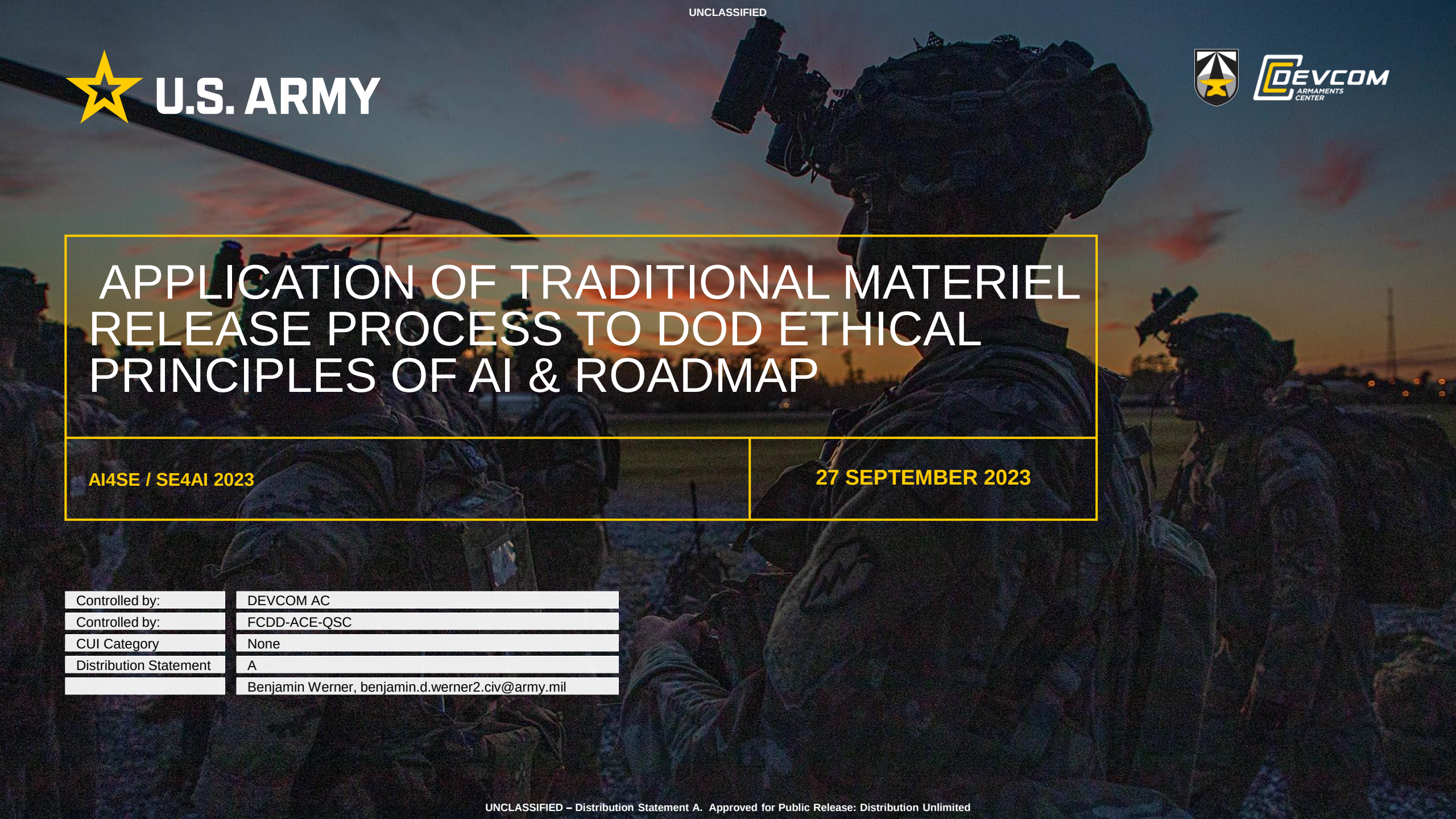


U.S. ARMY

DEVCOM ARMAMENTS CENTER

APPLICATION OF TRADITIONAL MATERIEL RELEASE PROCESS TO DOD ETHICAL PRINCIPLES OF AI & ROADMAP

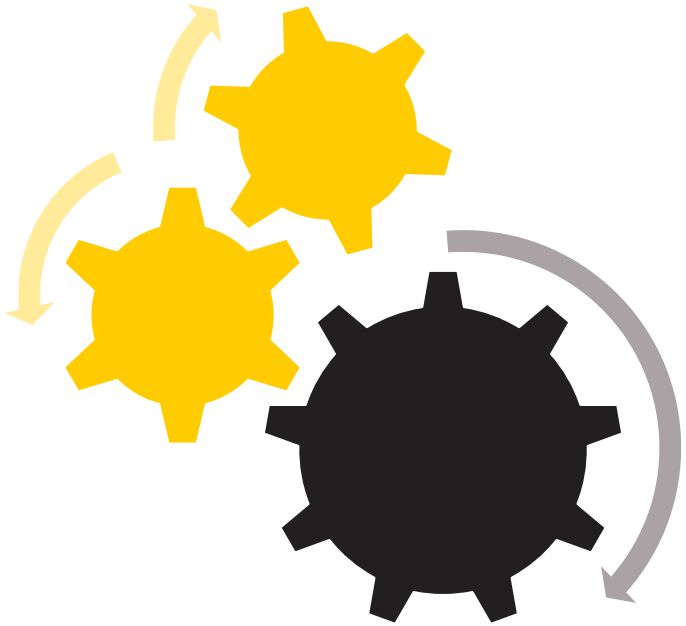
AI4SE / SE4AI 2023

27 SEPTEMBER 2023

Controlled by:	DEVCOM AC
Controlled by:	FCDD-ACE-QSC
CUI Category	None
Distribution Statement	A
	Benjamin Werner, benjamin.d.werner2.civ@army.mil

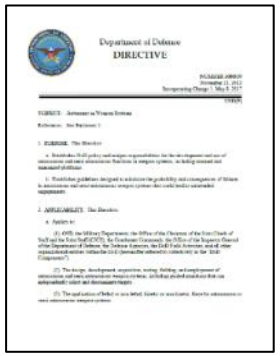
OVERVIEW AND OUTLINE

- Background
- Materiel Release
- Roadmap Development
- Materiel Release for AI Ethics
- Path Forward

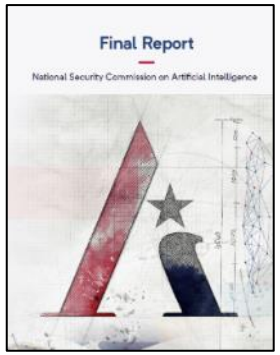


BACKGROUND (DOD)

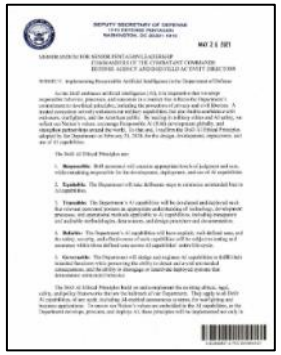
- Department of Defense has a well-established and proven record of accomplishment for developing and fielding battlefield technologies that are safe, thoroughly tested, and give our Warfighter an advantage over our adversaries
- Continuously developing systems that utilize novel technologies and methods for implementing new and complex mission requirements achieve this advantage
