

A green-tinted graphic featuring several soldiers in full combat gear. In the center, a large, glowing 'AI' logo is superimposed over a background of circuitry and data patterns. The overall aesthetic is futuristic and military-oriented.

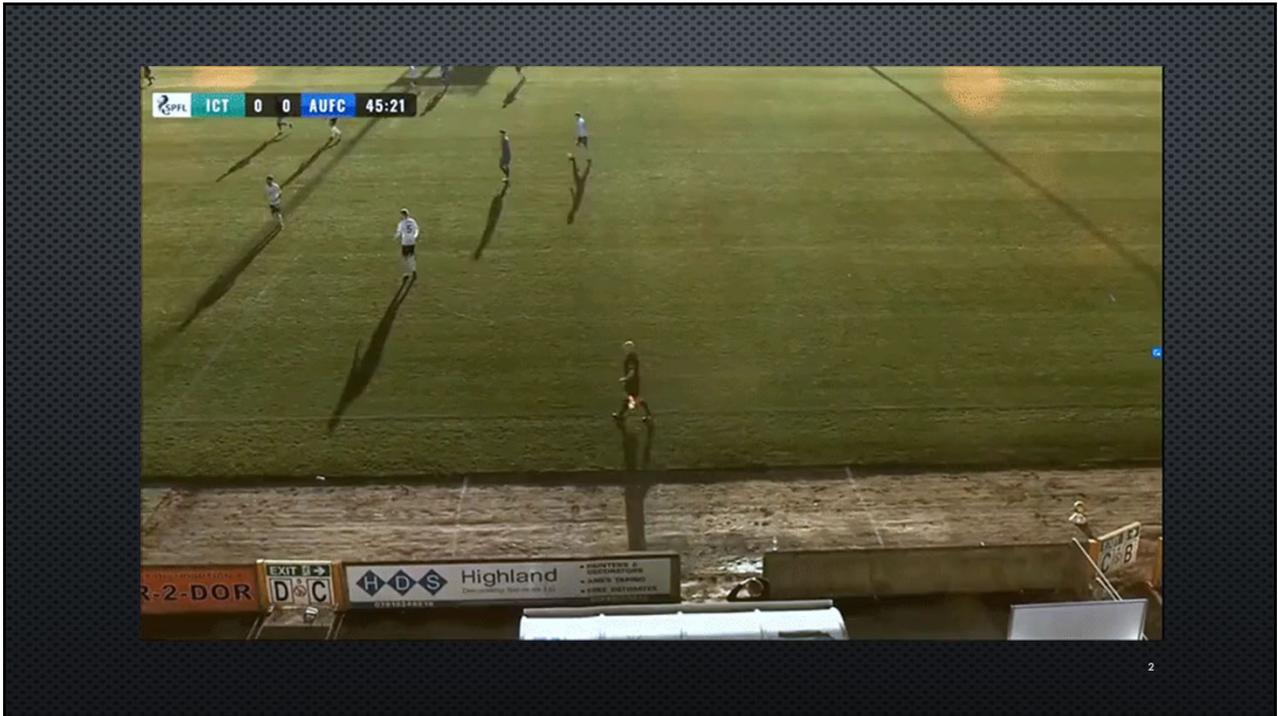
Some AI Considerations
Defense Health Board
5 March 2024

David Barnes, PhD
Brigadier General, U.S. Army (Ret.)
US Army AI Integration Center

