Unmanned System Safety Engineering Precepts Guide for DoD Acquisition

August 2021

Office of the Under Secretary of Defense for
Research and Engineering

and

Office of the Under Secretary of Defense for
Acquisition and Sustainment

Washington, D.C.

Unmanned System Safety Engineering Precepts Guide for DoD Acquisition

Office of the Under Secretary of Defense for Research and Engineering
Deputy Director for Engineering
3030 Defense Pentagon
Washington, DC 20301-3030
https://ac.cto.mil/engineering

Office of the Under Secretary of Defense for Acquisition and Sustainment
Deputy Assistant Secretary of Defense for Weapon and Platform Portfolio Management
3010 Defense Pentagon
Washington, DC 20301-3010
https://www.acq.osd.mil/

Contents

Figures

Tables

Preface

The role of unmanned systems (UxS) in Department of Defense (DoD) military operations has expanded rapidly and is expected to continue to do so as the Department develops new, more complex UxS deployment and engagement strategies. As system design and components grow more complex, external sensor suites expand, and data processing speeds up, these forces in turn will push the limits of safe human operational command and control.

In light of the UxS advances and their potential impact on safety, the DoD Offices of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) and OUSD for Acquisition and Sustainment (OUSD(A&S)) created this guide to update the *Unmanned Systems Safety Guide for DoD Acquisition* published in 2007. In this context, *safety* refers to freedom from conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment.

A team of UxS hardware, software, and system safety experts from across the Military Departments contributed to this guide. Organized according to "precepts" in three categories – programmatic, design, and operational – the guide is intended to support the development and design of safe UxS, associated safety significant software, support hardware and firmware, and Service safety reviews. The guide is directed toward UxS system safety engineers as well as UxS program managers (PMs), system designers, and test and evaluation (T&E) managers. The precepts are intended to be general, to be complemented by systems specific to a program office. The guide is intended to provide the PM with a point of initiation for precepts that can aid the development of a System Safety Engineering Program as required by DoD Instruction 5000.88, "Engineering of Defense Systems" (2020).

The guide includes a summary of the three types of safety precepts, an analysis of the major UxS safety concerns, and an assessment of the state of the art of artificial intelligence (AI) and autonomous capabilities, which, when integrated properly, can enable the desired performance of UxS autonomy, human-machine interaction, and command and control. The three appendices provide details on the three types of precepts, and the guide includes a revised glossary to help promote a common understanding of terms.

To ensure the proper and comprehensive application of the precepts, this guide includes background concepts, considerations, examples, and principles to aid system safety and design engineers to understand the intent of each precept as well as how engineers might extend the application of a precept to a unique UxS, a new technology, or various performance scenarios.

As more autonomous capabilities are implemented in UxSs, confidence in system operation and performance will be pivotal to operational use. System safety, software system safety, and T&E techniques will need to advance to facilitate the means to assess system safety, performance, and predictability and to characterize system behavior and mission capability across selected factors.

Confidence in system operation begins with requirements, which should be reflected in the design guidelines and development approaches. The requirements following T&E need to include operator training on system capabilities and limitations. All design and development efforts must be considered to ensure safe operation of the UxS.

This guide will aid the PM's team, the operational commander, and the systems engineer to recognize and mitigate system hazards unique to unmanned design capabilities. It augments the tasks within Military Standard 882E, "Department of Defense Standard Practice for System Safety," with additional details to address UxSs and the incorporation of greater levels of autonomy and AI. The precepts that apply to AI or autonomous technology comply with and support multiple principles within Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence."

While this guide highlights system safety in its application to UxSs, it also addresses how the systems engineering team operationalizes the *Ethical Principles for Artificial Intelligence* released by the Joint Artificial Intelligence Center (JAIC) (February 2020). Specifically, many of the concepts and precepts in this guide are tangible ways in which to instantiate the Department's AI Ethical Principles, including "reliable" and "governable." For example, to support the principle of "governable," this guide can assist the systems engineering team to detect and avoid unintended consequences within a system design and operation and can assist the operator to understand the process to disengage or deactivate deployed systems that could present unintended behavior.

This guide does not alter or supersede existing authorities and policies of the Under Secretary of Defense for Policy regarding autonomy in weapon systems as directed by DoD Directive (DoDD) 3000.09, "Autonomy in Weapons Systems" (2012), or other laws and regulations.

As this guide focuses specifically on system safety engineering, it does not provide comprehensive advice regarding programmatic, operational, or design precepts for environmental, occupational health, or human systems integration disciplines for UxS, nor does it discuss the specific UxS risks associated with those areas.

