Department of Defense
Joint Artificial Intelligence Center
Responsible AI Champions Pilot

Ethics is a foundational element of the Department of Defense's Artificial Intelligence (AI) strategy. To guide the Department's design, development, deployment and use of AI-enabled applications, Secretary Esper adopted AI Ethical Principles for the DoD ("Principles") in February 2020. The Joint Artificial Intelligence Center (JAIC) piloted a Responsible AI Champions program to bring these Principles to life. The pilot led a cross-functional group of individuals through an experiential learning journey to develop an understanding of the Principles, consider ways to operationalize them and to create a community of responsible AI ambassadors.

> *"This accomplishment marks an inflection point for the JAIC and DoD AI adoption. The Department is taking firm steps to take AI Ethics as a concept to AI Ethics as actionable and concrete components of how we manage this transformative technology – from first thoughts, to acquisition, to development and deployment, through sustainment and eventual retirement."*
>
> *- Nand Mulchandani, Acting Director of the JAIC*

Overview

The Department derived the five Ethical Principles—**Responsible**, **Equitable**, **Traceable**, **Reliable**, and **Governable**—from a robust and comprehensive 15-month study undertaken by the Defense Innovation Board, which included public consultation with a multidisciplinary group of experts from academia and industry, as well as current and former DoD leaders, and the public. These Principles align with existing ethical, legal and policy mechanisms, and national strategies. They also memorialize the Department's commitment to be a responsible leader in AI.

Responsible AI Champions Pilot

With an eye on action and understanding that ethics live and manifest in organizational operations, then director of the JAIC, Lt Gen Jack Shanahan, kicked off the Responsible Artificial Intelligence Champions pilot in April with a 15-member cross-functional cohort across the JAIC. The virtual pilot spanned over 10 weeks and included a formal kick-off, multi-day training sessions, eight weekly 90 minutes interactive sessions (with pre-reading requirements), a 3-hour final workshop and a concluding ceremony.

The inaugural cohort included a group of individuals embedded within strategic areas of the JAIC who represent the AI product lifecycle, such as product design and development, testing and evaluation/verification and validation, and acquisition teams as well as policy, plans, and performance. The interactive weekly sessions aimed to provide an experiential learning environment for the cohort by utilizing a mix of instruction (which was provided by a multi-disciplinary group of instructors), breakout discussions, and case studies.

> *"…. The case studies, the discussions with my brilliant colleagues, the guest speakers, and the progressively complex, intentional approach to an incredibly challenging but arguably no-fail element of AI development and adoption for the Department, resulted in my coming out of this Cohort with way more questions than answers; way more aware of our limitations in this space – both technologically and in terms of strategic forethought – but also resolute in my professional obligation, and strong personal desire, to be a part of the thought leadership and any proposed solutions."*
>
> *- Participant*

Objectives:

This pilot was designed with three primary objectives. The first was for the cohort to establish a common vocabulary around ethics and AI and gain an understanding of the DoD AI Ethical Principles, their inception, their necessity, and their impact. Another objective was to leverage the multi-disciplinary experiences of the participants to gain insights into what an implementation or operationalization framework of the DoD AI Ethical Principles may look like (throughout the design, development, deployment and use of AI-enabled technologies) from the perspective of each of their respective functional roles. The final objective was directed to building a community of "champions" that will serve as ambassadors for responsible AI and the Principles across the JAIC and the DoD.

Structure:

During the weekly sessions, the cohort engaged in deep dives into each of the five DoD AI Ethical Principles and other related topics, such as civil liberties, privacy, and security. They focused on the need to identify and ensure these values in the product lifecycle, and reconciling tradeoffs, if any. A multidisciplinary group of individuals from Carnegie Mellon University, Defense Innovation Unit, JAIC, The MITRE Corporation, and West Point provided instruction for the pilot.

The JAIC designed the cohort to ensure representation of multidisciplinary backgrounds as well as representation of each core functional area of the JAIC for a number of reasons. For example, this allowed for the breakout discussions to be curated on a weekly basis to either reinforce the desired learning objective or elicit a more robust dialogue and exchange of ideas or, alternatively, challenging of assumptions and creation of new ideas. As such, some breakout sessions included groupings of similar backgrounds/functional areas while in other sessions the groups were designed to have a mix of technical and non-technical participants to allow for peer-to-peer learning. The outcome of the discussions allowed for maximum impact across the organization.

> *"The RAIC pilot has been a catalyst for developing an ethical AI lens. This forum has empowered the cohort to discuss the complexities surrounding the development/use of ethical AI by facilitating cross-cutting team collaboration, dissecting the ethical principles, and strategizing ethical implementing tactics, techniques, and procedures."*
>
> *- Participant*

Discussion Insights:

The following illustrates a few observations from the robust discussions over the many sessions of the pilot.

On human-centered AI:

- AI should be used to support human decision making, rather than be used as the decision-making tool.
- Human choices, assumptions, simplifications, and trade-offs made along the AI product lifecycle shape the outcomes of the algorithm. Though we do not have to understand every aspect of the AI system (use case specific), we do need to understand the choices, assumptions, simplifications, and trade-offs of the people who designed the system.
- The transition from development to the end user is critical; the end user needs to understand how the system operates, how to interpret and utilize the system's output, and how to recognize when the system is not functioning as intended.

