U.S. DEPARTMENT OF DEFENSE RESPONSIBLE ARTIFICIAL INTELLIGENCE STRATEGY AND IMPLEMENTATION PATHWAY

Prepared by the DoD Responsible AI Working Council in accordance with the memorandum issued by Deputy Secretary of Defense Kathleen Hicks on May 26, 2021, Implementing Responsible Artificial Intelligence in the Department of Defense.

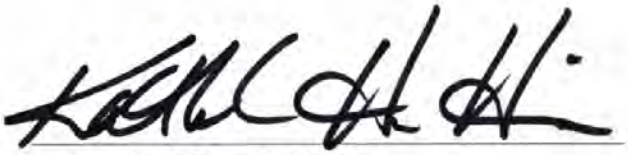
[Publish Date]

FOREWORD

While artificial intelligence (AI) is not new, technological breakthroughs in the last decade have drastically changed the national security landscape. Our adversaries and competitors are investing heavily in AI and AI-enabled capabilities in ways that threaten global security, peace, and stability. To maintain our military advantage in a digitally competitive world, the United States Department of Defense (DoD) must embrace AI technologies to keep pace with these evolving threats. Harnessing new technology in lawful, ethical, responsible, and accountable ways is core to our ethos. Those who depend on us will accept nothing less.

To ensure that our citizens, warfighters, and leaders can trust the outputs of DoD AI capabilities, DoD must demonstrate that our military's steadfast commitment to lawful and ethical behavior apply when designing, developing, testing, procuring, deploying, and using AI. The Responsible AI (RAI) Strategy and Implementation (S&I) Pathway illuminates our path forward by defining and communicating our framework for harnessing AI. It helps to eliminate uncertainty and hesitancy – and enables us to move faster. Integrating ethics from the start also empowers the DoD to maintain the trust of our allies and coalition partners as we work alongside them to promote democratic norms and international standards.

The RAI S&I Pathway makes our RAI policy tractable for implementation. It directs the Department's strategic approach for operationalizing the DoD AI Ethical Principles and, more broadly, advancing RAI—all while ensuring operational agility, maintaining speed of capability deployment, providing scalability, and prioritizing the efficient allocation of resources. This document is a critical step in our journey towards accelerating RAI and furthers the Department's commitment to responsible behavior, processes, and outcomes in the pursuit of AI technology.



Kathleen H. Hicks
Deputy Secretary of Defense

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Advancements in AI have demonstrated the ability to transform every sector of modern society. These impacts extend to business, finance, production, and social behaviors. As the DoD embraces AI, it remains focused on the imperative of harnessing this technology in a manner consistent with our national values, shared democratic ideals, and our military's steadfast commitment to lawful and ethical behavior.

In May 2021, the Deputy Secretary of Defense issued a memorandum ("RAI Memo") that established and directed the Department's holistic, integrated, and disciplined approach to RAI. This RAI Memo introduced the following foundational tenets that serve as priority areas to guide the implementation of RAI across the Department: RAI Governance, Warfighter Trust, AI Product and Acquisition Lifecycle, Requirements Validation, Responsible AI Ecosystem, and AI Workforce.

This resulting DoD RAI S&I Pathway is organized around the six tenets and identifies lines of effort to:

- Modernize governance structures and processes that allow for continuous oversight of DoD use of AI, taking into account the context in which the technology will be used;
- Achieve a standard level of technological familiarity and proficiency for system operators to achieve justified confidence in AI and AI-enabled systems;
- Exercise appropriate care in the AI product and acquisition lifecycle to ensure potential AI risks are considered from the outset of an AI project, and efforts are taken to mitigate or ameliorate such risks and reduce unintended consequences, while enabling AI development at the pace the Department needs to meet the National Defense Strategy;
- Use the requirements validation process to ensure that capabilities that leverage AI are aligned with operational needs while addressing relevant AI risks;
- Promote a shared understanding of RAI design, development, deployment, and use through domestic and international engagements; and
- Ensure that all DoD AI workforce members possess an appropriate understanding of the technology, its development process, and the operational methods applicable to implementing RAI commensurate with their duties within the archetype roles outlined in the 2020 DoD AI Education Strategy.

By leading in military ethics and AI safety, the DoD will earn the trust of our Service members, civilian personnel, and citizens. Our leadership here also encourages RAI development and use globally and strengthens our ability to solve modern defense challenges with allies and partners around the world.

DEPARTMENT OF DEFENSE RESPONSIBLE ARTIFICIAL INTELLIGENCE STRATEGY

BACKGROUND

In the past several years, the DoD has made significant progress in establishing policy and strategic guidance for the adoption of AI technology, as depicted in the figure below. As part of that effort, the Department has also matured its ethics framework to account for AI's unique characteristics and the potential for unintended consequences. This is most clearly articulated in the DoD's adoption of its AI Ethical Principles.

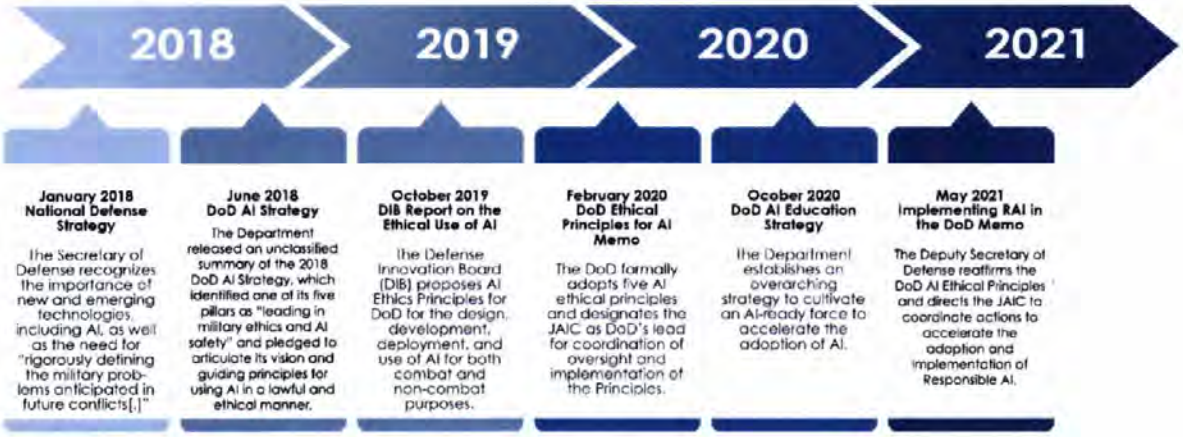


Figure 1: Advancements in DoD AI Strategy and Policy

These Principles were developed as part of a robust, inclusive, and transparent study conducted by the Defense Innovation Board (DIB) and are based on the existing ethical, legal, and policy framework under which the DoD has operated for decades. Essential foundations for the DoD AI Ethical Principles include the U.S. Constitution, Title 10 of the U.S. Code, the Law of War, privacy and civil liberties protections for individuals, and long-standing international norms and values. The DoD AI Ethical Principles do not substitute or deviate from the Department's existing framework. Rather, these Principles complement DoD's existing framework by offering AI-specific guidance and seeking to outline appropriate safeguards for a technology that continues to be subject to rapid developments.

Recognizing the need to address potential unintended consequences from evolving AI technology, the DoD was the first military in the world to publish AI ethics principles and continues to lead in the promotion of global AI standards and norms through the implementation of RAI. Shortly thereafter, the DoD released its RAI Memo, which outlined "the Department's holistic, integrated, and disciplined approach for RAI" and directed specific near-term steps for implementation. This memo reaffirmed the Department's commitment to the DoD AI Ethical Principles, established the RAI Foundational Tenets, and directed the creation of the DoD's RAI Strategy and Implementation Pathway.

DOD AI ETHICAL PRINCIPLES

These principles apply to all DoD AI capabilities, encompassing both combat and non-combat applications.

RESPONSIBLE: DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible 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.

Source: DoD Memorandum, "Artificial Intelligence Ethical Principles for the Department of Defense" (Feb 2020)

RESPONSIBLE AI IN DEFENSE

The DoD has decades of experience integrating new technology into military operations, for example, from the early passive acoustic homing torpedoes and computers used during World War II to the multitude of sophisticated platforms, sensors, and weapon systems of today. In doing so, the Department has consistently followed robust processes to develop and incorporate new technologies in a safe and responsible manner. Our approach to AI must be no different.

WHAT DOES A RESPONSIBLE AI APPROACH MEAN?

RAI is a journey to trust. It is an approach to design, development, deployment, and use that ensures the safety of our systems and their ethical employment. RAI manifests itself in ethical guidelines, testing standards, accountability checks, employment guidance, human systems integration, and safety considerations.

