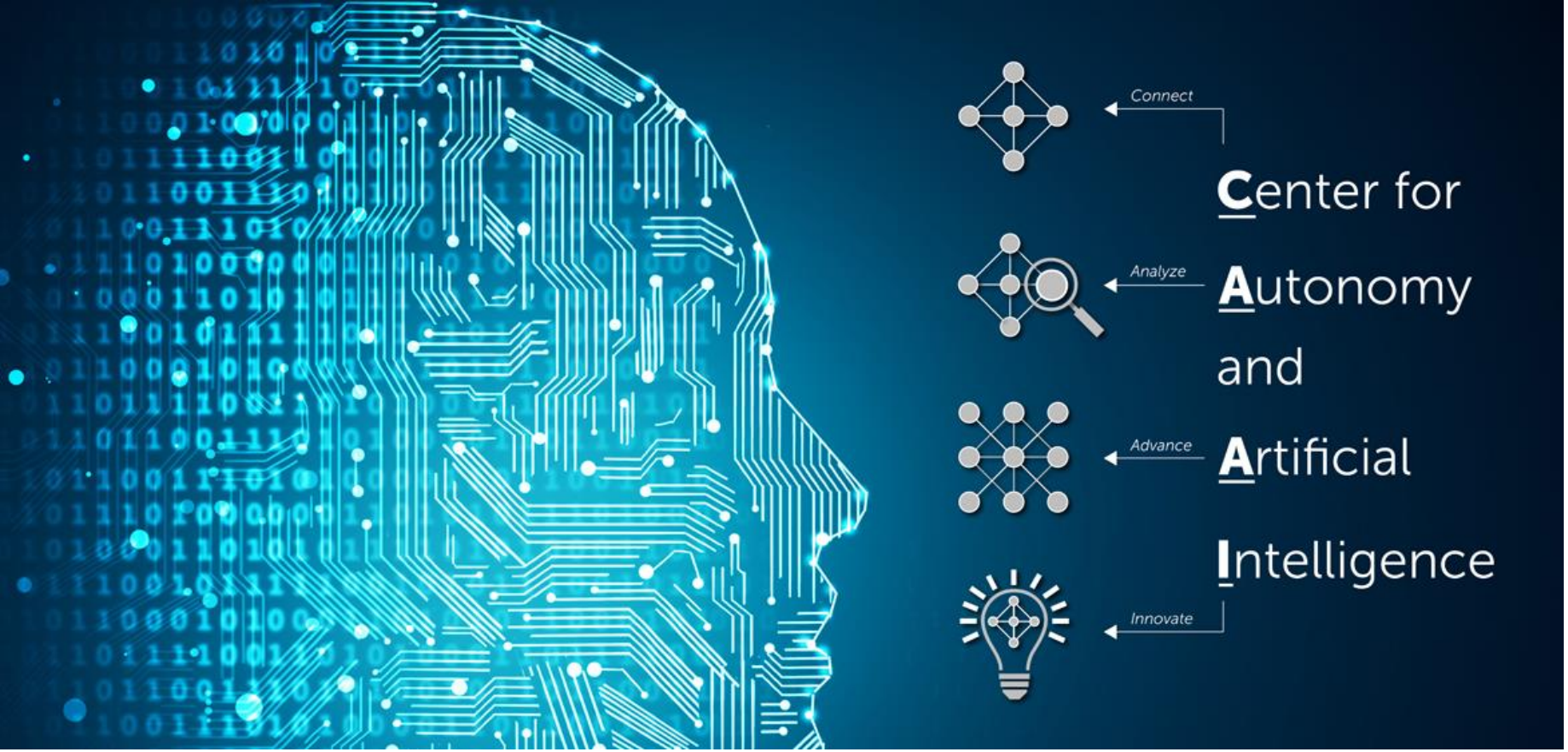


Autonomy and AI: Study Overview

Dr. Larry Lewis

March 2021

CNA ANALYSIS & SOLUTIONS

Approved for public release; distribution unlimited

CNA's Program on AI and Autonomy

- CNA began a dedicated program in 2017 to pull different parts of CNA together for a more comprehensive approach
  - Work across divisions to combine skill sets and areas of expertise
- Areas of focus include:
  - Overall strategy, approach, and priorities to move fast and effectively
  - Developing and refining specific cases for applying these technologies
  - Understanding and prioritizing needed enablers
  - Human-machine teaming
  - AI ethics and safety (including "AI for good")
  - Using AL/ML for practical applications (Data Science Division)