- One of the identified technologies with high impact and benefit to the Warfighter is the concept of Artificial Intelligence (AI) and Machine Learning (ML)
- Regardless of the type of applications for AI technology within the DoD, the technology implementation must be verified, validated, and ultimately risk accepted



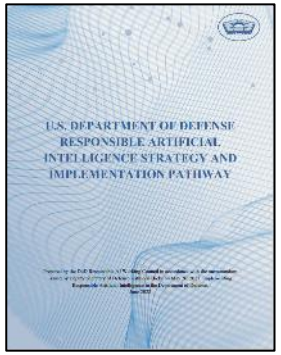
DoDD 3000.09 (2017)



NSCAI Final Report (2021)



RAI Memorandum (2021)



RAI Strategy & Implementation (2022)

BACKGROUND (DEVCOM AC)

Quality Engineering and System Assurance

Plan, direct, and execute quality engineering and system assurance programs for all Armaments systems/programs

Quality | **Reliability** | **Safety** | **Data Science**

Verification & Validation | **Assurance** | **Armaments Human System Integration**

Artificial Intelligence and Machine Learning (AI/ML) can revolutionize existing and future technologies developed by DEVCOM Armaments Center

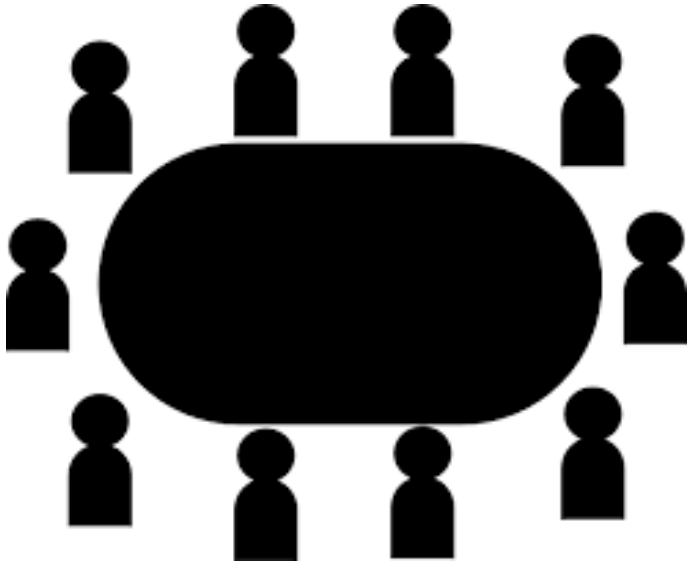
AI/ML Presents Unique Challenges and Disruptions