POSSEHL.STEPHANIE.LYNN.1229383805
Digitally signed by POSSEHL.STEPHANIE.LYNN.1229383805
Date: 2021.11.04 08:49:12 -04'00'

Deputy Director, Engineering
Office of the Under Secretary of Defense for Research and Engineering

RUOCCO.JAMES.A.1013973802
Digitally signed by RUOCCO.JAMES.A.1013973802
Date: 2021.11.22 11:20:19 -05'00'

Deputy Assistant Secretary of Defense for Platform and Weapon Portfolio Management
Office of the Under Secretary of Defense for Acquisition and Sustainment

1 Introduction to Unmanned System Safety Precepts

This guide presents programmatic, design, and operational precepts to ensure the safety of Department of Defense (DoD) unmanned systems (UxSs), regardless of domain or technology. This guide provides generic programmatic, operational, and design safety precepts to assist the program manager's (PM) team, the operational commander, test planners, and the systems engineer responsible for oversight of the design of the UxS. The safety precepts provide a baseline and represent agreed-upon best practice for the program management, operation, and design of safe UxSs.

The precepts should be implemented as early in the systems development as possible and as part of the program's system safety management strategy. Each precept recommends certain actions, operational controls, or design considerations. Programs should document adherence to and any deviation from the recommended safety precepts provided in Tables 1-1, 1-2, and 1-3 with appropriate engineering rationale. The appropriate risk acceptance authority should review and approve contributions to system risk when safety precepts are not applied to the management, operation, and design of UxSs. Safety precepts are intended to aid the PM's team, operational commanders, and system and design engineers in their responsibilities but not to dictate specific solutions. In addition, each UxS design program, upon identification of potential safety hazards, should define unique new precepts (i.e. tailored) to address system safety as the program design matures.

Each precept description should provide a basic truth, or presumption to enhance the safety of the system as discussed further in this section and in appendices A, B, and C. Use of these precepts complement the system safety program plan (SSPP) effort of identifying, analyzing, and mitigating system hazards found within the subject UxS. The process used to execute an SSPP is found in Military Standard 882E (MIL-STD-882E), "Department of Defense Standard Practice for System Safety."

Advances in technologies and new energy sources, increasing sophistication of system performance, complexity of system designs and components, expansion of external sensor suites, and greater and faster data processing all will require parallel advances in design engineering and analytical techniques and processes. Emerging and rapidly maturing autonomous technologies are spurring test and evaluation (T&E) and system safety engineering authorities to consider new techniques and methodologies to assess system performance and operational safety. Continued growth in unpredictable, constantly changing operational profiles that will manifest from these new technologies creates safety concerns that challenge both management and technical aspects of system acquisition.

All test infrastructure and tools (e.g., models, simulations, automated tools, synthetic environments) that support acquisition decisions must be verified, validated, and accredited by the intended user or appropriate agency as required in DoD Instruction (DoDI) 5000.89, "Test and Evaluation" (2020). In addition, DoDI 5000.80, "Operation of the Middle Tier of Acquisition (MTA)" (2019), and DoDI 5000.87, "Operation of the Software Acquisition Pathway" (2020), require a test strategy. The program's chief developmental tester or T&E lead, in collaboration with other T&E stakeholders including system safety engineers, should develop

the test strategy and discuss the approach to developing measurable criteria derived from requirements (e.g., user features, user stories, use cases).

Advances in technology and design include those that allow humans or machines to have oversight of one or more UxSs and the commander to decide whether the control entity will be a human or a machine. In this guide, the term "command entity" refers to a human in a position of authority to command and control the UxS, who can authorize a control entity to execute commands, and who can rescind authorization or transfer it. The term "control entity" refers to a human or software, when authorized to command and control the UxS, who can authorize a control entity to execute commands, and who can rescind authorization or transfer it to a different control entity at any time. An "operator" refers specifically to a human control entity.

Autonomous capabilities create unique safety challenges beyond those addressed in other safety guidance. A proposed set of system engineering issues for autonomous systems that could lead to or introduce associated safety challenges and considerations to mitigate those issues include:

- Synchronizing technology development with life-cycle planning – see Sections 1.1 and 2
- Understanding and managing human-machine interaction – see Section 1.2
- Refined requirements development – see Section 2 and Precepts
- Understanding consequences of self-learning systems – see Section 2.2
- Enhancing analysis, evaluation, and certification – see Section 3.1
- Unpredictable behavior of tightly coupled complex systems – see Sections 2.1, 2.2.2, 3.1, 3.3, 3.3.1, and Precepts
- Human-machine trust – see Sections 2, 3.1, 3.3, 3.3.1, 3.3.2, and 3.4
- Possibility of perverse instantiation or an artificial intelligence (AI)/autonomous system causing unintended harm to complete its mission – see Table 2-1 and Precepts
- Promoting the responsible design and development of self-learning systems