RAI is a dynamic approach to the design, development, deployment, and use of AI capabilities that implements DoD AI Ethical Principles to advance the trustworthiness of AI capabilities. RAI emphasizes the necessity for technical maturity to build effective, resilient, robust, reliable, and explainableAI, while recognizing the value of multidisciplinary teams to advise on ethics, accountability, and risk. At the same time, it supports AI development at the speed the necessary to meet the National Defense Strategy. RAI is the approach for *how* the Department must conduct AI design, development, deployment, and use.

RAI should not be viewed as a static end-state where the use of an AI capability is designated as "responsible" and never revisited. Instead, RAI centers around continuous oversight, moving beyond traditional performance metrics to include the aspects of workforce, culture, organization, and governance that affect how AI is implemented throughout the product lifecycle. RAI may look different at each stage but manifests throughout, presenting guidance and standards from prototype to production to use.

Effective RAI adoption requires an organizational culture that implements RAI as an enabler for AI adoption, rather than a set of barriers. For example, when it comes to AI for military applications, a frequently repeated concern is that we must move quickly or risk losing on the battlefield. Assessing ethical impacts of employing AI technology, along with evaluations on the trustworthiness of AI systems results in thoroughly tested and justifiably trusted systems; development and fielding strategies must account for these attributes. This cultural approach should enable program managers to view RAI as an integral, iterative, and enabling part of AI development.

With RAI, the DoD is able to guard against AI capabilities that are applied unethically or irresponsibly, including in combat scenarios. This approach enables developers and users to have appropriate levels of trust in the AI system. This trust in turn enables rapid adoption and operationalization of new technology, strengthening the Department's competitive edge.

DESIRED END STATE

The Department's desired end state for RAI is trust. Trust in DoD AI will enable the Department to modernize its warfighting capability across a range of combat and non-combat applications, taking into account the needs of those internal and external to the DoD. Without trust, warfighters and leaders will not employ AI effectively and the American people will not support the continued use and adoption of such technology. The DoD is taking its commitment to RAI seriously and is actively pursuing methods to make AI implementation safer and more effective.

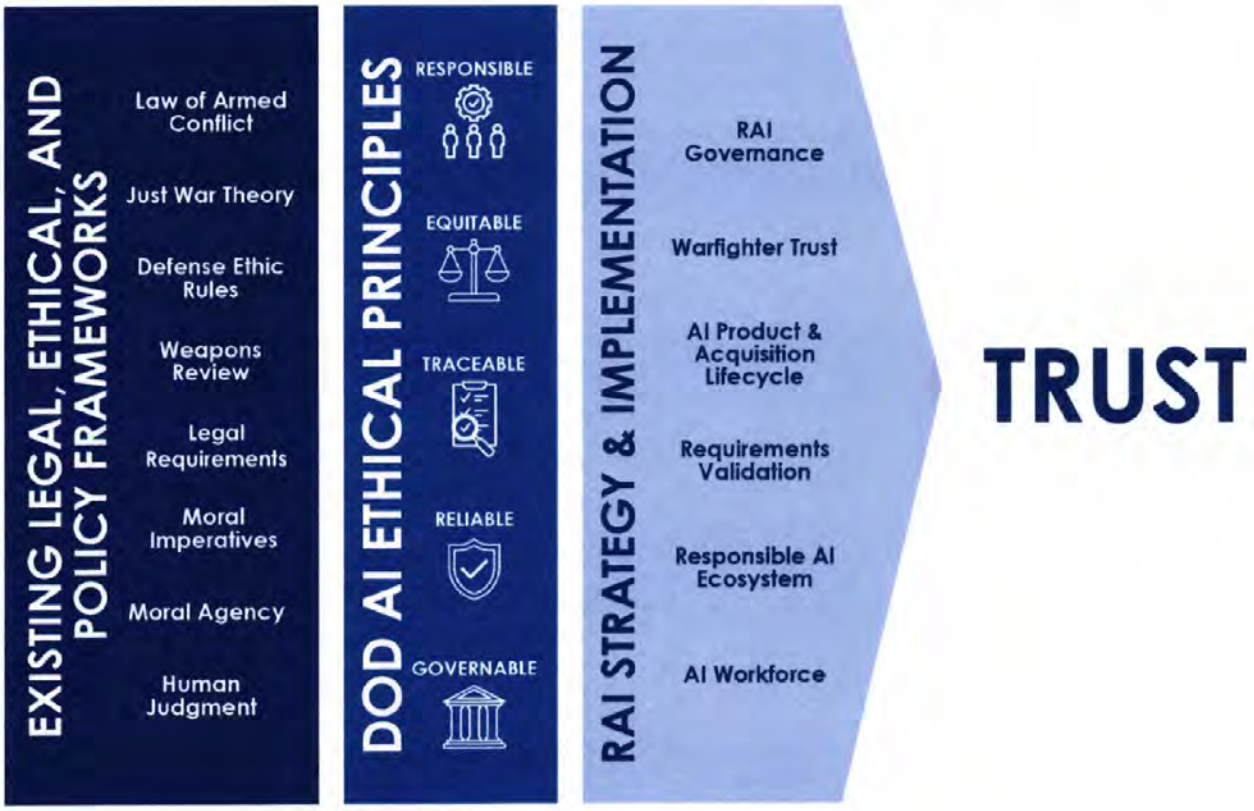


Figure 2: Overview Depicting RAI's Journey to Trust

Through RAI, the DoD will work to ensure that trust in an AI capability is appropriate, taking into account the conditions in which it is to be deployed, among other relevant factors. In doing so, a comprehensive risk management approach will be employed that addresses system-level, institutional, and socio-technical risks. This provides for multidimensional and contextual assessment of risk in the design, deployment, development, and use of AI capabilities across a wide range of scenarios.

To achieve the desired end state, the DoD cannot rely solely on technological advancements. Key factors of trustworthiness also include the ability to demonstrate a reliable governance

structure, as well as the provision of adequate training and education of the workforce. These efforts will help foster appropriate levels of trust, enabling the workforce to move from viewing AI as an enigmatic and incomprehensible technology to understanding the capabilities and limitations of this widely adopted and accepted technology. Additionally, DoD AI developers and users will have confidence that measures are in place to implement the DoD AI Ethical Principles and to report potential concerns.

Trust is also critical to our relationships with like-minded nations. The DoD is expanding its partnerships to set new international norms for AI usage that respect democratic values such as privacy and civil liberties, while defending against aggression. Utilizing the RAI Tenets and DoD AI Ethical Principles promotes open dialogue and proper governance of the use of AI capabilities, allowing for easier integration and minimization of any potential issues between allies.

Developing or employing AI irresponsibly would also result in tangible risks. Adversaries may seek to exploit supply chain vulnerabilities and inject themselves into critical AI training, testing, and update cycles, potentially introducing flawed or exploitable capabilities into consequential systems. For example, relying on external parties to generate, clean, and update DoD's foundational data pipelines without rigorous oversight introduces vulnerabilities that must be systematically considered in the supply chain risk assessment process. If developed without appropriate safeguards, even seemingly benign AI capabilities, like algorithms trained to inform decisions that affect warfighters' fitness or promotion, can lead to adverse outcomes.

Ultimately, DoD cannot maintain its competitive advantage without transforming itself into an AI-ready and data-centric organization, with RAI as a prominent feature. The United States must continue to demonstrate that a principled approach to AI, rooted in democratic values, represents a path to peace, security, and societal progress. While the Department has recognized the need to make key investments in our digital infrastructure to advance AI, special consideration must be paid to non-technical enablers such as our guiding policies and principles. It is imperative that the DoD adopts responsible behavior, processes, and objectives and implements them in a manner that reflects the Department's commitment to its AI Ethical Principles. Failure to adopt AI responsibly puts our warfighters, the public, and our partnerships at risk.

DOD RAI FOUNDATIONAL TENETS

As directed in the Deputy Secretary of Defense's RAI memorandum, the Department will implement RAI in accordance with the following foundational tenets:

- RAI Governance;
- Warfighter Trust;
- AI Product and Acquisition Lifecycle;
- Requirements Validation;
- Responsible AI Ecosystem; and
- AI Workforce.

Descriptions of each RAI Foundational Tenet are taken from the DoD RAI Memo and goals have been added to communicate the desired result in each priority area.

TENET 1: RAI GOVERNANCE

Description: 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.

Goal: Modernize governance structures and processes that allow for continuous oversight of DoD use of AI, taking into account the context in which the technology will be used. Governance structures and processes will enable the appropriate assessment of risks and the mitigation of unintended consequences or bias in AI capabilities. Users or developers will also have clear mechanisms to implement the DoD AI Ethical Principles and to report potential concerns.