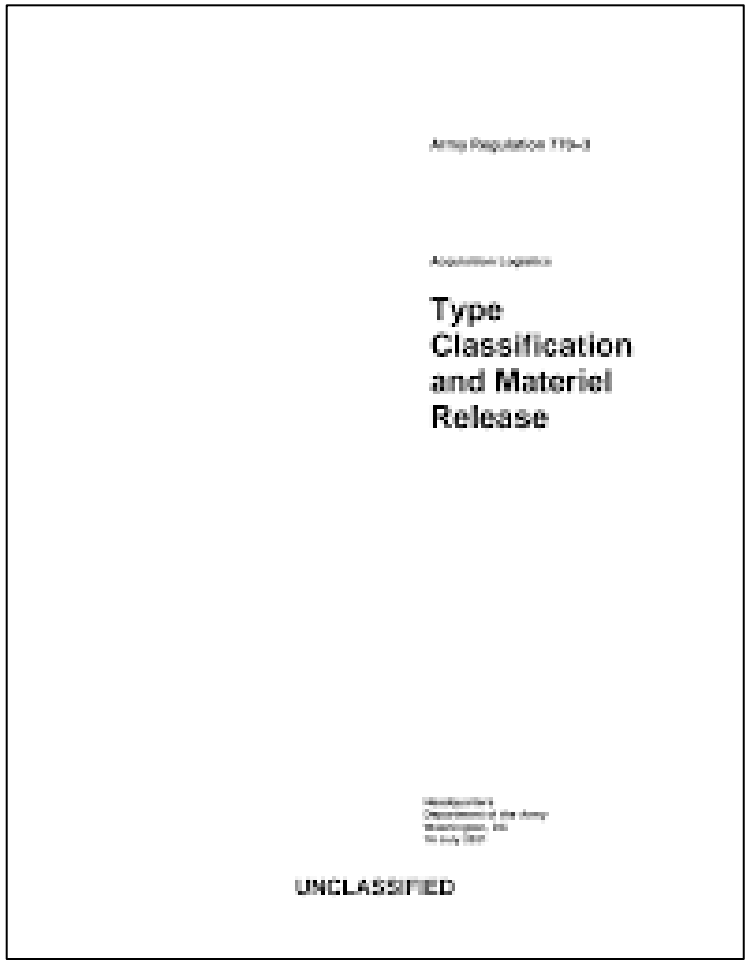
- Uncertainty in continuous learning, complex logic, and configuration management
- Novel methods needed for analysis and certification of AI training data sets
- Critical assessments required of essential enabling sensors/systems
- Unknown user buy-in, trust, and confidence of full system capabilities
- New methods for T&E/V&V to ensure AI is reliable, ethical, safe, and robust
- Possible impacts to existing development methodologies

Army Materiel Release Process is Critical Path for Deployment

MATERIEL RELEASE

- Convergence of Materiel Release Review Board
- Stakeholders across organizations review artifacts
- Determine system presents acceptable risk to field
- Safe / Suitable / Supportable
- DEVCOM Armaments Center MR Coordinator





MATERIEL RELEASE QUESTIONS & ARTIFACTS

PROCESS THAT CERTIFIES THAT ARMY MATERIEL
IS **SAFE, SUITABLE AND SUPPORTABLE** BEFORE ISSUED TO THE FIELD

SAFETY

Questions:

- Is the system safe?
- Have hazards to Soldiers, civilians, and equipment been identified and mitigated or accepted?
- Has AEC confirmed the system is safe?
- Have hazards related to health, EOD, energetics, or environment been identified and mitigated or accepted?

Artifacts:

- Safety Certification & Safety Data Package or Safety & Health Data Sheet
- AEC Safety Confirmation
- Mishap Risk Acceptance or System Safety Risk Assessment (SSRA)
- Health Hazard Assessment
- Surface Danger Zone
- ATEC Assessment or Evaluation
- Final Hazard Classification
- Army Fuze Safety Review Board Certification
- Energetic Materials Qualification Board Statement
- EOD Support Statement
- Environmental Support Statement
- Nuclear Regulatory Commission Licensing
- Air Worthiness Release
- Ignition System Safety Review Board Certification
- Hazards of Electromagnetic Radiation to Ordnance (HERO) Certification

SUITABILITY

Questions:

- Is the system suitable?
- Does the system meet requirements?
- Has the system been evaluated by ATEC? Do they concur?
- How will it function in operational setting?
- Does the system have sufficient reliability for intended missions?
- Have cyber security vulnerabilities been identified and mitigated?
- Has the software been assessed?
- Can the system be used on the network and interface?
- Are TIR/PCRs documented and resolutions effective?
- Have physical and functional configuration audits been conducted?

Artifacts:

- ATEC Assessment or Evaluation
- Quality and Reliability Statement
- Army Interoperability Certification
- Risk Management Framework
- Software Quality Statement
- Human Systems Integration (HSI) Assessment

SUPPORTABILITY

Questions:

- Is the system supportable?
- Has the sustaining command approved of the plan?
- How will software be supported?
- Has test and diagnostic equipment been identified?
- Has training been developed and approved?
- What is the fielding plan?
- Have the Gaining Commands been notified of the system that will be fielded?