- Today I will talk about a new study—and give a few thoughts from existing studies

CNA

New Study: Optimal Classification of IAS

- How can the Navy consider the advantages and disadvantages of different possibilities of the classification of an IAS to maximize military advantages and reduce risks?
- Background:
  - IAS may receive a certain classification based on policy and/or legal considerations, which will be determined by many variables
  - The options for legal and policy classification of IAS (a collective “classification” of an IAS) come with advantages and disadvantages that may prove to be more or less important for specific cases

Study Methodology

- Step 1: characterize the legal and policy-based constraints and considerations that determine legal and policy classification for a representative range of potential IAS, given many variables

CNA

Study Methodology

- Step 1: characterize the legal and policy-based constraints and considerations that determine legal and policy classification for a representative range of potential IAS, given many variables
- Step 2: identify and characterize other factors relevant to decisions regarding the employment and use of IAS

CNA

Study Methodology

- Step 1: characterize the legal and policy-based constraints and considerations that determine legal and policy classification for a representative range of potential IAS, given many variables
- Step 2: identify and characterize other factors relevant to decisions regarding the employment and use of IAS
- Step 3: map the many connections among legal, policy, and other considerations and use this mapping to develop a framework to help guide decisions and trade-offs regarding the optimal classification of IAS
  - We will validate and refine this standing framework through a workshop

Study Methodology

- Step 1: characterize the legal and policy-based constraints and considerations that determine legal and policy classification for a representative range of potential IAS, given many variables
- Step 2: identify and characterize other factors relevant to decisions regarding the employment and use of IAS
- Step 3: map the many connections among legal, policy, and other considerations and use this mapping to develop a framework to help guide decisions and trade-offs regarding the optimal classification of IAS
- For these steps, the CNA team will use a combination of literature review, consultations with subject matter experts (including from the U.S. government and allies, think tanks and academia, international organizations, and civil society), and a review of historical incidents that help illustrate the various risks and trade-offs in practice

CNA

Study Outcomes

- Primary outcome: develop a framework to help guide decisions and trade-offs regarding the optimal classification of IAS to maximize military advantages and reduce risks
- We will also look at considerations for the appropriate and responsible use of AI-enabled and autonomous capabilities:
  - What are the impacts of these trade-offs on responsibility and accountability for decisions regarding the use of force under the law of armed conflict?
  - What are the considerations regarding liability, negligence, or potential settlements for at-sea accidents?

