U.S. ARMY

DEVCOM ARMAMENTS CENTER

U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND ARMAMENTS CENTER

Summary of DEVCOM AC Work on Assurance of AI-Enabled Systems

OCT 2023

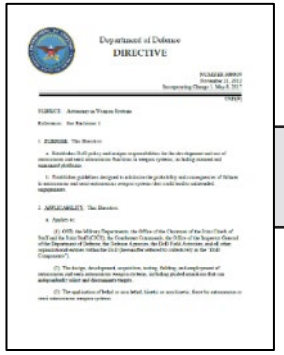
MR. BENJAMIN SCHUMEG, DEVCOM ARMAMENTS CENTER

MR. BENJAMIN WERNER, DEVCOM ARMAMENTS CENTER

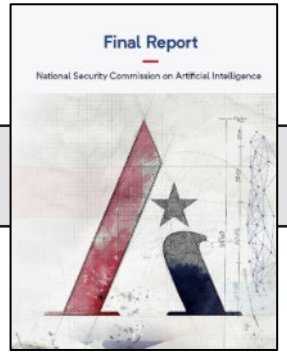
PATH TO TRUSTED & ASSURED AI/ML

… OR “REALLY WELL”

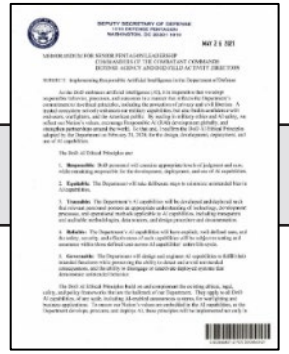
Reports, policies and strategies all point to increased need for assurance, integration, and trust of AI enabled systems fielded by DoD



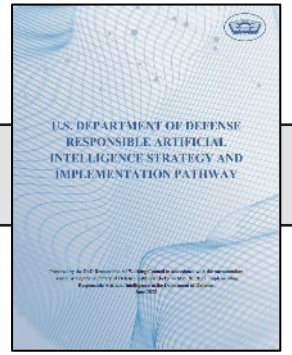
DoDD 3000.09
(2017)



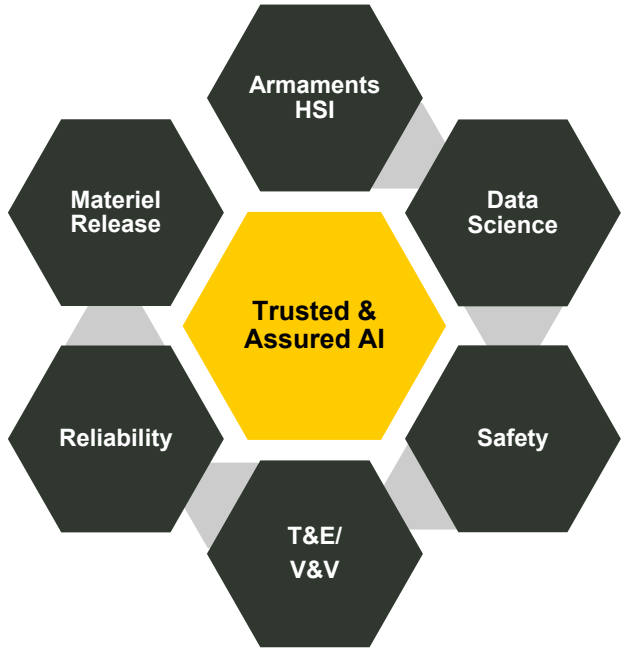
NSCAI Final Report
(2021)



RAI Memorandum
(2021)



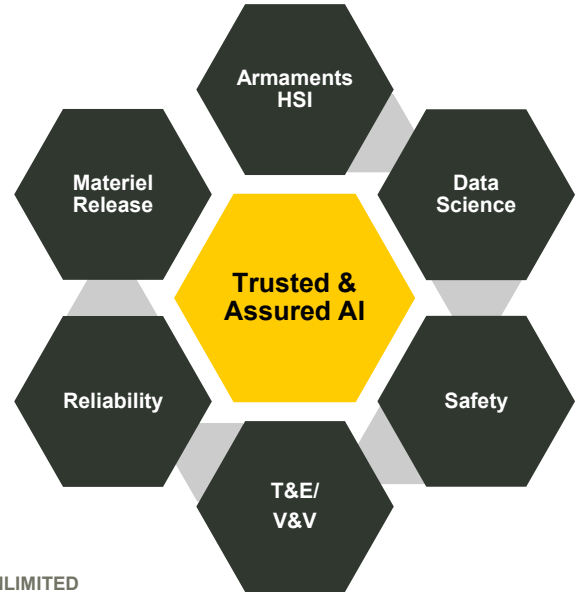
RAI Strategy &
Implementation
(2022)