Artifacts:

- Proof of TC-STD
- Logistics Certification from Sustainment Organization
- Software Supportability Statement
- Test, Measurement and Diagnostic Equipment (TMDE) Support Statement
- Signed Materiel Fielding Agreement (MFA)/Materiel Fielding Plan (MFP)/ Memorandum of Notification (MON)
- Training Assessment from Capability Developer

PATH TO TRUSTED & ASSURED AI/ML

Armaments Human System Integration (HSI)
- Development of appropriate mental models
- Interfaces optimized to convey the right information

Data Science
- Acknowledge criticality of data to AI/ML
- Identify way and means to evaluate data sets for risk and readiness for AI/ML application

Safety
- Identify unique hazards presented by AI/ML
- Define appropriate design criteria and mitigations to ensure safety

T&E/V&V
- Develop framework for T&E/V&V of AI/ML
- Establish procedures and measures for AI/ML performance and reliability

Reliability
- Identify potential failure modes of AI/ML
- Ensure enabling systems and sensors can meet needs

Materiel Release
- Coordinate across stakeholders to reduce risk
- Adapt and develop necessary deliverables to ensure safe/suitable/supportable

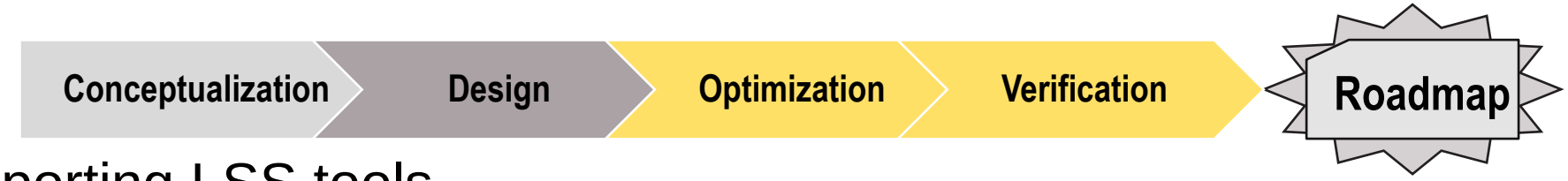
Trusted AI: Product that the warfighter trusts to deliver desired capability

Assured AI: Product can be released and fielded with confidence that it is robust and resilient after rigorous application of best practices and risk mitigation

ROADMAP DEVELOPMENT TO REDUCE RISK ASSOCIATED WITH THE DEPLOYMENT OF ARTIFICIAL INTELLIGENCE ENABLED SYSTEMS

ROADMAP DEVELOPMENT

- DEVCOM AC identified opportunity to mitigate future risks to MR through development of a process roadmap
- Utilized Lean Six Sigma (LSS) processes and phases



- Identified supporting LSS tools
  - Voice of Customer / Voice of Business
  - Ishikawa fishbone diagram
  - Quality Function Deployment
  - Failure Mode, Effects & Criticality Analysis
  - Pareto
  - … and others

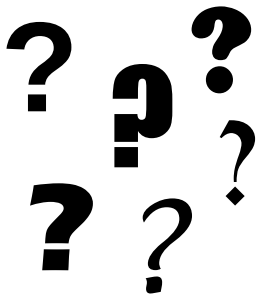
CONCEPTUALIZATION

- Listed AI fundamentals

- Identified MR stakeholders

- Brainstorming session
  - Baselined stakeholder understanding, knowledge, and expectations of AI
  - Stakeholder analysis
  - Voice of customer / voice of business

- Quality Functional Deployment
  - Transformed voice of customer/business to needs



