



# DoD AI Ethics Principles & Autonomy

## Alka Patel

### Chief, Responsible AI

Department of Defense  
Joint Artificial Intelligence Center

March 2, 2021

Controlled by: Joint Artificial Intelligence Center  
Controlled by: Responsible AI Area  
CUI Category(ies): OPSEC  
Distribution/Dissemination Control: DistroA  
POC: [alka.a.patel2.civ@mail.mil](mailto:alka.a.patel2.civ@mail.mil)



# Objectives

- 1 **History:** Path to Principles
- 2 DoD AI Ethics Principles & Responsible AI Framework
- 3 *DoD AI Ethics Principles & Autonomy*
- 4 **Operationalization:** Principles to Practice

## 2 Year Journey

**2018 DoD  
AI Strategy**  
*Jan 2018*

Identified **leading in military ethics and AI safety** as one of 5 strategic focus areas as DoD adopts AI to advance security, peace, and stability

**NDAA 2019**  
*Aug 2018*

Called on the DoD to develop appropriate **ethical, legal, and other policies governing the development and use of artificial intelligence**

**Defense  
Innovation  
Board  
Report**  
*Oct 2019*

Completed 15-month study on ethical principles for AI with **public listening sessions, expert roundtables, and 100+ interviews**

**SecDef  
Memo  
adopting  
Principles**  
*Feb 2020*

Informed by DIB recommendations, **Ethical Principles for AI are adopted** and **JAIC is designated as lead** for coordinating implementation and oversight

# Department of Defense AI Ethical Principles Apply Across AI Product Lifecycle

	Principle	Description
	<b>Responsible</b>	DoD personnel will exercise <b>appropriate levels of judgment and care</b> , while remaining responsible for the development, deployment, and use of AI capabilities.
	<b>Equitable</b>	The Department will take deliberate steps to <b>minimize unintended bias</b> in AI capabilities.
	<b>Traceable</b>	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 <b>transparent and auditable methodologies, data sources, and design procedure and documentation</b> .
	<b>Reliable</b>	The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be <b>subject to testing and assurance within those defined uses across their entire life-cycles</b> .
	<b>Governable</b>	The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the <b>ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems</b> that demonstrate unintended behavior.



UNCLASSIFIED

"Ethical principles are the **foundation and guide posts** for everything we do with artificial intelligence in the Department ... **our ethical baseline is the core of our trusted-AI ecosystem...**"

- **Lt. Gen. Michael Groen**  
Director of the JAIC

Source: JAIC Blog (26 February 2021)  
AI Ethical Principles - Highlighting the  
Progress and Future of Responsible AI in the DoD

# Responsible AI at the Department

## DoD Ethical Principles for AI



## What is Responsible AI?

**Responsible AI** combines *technology*, *organizational culture* and *operating structures* to ensure AI systems are designed, developed, deployed, and used in a manner that:

- **Increases the effectiveness and efficiency** of our forces
- **Empowers everyone** to apply the Ethical Principles for AI
- Adheres to the Department's **core values**, U.S. laws, **Law of War**, and the nation's **democratic values**
- Transparently promotes security, peace, and stability to the **benefit of our society**

# Implementing Responsible AI at the DoD

## Responsible AI Framework

### People

- Roles & Responsibilities
- Training and Education
- Escalation Path

### Processes

- Tools/frameworks (e.g., use case triage, data cards)
- Integration into existing workflows & platforms
- Governance

### Partnerships

- Build/manage partnerships across industry, academia, & allies
- Collaborate with acquisition partners