david@davidbarnes.io

© 2024 DAVID BARNES, LLC

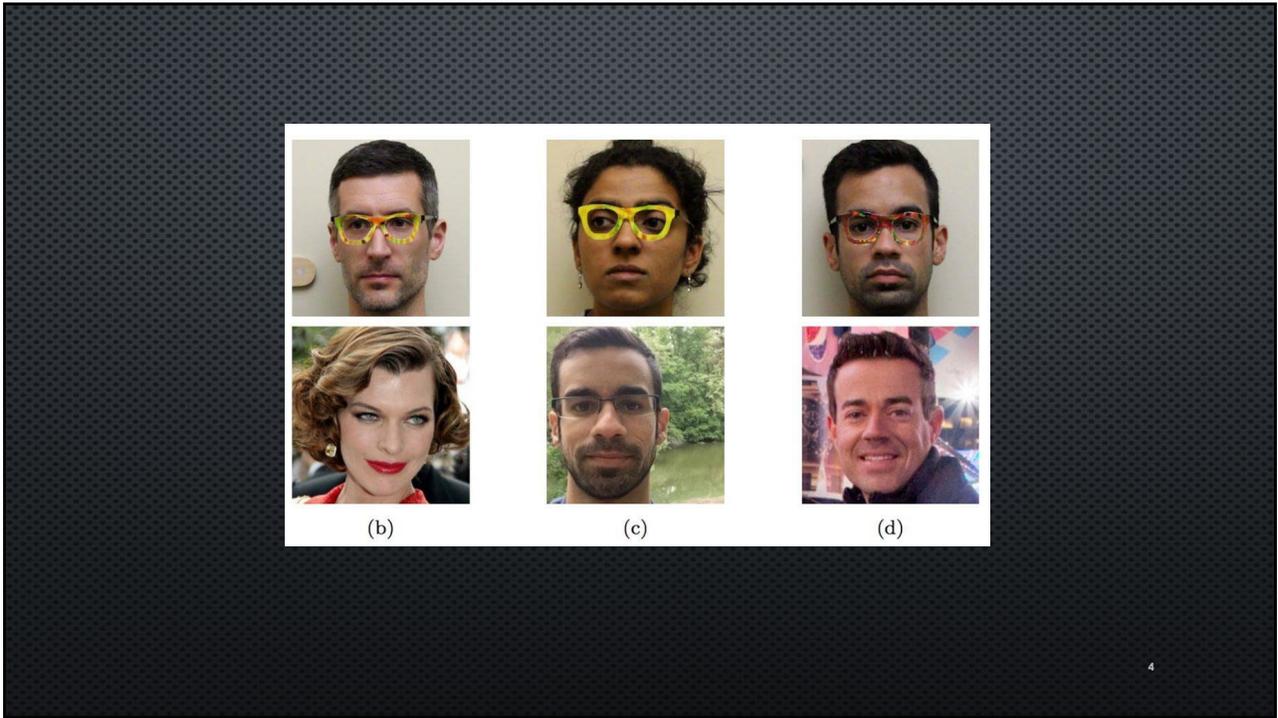
1



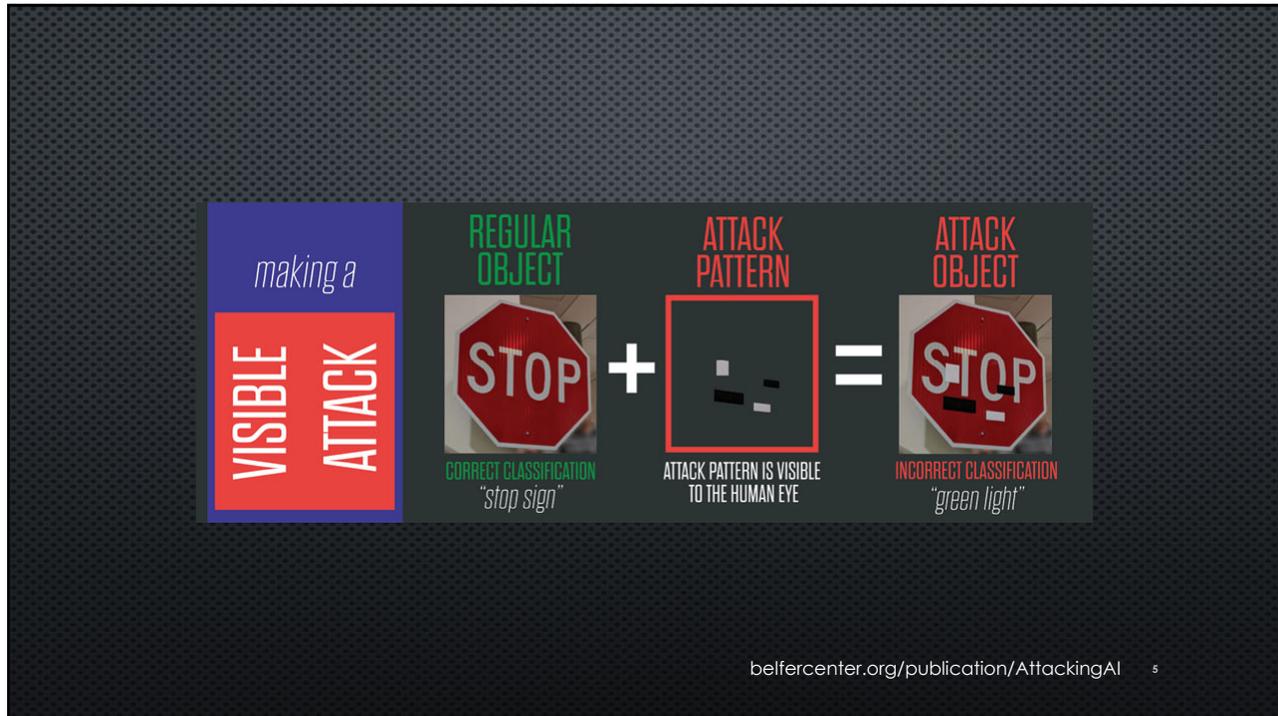
2



3



4



5



6

SOME RISKS AND CHALLENGES

- A. **ADVERSARIAL-AI; COUNTER-AI**
- B. **ALGORITHMIC BIAS**
- C. **DATA SECURITY; PRIVACY & FAIRNESS**
- D. **INSUFFICIENT AI TEV&V PROTOCOLS**

- E. **INTERACTIONS BETWEEN LEGACY, HYBRID, AND AI-EMERGENT SYSTEMS**
- F. **AI DEMOCRATIZATION & PROLIFERATION**
- G. **UNPREDICTABLE AI SYSTEM BEHAVIOR**
- H. **UNINTENDED INFLUENCE ON HUMAN BEHAVIOR (INCL. AUTOMATION BIAS)**
- I. **FEARFUL PUBLIC PERCEPTION OF AI.**

9

DATA DOES NOT ALWAYS REFLECT THE REAL WORLD

AI systems are as good as the data they have been trained on.

MANY PHENOMENA THAT PEOPLE ARE INTERESTED IN PREDICTING DO NOT HAVE SIMPLE DEFINITIONS CORRESPONDING TO MEASUREMENTS PRESENT IN DATASETS

Proxy Variables

10

HUMAN DECISIONS REGARDING THE DESIGN AND OPTIMIZATION OF WHOLE SYSTEM

Responsible AI design is not a one-and-done



ITERATE: RETHINK NOT ONLY MODEL, BUT ALSO DATA, GOALS, ASSUMPTIONS AS WELL.

DOCUMENTATION IS KEY!

NEED: **INTERDISCIPLINARY** EXPERTISE.

Do not let decisions be left to the intuitions of people without relevant expertise

WE DO NOT HAVE TO UNDERSTAND EVERY STEP IN THE ML PROCESS. WHAT WE DO HAVE TO UNDERSTAND, HOWEVER, ARE THE **CHOICES**, **ASSUMPTIONS**, **SIMPLIFICATIONS**, AND **TRADE-OFFS** MADE BY THE PEOPLE WHO DESIGNED THIS SYSTEM.