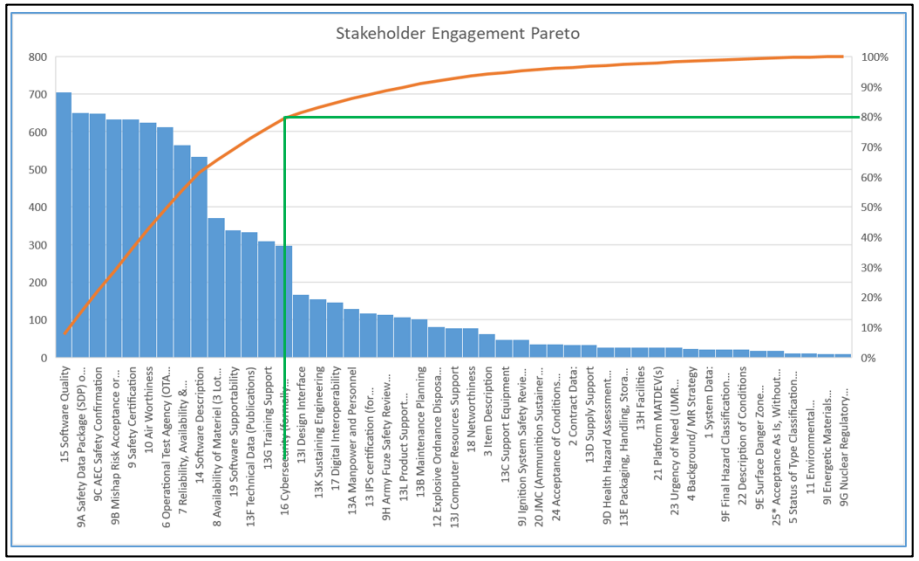
DESIGN

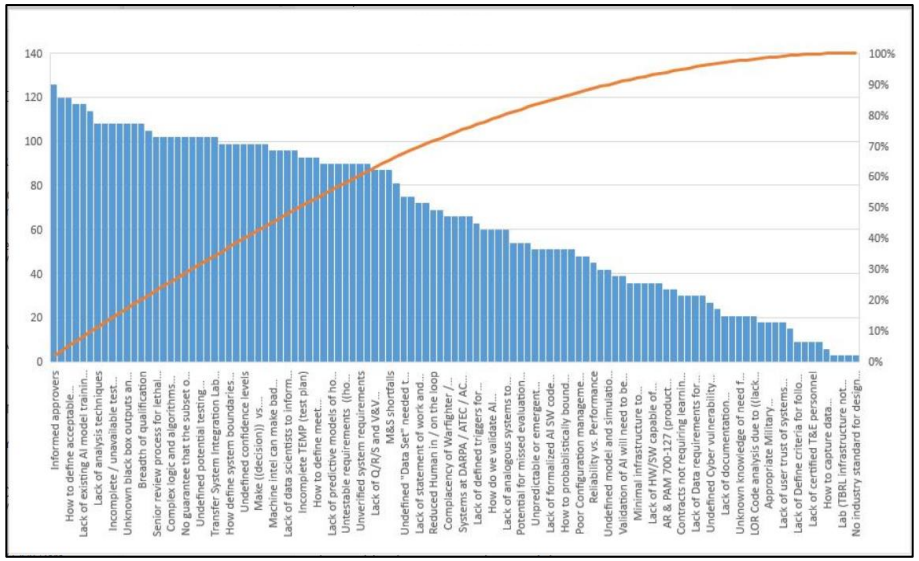
- Ishikawa fishbone diagram to identify impacts
- Quantified risks through Failure Mode, Effects & Criticality Analysis
- Interrelationship matrix compared Customer Needs to 12 critical causes
- Utilized QFD to transform needs to system design requirements

Cause (what can go wrong)	total	1 System Data:	2 Contract Data:	3 Item Description	4 Background/ MR Strategy	5 Status of Type Classification (TC) and Resultant Actions	6 Operational Test Agency (OTA) Milestone Assessment Report (OMAR) or OTA Evaluation Report (OER), and ATEC	Supportability	Quality Provisions, including Service/Shelf Life Assurance	8 Availability & Maintainability (RAM) Status and	9A Safety Data Package (SDP) or	(SDRHT)	9 Safety Certification	Mishap Risk Acce	Assessment
	###	24	36	66	27	4	605	558	364	626	645	627	642	20	11
Informed approvers	369	0	0	0	0	9	9	9	9	9	9	9	9	9	
Lack of training/certification/capability requirements for personnel in **all** parts of AI V&V/fielding process	234	0	0	0	0	0	9	9	3	9	9	9	9	0	0
Lack of analysis techniques	207	0	0	0	3	0	9	9	9	9	9	9	9	0	9
Lack of numbering schematics - lack of understanding on what triggers revision (LRED engagement) ((schematics relate to training - whate environments/use cases))	192	9	9	9	3	0	3	3	0	9	9	9	9	0	0
How to define acceptable (confidence levels)	165	0	0	0	0	0	9	9	3	9	9	9	9	0	0
No true understanding of the risks pertaining to machine intelligence	159	0	0	0	0	0	9	9	0	9	9	9	9	0	0
Lack of statement of work and specification inputs	156	0	9	9	9	3	0	9	0	9	9	0	3	0	0
Lack of who determines AI requirement needs ((design guidelines/spec))	141	0	0	0	0	0	9	9	9	9	9	9	9	0	0
Can you bound the suitability of a system. ((limitations?))	138	0	0	3	0	0	9	9	9	9	9	9	9	0	0

OPTIMIZATION

- Scored relationship of Causes to stakeholders
  - – Identified most impacted stakeholders
  - – Prioritized further engagements
- Summed Causes across the board
  - – Determined Cause with highest impact
- Proposed mitigations
- Reengaged with identified stakeholders
  - – Incorporated updated feedback
- Utilized QFD to transform design to architecture needs