### Policy

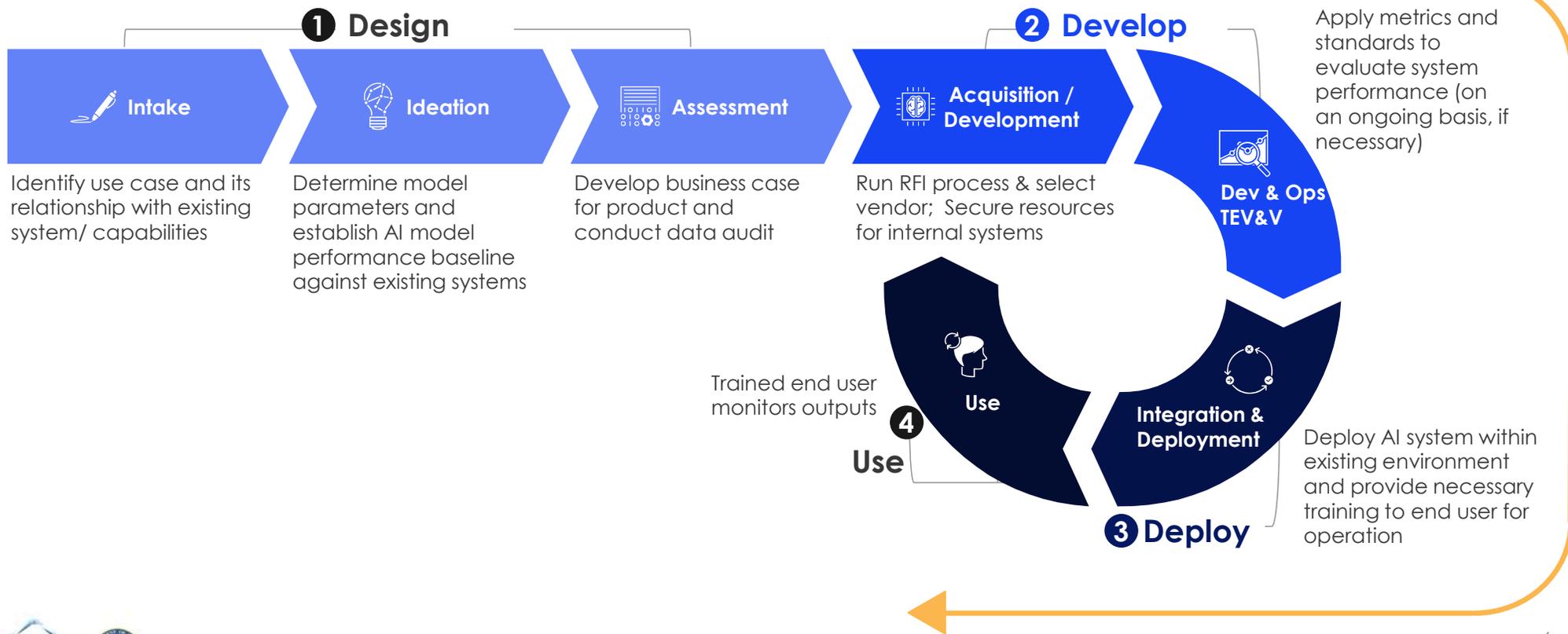
- Develop RAI implementation recommendations for Policy
- Consider Policy barriers for AI adoption



UNCLASSIFIED

# Responsible AI From the Start and Throughout

## Ethics Embedded Through the Entire AI Product Lifecycle



## How do DoD Ethical Principles for AI relate to Autonomy?

- The principles are a **reflection of the Department's core principles, the Law of War, and the nation's democratic values** that apply to both AI and autonomy.
- **Operationalization efforts are key** and require different tools and processes depending on the use case.
- The **implementation plan** for AI ethical principles should include the unique characteristics of autonomous systems and tailor specific solutions.
- AI is an **enabling technology** that drives autonomy. So **how** can the AI principles apply to autonomy?