- Study expected to last 12 months, currently going through approval process

CNA

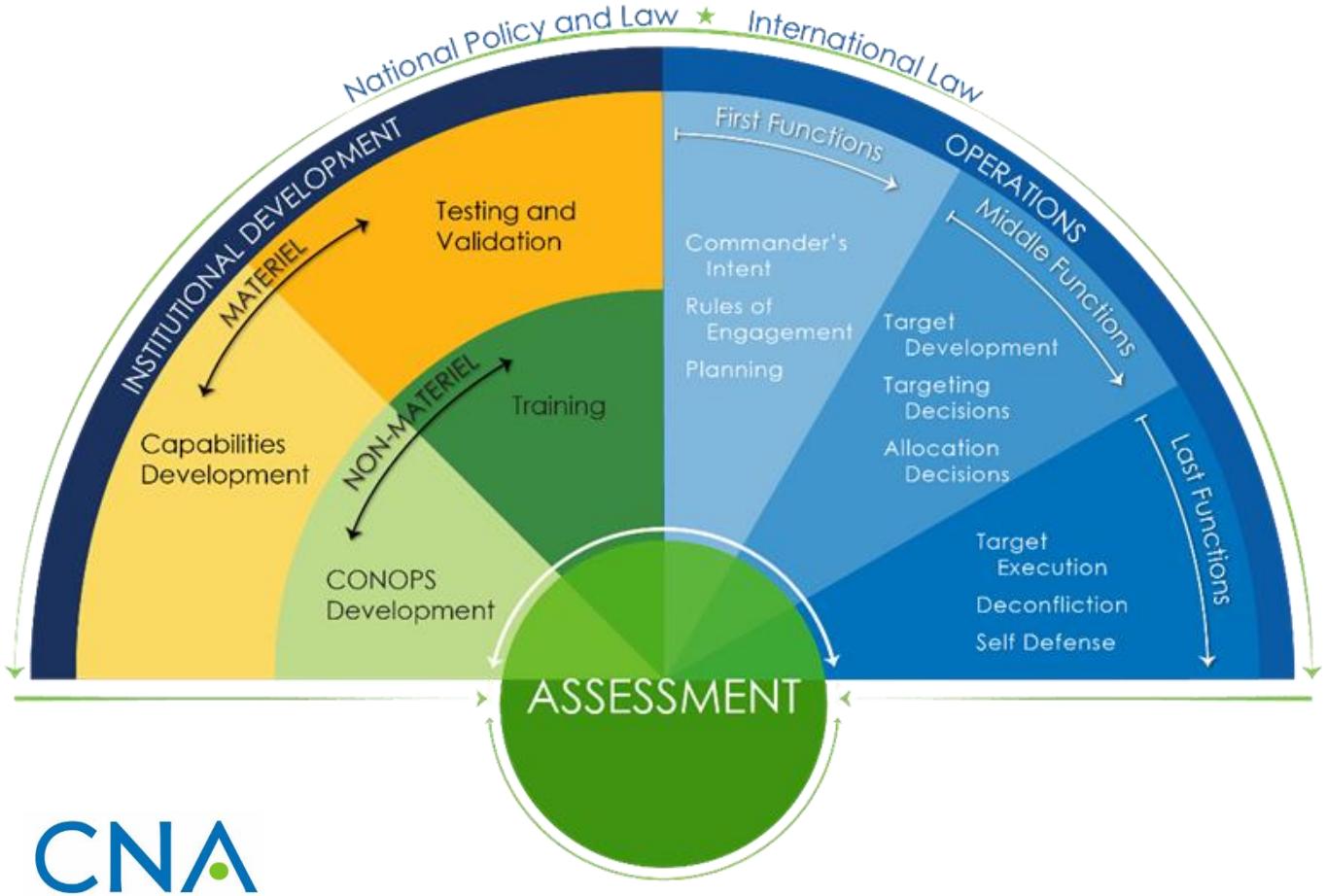
Human Control vs Appropriate Human Judgment

- How can we think about the call for “human control” of AI and autonomy?
- Control is broad in nature: U.S. military doctrine defines specific functions of control, including: *“planning, direction, prioritization, synchronization, integration, and deconfliction”*
  - Consistent with the notion of “appropriate human judgment”
- Military operations show that broad human involvement promotes safety better than narrow control over the trigger-pull decision:
  - ***Humans are fallible***
  - ***Better to get ahead of the many demands of the engagement decision***
  - ***In Afghanistan, a safety-net approach used broad human involvement to avoid patterns of civilian harm***
- What does this broader human involvement look like?