TENET 2: WARFIGHTER TRUST

Description: Ensure warfighter trust by providing education and training, establishing a test and evaluation and verification and validation (TEVV) framework that integrates real-time monitoring, algorithm confidence metrics, and user feedback to ensure trusted and trustworthy AI capabilities.

Goal: Achieve a standard level of technological familiarity and proficiency for system operators to achieve justified confidence in AI capabilities and AI-enabled systems. Trustworthiness is bolstered by the application of TEVV frameworks that allow for the monitoring of system performance, reliability, unintended behavior, and failure modes before fielding the system and during operation. The combination of these factors contributes to a greater understanding of an AI's capabilities and limitations, which will be critical for the development of an AI-ready force.

TENET 3: AI PRODUCT AND ACQUISITION LIFECYCLE

Description: 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.

Goal: Exercise appropriate care in the AI product and acquisition lifecycle to ensure potential AI risks are considered from the outset of an AI project, and efforts are taken to mitigate or ameliorate such risks and reduce the likelihood of unintended consequences while enabling AI development at the pace the Department needs to meet the National Defense Strategy. This includes robust documentation to understand, test, and act on informed risk assessments, recognizing that needs will vary based on the level of technical maturity, sensitivity, and context in which the AI capability will be used.

TENET 4: REQUIREMENTS VALIDATION

Description: 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.

Goal: Use the requirements validation process to ensure that capabilities that leverage AI are aligned with operational needs while addressing relevant AI risks. System performance requirements validation increases the reliability and safety of systems prior to and during deployment. A formalized requirements validation process also provides for better traceability, accountability, and both internal and external oversight.

TENET 5: RESPONSIBLE AI ECOSYSTEM

Description: 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.

Goal: Promote a shared understanding of responsible AI design, development, deployment, and use through domestic and international engagements. Such engagements will facilitate knowledge-sharing exchanges with intergovernmental stakeholders as well as partners in industry, academic institutions, and civil society. Through this, the DoD will collaborate on common challenges, advance shared interests, promote democratic norms and values, and increase interoperability with partners.

TENET 6: AI WORKFORCE

Description: 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.

Goal: Ensure that all DoD AI workforce members possess an appropriate understanding of the technology, its development process, and the operational methods applicable to implementing RAI commensurate with their duties within the archetype roles outlined in the 2020 DoD AI Education Strategy. DoD AI workforce education and training should promote consistent understanding across all DoD stakeholders and build a culture within the DoD that enables RAI. Proper training and education must be accompanied with strategies to recruit and retain the personnel whom the DoD trains and educates.

"We have a principled approach to AI that anchors everything that this Department does; we call this Responsible AI, and that is the only kind of AI that we do."

- Secretary of Defense Lloyd Austin (2021)

DOD RESPONSIBLE ARTIFICIAL INTELLIGENCE IMPLEMENTATION PATHWAY

OVERVIEW

As set forth in the RAI Memo, the Department will implement the DoD Responsible AI Strategy in accordance with the following RAI Foundational Tenets: RAI Governance, Warfighter Trust, AI Product and Acquisition Lifecycle, Requirements Validation, Responsible AI Ecosystem, and AI Workforce.

Many of the actions directed in this Pathway build upon the DoD's existing infrastructure for technology development, acquisition, governance, and legal review. Implementing RAI is not possible without sound software engineering practices or robust data management, for instance. This document offers AI-specific guidance to build upon the Department's existing infrastructure by leveraging best practices in industry and academia and lays out a comprehensive roadmap to accelerate RAI.

Implementing RAI in the DoD will not succeed with a set of rigid, one-size-fits-all requirements. A flexible approach is required to foster innovative thinking, as needs and complexity will vary based on factors such as technical maturity and context in which AI will be used. For example, project needs will change as DoD Components progress through the AI Product Lifecycle's four phases: design, develop, deploy, and use (as shown in the figure below). It may not be necessary or possible to conduct a full risk analysis for basic AI research projects without a proposed operational use case; however, an AI capability that is ready to deploy in an operational system must undergo processes to demonstrate that it meets DoD standards for safety, security, and more. The Department should continually strive to have the right balance between responsibility, speed, and ease of implementation of RAI while removing barriers to adoption and access to the data, talent, and compute/infrastructure required.

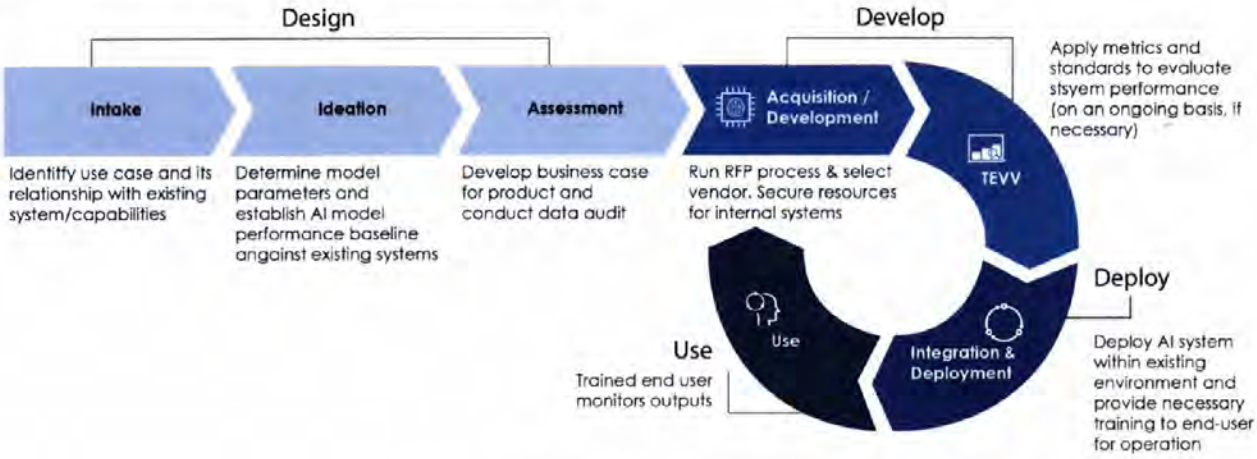


Figure 3: AI Product Lifecycle

For each RAI Foundational Tenet, lines of effort (LOEs) accompany each with overarching goals, corresponding Office(s) of Primary Responsibility (OPRs), and the estimated time horizon for implementation where known. Where not already indicated, OPRs will be responsible for

designating such implementation deadlines. Together, these LOEs direct actions to implement industry best practices and standards for AI development and where new approaches are necessary, task the Department to advance the enterprise RAI implementation in accordance with the RAI Tenets.

Finally, this document reflects an enterprise-wide approach and therefore applies to the Office of the Secretary of Defense (OSD), the Military Departments, the Office of the Chairman of the Joint Chiefs of Staff and the Joint Staff, the Combatant Commands, the Office of the Inspector General of the DoD, the Defense Agencies, the DoD Field Activities, and all other organizational entities within the DoD (hereafter referred to collectively as the "DoD Components").

ROLE OF CDAO, DOD COMPONENTS, DOD COMPONENT RAI LEADS, AND RAI WORKING COUNCIL

As outlined in the RAI Memo, the Office of the Chief Digital and Artificial Intelligence Officer (CDAO), as the successor organization to the Joint Artificial Intelligence Center (JAIC), serves as the Department's lead for coordinating the implementation and oversight of guidance and policy on AI, including RAI and the DoD AI Ethical Principles. The CDAO is responsible for enabling, assessing, and tracking the implementation of a DoD RAI ecosystem, with support from the DoD Components. This includes the Office of the Assistant to the Secretary of Defense for Privacy, Civil Liberties, and Transparency (ATSD(PCLT)), and the Joint Staff, and Military Departments, all of whom are represented through the RAI Working Council.

The RAI Working Council was created via the RAI Memo as an initial working body to ensure Department-wide input and coordination on the development of this RAI S&I Pathway. With the chartering of the CDAO and its governing processes, the RAI Working Council will be a permanent working group that reports to the CDAO's governing council, a 4-star level governance body run by the CDAO to oversee all aspects of data, analytics, and AI for the Department. This Pathway directs member organizations that comprise the current RAI Working Council to designate or hire RAI Leads who will be responsible for implementing this DoD RAI S&I Pathway in their respective organizations and provide reporting on such progress, and relevant barriers for removal, to the RAI Working Council, and up to the Deputy Secretary of Defense, when necessary, through the CDAO governing council. The RAI Working Council should update its membership to reflect the designated RAI Leads, formalize its role through a charter as approved by the CDAO, and assist the CDAO in reporting on the progress of DoD-wide RAI implementation.