PATH TO TRUSTED & ASSURED AI/ML

… OR "REALLY WELL"

- **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



MATERIEL RELEASE QUESTIONS & ARTIFACTS

PROCESS THAT CERTIFIES THAT ARMY MATERIEL IS

SAFE
SUITABLE
SUPPORTABLE

BEFORE ISSUED TO THE FIELD

AR 770-3

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

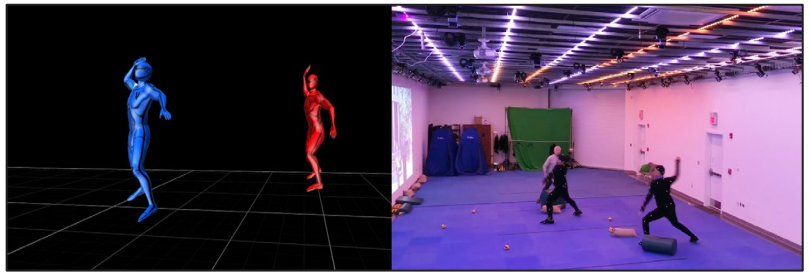
Questions:

- 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

HUMAN SYSTEM INTEGRATION VERIFICATION AND VALIDATION

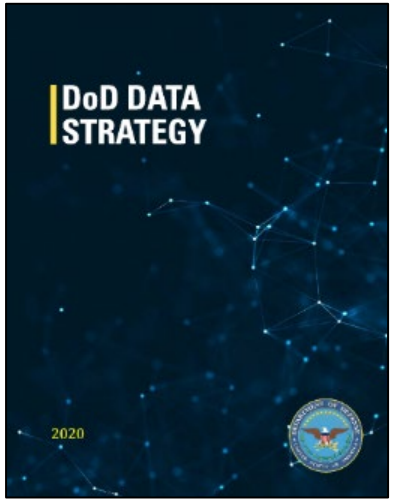


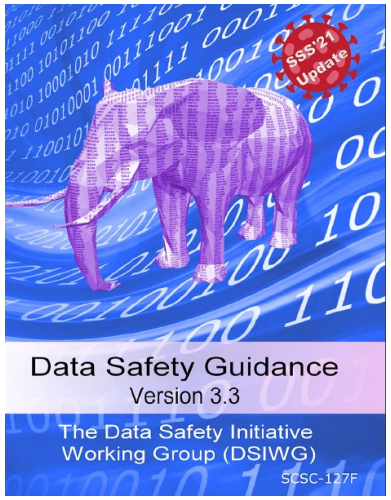
- **Verification and Validation Approach for AI/ML Evaluation**
  - Verification
    - Does the algorithm or robot function as intended with operators?
    - Does the AI identify/process the features that are critical to operators?
  - Validation
    - How do researchers incorporate the context and operational realities faced by operators?
    - Do Soldiers use the AI or robot as intended?
    - What performance components are enhanced by the AI?
    - What performance components are degraded by the AI?
- **Human - AI/ML Performance Data (Verification)**
  - How well do operators versus autonomous AI/ML perform a given set of tasks (Success/Error Rate)?
- **Human - AI/ML Teaming Experimentation (Validation)**
  - Soldier performance impacts during employment of AI/ML system
  - Measurement of Soldier trust across AI/ML reliability levels
- **Results provided used in design, development, and testing**
  - Refine AI training, human interfaces, and overall interaction with systems

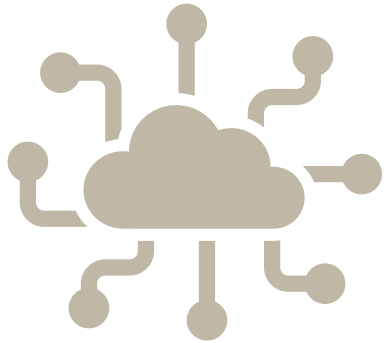
https://www.pica.army.mil/tbrl/index.html

DATA CRITICALITY

- **Data Critical for AI Technologies**
  - Training
  - Safety Assessment
  - Verification/Validation
- **Challenges**
  - Identifying and sourcing certified data
  - Appropriate, secure, and managed
  - Context and mission applicable
  - Clean, labeled, and trusted
- **Data Sets**
  - Data sets may be sparse and unable to meet the requirements
    - Define needs based on current or near-term data capabilities
  - Data set must be complete and certified prior to an AI model training
    - Data Readiness Levels
  - Future data needs may not be known, or the methods to collect them
  - Lack of sufficient data can result in insufficient or detrimental AI
  - Working to establish fundamental methodologies to verify and validate data for AI/ML models required in safety-critical functions of armaments systems







STTR-A22B-T002

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

- OptTek with UAH and AriAcoustics with ASU 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 Awarded

Safety Score

$$\frac{w_{tp}tp + w_{tn}tn}{w_{fp}fp + w_{fn}fn + w_{tp}tp + w_{tn}tn}$$

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, the outcome weights *wtp*, *wtn*, *wfp*, and *wfn* are application-specific.

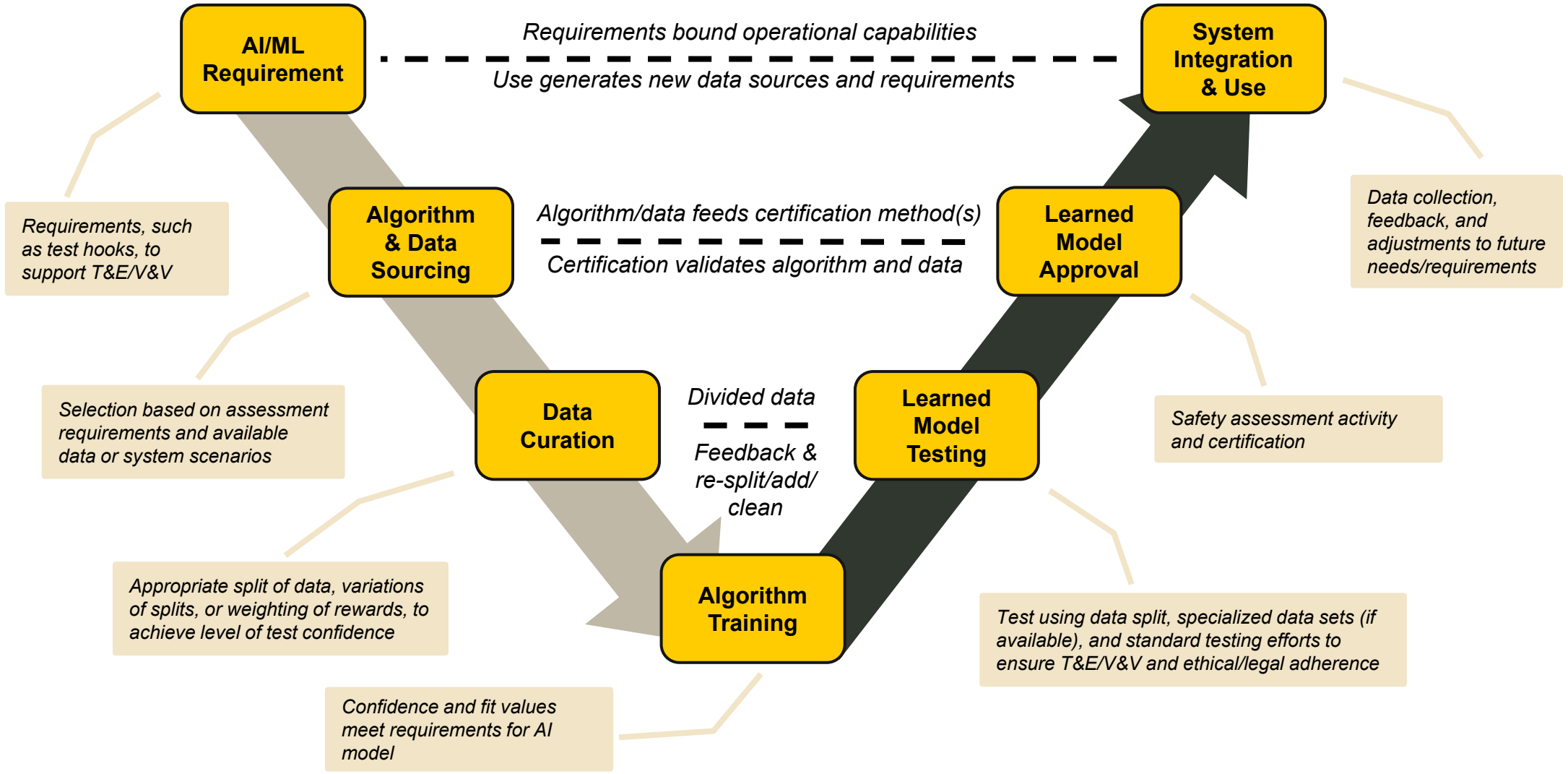
Data Quality Measures and Dimensions

- **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

SAFETY ENGINEERING

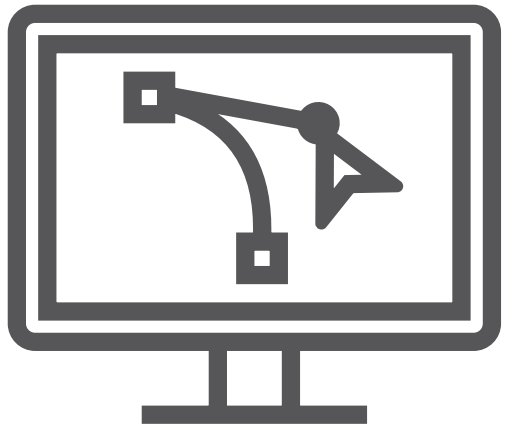
- **Safety challenges are significant**
  - Complexity of the design and architecture
  - Changing operational environments
  - Interactions with human in/on the loop
  - Perceived changing and adapting behavior
  - Adaptation for AI development pipeline
- **Document System Safety and Software System Safety Plan with AI Safety Approach**
  - Apply/Adapt current Safety methodologies/precepts