# How can AI Ethics Principles address Risks/Concerns of Autonomy?

ARTIFICIAL INTELLIGENCE, EMERGING TECHNOLOGY

## Who Is Responsible When Autonomous Systems Fail?

**THE BLACK BOX, UNLOCKED**  
PREDICTABILITY AND UNDERSTANDABILITY IN MILITARY AI

*Algorithmic Bias in Autonomous Systems*

As Killer Robots Loom, A Push to Keep Humans in Control of Use of Force



UNCLASSIFIED

# Illustrative Example of Applying DoD AI Ethics Principles

## Concerns with Autonomy

**Unintended Behavior  
(Uncontrollable)**



### 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.

**Unpredictability  
Lack of Explainability**



### 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 procedure and documentation.

**Appropriate Trust**



### 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.

**Accountability Gap**

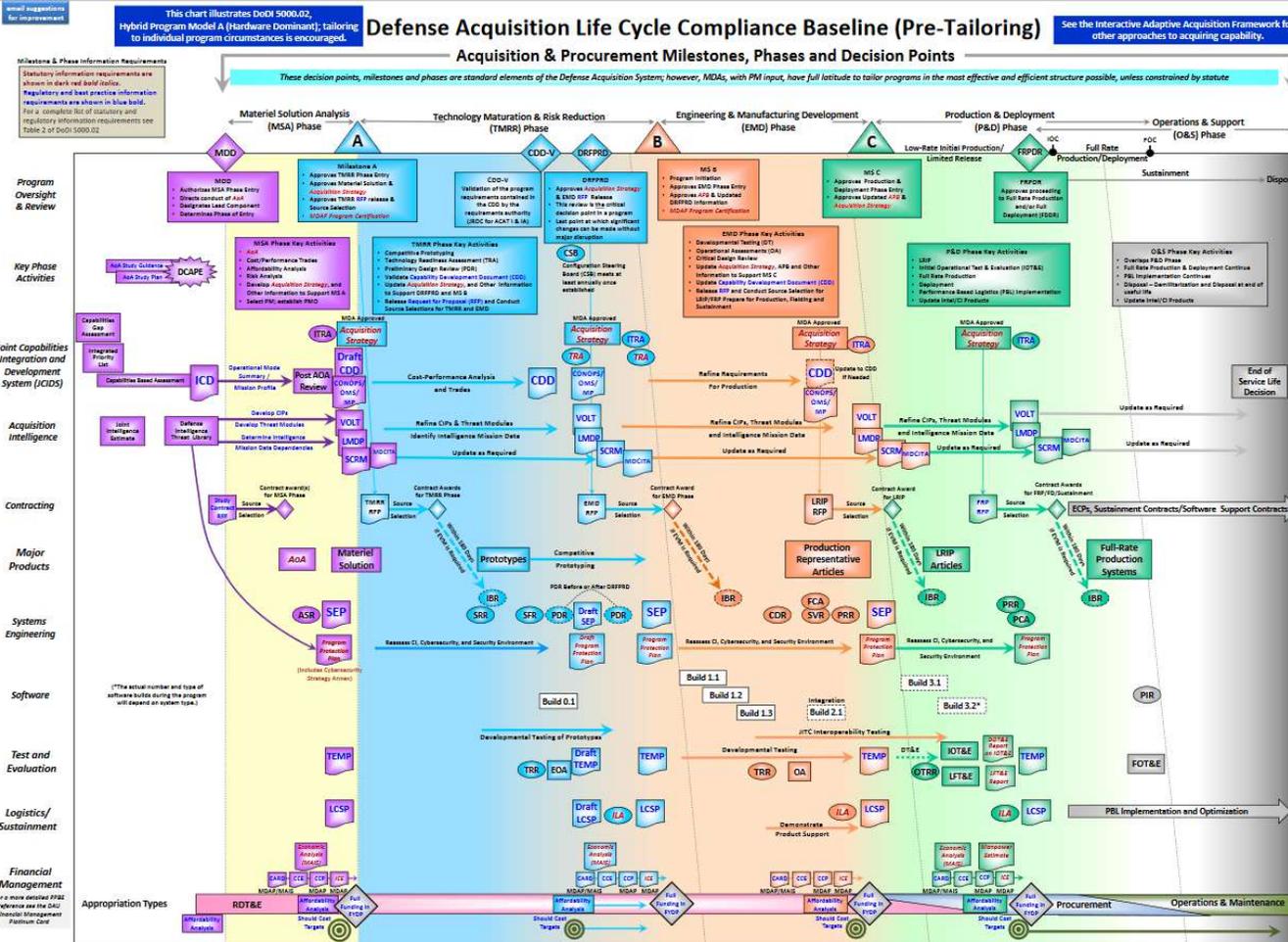


### RESPONSIBLE

DoD personnel will exercise appropriate levels of judgement and care, while remaining responsible for the development, deployment, and use of AI capabilities.



# DoD's Existing Frameworks and Reviews!



UNCLASSIFIED

## Food for Thought...

- 1 How do existing legal, ethical and policy frameworks, including the AI Ethics Principles, support autonomy?
- 2 What gaps might remain in “operationalizing” the AI Ethics Principles for autonomy?
- 3 How might we need to be thinking about organizational operating structures and practices (e.g. escalation, reporting governance, etc...) for autonomy?