As the DoD's lead for AI, the CDAO is designated within this Pathway as the OPR for many Department-wide AI activities, especially those that involve the creation of AI-enabling tools to be used across the DoD in support of its RAI approach. The CDAO should leverage the existing technical foundation set by the DoD's R&D enterprise and the commercial sector to the greatest

extent possible as it continues to coordinate the development of these tools with DoD Component end users.

All DoD Components must ensure that their AI capabilities are in alignment with the DoD Ethical Principles, and that their policies and practices enable RAI implementation. The outputs of this RAI S&I Pathway will offer tools and guidance for DoD Components to accomplish this task. For the purpose of this Pathway, select DoD Components have been designated as OPRs to serve as a lead coordinating entity for their respective LOEs. Many of the LOEs are complex, cross-cutting, and involve multiple external stakeholders and effectively completing them will require Department-wide input based on existing policy authorities, staffing requirements, and well-founded subject matter expertise. Those LOEs have been designated to OPRs believed appropriate to execute such coordination.

IMPLEMENTATION APPROACH

CONDUCT CENTRALIZED COORDINATION OF RAI POLICIES AND GUIDANCE WITH DECENTRALIZED EXECUTION

Centralized, DoD-wide coordination of DoD RAI policies and guidance is critical to establishing interoperability, consistent guidelines and approaches, sharing best practices, and identifying opportunities for collaboration. At the same time, DoD Components must optimize integration of RAI within the context of their existing workflows, structures, and processes. This approach of top-down policy and coordination with bottom-up execution and innovation allows for tailoring RAI integration to the uses and needs of each DoD Component while increasing overall adoption rate. The actions recommended in this Pathway identify OPRs to ensure adoption across the Department at all levels.

UTLIZE A COMPREHENSIVE RISK MANAGEMENT APPROACH

Due to continuous advancements in AI research and the dynamic nature of emerging AI strategies, an approach to AI development that incorporates existing risk management best practices affords the DoD the flexibility to leverage cutting-edge technology while adhering to our standards for safety, reliability, and ethics. This includes the continuous identification, evaluation, and mitigation of risks—including risks from inaction or opportunity costs—across the entire product lifecycle and well beyond deployment. Notably, this Pathway also recognizes the importance of context when tailoring risk mitigation actions and requesting accompanying documentation. Data and model documentation, product risk reviews, and post-deployment monitoring and training are critical factors of RAI; however, the level of attention and risk mitigation efforts must be scalable based on the level of technical maturity, sensitivity, or risk associated with AI projects.

FOCUS ON RESOURCING

DoD Components will identify the resources as well as the appropriate manpower—personnel, expertise, and experience—to carry out fully the RAI activities identified in this Pathway. In order to achieve enterprise-wide implementation, DoD Components are responsible for establishing and maintaining the resources and manpower required to comply with these policies and processes. A balanced approach is needed to ensure the long-term vision is achieved, the near-term impact is delivered, and time is provided to program appropriate manpower and financial resources. This approach will enable alignment across the DoD, while increasing the RAI literacy of the total force, decreasing adoption barriers, and catalyzing a broader cultural change.

MAINTAIN AN ITERATIVE APPROACH IN LINE WITH EMERGING RESEARCH

This RAI Strategy and Implementation Pathway will require updating tools, methods, and best practices to keep pace with state-of-the-art practices for AI development. As progress is made along the identified lines of effort, iterations of such resources should be expected in order for the DoD to advance effective implementation of RAI and maintain its status as a global democratic norm-setter. The Department will continue its AI research agenda, drawing upon advancements from industry and academic partners, to produce well-tested tools and methodologies that actively advance RAI. Updates should be expected as advancements in research and technology emerge, changes in department structures occur, and other developments dictate.

Figure 4: Summary of the DoD Ethical Principles for AI and Responsible AI Tenets

IMPLEMENTATION LINES OF EFFORT

TENET 1: RAI GOVERNANCE

Description: 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 RAI adoption within the DoD.

LOE 1.1: BUILD ORGANIZATIONAL CAPACITY TO IMPROVE OVERSIGHT AND ACCOUNTABILITY OF DOD AI.

Lines of Effort	OPR
LOE 1.1.1: Fully staff the CDAO RAI Office with expertise in AI technology, policy, acquisition, workforce, and governance.	CDAO
LOE 1.1.2: RAI Working Council member organizations will designate or hire DoD Component RAI Leads to lead the implementation of this DoD RAI S&I Pathway within their respective Components. DoD Component RAI Leads will be provided with the adequate authorities, staffing, and resources to fulfill all relevant duties.	RAI Working Council Members
LOE 1.1.3: As part of DoD-wide metrics to measure AI transformation, develop metrics to measure RAI adoption, including progress on individual LOEs identified in this DoD RAI S&I Pathway, and report progress as required to the CDAO governing council.	CDAO
LOE 1.1.4: Develop DoD Component-specific metrics as informed by LOE 1.1.3 to measure RAI adoption, including progress on individual LOEs identified in this DoD RAI S&I Pathway, and establish reporting mechanisms.	DoD Component RAI Leads
LOE 1.1.5: Appropriately staff DoD- and Component-level internal oversight bodies to ensure appropriate expertise is in place to conduct robust and effective oversight of DoD's use of AI under their roles and authorities.	DoD Inspector General; ATSD(PCLT); DOT&E; D,DTE&A
LOE 1.1.6: Identify methods for users and developers to report concerns about the implementation of the DoD AI Ethical Principles. Ensure such methods are clearly communicated to users and developers.	DoD Component RAI Leads
LOE 1.1.7: Update the Department's existing governance framework for AI development and delivery.	CDAO

LOE 1.2: PROVIDE TOOLS AND RESOURCES TO SUPPORT A DOD-WIDE RAI GOVERNANCE STRUCTURE, INCLUDING THROUGH COORDINATED AND REGULAR KNOWLEDGE-SHARING.

Lines of Effort	OPR
LOE 1.2.1: Report to the CDAO (no less than once per year) exemplary AI use cases, identifying best practices, failure modes, and risk mitigation strategies, including after-action reports as appropriate. Identify capability use cases, training in system capabilities, and documentation on how to employ and retire systems responsibly, in order to support the safe and responsible employment of AI capabilities by operators. Include any reports of concerns per LOE 1.1.6.	DoD Component RAI Leads
LOE 1.2.2: Create and maintain a DoD-wide central repository of exemplary AI use cases and any supplementary information to support coordinated and regular knowledge-sharing of best practices and risk mitigation.	CDAO
LOE 1.2.3: Within six months of this Pathway's approval, report to the CDAO any perceived significant barriers – including gaps in infrastructure required to support traceability, auditability, risk analysis, and forensics, and recommended hardware, software, or other infrastructure needs – necessary to fulfill RAI and AI requirements and best practices. Update annually.	DoD Component RAI Leads
LOE 1.2.4: To maintain the DoD AI Inventory in accordance with Public Law 116-260 House Report Division C, identify and report all DoD AI activities to the CDAO, including program's appropriation, project, and budget line number; current and future years' defense program funding; names of academic or industry mission partners, if applicable; and any planned transition partners, if applicable.	DoD Component RAI Leads in coordination with CDAO, OUSD(A&S), and OUSD(R&E)

LOE 1.3: ENSURE THAT RAI IS INCORPORATED IN DOD'S STRATEGIC PLANNING EFFORTS.

Lines of Effort	OPR
LOE 1.3.1: Incorporate RAI and elements of the DoD RAI S&I Pathway as appropriate into the DoD strategies for data, analytics, and AI adoption.	CDAO

LOE 1.3.2: Ensure RAI is incorporated into the Defense Planning Guidance for resourcing planning purposes including but not limited to manpower, tools, education and training, and post-deployment monitoring and retraining.	OUSD(P)
LOE 1.3.3: Identify implementing RAI as a priority in DoD Military Department AI strategies and plans and incorporate Service-specific actions as appropriate.	Military Department RAI Leads

LOE 1.4: UPDATE DOD'S REVIEW PROCESSES FOR WARFIGHTING CAPABILITIES TO SUPPORT IMPLEMENTATION OF THE DOD AI ETHICAL PRINCIPLES.