CNA

Mitigating Risks: a Broad Approach

Broad human involvement in the use of force

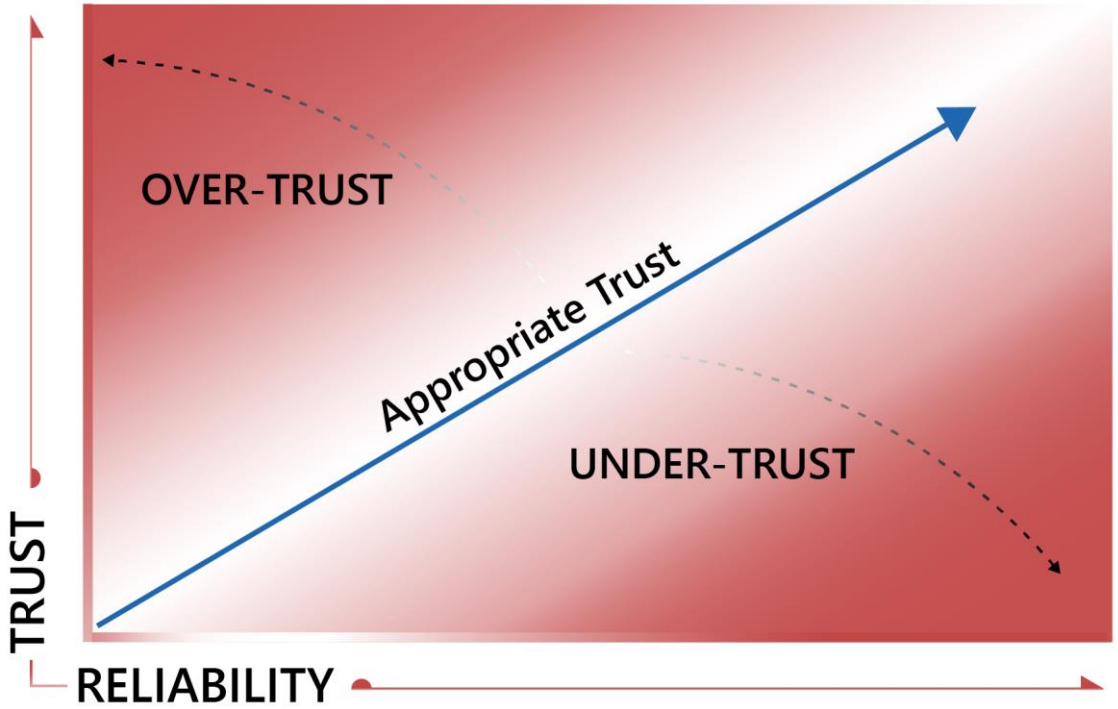


Military element	Ways to mitigate the risks of AI and autonomy
Operations	• Military force should make operational adjustments to mitigate risks and leverage specific strengths of AI-driven and autonomous systems to improve operational outcomes
Institutional development: Capability development	• Build in protections to mitigate potentially fewer communication opportunities for autonomous systems operating in communications-denied and covert modes • Develop processes and protocols using data available to multiple systems to override and preempt potential problems associated with the lack of a human operator • Address and mitigate potential biases in training data for AI • Update intelligence and intelligence requirements to support the development of training data for planned AI applications
Institutional development: Test and evaluation	• Develop test and evaluation processes appropriate for non-deterministic and adaptive systems
Institutional development: CONOPS development	• Ensure that planned use of AI-driven and autonomous systems are consistent with their capabilities and limitations
Institutional development: training	• Train operators regarding the correct and appropriate operation of systems employing AI and autonomy • Cultivate appropriate trust, based on knowledge of system capabilities and limitations, specific to the operating environment and intended purpose
Law and policy	• Conduct legal weapon reviews (e.g., Article 36 reviews) to help ensure developed systems comply with IHL in their intended applications • Review CONOPS, doctrine, and training for use with AI and autonomy with respect to IHL • Develop and maintain policy for autonomy in weapon systems, including safeguards and limits • Develop policy for AI and its potential role in operations, including safety measures and ways to leverage its strengths • Ensure ethical and legal issues regarding the collection of training data are sufficiently addressed

