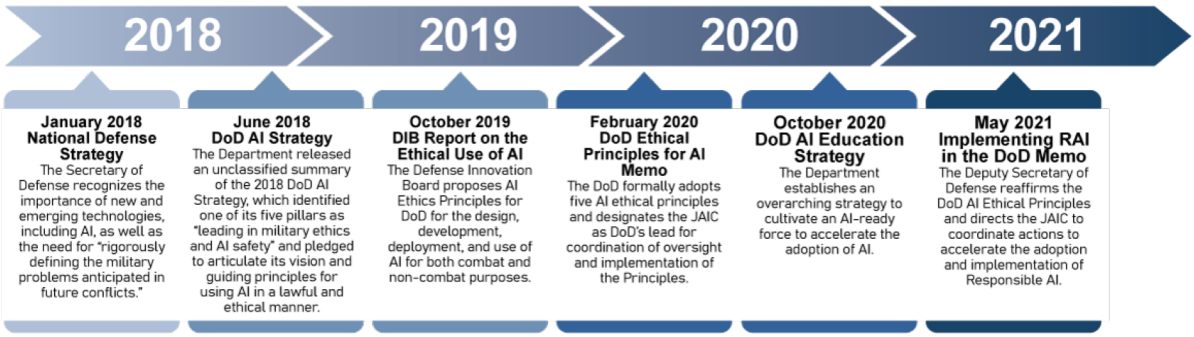
22 July 2022

For more information:
Diane Staheli | diane.p.staheli.civ@mail.mil

U.S. Department of Defense

ai.mil

Advancing AI in the Department



RAI Advancements

The following are exemplary efforts in the Department's RAI implementation journey:

Training & Education

In accordance with the DoD AI Education Strategy, the JAIC rapidly piloted a set of courses customized to DoD workforce archetypes. The learners represent communities across the DoD, including personnel from the Services, Combatant Commands, and DoD agencies, along with a host of served professionals to include acquisition professionals, General and Flag Officers and SES personnel, and data scientists.

The Strategy also includes an explicit RAI training component for each functional role/archetype outlined in the Strategy as part of its learning journey. The integrated RAI learning pilots are customized (varying in breadth and depth) for each functional role.

Acquisition

At the 2021 DoD AI Symposium, the JAIC announced an RAI Request for Information (RFI), through its AI procurement ecosystem, Tradewind. The RFI called on RAI subject matter experts and organizations to provide input on knowledge, tools, services, and solutions to aid in advancing the six RAI Implementation Tenets outlined in the DoD's May 2021 memo. The JAIC is also working to ensure that AI acquired via Tradewind is aligned with the Principles by identifying entry points for RAI integration into the AI acquisition process with a combination of tools and reviews for oversight and accountability.

Engagement with Allies

September 2020 marked the inaugural convening of the AI Partnership for Defense hosted by the JAIC with like-minded military and defense forces with now a group of 16 other nations. The Partnership aims to promote the responsible design, development, and use of AI by engaging in coordination and collaboration on AI technologies, governance, and policy to drive AI-enabled interoperability. While initial convening focused discussion solely on RAI, subsequent meetings ensure a sustained integration and connection to RAI.

RAI IMPLEMENTATION TENETS

On May 26, 2021, Deputy Secretary of Defense Kathleen Hicks signed a memo identifying the following six tenets for RAI implementation:

1. **RAI Governance:** Ensure disciplined governance structure and processes at the Component and DoD-wide levels for oversight and accountability and clearly articulate DoD guidelines and policies on RAI and associated incentives to accelerate adoption of RAI within the DoD.
2. **Warfighter Trust:** Ensure warfighter trust by providing education and training, establishing a test and evaluation and verification and validation (TE/VV) framework that integrates real-time monitoring, algorithm confidence metrics, and user feedback to ensure trusted and trustworthy AI capabilities.
3. **AI Product and Acquisition Lifecycle:** Develop tools, policies, processes, systems, and guidance to synchronize enterprise RAI implementation for the AI product throughout the acquisition lifecycle through a systems engineering and risk management approach.
4. **Requirements Validation:** Incorporate RAI into all applicable AI requirements, including joint performance requirements established and approved by the Joint Requirements Oversight Council, to ensure RAI inclusion in appropriate DoD AI capabilities.
5. **Responsible AI Ecosystem:** Build a robust national and global RAI ecosystem to improve intergovernmental, academic, industry, and stakeholder collaboration, including cooperation with allies and coalition partners, and to advance global norms grounded in shared values.
6. **AI Workforce:** Build, train, equip, and retain an RAI-ready workforce to ensure robust talent planning, recruitment, and capacity-building measures, including workforce education and training on RAI.

For more information:
Diane Staheli | diane.p.staheli.civ@mail.mil

U.S. Department of Defense

ai.mil