Lines of Effort	OPR
LOE 1.4.1: Explore whether a review procedure is needed to ensure DoD warfighting capabilities will be consistent with the DoD AI Ethical Principles. Provide a recommendation to the Deputy Secretary of Defense on whether such a review procedure should be created, identifying any gaps in existing processes, gaps in requisite policies supplementing the DoD AI Ethical Principles, and potential impacts on development and fielding timelines if such review is required. Explore whether or how legal review processes can support implementation of the DoD AI Ethical Principles, including the review of the legality of weapons per DoDDs 2311.01, 5000.01, 3000.03E, and 3000.09.	CDAO in coordination with the Military Department RAI Leads, OUSD(P), and DoD OGC
LOE 1.4.2: Update or supplement DoDD 3000.09 with guidance on the DoD AI Ethical Principles. Include any additional information requirements to fulfill the DoD AI Ethical Principles as part of the senior review package.	OUSD(P) in coordination with CDAO

TENET 2: WARFIGHTER TRUST

Description: Ensure warfighter trust by providing education and training, establishing a test and evaluation and verification and validation framework that integrates real-time monitoring, algorithm confidence metrics, and user feedback to ensure trusted and trustworthy AI capabilities.

LOE 2.1: BUILD A ROBUST TEVV ECOSYSTEM AND ACCOMPANYING INFRASTRUCTURE TO DEVELOP AND FIELD AI CAPABILITIES SAFELY AND SECURELY.

Lines of Effort	OPR
LOE 2.1.1: Develop a TEVV framework to articulate how test and evaluation (T&E) should be intertwined across an AI capability's lifecycle and pathways for continuous testing and standards for documentation and reporting. Identify synergies between AI T&E and traditional T&E (e.g. effectiveness, suitability, security, safety) to empower programs to streamline T&E efforts. Include guidance for operationalizing RAI principles into testable conjectures for common technologies, mission domains, and uses cases.	CDAO in coordination with OUSD(R&E) and DOT&E
LOE 2.1.2: Develop or acquire AI-related T&E tools to be used as a resource for AI developers and testers. This AI T&E Toolkit shall draw upon best practices and innovative research from industry and the academic community, as well as commercially available technology where appropriate. The Toolkit will be made widely available to DoD users and shall include: a) Tangible, concrete guidance for PMs, testers, and other relevant T&E stakeholders about how to implement RAI T&E throughout a capability's lifecycle; b) A T&E Master Plan template for AI and a set of templates for test plans; c) A library of T&E metrics for AI systems, including metrics for trustworthiness and confidence; and d) Necessary tools and technologies required to detect both natural degradation of and adversarial attacks on AI, including those to detect various attacks on AI systems, and that can notify testers or operators when such attacks are occurring.	CDAO in coordination with OUSD(R&E) and DOT&E

LOE 2.1.3: Create a test range environment and central repository of tools for T&E of AI, linking to existing and emerging equivalent DoD Component environments, that enables easy and continuous testing for DoD testers. Where appropriate, tools that are housed in this environment should comply with DoD Enterprise DevSecOps Reference Designs for portability across the Department.	CDAO in coordination with the DoD Test Resource Management Center
LOE 2.1.4: Establish best practices and requirements for utilization of the T&E Master Plan and test plan templates for AI and update appropriate DoD issuances and Military Standards (MIL-STDs).	OUSD(R&E) in coordination with CDAO; DOT&E
LOE 2.1.5: Regularly update the CDAO's Human Systems Integration (HSI) Framework based on user feedback and provide clear guidance on how and when the framework should be used to help address system design, system performance, User Experience/User Interface (UX/UI), and user training.	CDAO in coordination with the Joint HSI Steering Committee
LOE 2.1.6: Continue research into emerging AI topics, such as: a) HSI practices to the inform the design and developments of AI capabilities that can be used consistent with the DoD AI Ethical Principles; b) New methods for TEVV of AI; and c) AI security and defense to protect against adversarial attacks.	OUSD(R&E) in coordination with CDAO and Military Service Labs
LOE 2.1.7: Develop and distribute DoD-wide guidance for AI security, leveraging existing best practices for risk management, supply chain security, and cybersecurity. This guidance should be updated as the field matures.	CDAO in coordination with DoD CIO

LOE 2.2: DEVELOP BEST PRACTICES TO INCORPORATE OPERATOR AND SYSTEM FEEDBACK THROUGHOUT THE AI LIFECYCLE, ENSURE ADEQUATE DOCUMENTATION AND TRAINING, AND DELINEATE CLEAR ROLES AND RESPONSIBILITIES FOR DEVELOPERS AND USERS OF THE AI CAPABILITY.

Lines of Effort	OPR
LOE 2.2.1: Require AI vendors and developers to plan for, resource, and provide appropriate training and documentation, such as user manuals, to be used prior to capability fielding in order to ensure warfighters are equipped with an appropriate understanding of the capability's function, risks, performance expectations, and potential harms.	DoD Component RAI Leads
LOE 2.2.2: Require AI vendors and developers who design systems and create features with human-facing interfaces to provide traceable feedback on systems status and clear procedures to trained operators to activate and deactivate system functions. This information should be used to update training, documentation, interfaces, and/or other components as appropriate.	DoD Component RAI Leads
LOE 2.2.3: Promulgate guidance in accordance with LOE 3.1.2 and 3.2.2 delineating responsibilities and authorities for Program Offices to: a) Monitor performance of their AI systems after fielding (including guidance on metrics for monitoring and system instrumentation tools to support this); and b) Establish processes for operators of AI-enabled capabilities to notify and report changes in capability performance, outcomes, emergent behavior, and/or disengagement in accordance with existing DoD processes.	PEOs; Program Offices; DoD Components

TENET 3: AI PRODUCT AND ACQUISITION LIFECYCLE

Description: 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, while enabling AI development at the pace the Department needs to meet the National Defense Strategy.

LOE 3.1: DEVELOP RAI-RELATED ACQUISITION RESOURCES AND TOOLS, SUCH AS STANDARD LANGUAGE FOR ANNOUNCEMENTS AND CONTRACTS, AND BEST PRACTICES THAT LEVERAGE THE DOD'S ADAPTIVE ACQUISITION PATHWAYS AND ALLOW THE DOD TO IDENTIFY AND MITIGATE RISKS THROUGHOUT THE ACQUISITION AND SUSTAINMENT PROCESS.

Lines of Effort	OPR
LOE 3.1.1: Develop an Acquisition Toolkit that draws upon best practices and innovative research from the DoD enterprise, industry, and the academic community, as well as commercially available technology where appropriate. The Toolkit will be made widely available to DoD users and shall include: a) Standard language in the initial announcement, request for proposal (RFP), and request for information (RFI) for AI capabilities to provide guidance on how vendors and developers can meet the DoD AI Ethical Principles; b) A set of RAI-related evaluation criteria that are testable and operationally relevant; c) Standard AI contract language that provides clauses for: independent government T&E of AI capabilities, methods of immediate remediation when the vendor-provided AI capabilities cannot be used in accordance with the DoD AI Ethical Principles, requesting training and documentation from vendors, performance monitoring of AI capabilities, and appropriate data deliverables and rights; and d) Any other resources as appropriate.	CDAO in coordination with OUSD(A&S) and OUSD(R&E)

LOE 3.1.2: Identify and develop approaches for continuous engagement of RAI expertise within the DoD's six adaptive acquisition pathways (Urgent Capability Acquisition, Middle Tier of Acquisition, Major Capability Acquisition, Software Acquisition, Defense Business Systems, and Acquisition of Services) to address RAI risk considerations.	OUSD(A&S) in coordination with CDAO
LOE 3.1.3: Identify best practices related to strategies to preserve government intellectual property (IP) in the acquisition of AI and AI-enabled systems in order to ensure open architecture of secure data deliverables and rights that support the protection of government IP, best-value acquisition, avoidance of proprietary lock-in, and oversight of DoD use of AI capabilities and adherence to the DoD AI Ethical Principles.	OUSD(A&S) in coordination with CDAO

LOE 3.2: ADOPT INDUSTRY BEST PRACTICES FOR AI DEVELOPMENT BY DEVELOPING AND APPLYING TOOLS, TECHNOLOGIES, AND BEST PRACTICES TO IDENTIFY AND MITIGATE RISKS AS THEY RELATE TO THE AI CAPABILITY THROUGHOUT THE AI PRODUCT LIFECYCLE.