11

DOD AI ETHICAL PRINCIPLES

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

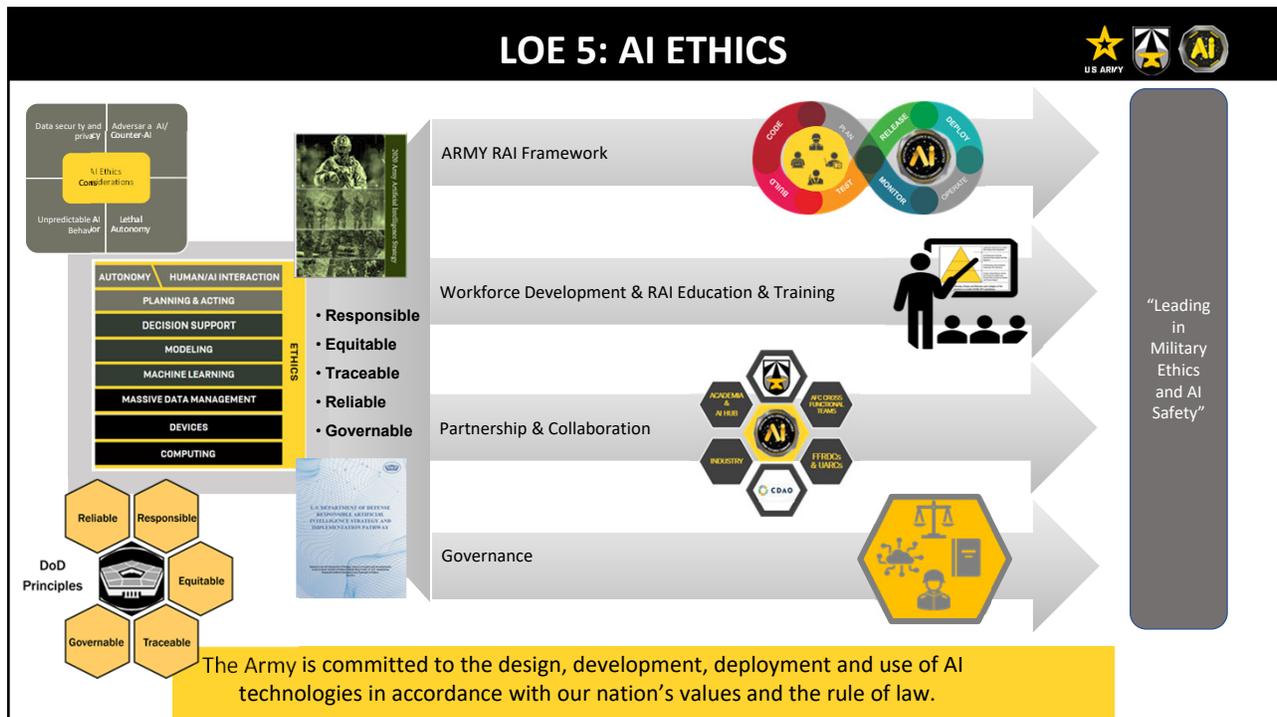
TRACEABLE: THE DEPARTMENT'S AI CAPABILITIES WILL BE DEVELOPED AND DEPLOYED SUCH THAT RELEVANT PERSONNEL POSSESS AN APPROPRIATE UNDERSTANDING OF THE TECHNOLOGY, DEVELOPMENT PROCESSES, AND OPERATIONAL METHODS APPLICABLE TO AI CAPABILITIES, INCLUDING WITH TRANSPARENT AND AUDITABLE METHODOLOGIES, DATA SOURCES, AND DESIGN PROCEDURE 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.

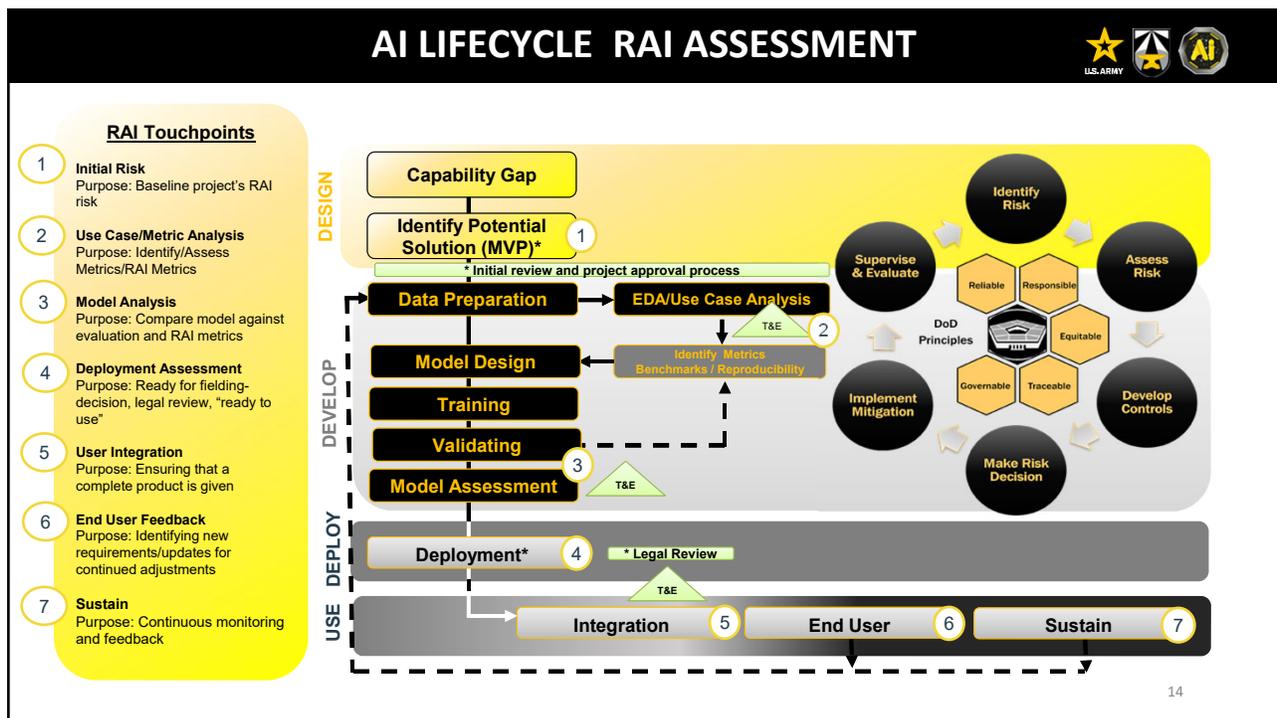
GOVERNABLE: THE DEPARTMENT WILL DESIGN AND ENGINEER AI CAPABILITIES TO FULFILL THEIR INTENDED FUNCTIONS WHILE POSSESSING THE ABILITY TO DETECT AND AVOID UNINTENDED CONSEQUENCES, AND THE ABILITY TO DISENGAGE OR DEACTIVATE DEPLOYED SYSTEMS THAT DEMONSTRATE UNINTENDED BEHAVIOR.