To be deployed, a UxS must overcome these challenges, especially with a focus on the higher probability and more serious safety risks. UxS challenges will grow rapidly as more complex and functionally sophisticated autonomous systems are introduced, for example, the complex autonomous systems addressed in the *Technology Investment Strategy 2015-2018* (Autonomy Community of Interest (COI) Test and Evaluation, Verification and Validation (TEVV) Working Group 2015) and *DoD Priorities for Autonomy Research and Development* (Stone 2011). The safety precepts facilitate addressing the aforementioned system engineering issues associated with autonomous capabilities that potentially lead to or introduce associated safety challenges.

The precepts discussed in this section and in more detail in appendices A, B, and C facilitate effective system safety analyses for UxSs and outline techniques to mitigate the challenges. Use of the precepts can help manage the magnitude of the safety aspects of the challenge and other aspects of the overall acquisition effort. The precepts recommend how to operationalize the DoD's AI Ethical Principles, especially "reliable" and "governable."

Although the precepts do not create requirements regarding how test, evaluation, verification, and validation (TEV&V) should be performed, they do support an adequacy of safety and safe operation of the system during the T&E and V&V processes. Designs that apply the precepts will generally have improved capability to support TEV&V processes being evolved to meet the above listed set of challenges while complying with DoD policy.

The three types of safety precepts are as follows:

1. Programmatic safety precept (PSP) – directed specifically at program management principles and guidance that will help ensure safety is adequately addressed throughout the life-cycle process.

2. Operational safety precept (OSP) – directed specifically at system operation and include operational rules that should be adhered to during system operation. These safety precepts may generate the need for design safety precepts (DSP).

3. DSP – directed specifically as general design guidance intended to facilitate safety of the system and minimize hazards. These safety precepts are intended to influence, but not dictate, specific design solutions.

The appendices provide additional information about each precept, including the following:

- Scope – A statement addressing the applicability of each safety precept.
- Rationale – A statement explaining why each safety precept is required.
- Examples – Sample system functions or operational events germane to the intent of each safety precept.
- Detailed considerations – Information to assist in implementation of the safety precept.

There are technical and ethical interrelationships among the three precept types and with standing policy as well as the DoD's AI Ethical Principles, which apply to the design, development, deployment, and use phases of AI-enabled capabilities. Therefore, it is important to consider the entirety of this guide when assessing the applicability of a precept as well as understanding the detailed considerations prudent to implementing certain precepts.

1.1 Programmatic Safety Precepts

The PSPs described in Table 1-1 are directed specifically at program management. These precepts are designed to ensure safety is adequately addressed throughout the UxS life-cycle (e.g., development, acquisition, and sustainment). For a program to be successful in developing an optimized system, it is incumbent upon the program office to establish system safety engineering management early in life-cycle planning to instill a robust system safety culture in the program.

The PSPs listed in Table 1-1 provide the programmatic guideposts necessary to ensure the primary tenets of policy, DoD's AI Ethical Principles, and UxS design, development, and operational best practices are considered as part of the program management strategy.

Table 1-1. Programmatic Safety Precepts

PSP-1	The program office should integrate into their MIL-STD-882E safety program the UxS safety precepts and the DoD's AI Ethical Principles.
PSP-2	The program office should ensure the UxS complies with current safety policy, standards, and design requirements and the DoD's AI Ethical Principles.
PSP-3	The program office should ensure adherence to and any deviation from the UxS safety precepts are addressed during program reviews.
PSP-4	The program office should ensure off-the-shelf items, reuse items, original-use items, design changes, technology refresh, and technology upgrades within the system are assessed for safety.
PSP-5	The program office should ensure that the UxS, by design and operation, does not allow subversion of an authorized human command or control of the UxS.
PSP-6	The program office should ensure that the UxS safety significant functions and components are not compromised when utilizing flexible autonomy.
PSP-7	The program office should prioritize personnel safety in UxS intended to team with or operate alongside manned systems.
PSP-8	The program office should ensure authorized and secure control (integrity) between platform and controller to minimize potential UxS mishaps and unauthorized command and control.
PSP-9	The program office should ensure that software systems exhibiting unpredictable, nondeterministic behavior are employed safely and comply with current policy.