Lines of Effort	OPR
LOE 3.2.1: Continue developing a Product Toolkit that draws upon best practices and innovative research from the DoD enterprise, industry, and the academic community, as well as commercially available technology where appropriate. The Toolkit will be made widely available to DoD users and shall include: a) Defense Innovation Unit's (DIU) RAI Guidelines; b) Templates for AI project management with an emphasis on ensuring that developers have an understanding of user needs and operational context; c) AI Data Cards and Model Cards, and corresponding catalogs, with detailed instructions; and d) Any other tools, guidance for system and project documentation, or risk assessment frameworks as appropriate.	CDAO; OUSD(R&E) (for subtask a)

LOE 3.2.2: Provide guidance on how and when the RAI Tools in LOE 3.2.1 should be used across the AI Product Lifecycle, based on an assessment of the AI technology's level of technical maturity, project sensitivity, and overall risk. Such assessments of the AI technology will be determined by DoD Component RAI Leads.	CDAO in coordination with DoD Component RAI Leads
LOE 3.2.3: Use AI Data and Model Cards to publish AI data assets in the DoD federated data catalog, in accordance with the Department's *Creating Data Advantage* memorandum (May 5, 2021).	CDAO; DoD Component CDOs
LOE 3.2.4: Fund the development and piloting of new resources and tools that augment the RAI Toolkit.	CDAO; Military Departments
LOE 3.2.5: Publish best practices to preserve privacy and civil liberties and to avoid unintended bias in the design and development of AI capabilities that involve the use of personal information.	ATSD(PCLT)
LOE 3.2.6: Develop additional guidance for applying the RAI Foundational Tenets to the early-stage AI research and engineering projects, such as those funded by Budget Activities 1, 2, and 3.	OUSD(R&E) and CDAO in coordination with DoD Component RAI Leads

TENET 4: REQUIREMENTS VALIDATION

Description: Incorporate RAI into all applicable AI requirements, including joint performance requirements established and approved by the Joint Requirements Oversight Council, to ensure RAI inclusion and appropriate DoD AI capabilities.

LOE 4.1: INTEGRATE AI RISK CONSIDERATIONS INTO THE DOD JOINT PERFORMANCE REQUIREMENTS PROCESS BY IDENTIFYING RESPONSIBILITIES, AUTHORITIES, AND RESOURCES WHEN UPDATING OR DEVELOPING POLICIES AND PLANS.

Lines of Effort	OPR
LOE 4.1.1: Draft a JROC Memorandum (JROCM) with changes that need to be made in requirement-setting processes to implement the LOEs in Tenets 2 and 3. Make recommendations to the Vice Chairman of the Joint Chiefs of Staff (VCJCS) and the Deputy Secretary of Defense (DSD), as appropriate.	Joint Staff in coordination with USD(A&S), CDAO

LOE 4.2: DEVELOP A TAILORABLE PROCESS THAT PROGRAMS CAN FOLLOW TO WRITE AI-RELATED REQUIREMENTS THAT ARE TESTABLE AND OPERATIONALLY RELEVANT.

Lines of Effort	OPR
LOE 4.2.1: Create a repository of AI-related requirements for common use cases, mission domains, and system architectures to facilitate reusability.	CDAO in coordination with DoD Component RAI Leads
LOE 4.2.2: Integrate and coordinate the AI requirements captured in LOE 4.2.1 with the TEVV strategy to ensure that adequate methods exist for continuous testing and validation of capabilities developed, consistent with existing policies for iterative acquisition such as the DoDI 5000.87.	CDAO in coordination with OUSD(R&E) and DoD Component RAI Leads
LOE 4.2.3: Produce guidance for Program Executive Offices and Program Offices to apply the resources from LOE 3.1 and write RAI requirements into future contracts, including for non-ACAT and non-MDAP systems.	OUSD(A&S) in coordination with CDAO; DOT&E; Operational Test Agencies (OTAs)

TENET 5: RESPONSIBLE AI ECOSYSTEM

Description: 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.

LOE 5.1 COORDINATE AND COLLABORATE ACROSS DOD, THE INTELLIGENCE COMMUNITY, AND OTHER U.S. GOVERNMENT DEPARTMENTS AND AGENCIES ON THE RAI FOUNDATIONAL TENETS AS WELL AS ON THE DOD AI ETHICAL PRINCIPLES. ENGAGE WITH CONGRESS ON THE TENETS AND PRINCIPLES TO RAISE AWARENESS AND SUPPORT APPROPRIATE CONGRESSIONAL OVERSIGHT OF DOD'S RESPONSIBLE AI EFFORTS.

Lines of Effort	OPR
LOE 5.1.1: Identify any gaps in participation in CDAO-led governance bodies from DoD and IC Components and provide a plan to recruit and maintain new members.	CDAO
LOE 5.1.2: Coordinate regularly with Office of the Director of National Intelligence (ODNI) AI ethics team to ensure interoperability and alignment on the operationalization of the Intelligence Community (IC) AI Ethics Principles and the DoD AI Ethical Principles.	CDAO; DoD Component RAI Leads; ATSD(PCLT)
LOE 5.1.3: Coordinate regularly with appropriate Federal interagency bodies such as those housed in the Executive Office of the President, Office of Management and Budget, and the General Services Administration on RAI.	CDAO; DoD Component RAI Leads; ATSD(PCLT)
LOE 5.1.4: Develop a legislative strategy and ensure the strategy is clearly communicated across Department to ensure appropriate engagement with the CDAO and consistent messaging, technical assistance, and advocacy to Congress.	ASD(LA) in coordination with CDAO

LOE 5.2: BUILD ENDURING ENGAGEMENTS AND COLLABORATION ACROSS INDUSTRY, ACADEMIA, AND CIVIL SOCIETY TO PROMOTE DEVELOPMENT, ADOPTION, AND IMPLEMENTATION OF RAI.

Lines of Effort	OPR
LOE 5.2.1: Provide a prioritized list of research gaps in RAI-related fields to the White House National AI Initiative Office to encourage funding by Department of Education, National Science Foundation, and National Institute of Standards and Technology (NIST) as authorized by the National AI Initiative Act.	OUSD(R&E); CDAO
LOE 5.2.2: Ensure RAI expertise, including privacy, civil liberties, and data ethics expertise, is represented on the AI Advisory Board as authorized by FY21 National Defense Authorization Act, Section 233.	CDAO
LOE 5.2.3: Explore funding opportunities to establish a development program on RAI tools with industry, academia and the Department.	OUSD(R&E)
LOE 5.2.4: Develop and execute a public-affairs strategy for DoD RAI activities that is integrated across Department. This includes coordination of communications activities with the CDAO, regularly communicating DoD's RAI implementation progress as appropriate to through speaking engagements, press releases and conferences, blog posts, and op-eds.	ASD(PA) in coordination with CDAO
LOE 5.2.5: Produce guidance on the sharing or publication of DoD AI capabilities with entities outside the Department to preserve operational security and prevent unintended exposure.	ASD(PA) in coordination with OUSD(I&S) and CDAO

LOE 5.3: INTEGRATE RAI AS AN ELEMENT OF INTERNATIONAL ENGAGEMENTS, TO ADVANCE SHARED VALUES, LESSONS LEARNED, BEST PRACTICES, AND INTEROPERABILITY GLOBALLY.

Lines of Effort	OPR
LOE 5.3.1: Actively seek opportunities to engage allies and partners on RAI (including NATO, Five Eyes, Quadrilateral Security Dialogue, etc.) with particular emphasis on interoperability with partners and allies with data, compute, and storage systems, software, and schema.	OUSD(P) in coordination with CDAO

LOE 5.3.2: Continue to communicate, promote, and educate the DoD's AI Ethical Principles and RAI implementation to partners and allies through the DoD AI Partnership for Defense (AI PfD) as well as bilateral and multilateral engagements.	CDAO
LOE 5.3.3: Organize a workshop with representatives from the international community (academia, industry and government) on AI ethics, safety, and trust in defense in order to exchange best practices and promote shared values.	CDAO

TENET 6: AI WORKFORCE

Description: 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.

LOE 6.1: PARTICIPATE IN ONGOING EFFORTS TO DEVELOP THE DOD'S AI TALENT MANAGEMENT FRAMEWORK TO SUPPORT THE IDENTIFICATION, RECRUITMENT, ELEVATION, AND RETENTION OF MILITARY AND CIVILIAN PERSONNEL NEEDED FOR IMPLEMENTATION OF RAI.

Lines of Effort	OPR
LOE 6.1.1: Develop a mechanism to identify and track AI expertise across the Department and the Military Departments and Services, including the Acquisition Workforce, by a) leveraging existing coding efforts (e.g., DoD CIO's Cyber Workforce Framework expansion effort) and, if needed, developing new codes in coordination with the Office of Personnel Management (OPM); and b) developing standardized personnel coding mechanisms, recognizable by current and future civilian and military Service department personnel systems (e.g., Defense Civilian Human Resources Management System (DCHRMS), military integrated personnel systems).	OSD(P&R); DCPAS; Military Departments; OSD(A&S); DoD CIO; CDAO
LOE 6.1.2: Conduct a gap analysis to determine whether any additional knowledge, skills, abilities, and tasks are needed for the six archetypes captured in the 2020 DoD AI Education Strategy to successfully implement RAI.	CDAO in coordination with DoD Component RAI Leads
LOE 6.1.3: Develop career fields and pathways for military personnel who perform AI work as a major portion of their job, including promotion eligibility.	Military Department RAI Leads in coordination with

	CDAO and OUSD(P&R)
LOE 6.1.4: Recommend to OUSD(P&R) and OPM career fields and pathways for civilian personnel who perform AI work as a major portion of their job.	CDAO in coordination with DoD Component RAI Leads; OUSD(P&R)
LOE 6.1.5: Initiate workforce planning processes to attract, recruit, and maintain highly skilled experts to fill the gaps identified in LOE 6.1.2 to include (but not limited to) creating and reclassifying billets and adding AI positions in government performance of acquisition and requirements management functions.	DoD Component RAI Leads in coordination with OUSD(A&S) and OUSD(P&R)

LOE 6.2: SUPPLEMENT EXISTING DOD AI TRAINING EFFORTS WITH CURRICULA THAT WILL ENABLE RAI IMPLEMENTATION.