12

12



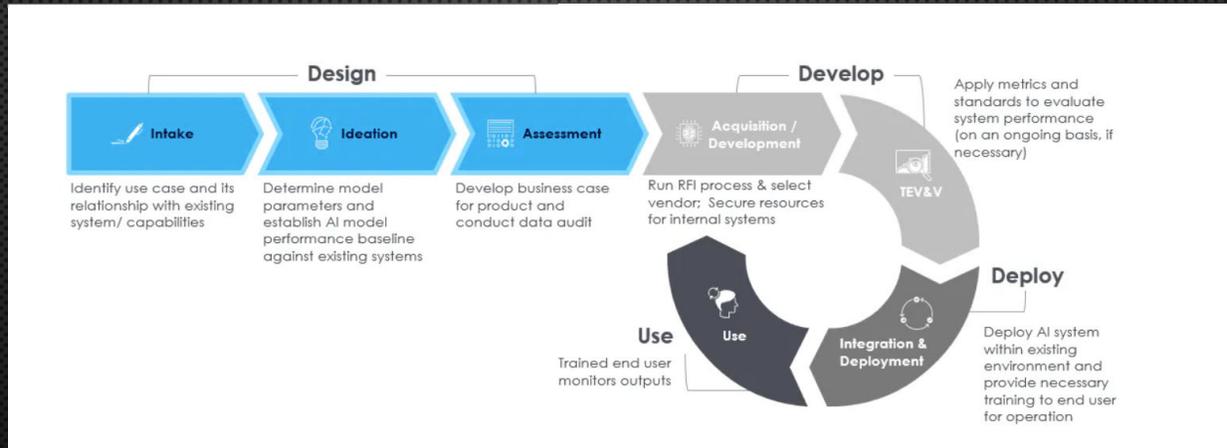
13



14

CDAO RAI TOOLKIT

RAI.TRADEWINDAI.COM



15

15

KEY LEARNINGS



- AI, AN ECOSYSTEM OF DEDUCTIVE INDUCTIVE AND HYBRID APPROACHES, TOOLS FOR NARROW, DATA-CENTRIC ENVIRONMENTS, IS HERE TO STAY, AND MILITARIES WANT TO INTEGRATE AI INTO ALL PROCESSES FROM THE BACK OFFICE TO THE BATTLEFIELD.
- GENERAL LACK OF EDUCATION ABOUT THE BENEFITS, LIMITATIONS, AND THE RISKS OF AI
- NEED INTERDISCIPLINARY, USER CENTRIC TEAMS, ITERATING AND DOCUMENTING THE CHOICES, ASSUMPTIONS, SIMPLIFICATIONS, AND TRADE-OFFS.
- TEV/V NEEDS TO EVOLVE, ACCOUNT FOR AI RISKS
- GENAI IS A SUBSET OF AI TECHNIQUES
- AI IS A SET OF TOOLS; WE ARE THE LIMFAC!

16

16