1.2 Operational Safety Precepts

OSPs described in Table 1-2 are directed specifically at system operation. These precepts contribute to operational rules that should be adhered to during system use and operation and may generate a need for DSPs.

Some OSPs listed in Table 1-2 mitigate risk associated with a command entity and operator actions that involve the human-machine interface, while others mitigate risk associated with sharing control of the UxS between the human and the autonomous machine. Some OSPs are tied tightly to DSPs due to the nature of the human-machine interface and the functions the operator needs to have at their disposal when a UxS deviates from its mission for any reason.

Table 1-2. Operational Safety Precepts

OSP-1	The control entity of the UxS should have adequate mission information to support safe operations.
OSP-2	The UxS should be considered unsafe until a safe state can be verified.
OSP-3	The control entity should verify the state of the UxS to ensure a known and intended state prior to performing operations or tasks.
OSP-4	The UxS weapons should be loaded and energized as late as possible in the operational sequence.
OSP-5	Only authorized, qualified, and trained personnel using approved procedures should operate or maintain the UxS.
OSP-6	The operator should be aware during all phases of the mission when autonomous behaviors are utilized.
OSP-7	The operator should be able to establish alternative recovery points prior to or during mission operations.
OSP-8	Weapons should only be fired or released with human consent, or with control entity consent in conjunction with preconfigured criteria established or verified by the operator.
OSP-9	The operator should have the ability to take control of the UxS, as appropriate and feasible.
OSP-10	The operator should have the ability to abort the mission of the UxS, if appropriate and feasible.
OSP-11	The operator should be able to disable learning mode.
OSP-12	One operator should maintain positive and active control of the UxS during transfer of control.

1.3 Design Safety Precepts

DSPs outlined in Table 1-3 provide detailed and specific design guidance to consider. This guidance will assist in addressing potential safety issues during the design and development of UxSs thus reducing the potential for a design related mishap from manifesting during operational use of the system. This guidance is the direct result of experience and lessons learned on both manned and UxSs.

DSPs should influence, but not dictate, specific design solutions to mitigate system hazards. The DSPs address, among other things, the functional partitioning of the UxS components or subsystems and hazard mitigations within some of the key subsystems such as weapons, command and control, human-machine interfaces, and states and modes. These safety precepts may generate a need for, or relate to, an OSP.

Table 1-3. Design Safety Precepts

DSP No.	Description
DSP-1	The UxS should be designed to minimize the mishap risk during all life-cycle phases.
DSP-2	The UxS should be designed to fulfill valid commands only from the control entity.
DSP-3	The UxS should be designed to provide command and control for safe operations.
DSP-4	The UxS should be designed to prevent unintended release and/or initiation and firing of lethal and nonlethal weapon systems or any other form of hazardous energy.
DSP-5	The UxS should be designed to prevent release and/or firing of weapons into the UxS structure itself or other friendly UxS/weapons.
DSP-6	The UxS should be designed to minimize the potential for releasing or firing of a weapon on a friendly or unintended target group selection.
DSP-7	The UxS should be designed to safely initialize in the intended state, safely and verifiably change modes and states, and prevent hazardous system mode combinations or transitions.
DSP-8	The UxS should be designed to include an abort function that transitions the system to a safe state.
DSP-9	Safety significant software should be appropriately physically and functionally partitioned.

DSP No.	Description
DSP-10	The UxS should be designed to minimize single-point, common-mode, or common-cause failures that result in high and/or serious risks.
DSP-11	The UxS should be designed to transition to a preconfigured safe state and mode in the event of safety significant failure.
DSP-12	The UxS should be designed for safe recovery if recovery is intended.
DSP-13	Use of the UxS newly learned behavior should not impact the UxS's safety functionality until the newly learned behavior has been validated.
DSP-14	Autonomy should only select and engage targets that have been predefined by an authorized human.
DSP-15	Common user controls and display status should be utilized for similar functions such as manual override, terminate mission, and learning mode.

2 Unmanned System Safety Analysis

This section provides information and concepts to consider by the system safety practitioners as they begin to bind the process for characterizing the mishap hazards and risks for their particular UxS.

When compared to manned systems, the fundamental techniques and methodologies in the overall system safety engineering analytical processes remain unchanged for UxS. In traditional acquisition design processes, software system safety synchronizes with the design effort to ensure a thorough analysis of the software functionality and an adequate understanding and characterization of the software levels of autonomy. System safety engineering is integrated into the overall systems engineering process. The processes detailed in MIL-STD-882E are used to address environment, safety and occupational health risks associated with system-related hazards. In addition to MIL-STD-882E, the guidance identified in the *DoD Joint Software Systems Safety Engineering Handbook* (JSSSEH), Joint Software Systems Safety Engineering Working Group (2010), and the *Software System Safety: Implementation Process and Tasks Supporting MIL-STD-882E*, Joint Services – Software Safety Authorities (JS-SSA) (2017), are used to assess software contribution to system level risk.