Lines of Effort	OPR
LOE 6.2.1: Develop and update DoD-wide standardized curricula (covering topics such as AI benefits and limitations, risk factors, and security) to be integrated into all AI education and training for the DoD AI workforce, and ensure that it is appropriate for all levels of both military and civilian seniority.	CDAO in coordination with DoD Component RAI Leads
LOE 6.2.2: Integrate the curricula developed in LOE 6.2.1 in Components' AI education and training programs (including initial and mission qualified training programs to build RAI skills capacity of current workforce and pipeline).	DoD Component RAI Leads
LOE 6.2.3: Establish education, training, and experience standards relevant to respective AI-related positions based on the level of complexity of duties for DoD Components.	DoD Component RAI Leads
LOE 6.2.4: Collaborate with the Defense Acquisition Workforce Functional Managers to reshape the training, education, and experience requirements for the Defense Acquisition Workforce to include AI curricula as appropriate. This includes: a) potentially updating Foundational and Practitioner Certifications; b) creating new AI credentials for RAI implementation roles; and c) leveraging curricula per LOE 6.2.1 to existing training curricula available for the Defense Acquisition Workforce.	OUSD(A&S); DAU

LOE 6.2.5: Explore the need of an annual AI Ethics Awareness training (similar to the annual Cyber Awareness training) requirement for DoD AI Workforce to promote awareness and application of the DoD AI Ethical Principles, as well as consistency of understanding.	CDAO

LOE 6.3: BUILD DOD CAPACITY FOR RAI THROUGH COMMUNITIES OF INTEREST/PRACTICE AND PROFESSIONAL DEVELOPMENT OPPORTUNITIES.

Lines of Effort	OPR
LOE 6.3.1: Deliver annually the RAI Champions Program Department-wide in order to build a network of RAI advocates and experts in DoD Components.	CDAO
LOE 6.3.2: Build RAI DoD-wide communities of interest or practice, leveraging existing bodies to accelerate their establishment, such as the RAI Subcommittee. Scale the RAI Champions Training Program through DoD (via Train the Trainer model) to create network of champions who can participate in these Communities of Interest/Communities of Practice (COIs/COPs).	CDAO
LOE 6.3.3: Identify funding or partnership opportunities for non-DoD training on AI ethics in other agencies, academia, and industry.	DoD Component RAI Leads

CONCLUSION

The Department's charge is to protect the American people through integrated deterrence and—when called upon—to fight and win our Nation's wars. The future security and prosperity of the United States depend on the warfighters' ability to effectively and responsibly wield AI-enabled technologies. Additionally, maintaining trust with those external to the Department is equally vital.

To achieve this, it is not enough to communicate our values through the DoD AI Ethical Principles; we must also live by them. The articulation of the Principles was an essential milestone for guiding the ethical use of AI for military applications and represents the beginning of a broader effort around scaling the implementation of RAI.

The implementation of DoD policies in the past has required tools and guides including manuals, doctrine, and technical resources. Similarly, the DoD AI Ethical Principles will also require implementation tools to provide concrete guidance on what the principles mean and how to apply them in context. Such guidance will enable successful implementation of RAI practices by providing tools and best practices—now and as they are developed.

Furthermore, incorporation of the Principles into the DoD's culture and operational execution does not rest upon any individual (e.g., a developer or user) or process (e.g., TEVV or acquisition) but is a collective effort that involves a multitude of personnel. Appropriate implementation of the Principles across the system's lifecycle will enable warfighters to achieve their missions responsibly, effectively, and efficiently.

Ultimately, if AI systems are not designed and developed responsibly, the Department stands to lose trust, at home and abroad. RAI must become a core component of DoD's AI transformation. The DoD will continue to lead by example and promote a vision for ethical and safe military use of AI, through the implementation of this RAI Strategy.

ADDITIONAL RESOURCES

The list of tools and additional resources that can be utilized and/or customized for use by the DoD to implement this RAI S&I Pathway is not static. As tools and resources are created, modified, or adopted by the Department, the RAI Toolkit will change.

The CDAO, in coordination DoD Component RAI Leads, shall make these resources widely and easily available.

The RAI toolkit includes the following available resources:

- DIU RAI Guidelines
- Templates for AI Project Management
- AI Data Card template
- AI Model Card template
- HSI Framework

Additional toolkit resources directed by this plan include:

- AI T&E Toolkit
  - T&E Master Plan template for AI
  - Library of T&E metrics, to include those used to assess trustworthiness and confidence
  - Tools and technologies to detect both adversarial attacks on AI and natural degradation of AI system performance
- Acquisition Toolkit
  - Standard language in the initial announcement, RFP, and RFI for AI capabilities
  - RAI-related evaluation criteria that are testable and operationally relevant
  - Standard AI contract language
- Repository of AI-related requirements for common use cases, mission domains, and system architectures to facilitate reusability

Neither of these lists is exhaustive; however, DoD Component RAI Leads should widely promulgate the existence of these resources within their DoD Components and provide easy-to-use mechanisms by which DoD personnel can gain access to any needed resources.

ACRONYMS

Acronym	Acronym for
ACAT	Acquisition Category
AI	Artificial Intelligence
AI/ML	Artificial Intelligence/Machine Learning
AI PfD	AI Partnership for Defense
ASD(LA)	Assistant Secretary of Defense for Legislative Affairs
ASD(PA)	Assistant Secretary of Defense for Public Affairs
ATSD(PCLT)	Assistant to the Secretary of Defense for Privacy, Civil Liberties, and Transparency
CAPE	Cost Assessment and Program Evaluation
CDAO	Chief Digital and Artificial Intelligence Officer
CIO	Chief Information Officer
COI	Community of Interest
COP	Community of Practice
DAU	Defense Acquisition University
DCHRMS	Defense Civilian Human Resources Management System
DCPAS	Defense Civilian Personnel Advisory Service
DIB	Defense Innovation Board
DIU	Defense innovation Unit
D,DTE&A	Director for Developmental Test, Evaluation, and Assessments
DoD	Department of Defense
DoD CDO	Department of Defense Chief Data Officer
DoD CIO	Department of Defense Chief Information Officer
DoD OGC	Department of Defense Office of General Counsel
DoDD	Department of Defense Directive
DOE	Department of Energy
DOT&E	Director, Operational Test and Evaluation
DSD	Deputy Secretary of Defense
FCB	Functional Capabilities Board
FY	Fiscal Year
FYDP	Future Years Defense Program
GG	General Government
HSI	Human Systems Integration
IC	Intelligence Community
IP	Intellectual Property
JCIDS	Joint Capabilities Integration and Development System
JCB	Joint Capabilities Board
JROC	Joint Requirements Oversight Council

JROCM	Joint Requirements Oversight Council Memorandum
KM/DS	Knowledge Management/Decision Support (KM/DS)
KPP	Key Performance Parameter
KSA	Key System Attributes
LOE	Line of Effort
MDAP	Major Defense Acquisition Program
MIL-STD	United States Military Standard
MPRD	Minimum Product Requirement Document
NATO	North Atlantic Treaty Organization
NIST	National Institute of Standards and Technology
NSF	National Science Foundation
OCDAO	Office of the Chief Digital and Artificial Intelligence Officer
ODNI	Office of the Director of National Intelligence
OPM	Office of Personnel Management
OPR	Office of Primary Responsibility
OSD	Office of the Secretary of Defense
OTA	Operational Test Agency
OUSD(A&S)	Office of the Under Secretary of Defense for Acquisition and Sustainment
OUSD(P)	Office of the Under Secretary of Defense for Policy
OUSD(P&R)	Office of the Under Secretary of Defense for Personnel and Readiness
OUSD(R&E)	Office of the Under Secretary of Defense for Research and Engineering
PEO	Program Executive Office
PM	Program Manager
PSA	Principal Staff Assistant
R&D	Research and Development
RAI	Responsible Artificial Intelligence
RDT&E	Research, Development, Test, and Evaluation
RFI	Request for Information
RFP	Request for Proposal
S&I	Strategy and Implementation
SBIR/STTR	Small Business Innovation Research/Small Business Technology Transfer
T&E	Test and Evaluation
TEMP	Test Evaluation Master Plan
TEVV	Test, Evaluation, Verification, and Validation
TRMC	Test Resource Management Center
U.S.	United States