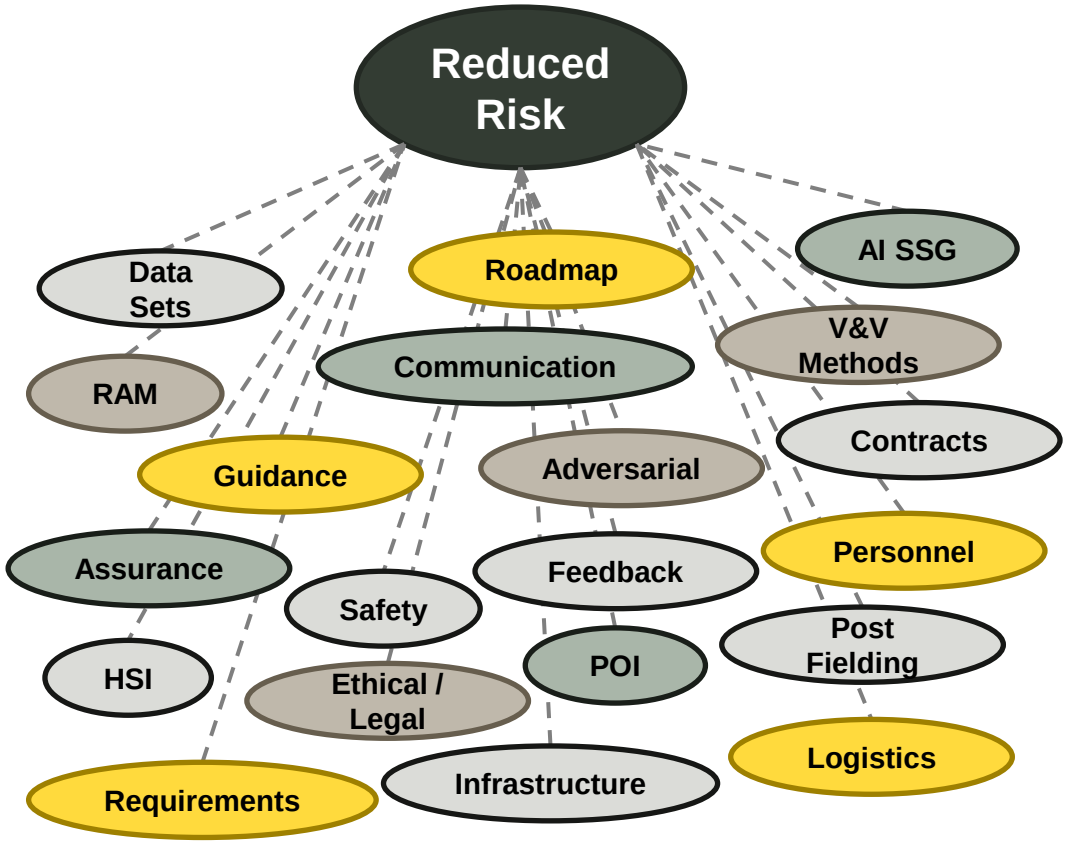
VERIFICATION

- Outlined mitigations into just over 100 initiatives
  - Assessed for partners, duration, impact, deliverables
- Categorized mitigations to align to domains
- Compiled mitigations into roadmap
  - Predecessor and successor deliverables identified
- Identified parallel and serial efforts
- Utilized QFD to establish controls for system architecture
- Briefed all stakeholders on final roadmap
  - Positive feedback and endorsements

Ongoing validation through engagements, briefs, products shared with stakeholders across DoD, Industry, & Academia continue to refine and strengthen the roadmap and associated initiatives

ROADMAP CATEGORIES

- **Identified 100+ activities/initiatives/elements required to achieve the final objective of mitigated risk for MR of AI enabled systems**
- **Affinitized roadmap elements to each other and to QFD outputs into categories**
- **Developed "monopoly cards" for each element to catalog information**
  - Predecessors and successors
  - Key Activities
  - Alignment to regulations, guidance
  - Personnel/Capability needed
  - Infrastructure needs
  - Stakeholders
  - Potential Partners
  - Complexity
  - Timeline
  - Next Steps



NEXT STEPS AND FUTURE WORK

- Continue execution on roadmap elements/initiatives
- Socialize roadmap with community
- Encourage continuous feedback from all stakeholders
- Assess Responsible AI Strategy elements for alignment to roadmap
- Execute roadmap initiatives and strengthen partnerships
  - Design for Reliability for AI
  - Data Assurance (STTR)
  - Assurance of Ethical Principles

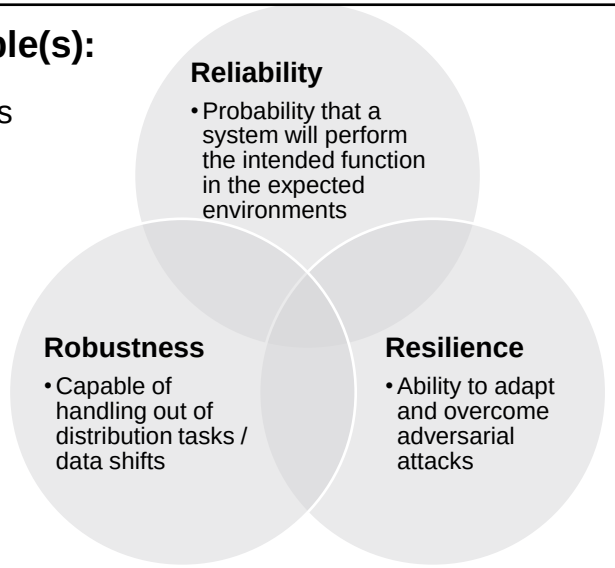
DESIGN FOR RELIABILITY FOR AI

Background Information:

- Establish best practices for Design for Reliability (DfR) tools and techniques for AI Enabled Systems (AIES)
- A strong DfR approach applied early in the system lifecycle has the capability to identify, investigate, and mitigate potential failure modes to improve system reliability, reducing future costs and accelerating a product to the field
- Critical to identify early on what are the intended functions and what conditions are expected
  - Not just intended functions but what would be considered failures
  - Not just nominal conditions but expected CONOPS
  - What enabling technologies (sensors, networks) are required and their reliability

Target Outcome/ Deliverable(s):

- Established methods and tools that can be leveraged to ensure AIES are designed with reliability in mind
- Reduce risk and grow reliability early through mitigations to identified failure modes



Concept:

- Tools such as FMECA and RBDs have direct application to identify potential failure modes, suggest corrective actions, and mitigate risk
- Other HW specific tools such as POF and Highly Accelerated Life Testing HALT may not be a directly applicable however analogous tools can be developed that leverage the same principles
- Leverage reliability best practices to other AI elements to include robust and resilient
- Define and develop analogous tools to aid in the development of AIES with high reliability

Design in Reliability → Design out Failures

Plan of Action and Milestones:

- Document concept and identify potential stakeholders
- Peer review/validation of proposal
- Seek support where possible from stakeholders
- Develop and distribute proposed concepts/tools
- Pilot/demonstrate with S&T efforts

Measures (Follow-up):

- Continue to refine and develop concepts to DfR for AI
- Continuous assessment and validation with stakeholders

STTR-A22B-T002

Metrics and Methods for Verification, Validation, Assurance and Trust of Machine Learning Models & Data for Safety-Critical Applications in Armaments Systems

- Two vendors selected for 6mo Phase I effort
- Two different approaches: one focused on data cards and model cards, the other explicit measurements

Example Products

Machine Learning Qualification Process

- Templates for Data Cards, Feature Cards, Model Cards
- Qualitative & Quantitative Metrics
- Relating Metrics to Measures of Risk

Phase II Planned

Safety Score

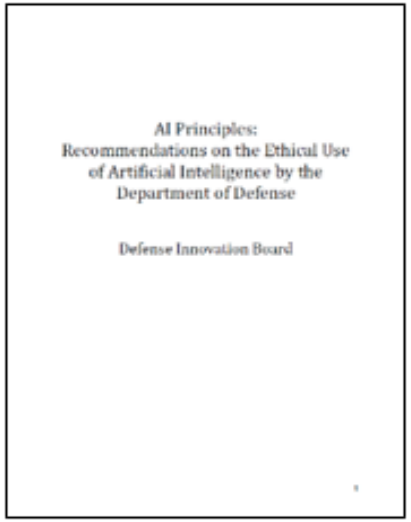
Safety Score Function: *tp*, *tn*, *fp*, and *fn* are the four possible outcomes in a binary classifier (true positive, true negative, false positive, and false negative). In the safety score formula, outcome weights are applied to each of the possible outcomes to assess the risk/impact

Data Quality Measures and Dimensions

- **Consistent representation:** degree to which features do not have multiple semantically equivalent values in the dataset
- **Completeness:** ratio of non-missing feature values to number of samples in the dataset
- **Feature accuracy:** deviation of feature values in the dataset from their true values
- **Target accuracy:** deviation of target feature values in the dataset from their true values
- **Uniqueness:** fraction of unique samples in the dataset
- **Target balance**: relative proportion of samples of each target class in the dataset

AN ASSURANCE CASE FOR THE DOD ETHICAL PRINCIPLES OF ARTIFICIAL INTELLIGENCE

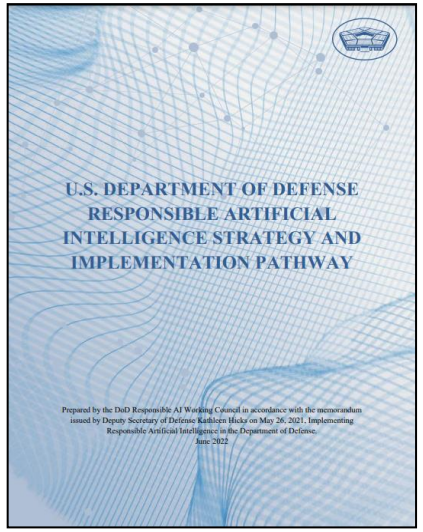
EVOLUTION OF ETHICAL PRINCIPLES



Defense Innovation Board Report (2019)
- Ethical Principles 1st Published

Responsible AI Memorandum (2021)
- Ethical Principles Championed
- Responsible AI Working Council

Responsible AI Strategy (2022)
- Tenets and Lines of Effort identified to embody the Ethical Principles