Software system safety practices and techniques are also used to accurately assess the software's level of criticality and ensure the appropriate level of rigor (LOR) and analytical techniques are applied prior to and during TEV&V activities in accordance with the DoD JSSSEH. The software test strategy as part of the acquisition strategy should discuss the approach to developing measurable criteria derived from requirements (e.g., user features, user stories, use cases) as well as the identification of test platforms and infrastructure.

Conversely, acquisition processes and schedules continue to compress, and the speed of design and development of new systems with new capabilities continues to increase – driving a more agile and even spiral acquisition development process. This, combined with increases in technical complexity and software-controlled or -invoked functionality, is driving more iterative software development processes.

Such iterative development processes are consequently driving rapid spinouts of software versions to address immediate or looming warfighter requirements or needed functionality. Software system safety processes must synchronize with these iterative and dynamic activities in order to effectively incorporate necessary software LORs and system safety mitigations into these rapidly designed and deployed systems and for responsible design, development and use of the systems.

As technologies continue to advance in the areas of AI and machine learning, safety assurance in the system becomes a much more prominent area of interest and analysis. These technological advances, coupled with a more iterative system and software development profile, also increase the system safety engineering workload for system safety and software safety analysis.

Of course, as the pace at which the occurrence of these processes increase and the need for the responsible design and development of these systems is heightened. The precepts provided

herein can assist the systems engineer in considering a menu of system safety requirements that may mitigate potential hazards from occurring despite the rapid acquisition profile, thus providing greater confidence of safety of the system.

2.1 Unmanned System Potential Mishaps and Safety Concerns

The five example mishaps presented in Table 2-1 may be used by any UxS PM to assist in identification of crucial safety areas for their particular program. While the list of mishaps presented in Table 2-1 is not exhaustive, the depth and breadth of the example mishaps facilitated the development of all safety precepts presented in this document. These example mishaps are presented in Table 2-1 to facilitate a better context when considering particular precepts. The mishaps listed have not been prioritized and are in random order. For guidance on mishaps which may result in environmental damage, consult Service-level environmental policy, requirements, and legal guidance.

Table 2-1. Example Potential Mishaps

Example Potential Mishaps	
Mishap-1	Unintended/abnormal system operation or performance degradation
Mishap-2	Inadvertent firing or release of weapons
Mishap-3	Engagement/firing upon unintended targets
Mishap-4	Self-damage of own system from weapon fire/release
Mishap-5	Vehicle collision

Manned and unmanned systems are hazardous to humans for many different reasons, ranging from unpredictable movements, to loss of control, to potential failures in either hardware or software. If the system is an armed weapon system, the range of potential dangers is that much greater. The hazards posed by a system could result from inadequate design, inappropriate use, or failures in the system's hardware and/or software.

Safety concerns or causal factors that may result in or lead to a higher level mishap for military UxSs that apply to semi-autonomous, human-supervised, and autonomous UxSs include, but are not limited to, the following:

1. Loss of command and control over the UxS.

2. Loss of necessary and intended communications with the UxS.

3. Loss of UxS weapons.

4. Unsafe UxS returns to base.
5. Indeterminate or erroneous state of UxS.
6. Knowledge of the state a UxS resides in (e.g., safe or unsafe).
7. Unexpected human interaction with the UxS.
8. Inadvertent firing of UxS weapons.
9. Erroneous firing of UxS weapons.
10. Erroneous target discrimination.
11. UxS injures operators, own troops, etc.
12. Loss of or inadequate situational awareness.
13. UxS exposure to radiation, biological contamination, etc.

As the use of autonomy and collaborative human-machine missions increase, additional safety and performance concerns will undoubtedly arise. Performance concerns include:

1. Differences between real mission data versus training data leads to inappropriate or undesirable behavior by the UxS.
2. UxS executes orders using an approach that is illegal and/or unethical for the situation.
3. UxS takes orders or commands from unauthorized parties (e.g., from other autonomous systems or from enemy operators).
4. UxS executes order based on autonomously derived criteria.

Safety analysis for any particular UxS can use these sets of concerns as a starting point to generate and categorize system-specific safety risks, causal factors, and mishaps.

2.2 States and Modes

States identify the conditions in which a system or subsystem can exist. A system or subsystem may be in only one state at a time. A safe state is a state in which the system poses an acceptable level of risk for the operational mode and environment. For example, “weapons armed” is not a safe state during logistics and pre-deployment modes, but “weapons armed” is a safe state when engaging a target (except to the enemy). Thus, safety analysis can be viewed as evaluating the level of risk for different states for different operating modes.

Used herein, the term "mode" refers to operational views of the system and the term "state" refers to design views of the system. For example, the overall system could be in learning mode, training mode, maintenance mode, transit mode, execution mode, etc. In addition, the system or individual functions could be in a degraded operational mode when the components are in a degraded state.

The overall safety of a system depends upon understanding its states and modes and the transitions among them; this is particularly true in UxSs. Each combination of operating modes that fully describes an operational configuration of the system maps to a set of states. The system will generally transition among states during execution of a task. Some combinations of modes may map to an unsafe set of states; others to a safe set of states. Yet other combinations of modes may map to a set of states in which some are safe and others are unsafe. Some states may occur for multiple combinations of modes.

Standard system safety practices use legacy concepts of states, modes, and transitions to or from a particular state or mode as part of the analytical process. Machines that learn may dynamically create new states and even new modes as they learn, which challenges the ability to do thorough safety analysis by standard processes. To address this and to assist in achieving the intent of the precepts, the concepts of deterministic checkpoints and learning mode are discussed below.

2.2.1 Deterministic Checkpoints

This guide coins the term "deterministic checkpoint" to characterize the use of an appropriate means of bounding the potentially unpredictable, non-deterministic behavior of the complex AI technologies that will increasingly be used to implement the capabilities and functions of UxS. This term is intended to cover the breadth of possible approaches that are available – from the implementation of human-in-the-loop oversight to the use of an appropriate run-time monitoring, run-time verification, safety monitoring, or safety kernel control component. For example, the run-time assurance architecture documented in ASTM F3269-17, "Standard Practice for Methods to Safely Bound Flight Behavior of Unmanned Aircraft Systems Containing Complex Functions," which documents the use of a "pedigreed" run-time framework including a Safety Monitor and one or more "deterministic" Recovery Control Functions to bound the behavior of a "non-pedigreed" complex function, which might implement a "non-deterministic", "intelligent", or "adaptive" algorithm.

A deterministic checkpoint is a point in the process at which the control entity, whether operator or software, may review the planned behavior chosen by the system's software and decide to intercede or not (for supervised autonomous), or give approval or not (for semi-autonomous) software. In some instances, the control entity will decide based on mission performance. In other instances, the control entity will decide based on safety.

Deterministic checkpoints provide an example approach to help enable safety analysis of those critical mission decision points that involve a safety significant decision. By implementing the checkpoints as deterministic software, the checkpoint functions can be verified and validated, which is important to the safety analysis. Their function can include checking and monitoring that the autonomous software's performance is restrained and in the specified operational mode.

Including deterministic checkpoints for a UxS provides a means for the mission control architecture to meet safety precepts for maintaining oversight by the control entity of AI or autonomous behaviors (OSP-6) to ensure safe operation (OSP-1, OSP-2, OSP-3, and DSP-3) and provide for safe recovery (OSP-7 and DSP-12).

As part of that function, reliability is also improved if the checkpoint confirms that the autonomous function's decision logic is reaching decision points at a tempo compatible with good execution of the assigned tasks, and if the deterministic checkpoint has alternative actions to apply when necessary. In addition to verification and validation (V&V) of system performance, as well as reliability of the system decision logic, these checkpoints can also aid in the review and assessment of the responsible design and development of the system.

2.2.2 Safe Machine Learning

For the foreseeable future the overwhelming majority of computational resource requirements in modern machine learning techniques are driven by off-line training or learning. Moreover, it is acknowledged that real-time on-line running of the decision engine during operations would consume resources that could otherwise be devoted to operations, and if care is not taken, this could introduce safety hazards of its own

Real-time on-line learning would be accomplished by replicating the human decision-making process and allowing the UxS systems to implement change in real-time. A subset of machine learning is also known as self-modifying software, where the system can learn and update its own software code and execute the new code and decision paths.

Machine learning enables machines to learn to do tasks without a need for task-specific programming. Thus, during training, when the machine is to learn how to perform a set of tasks, the machine's learning function must be enabled so that it can develop and refine its decision logic. However, if real-time learning is available, learning occurs after validation of the machine's learned decision logic, the resulting changes to the decision logic nullify that validation. Thus, for safety, learning must be validated prior to execution.

Until newly learned behavior has been validated, the UxS in question must not be able to implement those changes in behavior. A learning mode that allows enabling or disabling of machine learning is one means of achieving this. Another technique might be to conduct real-time V&V, as well as risk analysis on newly learned behavior. It is conceived that there will be interconnection between these two modes which may create safety hazards that need to be managed appropriately. Newly learned behaviors also require the need for an analysis of ethical considerations such as increased risk of unintended outcomes, performance reliability, system integrity, and traceability.

When off-line or real-time on-line learning mode is available but disabled during a mission, the system executes the mission using decision logic installed in the system when the mission was launched or initiated. Depending on issues other than safety, the system may collect and update data that is used together with the decision logic to select behaviors. It may develop new decision logic and log such developments for future evaluation and V&V. However, when

learning mode is disabled, it is prohibited from using the new decision logic. The new decision logic can be stored for future analysis and application.

When real-time learning mode is available and enabled during mission execution, the machine can implement new decision logic and the corresponding modified behaviors. The circumstances for which enabling learning mode would be beneficial during mission execution may be system-specific or mission-specific. The decision to have learning mode enabled during a mission depends on the potential mission value of the modified behaviors relative to the risks, including safety risks, associated with using non-validated decision logic. The learning mode could be enabled for all or part of the overall mission and limited to specific tasks.

During training of the machine, learning mode must be on so that it can learn; however certain characteristics of the overall mode might be different than they are during a mission. For example, any learning mode capability related to target engagement would probably be disabled during training and deployment operations. The system's weapons might be replaced by non-destructive equivalents, or other measure might be taken to ensure that the system would not be able to cause unintended damage while it undergoes training. During training of operators, learning mode would generally be 'off' for safety, except when operators are being trained in the use of the machine's learning mode. Additional and variant modes related to machine learning may be identified to augment this set as technologies mature that support the creation of autonomous systems.

3 Technical Discussion for Assurance of Autonomous Systems

This information should facilitate a better perspective on the precepts provided herein. Some portions introduce safety aspects extracted from detailed discussions available in the guide's references. Other portions introduce safety concepts that were developed by working groups that contributed to this guide.

This guide can help the system safety engineer to recognize the risks that may be associated with human-machine collaboration. It also provides precepts to mitigate some of those risks and address the responsible design, development, and use of these systems. Human-machine collaboration is not a new concept, but it will require a case-by-case analysis to determine how to best implement it. In addition, it will be incumbent on the system safety engineer to identify additional questions and topics to be analyzed for safety impacts. Moreover, the system safety engineer's challenge goes beyond safety for a single type of UxS; it includes identifying and mitigating potential risks at the unit level, for units comprised of people and heterogeneous sets of machines.

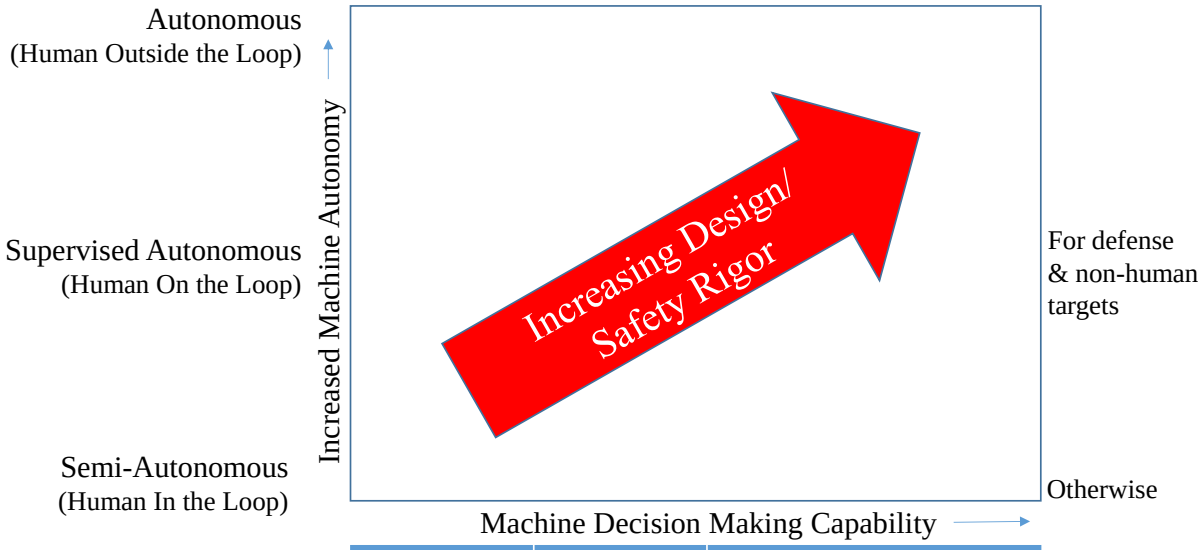
3.1 Verification and Validation, Test and Evaluation, and Safe Autonomy