USSOCOM	United States Special Operations Command
UX/UI	User Experience/User Interface
V&V	Verification and Validation
VCJCS	Vice Chairman of the Joint Chiefs of Staff

GLOSSARY

Term	Definition
Algorithm *source: DARPA*	A method or set of rules or instruction to be followed in calculations or other problem-solving operations, particularly by a computer.
Artificial Intelligence *source: DoD AI Strategy (2018)*	AI refers to the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems.
Autonomy *source: DARPA*	Autonomy refers to a system's ability to accomplish goals independently, or with minimal supervision from human operators in environments that are complex and unpredictable.
Autonomous Weapon System *source: DoDD 3000.09*	A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation.
Data Card *source: JAIC, now CDAO*	A document for a dataset that provides insight into collection, processing, usage, and security practices.
DoD Component Responsible AI Lead *source: CDAO*	Individuals designated or hired by the CDAO governing council member organizations to be responsible for ensuring the adoption, integration, and implementation of Responsible AI programs and processes within their respective Components.
Equitable	DoD AI Ethical Principles: The Department will take deliberate steps to minimize unintended bias in AI capabilities

source: DoD Memorandum, "Artificial Intelligence Ethical Principles for the Department of Defense" (Feb 2020)	
Explainability *source: NSCAI Final Report*	A characteristic of an AI system in which there is provision of accompanying evidence or reasons for system output in a manner that is meaningful or understandable to individual users (as well as to developers and auditors) and reflects the system's process for generating the output (e.g., what alternatives were considered, but not proposed, and why not).
Governable *source: DoD Memorandum, "Artificial Intelligence Ethical Principles for the Department of Defense" (Feb 2020)*	DoD AI Ethical Principles: 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
Harm *source: DARPA*	Suffering, misfortune, or wrongdoing (physical or otherwise) as done to or suffered by some person or thing; hurt, injury, damage, or mischief; to do harm to or injure (physically or otherwise); to hurt, damage.
Human Systems Integration *source:OUSD(R&E)*	A comprehensive, interdisciplinary management and technical approach applied to system development and integration as part of a wider systems engineering process to ensure that human performance is optimized to increase total system performance and minimize total system ownership costs.
Key Performance Parameter *source: JCIDS Manual*	Performance attribute of a system considered critical or essential to the development of an effective military capability. KPPs are contained in the Capability Development Document (CDD) and the updated CDD and are included verbatim in the Acquisition Program Baseline (APB). KPPs are expressed in term of parameters which reflect Measures of Performance (MOPs) using a threshold/objective format. KPPs must be

	measurable, testable, and support efficient and effective Test and Evaluation (T&E). Mandatory KPPs are specified in the JCIDS Manual.
Military Department and Service RAI Lead *source: JAIC, now CDAO*	A subset of DoD Component RAI Leads that includes the Military Departments (Department of the Air Force, Department of the Army, Department of the Navy) and the Military Services (Air Force, Space Force, Army, Marine Corps, Navy, and Coast Guard).
Model Card *source: JAIC, now CDAO*	A document that communicates the development processes and limitations of a model to enable developers, policymakers, and users to understand aspects of trained models.
Machine Learning *source: NSCAI Final Report*	The study or the application of computer algorithms that improve automatically through experience. Machine learning algorithms build a model based on training data in order to perform a specific task, like aiding in prediction or decision-making processes, without necessarily being explicitly programmed to do so
Reliable *source: DoD Memorandum, "Artificial Intelligence Ethical Principles for the Department of Defense" (Feb 2020)*	DoD AI Ethical Principles: 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
Responsible *source: DoD Memorandum, "Artificial Intelligence Ethical Principles for the Department of Defense" (Feb 2020)*	DoD AI Ethical Principles: DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities
Responsible AI *source: DoD Memorandum, "Implementing Responsible Artificial Intelligence in the Department of Defense" (May 2021)*	Responsible AI is a dynamic approach to the design, development, deployment, and use of artificial intelligence systems that implements the DoD AI Ethical Principles to advance the trustworthiness of such systems.
Responsible AI Champions *source: JAIC, now CDAO*	A community of individuals who are knowledgeable advocates of the DoD AI Ethical Principles who are able to help to

	operationally define the principles substantively and through engineering practices and who will engage in peer-to-peer teaching within their areas.
Risk *source: JAIC, now CDAO*	The potential for the use of a technology or system to result in negative outcomes due to an impact on the organization, individuals, or society.
Robust AI *source: NIST*	An AI system that is resilient in real-world settings, such as an object-recognition application that is robust to significant changes in lighting. The phrase also refers to resilience when it comes to adversarial attacks on AI components
Test and Evaluation, Verification and Validation *source: NSCAI Final Report*	A framework for assessing, incorporating methods and metrics to determine that a technology or system satisfactorily meets its design specifications and requirements, and that it is sufficient for its intended use.
Test & Evaluation *source: DAU*	Test & Evaluation (T&E) is the process by which a system or components are compared against requirements and specifications through testing. The results are evaluated to assess progress of design, performance, supportability, etc. Developmental test and evaluation (DT&E) is an engineering tool used to reduce risk throughout the acquisition cycle. Operational test and evaluation (OT&E) is the actual or simulated employment, by typical users, of a system under realistic operational conditions.
Traceable *source: DoD Memorandum, "Artificial Intelligence Ethical Principles for the Department of Defense" (Feb 2020)*	DoD AI Ethical Principles: 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 procedure and documentation

Trust *source: NIST, IEEE, ISO*	Trust is established by ensuring that AI systems are cognizant of and are built to align with core values in society, and in ways which minimize harms to individuals, groups, communities, and societies at large. Defining trustworthiness in meaningful, actionable, and testable ways remains a work in progress. In part, we rely on the practice of trustworthy computing as adopted by some in computer science and system engineering fields—"trustworthiness of a computer system such that reliance can be justifiably placed on the service it delivers (IEEE)"; "of an item, ability to perform as and when required (ISO/IEC/IEEE)". On other hand, the AI user trust decision, as other human trust decisions, is a psychological process. There is currently no method to measure user trust in AI or measure what factors influence the users' trust decisions.
Use Case *source: DARPA*	A use case is a specific situation in which a product or service could potentially be used.

REFERENCES

(A) Department of Defense. *Autonomy in Weapon Systems*. DoD Directive 3000.09. Washington, DC: Department of Defense, 2012. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf?ver=2019-02-25-104306-377.

(B) Department of Defense. *DoD Executive Agent for Non-Lethal Weapons (NLW), and NLW Policy*. DoD Directive 3000.03E. Washington, DC: Department of Defense, 2013. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300003p.pdf?ver=2018-10-24-112944-467.

(C) Department of Defense. *Operation of the Adaptive Acquisition Framework*. DoD Instruction 5000.02. Washington, DC: Department of Defense, 2020. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf.

(D) Department of Defense. *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*. Washington, DC: Department of Defense, 2018. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

(E) Department of Defense. *System Safety*. MIL-STD-882E. Washington, DC: Department of Defense, 2012. https://www.dau.edu/cop/armyesoh/DAU%20Sponsored%20Documents/MIL-STD-882E.pdf.

(F) Department of Defense. *The Defense Acquisition System*. DoD Directive 5000.01. Washington, DC: Department of Defense, 2003. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500001p.pdf.

(G) Department of Defense. *2020 Department of Defense Artificial Intelligence Education Strategy*. Washington, DC: Department of Defense, 2020. https://www.ai.mil/docs/2020_DoD_AI_Training_and_Education_Strategy_and_Infographic_10_27_20.pdf.

(H) Hicks, Kathleen. "Implementing Responsible Artificial Intelligence in the Department of Defense." Official memorandum. Washington, DC: Department of Defense, 2021. https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF.

(I) Hicks, Kathleen. "Establishment of the Chief Digital and Artificial Intelligence Officer." Official memorandum. Washington, DC: Department of Defense, 2021. https://media.defense.gov/2021/Dec/08/2002906075/-1/-1/1/MEMORANDUM-ON-ESTABLISHMENT-OF-THE-CHIEF-DIGITAL-AND-ARTIFICIAL-INTELLIGENCE-OFFICER.PDF.

(J) National Artificial Intelligence Initiative Act of 2020, Pub. L. No. 116-283 § 5001 (2021). https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210.

(K) William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116-283 (2021). https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf.