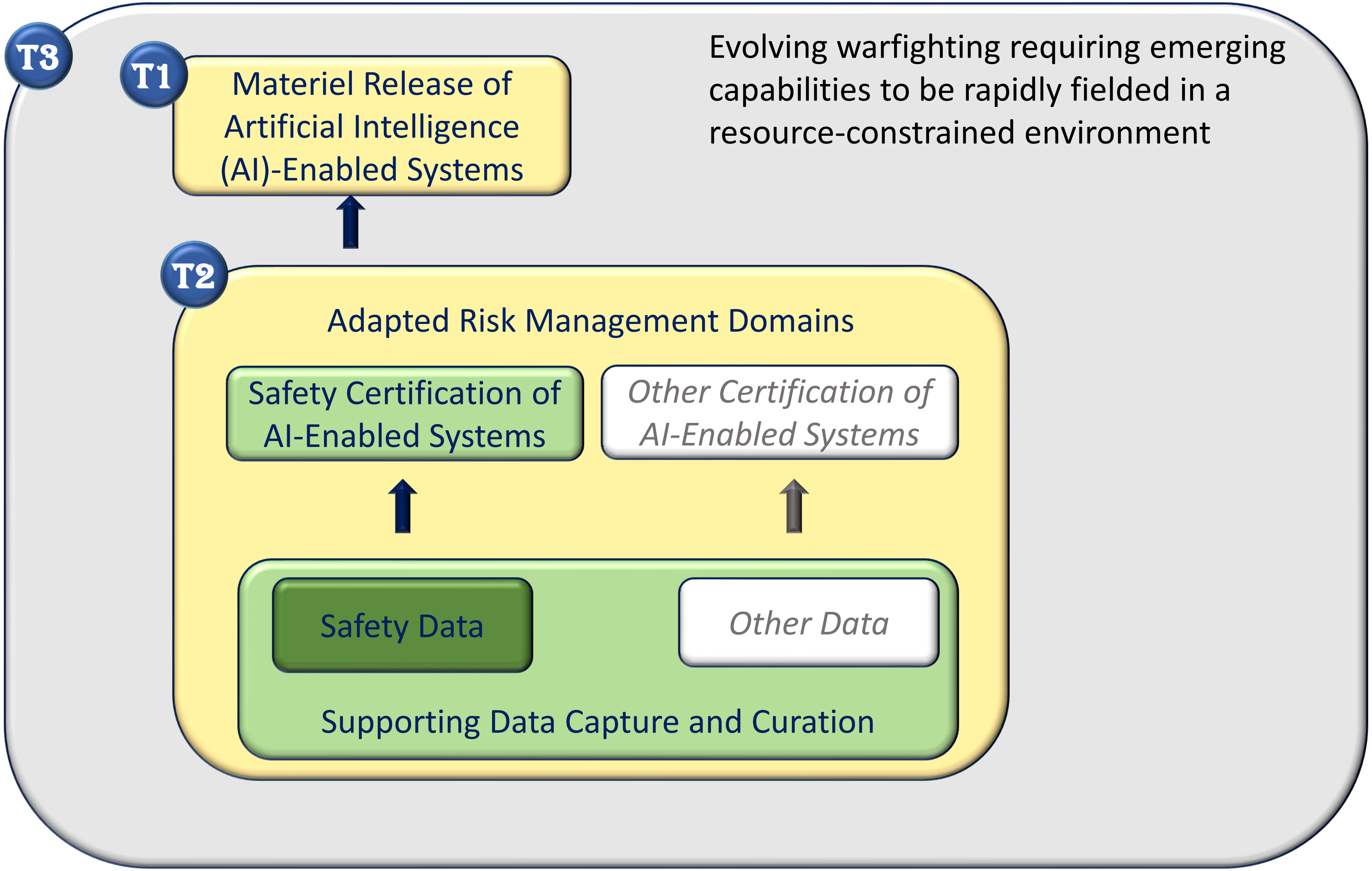
U.S. ARMY

Adapted Risk Assessment for Safety Certification of AI-Enabled Mission Software Applications

Briefer: Dr. Laurence H. Mutuel
AMCOM Safety Office
Aviation Systems Division

Roadmap



Aviation and Missile Command (AMCOM) Safety Office Initiatives

T4

Materiel Release Process Objectives

T1

Army Regulation (AR) 770-3,
Materiel Release

- For use by Soldiers
- For use in demonstration/testing
- For use in training
- For use in operations

Materiel Solution

Verified to be safe for Soldiers when operated in accordance with intended use and operational environment(s)

Safety Certification

AMCOM

Verified to meet its performance requirements (fulfills need)

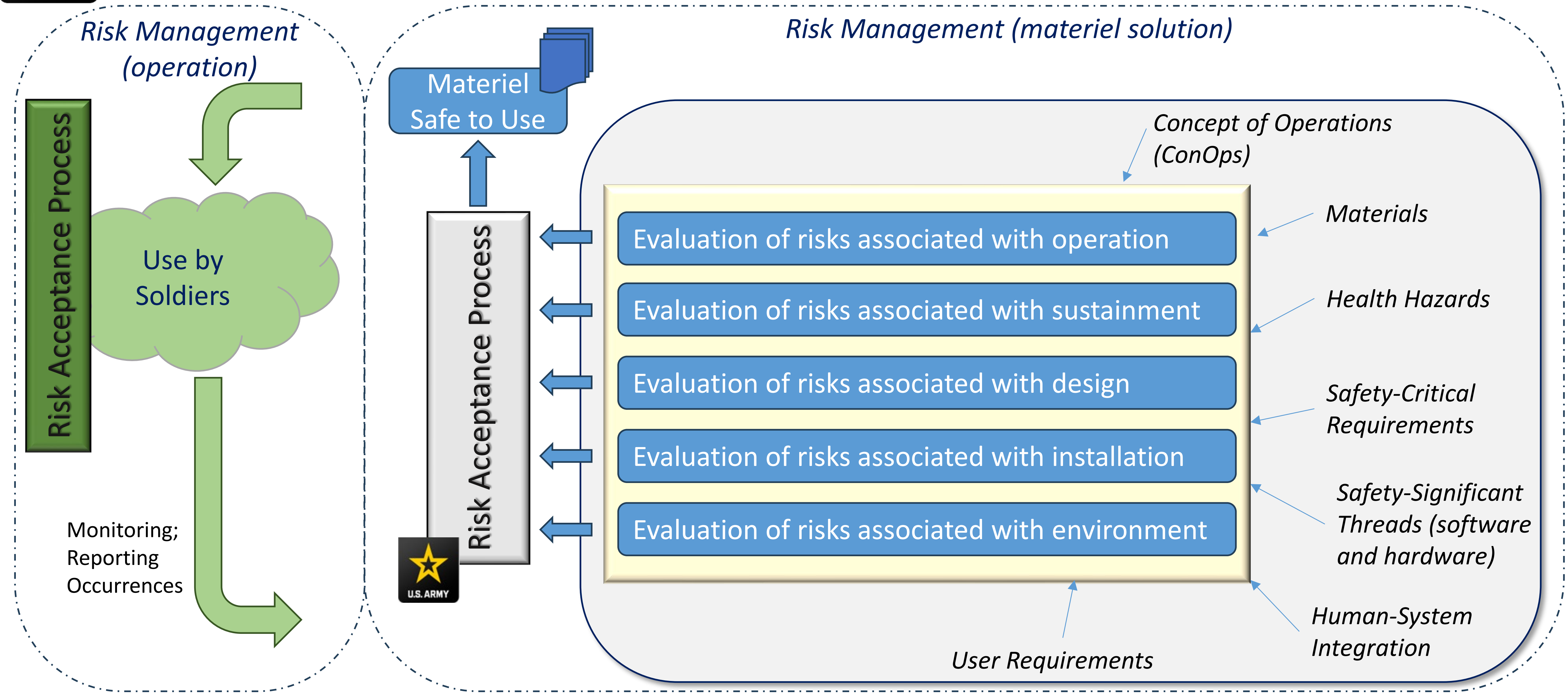
Suitability Certification

Verified to be supportable logistically

Supportability Certification

Current safety requirements likely not adapted to AI-enabled systems.

Risk Management

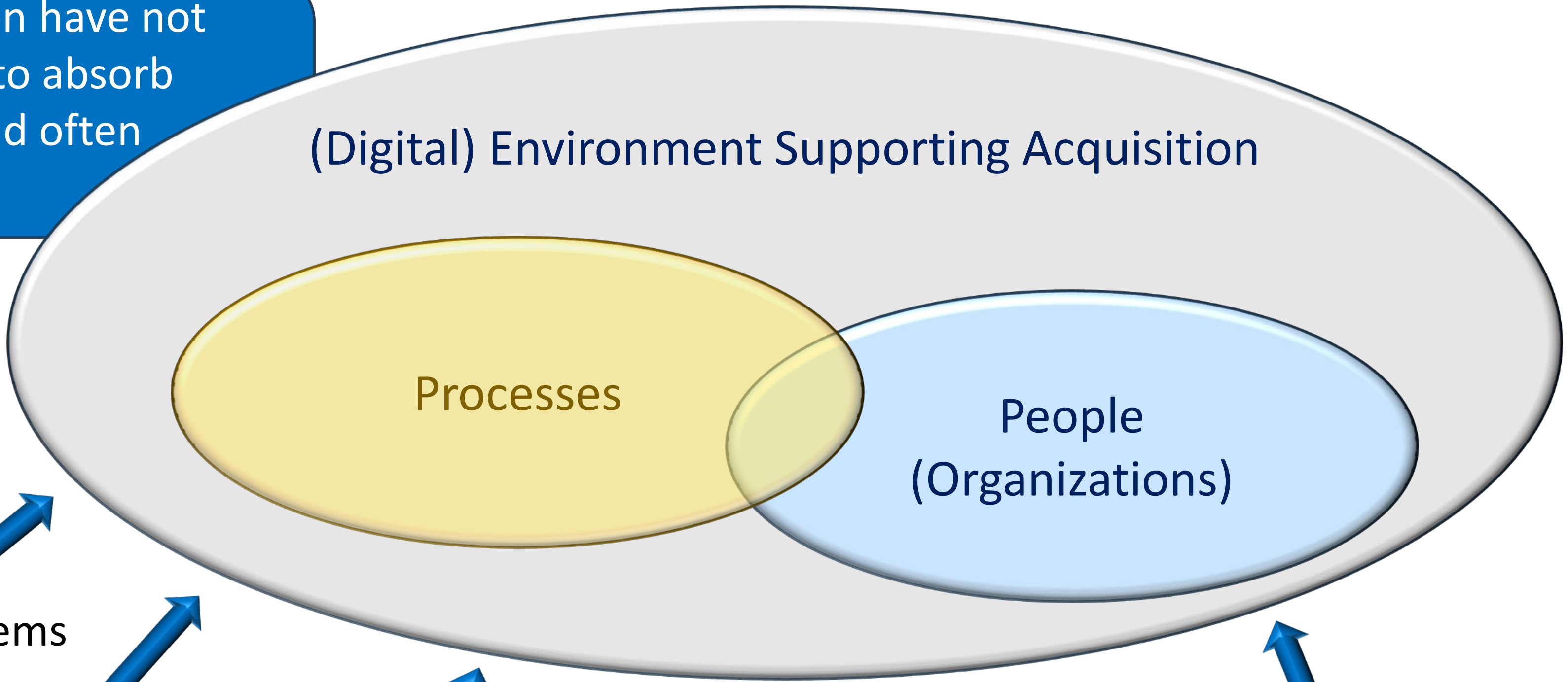


Evolution from distinct materiel and operational risk toward composite risk requires change to hazard analyses.

Modernization on Several Fronts

Objectives of safety certification have not changed. The <u>pathway</u> needs to absorb disjointed but simultaneous and often competing constraints.

Army degradation of "corporate" knowledge



Digital transformation

Accelerated acquisition of consumable, off-the-shelf systems

Software acquisition pathway with development, security, and operations (DevSecOps) and continuous integration, continuous deployment (CI/CD)

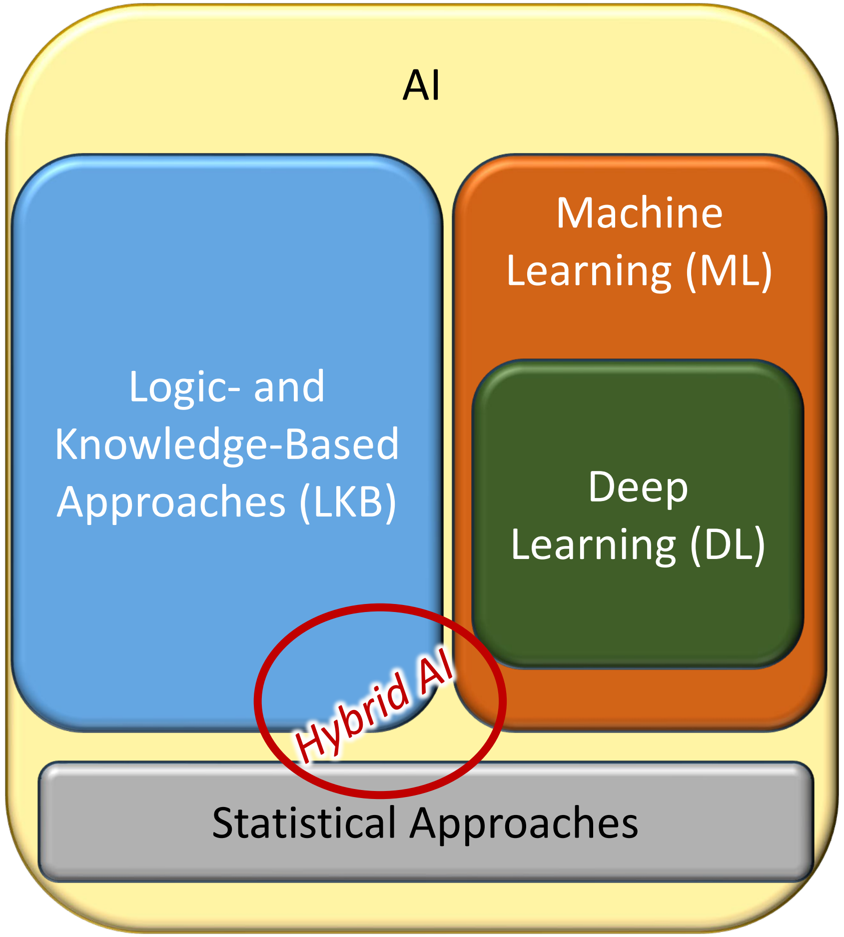
Separability of hardware and software, modular open system architecture principles

Integration of civilian harm mitigation and response (CHMR) action plan

Acceleration adoption of new technologies, including AI

Unprecedented multidisciplinary impact to policies and regulations, lacking detailed guidance and consideration of competing objectives. Needed modernization faces loss of knowledge.

AI Focus in AMCOM Aviation Systems



LKB: Problem solving by drawing inferences (e.g., expert systems)

ML: Algorithm performance improves with exposure to data

DL: Multilayered neural networks learning from vast amounts of data

Statistical Approaches: Predetermined equations used to determine how to fit data

Assistance to Human
- 1A: Human Augmentation
- 1B: Human Cognitive Assistance in Decision-Making

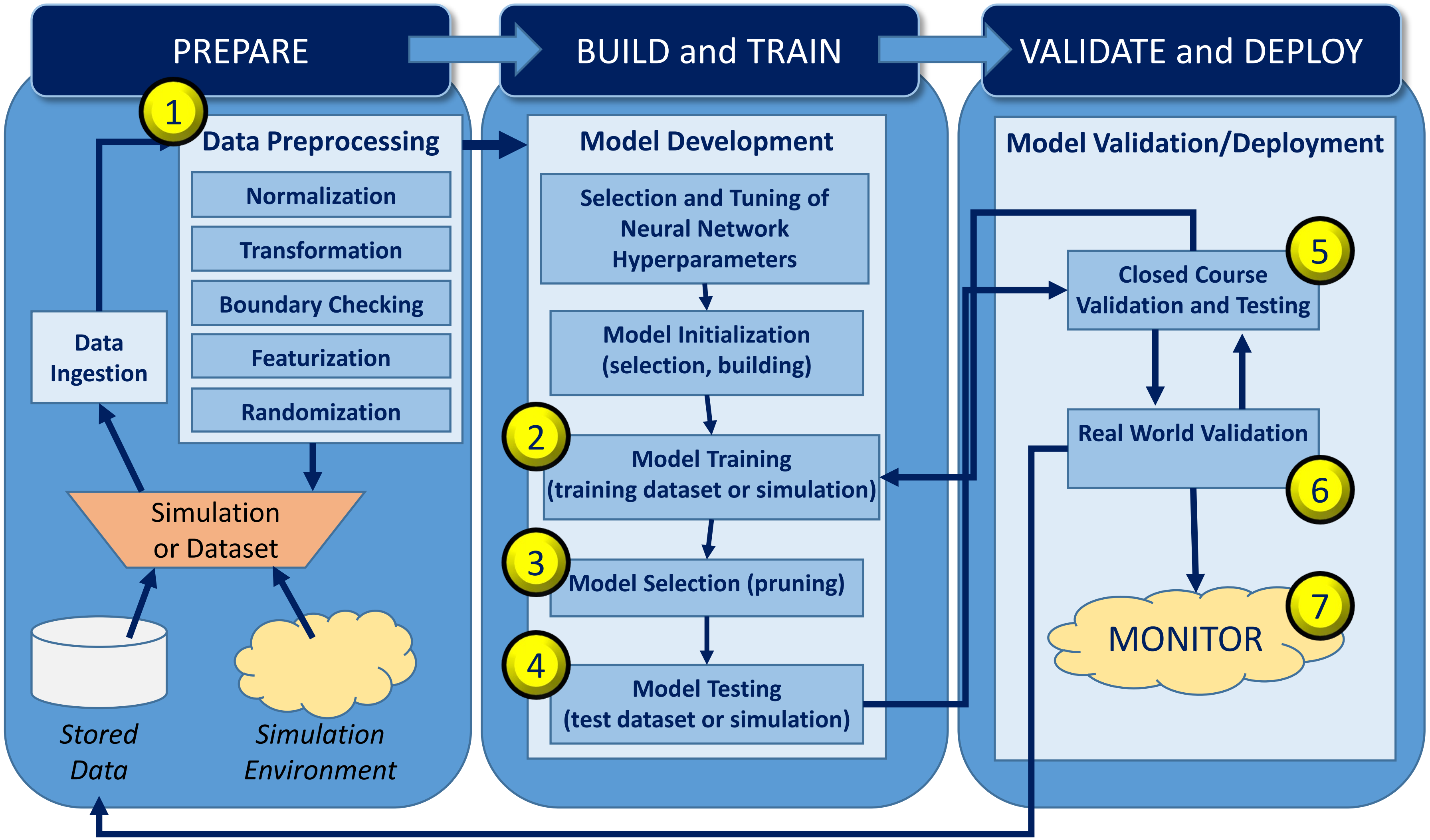
Human-AI Teaming
- 2A: Human-AI Cooperation *(CJADC2 Level 2*)*
- 2B: Human-AI Collaboration *(CJADC2 Level 3 AI Primary*)*

Level 3 Advanced Automation
- 3A: Overridable AI-Generated Decisions/Actions *(CJADC2 Level 4 Solo AI*)*
- 3B: Nonoverridable AI-Generated Decisions/Actions *(CJADC2 Level 5 Solo AI*)*

* CJADC2: Combined Joint All Domain Command and Control AI Engineering Handbook.

Expectation to see materiel release efforts for AI systems levels 1 through 3A (Joint All Domain Command and Control up to level 4).

Areas of Safety Concerns



AI System Lifecycle

1. Definition and hazard identification for safety data
2. Requirements for training models for safety coverage
3. Identification of constraints on model pruning
4. Independence of training and testing models
5. Comprehensiveness of validation model
6. Implementation of operational domain
7. Realizing a closed-loop risk management system

Evolution toward composite risk management (materiel risk and operational risk) with gaining commands responsible for monitoring AI system and supporting risk closed-loop system.

Adapting to a Global Scope of Analysis (Launched Effects)

Friendly Forces

Sustainment

Tactical Operational Center, Command Post

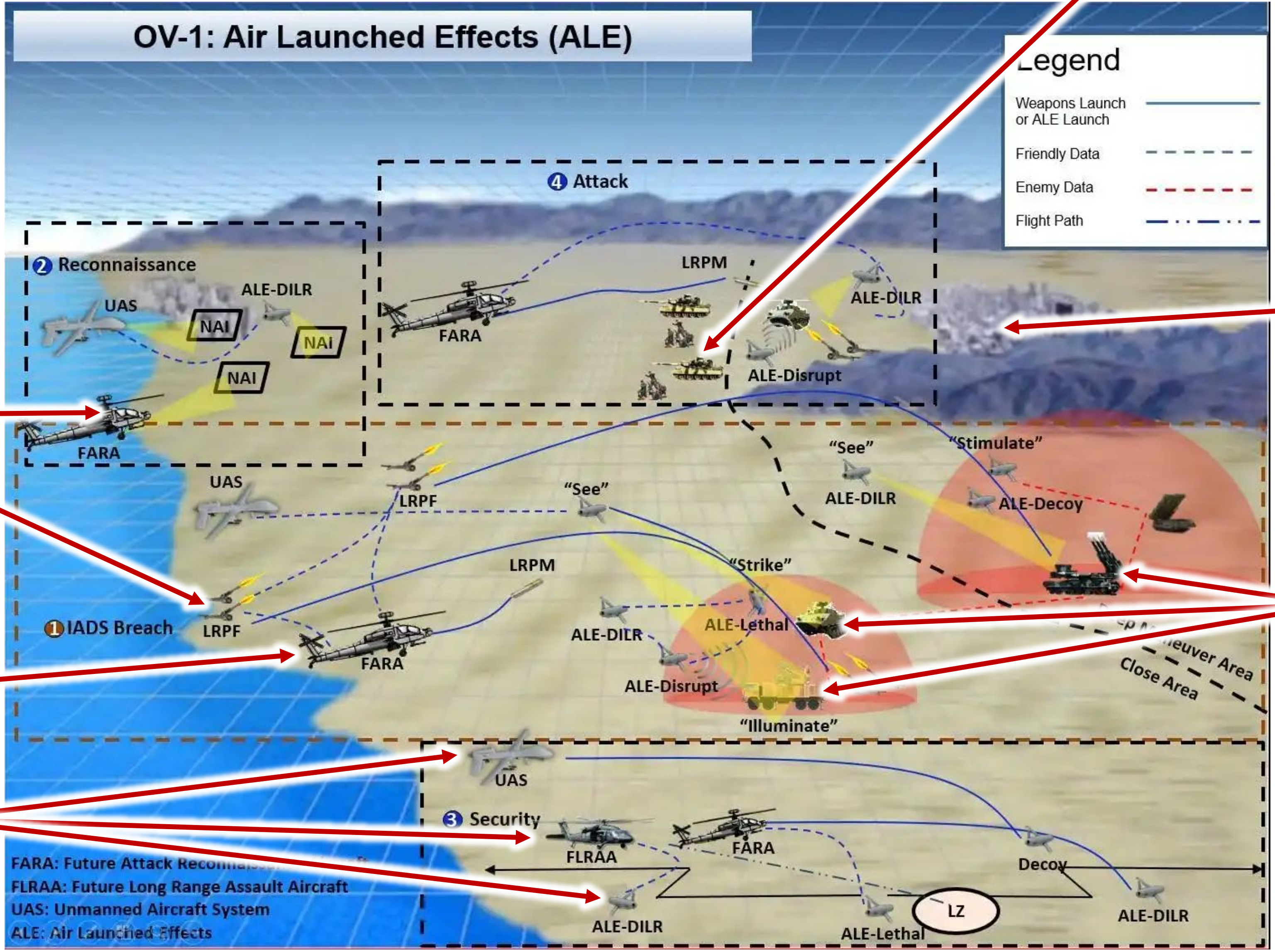
Operators

System

Other Systems

Noncombatants

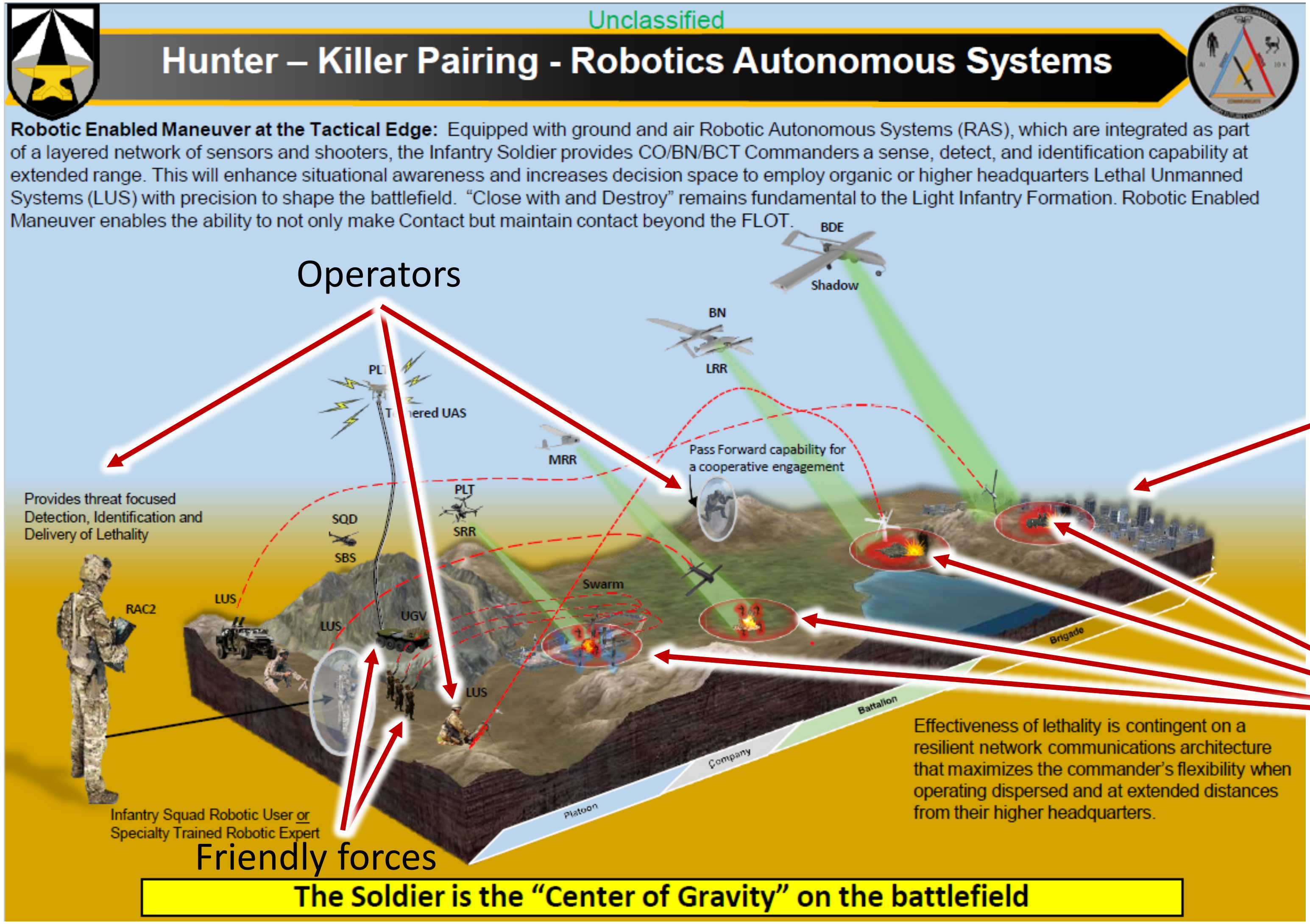
Threats



NOTE: The FARA program was cancelled by the Army in February 2024. Capabilities remain in inventory.

Source: SAM.GOV, https://sam.gov/opp/054e842814ab4d5ba351c84b713511cb/view, 2024.

Adapting to a Global Scope of Analysis (Robotics and Autonomous Systems [RAS] Controller)



Source: RDD Industry Day, April 2023, courtesy of U.S. Army DEVCOM

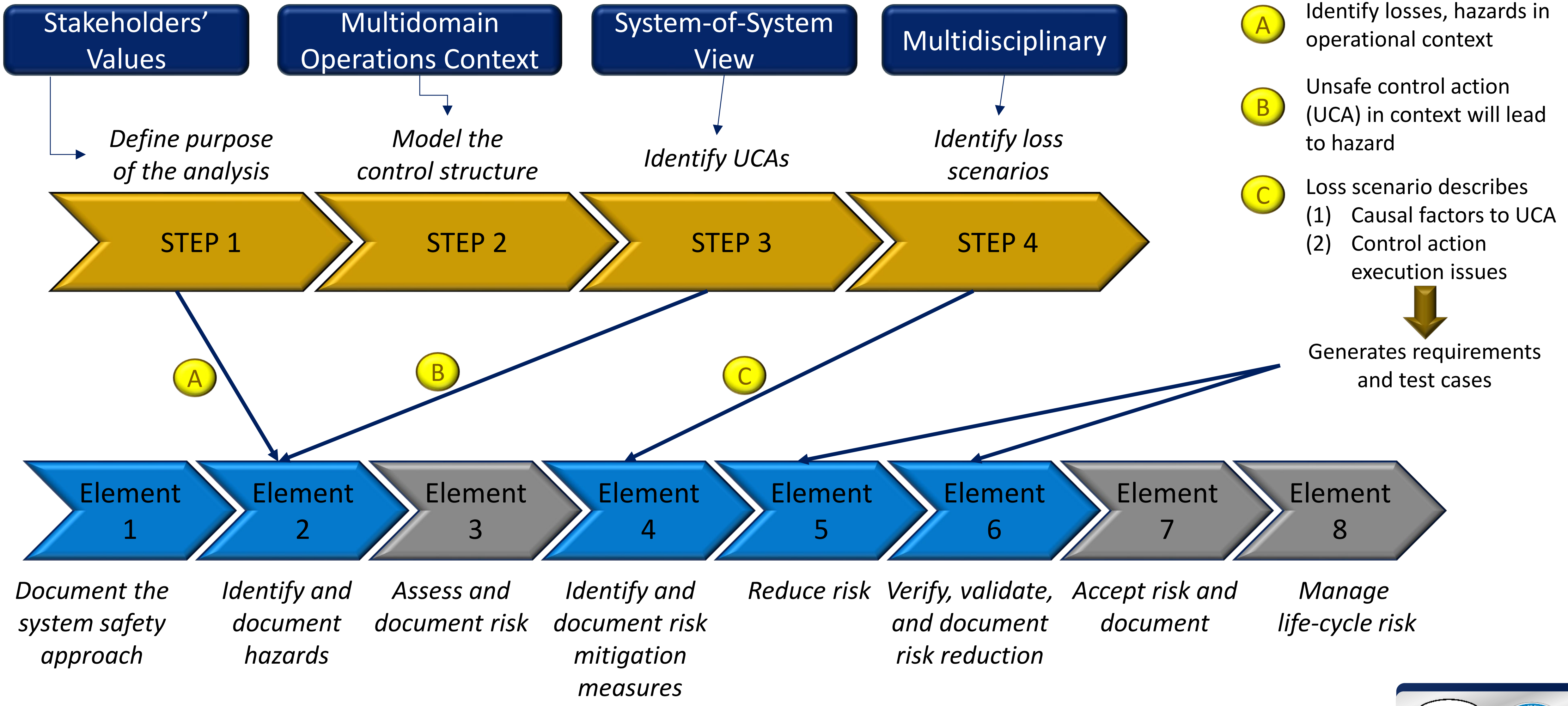
Where is the "system under analysis?"
Evolution toward system of systems.

Next Evolution

CURRENT STATE		NEXT EVOLUTION
Decoupled doctrine, organization, training, materiel, leadership, and education, personnel, and facilities (know as DOTMLPF) risk contributions Static capability growth	**Risk Management Line of Effort (LOE)** →	ConOps/concept of employment (ConEmp)-oriented composite risk Semi-static capability growth with AI
Hazard analyses boundary constraints Loose cross-discipline check	**Safety Practice LOE** →	Evolve system-of-systems analyses Establish multidisciplinary analysis framework
Missing guidance Missing requirements	**Regulatory LOE** →	Develop guidance/examples/pilot missions Develop requirements for fielding AI-enabled systems

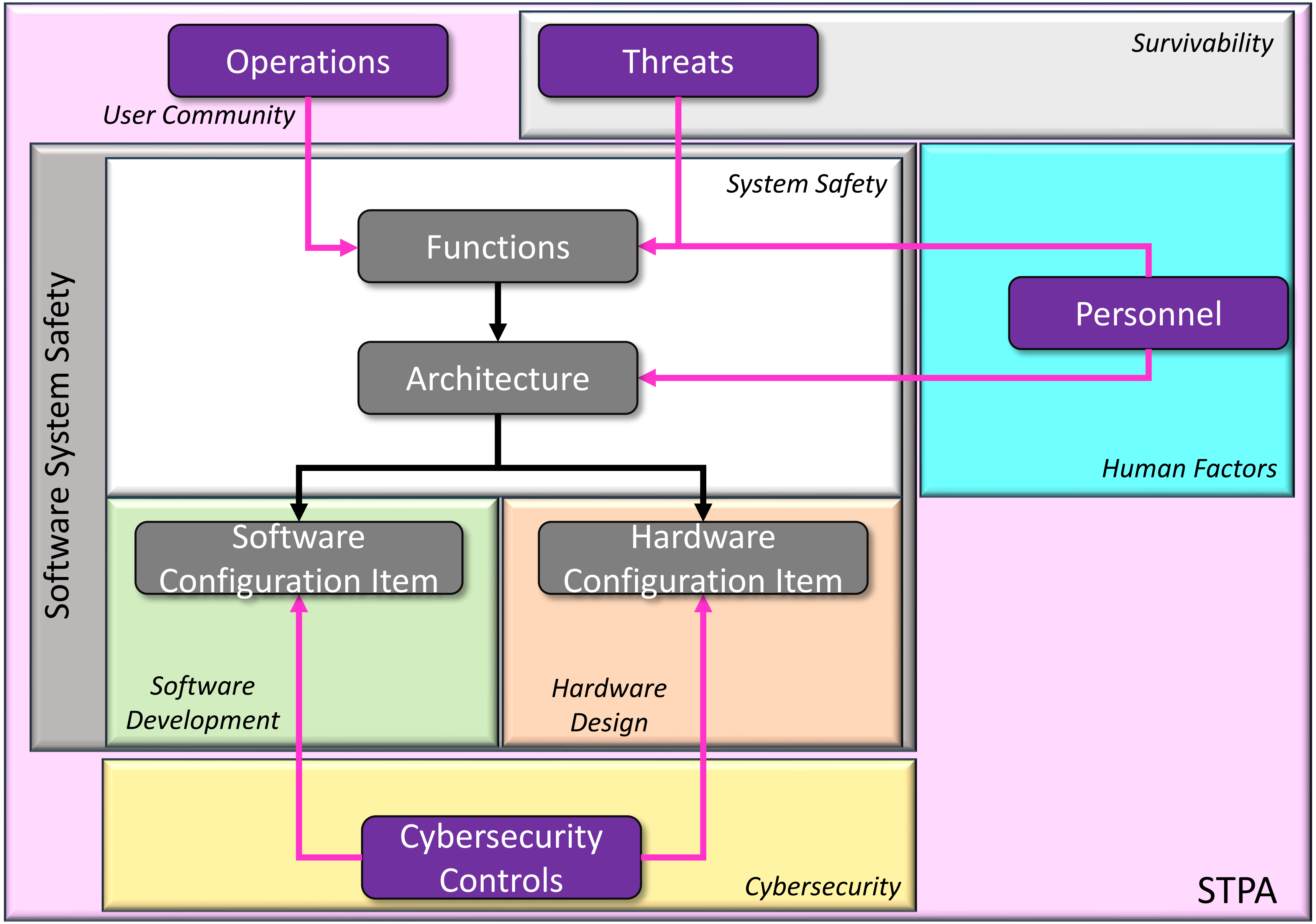
System theoretic process analysis (STPA) supports synchronous multidisciplinary evaluation of composite risk in operational environment and is identified as technique to support evaluation of AI systems.

STPA and Risk Assessment Processes



Source: AMCOM Safety Office, "Introduction to STPA," July 2023.

STPA vs. Traditional Swimlanes



Source: AMCOM Safety Office, "Introduction to STPA," July 2023.

Supplementing With STPA

✓ **STPA provides a structured approach to address:**

- Hazards that do not result from failures
- Human contribution to the occurrence of hazard, not necessarily related to design (e.g., doctrine, training, culture)
- Software contributions in complex systems
- Unintended effects resulting from complex interactions or system integration (e.g., software, hardware, operators, maintainers, engineers)
- Unintended effects resulting from interactions between safety, survivability, and cybermitigation of these risk sources
- Process deficiencies in design, training, operating, and supporting
- Design in early stages of defining ConOps or ConEmp or evolving ConEmp

Source: AMCOM Safety Office, "Introduction to STPA," July 2023.

U.S. ARMY

Example Stakeholders

Step 1

T4

Stakeholders are identified from analyzing operational concepts, applicable doctrine, and program's requirements. The stakeholders then identify their "loss" from what they value.

Source Information to identify stakeholders:

- Operational View-1 (OV-1)
- Joint Publication (JP) 3-0
- JP 3-30
- Army Techniques Publications 6-0.5
- The Law of Armed Conflict
- Capability Description Document
- Specifications

ID	Stakeholder
S1	**Commanders**
S1.1	Chief of mission in the command identified as higher mission authority (combatant commander and operational-level joint force commander)
S1.2	Tactical commander in the Tactical Operation Center (TOC) or Tactical Command Post (CP)
S2	**Aircraft Operation Personnel**
S2.1	Mission planners in the TOC supporting planning horizons shorter than long range
S2.2	Aircrew of all aircraft involved in a FARA-supported mission (all OVs) allocated to the CP
S2.3	Remote operators of all uncrewed systems involved in a FARA-supported mission (all OVs) located in ground control stations (as opposed to ground forces) and allocated to either the TOC or CP
S3	**Maintenance/Logistics Element** in the TOC and CP, representing the sustainment functional cell
S4	**Authorizing Officer and More Broadly Staff:** This category includes representatives of all functional cells (intelligence, movement and maneuver, fires, protection, etc.)
S5	**Combatant on the Ground, Ground Assault Forces**
S5.1	Ground scouts
S5.2	Inserted special operations forces
S6	**Noncombatants**
S7	**Pilot Instructors**

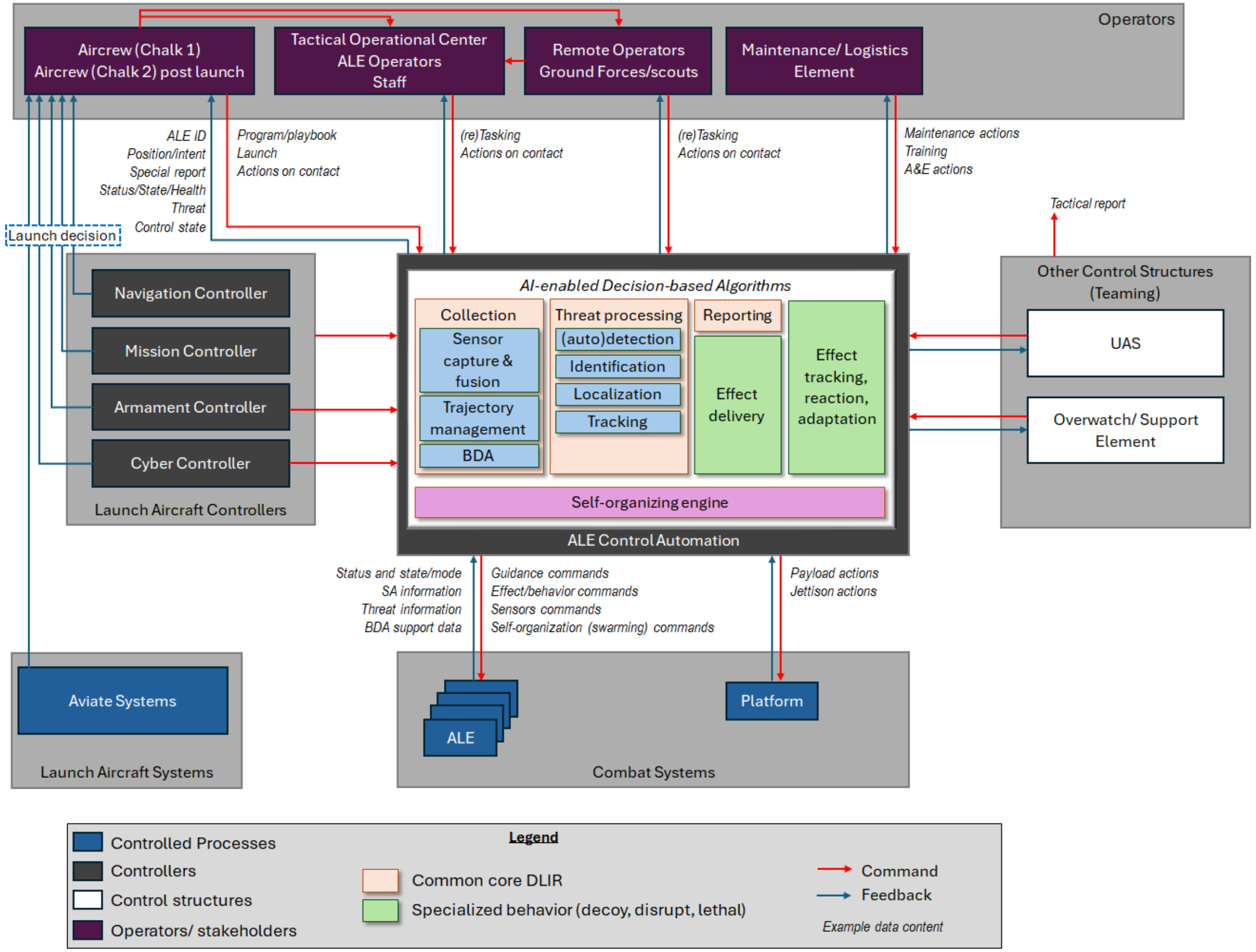
Bad actors, in theater or remote, using peer or near-peer threats or exploiting cybervulnerabilities, are not listed as stakeholders.

Source: AMCOM Safety Office, "Introduction to STPA," July 2023.

Launch Effect Hazard Analysis Structure (simplified)

Step 2

T4



Benefits from integrating hazard analysis structure derived from STPA:

- ❑ Supports composite risk management
- ❑ Is resilient against operational uncertainties
- ❑ Allows collaborative and simultaneous analysis for safety, cyber, software, and human-system integration
- ❑ Is compatible with agile methodologies, DevSecOps, CI/CD frameworks
- ❑ Seamlessly integrates CHMR assessments

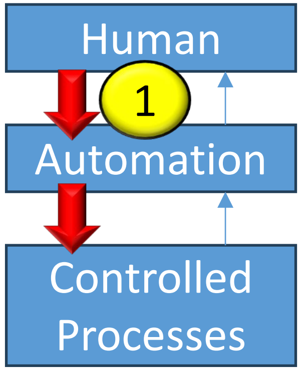
Implementation on future vertical lift ecosystem resulted in identifying missing interfaces and generated multidisciplinary test cases.

U.S. ARMY

Identify UCAs

Step 3

T4



- ✓ **UCA is a control action that, in a particular context and environment, leads to a hazard**
  - Ensures analysis is developed in context of intended (or foreseen) use
- ✓ **To structure UCA identification, there are four ways a control action can be unsafe:**
  1. Not providing the control action leads to a hazard
  2. Providing the control action leads to a hazard
  3. Providing a potentially safe control action but too early, too late, or out of sequence leads to a hazard
  4. The [continuous] control action lasting too long or stopping too soon leads to a hazard

  The UCA allows consideration of both failure-based and nonfailure based scenarios
- ✓ **UCAs generate constraints on the controller**

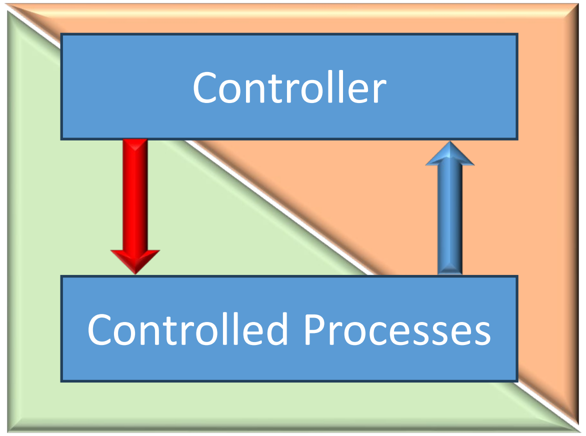
1 Human UCAs support HIS safety-cyberintegration

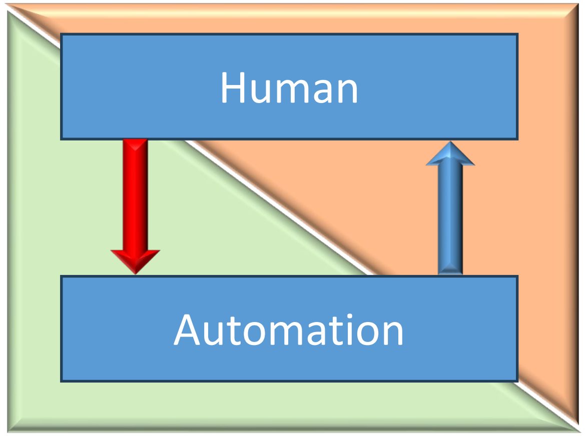
Source: AMCOM Safety Office, "Introduction to STPA," July 2023.

Identify Loss Scenarios

✓ **Loss scenario identifies the causal factors that can lead to the UCA (and, thus to, the hazard[s]).**

Why would UCA occur?





Why would control actions be improperly executed or not executed?

- Considers all risk sources, materiel, and nonmateriel (doctrine, process, training, etc.)
- Considers all stakeholders and all materiel controllers and the relationships/interactions to execute the mission
- Considers pathways for command and feedback separately
- Covers failure-based loss scenarios and nonfailure-based loss scenarios

Source: AMCOM Safety Office, "Introduction to STPA," July 2023.

Executing STPA—Lessons Learned

✓ The STPA Project Team Is Foundational to Success

Multidisciplinary	All stakeholders need representation (operation, sustainment, engineering, training).
Varied Depth of Expertise	Invitees need adjusted to the depth and scope of analysis for each step and within a step.
Facilitator	Key personnel for success need experience and training.
Dual-Loop Learners	All should think more deeply about own assumptions and beliefs.

STPA success is derived from
(1) the complexity of the problem to solve and (2) the selection of the team to assess it.

Source: AMCOM Safety Office, "Introduction to STPA," July 2023.

AMCOM Safety Office Initiatives

Legend
- Revision in planning
- Under major revision
- Revised
- **XX** AI related
- Current focus

Complexity of Implementation

Civilian Harm (822F)
AI/ML (822F)
Technologies
Hazard Analyses

Military Standard MIL-STD-882F System Safety Standard Practice (Joint Weapon Safety Working Group [JWSWG], Office of the Undersecretary of Defense for Research and Engineering [OUSD{R&E}])

Model-Based Systems Engineering Implementation Guide for System Safety (JWSWG, OUSD[R&E])

AR 385-10 Army Safety and Occupational Health Program

Crediting civil standards
Firmware
Nondevelopmental Items
*Use of tools and **models***
*Databases and **datasets***
*Integration: HSI, **AI/ML system of systems***
*Detailed guidance on **analyses***

AMCOM Regulation 385-17 Software System Safety Policy

AMCOM Regulation 385-10

Risk management, civilian harm
Materiel release type
AI-enabled software application, problem statement
Software acquisition pathway
Risk assessment
Operation of the Software System
Safety Technical Review Panel

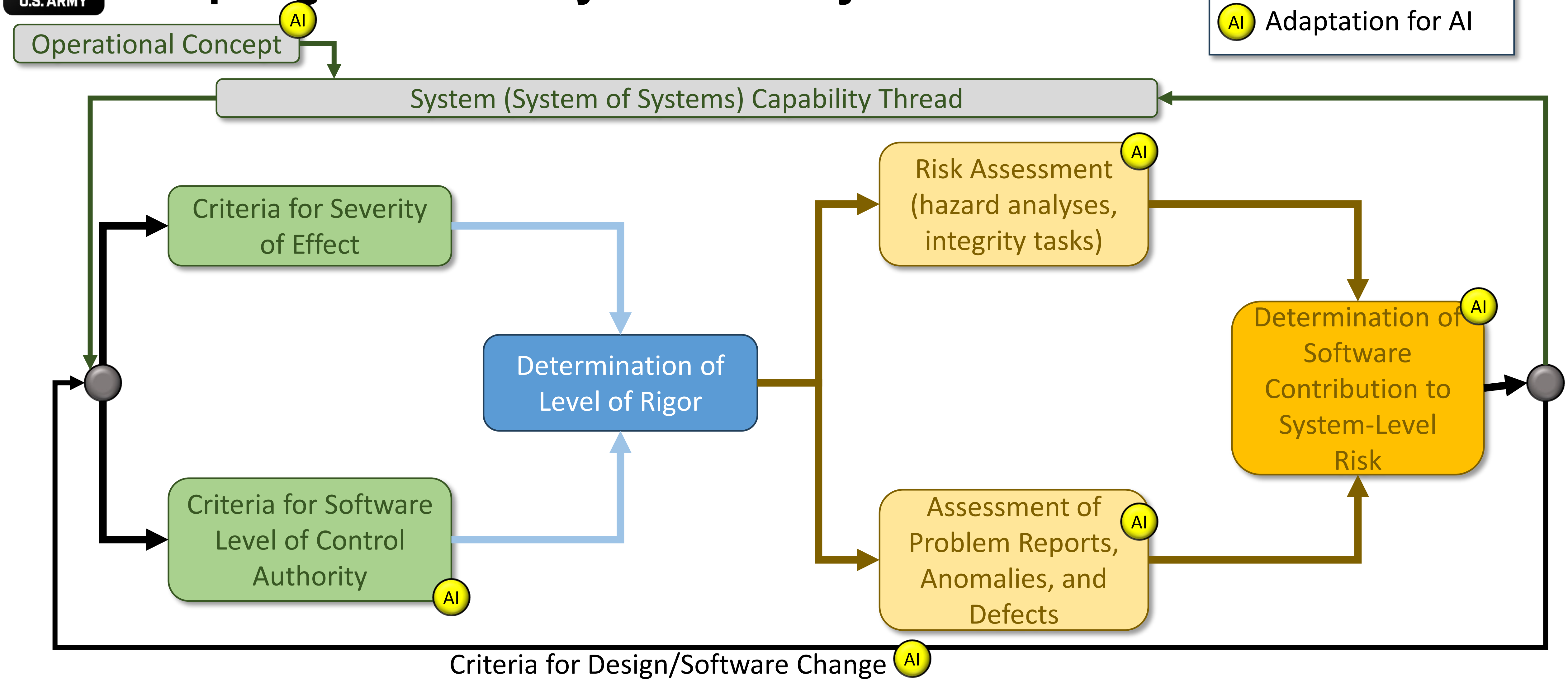
AMCOM Safety, Standard Operating Procedures

Activity Velocity

Adapting Software System Safety Processes

Legend

AI Adaptation for AI



AMCOM System Safety has initiated update of AMCOM Regulation 385-17 to address safety of applications embedding AI technologies for all impacted software integrity processes shown above.

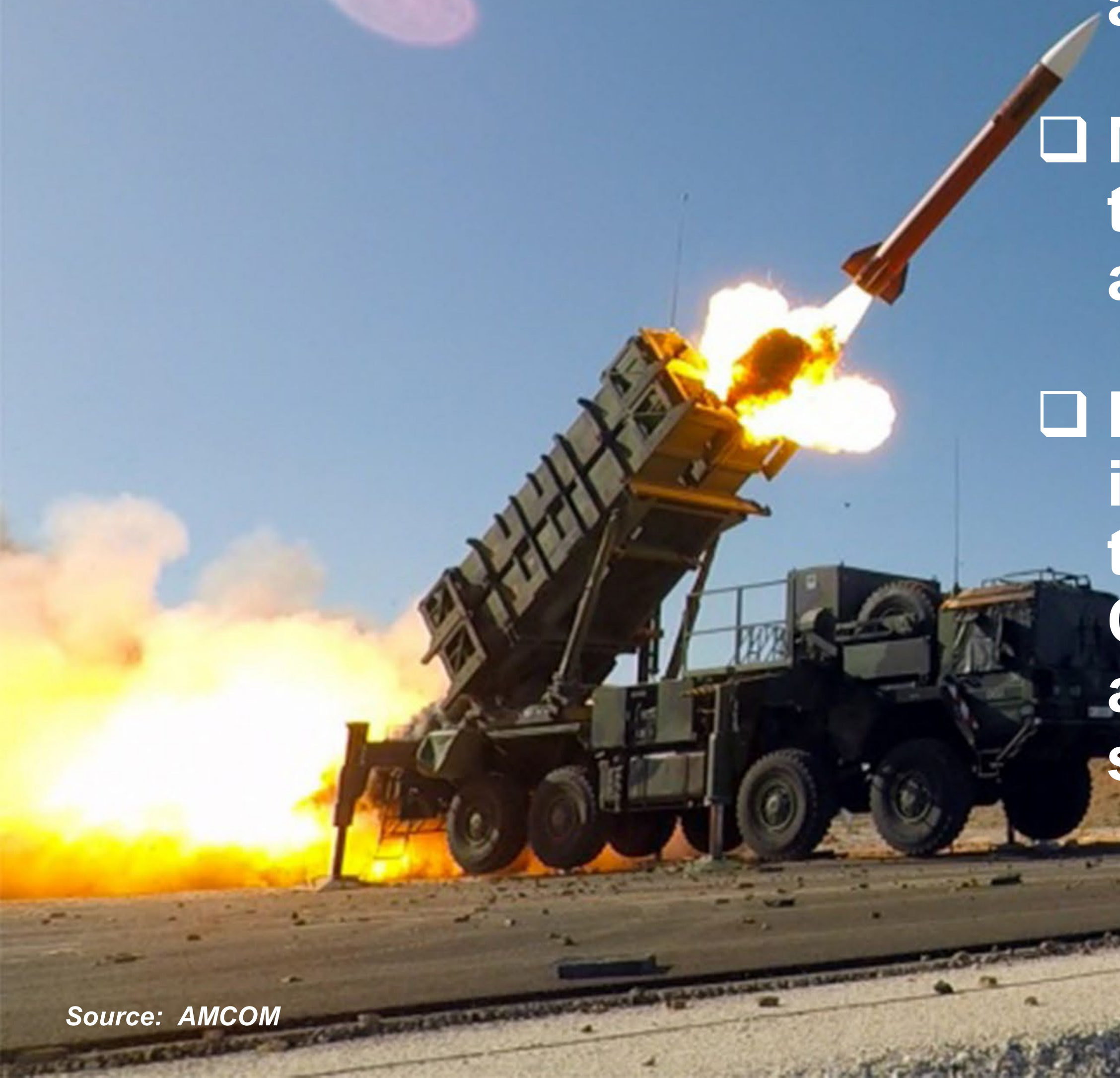
U.S. ARMY

Conclusions

- Pace of technology and ConOps challenge the deliberate update of regulations, policies, and guidance supporting materiel release to Soldiers; local initiatives are needed to provide a framework for developing guidance.

- Multiple and concurrent complex challenges to the safety certification should not be addressed in isolation or sequentially.

- Existing hazard analyses approaches may be ill suited for the level of flexibility/uncertainty that comes with some of the AI system ConEmp; multidisciplinary system-theoretic analyses carve an avenue but not a complete solution.



Source: AMCOM

THANK YOU

Point of contact: Laurence H. Mutuel, laurence.h.mutuel.civ@army.mil