On operationalizing the Principles:

- Ethics and responsible AI should be considered during the design phase of any AI-enabled project and throughout the project lifecycle rather than as an afterthought.
- Data is not neutral; we should include diverse experiences and backgrounds, to help identify different undesired outcomes; we should also start the design process with the assumption that bias exists.
- Training, documentation, and oversight can help stakeholders appropriately calibrate their trust to the system (not over-trust it or under-trust it) and better understand its intended and unintended uses.

> *"These ethical principles are a foundational component in the design of the JAIC's solutions."*
>
> *- Matthew Rose, JAIC Chief Design Officer*

Additionally, during weekly sessions that focused on the Principles, the discussions yielded specific insights (as shown in the graphic below) deconstructing each Principle into themes and further, examples of tactical steps used to instantiate each Principle.

DoD AI Principles	Themes	Tactics
Responsible "DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities."	• Create a culture where everyone feels responsible for questioning safety and ethics at all points of lifecycle • Build processes that allow for policy, legal and ethical reviews at the earliest stages of development (design or earlier) stage, and throughout lifecycle • Proactive approach in design, development, and deployment • (…)	• Define explicit roles and responsibilities • Create cross-functional/multidisciplinary project teams • Utilize checklists with checkpoint assessments (recognizing that answers are not binary) • (…)
Equitable "The Department will take deliberate steps to minimize unintended bias in AI capabilities"	• Take deliberate steps to minimize unintended bias • Frame problem statement from perspective of all stakeholders • Account for statistical, social, and human biases • (…)	• Identify entry points for bias and interject controls to measure/mitigate/test for bias throughout the lifecycle • Consider how other actors may use the data or model which may lead to unintended outcomes • Rigorous testing aimed at reducing risk, bias, and harm reduction • (…)
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."	• Ensure relevant personnel possess an appropriate understanding of technology/processes/methods • Transparent and auditable methodologies, data sources & design procedure across lifecycle • Documentation for traceability and knowledge transfer • (…)	• Required training and use of multidisciplinary teams of domain experts and technical SMEs • Policies, instructions and responsible authorities to ensure consistency of understanding and interpretation • Use of tools (e.g. data cards, model cards, etc.…) and multi-disciplinary assessments • (…)
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."	• Ensure explicit well-defined uses • Consider robustness of system including safety, security and effectiveness • Identify performance parameters/benchmarks for testing and assurance within uses and across their life-cycle • (…)	• Incorporate feedback loops for operators on status/robustness of system • Implement continuous performance monitoring • Establish and document safety requirements, processes, and system safety approach • (…)
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."	• Design AI capabilities to fulfill their intended functions • Ability to detect and avoid unintended consequences • Ability to disengage or deactivate systems that demonstrate unintended behavior • (...)	• Identify ontology of potential failures • Design system with potential failure as an assumption • Design human or automated disengagement, kill switches and/or emergency stops where appropriate • (…)

Outcomes:

In addition to the objectives described above, the pilot more significantly, provided a vehicle by which to create a community of Responsible AI Champions. These Champions will serve as ambassadors where they will advise, educate, and inform. They will advise by utilizing their newfound knowledge to identify ways to interject processes and concepts to operationalize the Principles within their own roles as well as within their functional areas. Concurrently by doing so, their respective teams the benefit from peer-to-peer learning provided by the Champions thereby scaling awareness and creating impact across the JAIC. In addition to advising and educating, the Champions will also inform by becoming the organization's eyes and ears on the ground to identify new implementation strategies and escalate concerns. The cumulative effect of the pilot is to create culture change and increase muscle memory around the DoD AI Ethical Principles and responsible AI.

> *"The RAIC program changed the DoD's Ethical AI Principles from being a policy document I needed to read for situational awareness, to instead be an invaluable set of guideposts that influence my thinking about my work and about the AI technologies and the media pieces I encounter daily."*
>
> *- Participant*

What Comes Next

Just as AI/ML systems continually learn, require a feedback loop, as well as continuous monitoring, so to must our efforts around implementation of the DoD AI Ethical Principles and responsible AI. This pilot was one step in the DoD's implementation journey. And as we look to refine and expand our efforts of shifting from *principles to practice* as the DoD community continues to adopt AI, we know that the Department will ever increasingly be looking to ensure that the design, development, deployment and use of AI is responsible and aligned with the five DoD AI Ethical Principles. The JAIC will be looking to build off the momentum of this pilot as it further builds out the curriculum and looks to scale the effort across the DoD.

About the JAIC:

The mission of the Joint Artificial Intelligence Center is to accelerate the adoption and integration of artificial intelligence (AI) in the Department of Defense to achieve mission impact - at scale. The JAIC serves as the focal point for the execution of the DoD's AI Strategy that supports the U.S. National Defense Strategy (2018). As the DoD center excellence for AI, the JAIC continues to attract the best and brightest people from government, industry, and academia to carry out its vision to transform the U.S. military through AI.

The JAIC acknowledges The MITRE Corporation for its contributions and support of this pilot program.

For more information on the Responsible AI Champions Pilot or the DoD AI Ethical Principles, please contact Alka Patel, Head of AI Ethics Policy, JAIC.