  - Develop/Modify Level of Rigor (LOR) tasks and metrics
  - Develop hazard mitigation guidance for AI technologies, incorporated into Safety Confirmations and Safety Releases
  - Establish Risk Assessment Approach for AI: Level of Autonomy, Criticality Index, LORs
  - Engaged in Level of Rigor Workshop
  - Co-Authored Army Safety AI Requirement Guidelines and Precepts White Paper
  - Co-Leading Army AI Safety Working Group on Risk Management with ATEC

PROPOSED V-MODEL



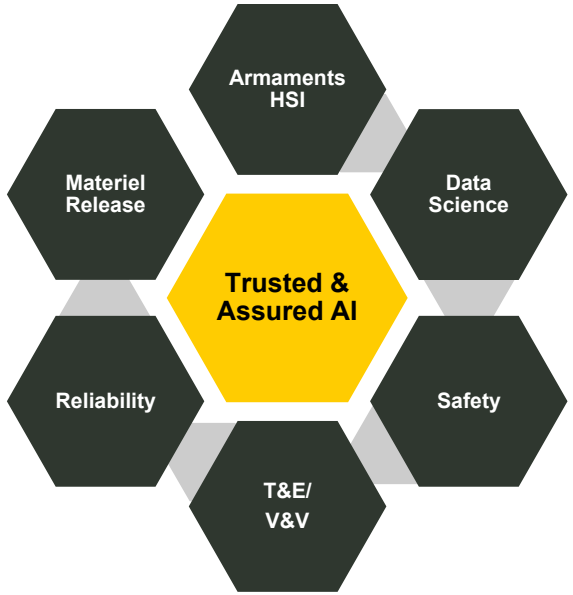
DESIGN FOR ASSURANCE CONSIDERATIONS

- **How do we demonstrate the program is safe/suitable/supportable?**
- **How do we design for assurance, safety, and testability?**
  - Consider the training data early for applications
  - What are the safety implications?
  - Do the test technologies and capabilities meet the intended design?
  - Is the system suitable for the intended operations?
  - How will the system be sustained?
- **Considering the Materiel Release requirements early in the design process allows for optimized path for fielding**
- **Integration early in design develops the framework for artifacts to mitigate risk**



CONCLUSION – REALLY WELL?

- Identified common concerns within DEVCOM AC and Army communities
  - Used AR 770-3 as a holistic view to determine system is ready for release and deployment
- Compare existing processes to identify gaps and challenges
  - Identified possible areas for analysis
  - Developed initiatives, goals, and possible mitigations
- Work with Government agencies, academia, and industry to mitigate risks and overcome challenges
- Continue evolving thoughts and process as technology continues to grow and mature
  - Maintain engagement with stakeholders within Army and DoD to ensure “really well” remains the goal




U.S. ARMY