DOD ETHICAL PRINCIPLES OF AI

- ***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.*

RESPONSIBLE

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

- Materiel Release process ensures responsible risk mitigation across domains
- Materiel Release is a requirement for Army systems prior to fielding or deployment
- Upon deployment systems may only be used for the intended form, fit, and function, use outside of these bounds would be cause for a new evaluation

Materiel Release Artifacts

Safety

- Safety Certification & Safety Data Package or Safety & Health Data Sheet
- AEC Safety Confirmation
- Mishap Risk Acceptance or System Safety Risk Assessment (SSRA)
- Health Hazard Assessment
- Surface Danger Zone
- Final Hazard Classification
- Army Fuze Safety Review Board Certification
- Energetic Materials Qualification Board Statement
- EOD Support Statement
- Environmental Support Statement
- Nuclear Regulatory Commission Licensing
- Air Worthiness Release
- Ignition System Safety Review Board Certification

Supportability

- Proof of TC-STD
- Army Logistician Assessment
- Logistics Certification from Sustainment Organization
- Software Supportability Statement
- Test, Measurement and Diagnostic Equipment (TMDE) Support Statement
- Materiel Fielding Agreement /Materiel Fielding Plan /Memorandum of Notification Training Assessment from Capability Developer

Suitability

- ATEC OMAR/OER Support Statement
- Quality and Reliability Statement
- Army Interoperability Certification
- Risk Management Framework
- Networthiness Certification
- Software Quality Statement

EQUITABLE

Equitable. *The Department will take deliberate steps to minimize unintended bias in AI capabilities.*

- There are multiple artifacts and deliverables to measure system bias and performance
  - Operational Test Agency (OTA) Milestone Assessment Report, OTA Evaluation Report
    - Test and Evaluation community report on operational performance of the system, would capture observed biases
  - Quality Assurance / Reliability Availability Maintainability (QA/RAM) Statement
    - Documents system reliability across the spectrum of expected operating conditions
    - Would identify and measurable bias in performance
    - Bias can be interpreted as a failure, intent of reliability engineering is to mitigate failures
- Design for Assurance tools and practices promote building in equitability
  - Design of AI/ML is predicated on the training data sets
  - Risk and readiness analysis of data sets to include diversity, distribution and completeness

TRACEABLE

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.*

- Materiel Release elements include Manpower/Personnel evaluations
  - Correct Military Occupational Series (MOS)
  - Identify proper skillsets, training, knowledge
- Program of instruction for all stakeholders
- Human System Integration (HSI) assessment completed
  - Understand how and where system is capable or limited
  - Develop mental models to make clear to users the data and methodology used
  - Measures/metrics of human interaction, capabilities
- Training evaluated and exercised in a Logistics Demonstration

RELIABLE

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.*

- Reliability: probability that the system will perform without failure over a specific interval, under specified conditions
  - Documented in OMAR/OER and QA/RAM Statement
  - Measured against specific use cases for testing and assurance
- Safety: ensuring that hazards to human, equipment and environment have been mitigated to acceptable levels
  - Documented through Safety Certification, System Safety Risk Assessment, Safety Confirmation
  - Critical requirement for armaments systems
- Security: cybersecurity is a critical element of Materiel Release
- Effectiveness, testing and assurance
  - Testing for effectiveness is documented through the OMAR/OER
  - Assurance in reliability/safety/security throughout the lifecycle

GOVERNABLE

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 - Design for Assurance / Reliability / Safety
- Intended functions – high reliability
- Detect and Avoid – leverage reliable, safe, robust design practices
- Unintended consequences – reliability failures, safety hazards
- Disengage or deactivate – leverage redundancies, guardrails, governors, mitigations
- Unintended behavior – reliability failures, safety hazards

SUMMARY

- DoD Ethical Principles provides design guidance

- This design guidance can be embodied through the activities and deliverables required for Materiel Release

- The Materiel Release process presents an assurance case for the Ethical Principles

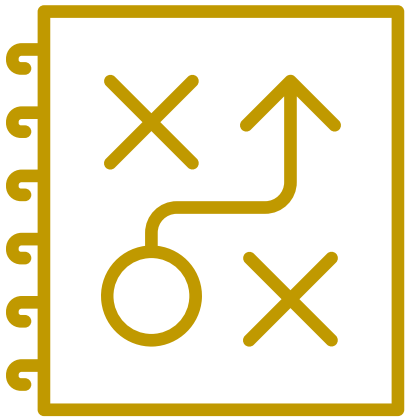
  – Processes and artifacts already in place to ensure Army develops responsibly
  – Identification of gaps can garner new initiatives and capabilities (ie data quality)

- Align assurance case concept to the new DoD Responsible AI Strategy

– Current activities and deliverables may present solutions to the documented tenets and lines of effort

PATH FORWARD

- Continue to engage with stakeholders across DoD
- Refine roadmap and initiatives
- Develop data assurance/qualification metrics
- Execute STTR Phase II
- Align roadmap and MR elements to RAI Implementation Pathways
- Lead cross functional efforts to advance TEVV and Assurance of AI Enabled Systems to enable Trust




U.S. ARMY