CNA

Cultivating Appropriate Trust

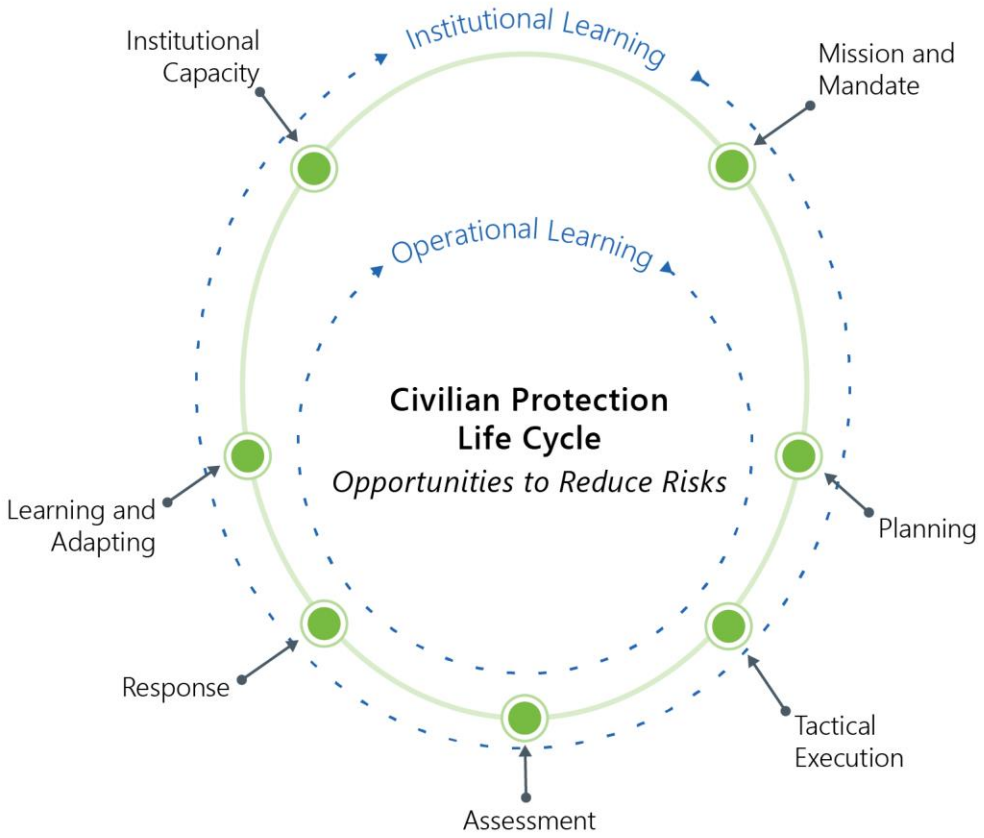
- Commonly expressed concern about AI and autonomy: how do we get operators to trust it and use it?
- We need more than to make up a deficit of trust: we need to cultivate *appropriate trust*
  - A “Goldilocks” trust: not too much, not too little, but the trust suited to the system’s reliability
- We use case studies from past operations to explore the contributions to trust, the risks of under- and over-trust, and ways to calibrate appropriate trust



CNA

AI for Humanitarian Benefits

- Thesis: AI can be used specifically to obtain humanitarian benefits in armed conflict
- Starting point: what are pressing, specific challenges to civilian protection? Which ones could be amenable to AI-enabled solutions?
- Anchored in 10+ years of analysis of civilian harm for DoD and others
  - Overcoming misconceptions about causes of civilian harm
  - Identifying specific, actionable steps to mitigate harm



CNA

Contact:

Dr. Larry Lewis
Director, Center for Autonomy and Artificial Intelligence
CNA

lewisl@cna.org
Lawrence.lewis@cna.navy.smil.mil
703.725.3633 (mobile)

For more information, see:
https://www.cna.org/CAAI

CNA
ANALYSIS & SOLUTIONS