To delineate the V&V and T&E challenge, it is important to consider the amount of authority the UxS is given to execute the decisions that it makes and under what circumstances. Figure 3-1 introduces the relationship between the V&V and T&E challenge and the UxS's decision-making capabilities. Further, the decision-making capabilities of the UxS depend on the system designer. For systems that learn, the decision-making capabilities also depend on the training process that the UxS undergoes. The autonomy afforded the UxSs depends on the intended operational environment and use for the system, the system designer, the operator, DoD policy, the assurance developed during V&V and T&E and the confidence developed during deployed use of the system.

Figure 3-1 summarizes some of the complex relationships between the V&V and T&E challenge, the systems capabilities, and the degree of autonomy afforded it. The degree of autonomy afforded the UxSs depends on intended operational environment and use for the system, the system design, the operator, DoD policy, the trust developed during TEV&V, and the trust developed during deployed use of the system. In Figure 3-1, the horizontal axis represents the increasing complexity of decision-making required to execute operationally useful and safe behaviors. Associated with this complexity will be an increasing difficulty in predicting system behaviors. The vertical axis represents the amount of autonomy that the machines may be allowed to execute those decisions.

As UxSs become more capable of making autonomous decisions, system safety must determine the level of risk associated with allowing those UxSs to act on those decisions, apply UxS safety precepts to mitigate those risks, and determine how much autonomy the UxS can be safely allowed. The challenge to make the system capable and safe while meeting policy and passing the V&V and T&E portion of the acquisition process increases both as the machines decision-making capabilities increase and as the levels of human control that it is provided decreases.

This is represented by the large arrow in Figure 3-1. The magnitude of the challenge as UxS systems evolve in the direction of the arrow may increase dramatically.



	Automatic	Automated	Autonomy	
			Behavioral	Cognitive
V&V	Verifiable	Formally Verifiable	Partially Verifiable	TBD
Technology	Data Processing	Expert System	Artificial Intelligence	
Complexity	Simple	Complex	Highly Complex	TBD
Outputs	Deterministic	Deterministic	Chaotic	TBD

Figure 3-1. Test and Evaluation & Verification and Validation Complexity Challenge

In Figure 3-1, at the "automatic" end of the horizontal axis, the machine's performance of a function is fully deterministic in its operation (always provides the same output for the same input). It is verifiable and programmed entirely by a person or people. At the "autonomy" end of the scale, machines perform decisions through the use of AI technologies such as machine learning to first learn tasks and then to improve performance over time. Between those two endpoints are "automated" functions.

Automated functions may reproduce human-level decisions and eventually, in part, may use some AI technologies to do so. However, autonomous functions are generally implemented with complex software in which pre-determined system operator actions have been explicitly and accurately programmed to repeat the actions by the autonomous system for a specific problem set. While an automated system may include some learning-like behaviors, those too are generally performed per human programming to reproduce actions specific to the task.

The vertical axis enumerates three levels of autonomy that may be granted to the machine for execution of a particular function. These functions include:

- Semi-autonomous (human-in-the-loop), a mode of operation whereby the system is permitted to perform only selected decisional functions without further human interaction/intervention. This means the operator gets the opportunity to approve important machine decisions before the machine executes, which can be pre-approval of target sets that the machine is allowed to engage or real time approval of decisions;
- Supervised Autonomous (human-on-the-loop), a UxS operating mode that enables a human operator to intervene and terminate selected operations or activities. This means that the machine can proceed to execute its decisions, while the operator can terminate that execution; and
- Autonomous (human-outside-the-loop), a mode of operation whereby the system is permitted to perform all designed functions without further human interaction / intervention necessary to safely execute a specified task. This means that the machine can execute its mission without human intervention.

In addition to the above categories, the machine can also be influencing a safety decision by providing information that is of a safety significant nature used by the operator to make decisions.

Autonomous software functions that exhibit probabilistic behavior increase the V&V challenge. Probabilistic software is a subtype of nondeterministic software. Nondeterministic software can give different answers on different executions of the same scenario. Weapon systems with AI control of functionality provide an additional concern. DoD policy regarding autonomy in weapon systems is established in DoDD 3000.09. The Directive establishes policy and assigns responsibilities for development and use of autonomous and semi-autonomous functions in weapon systems. It requires systems to undergo rigorous hardware and software V&V and realistic system developmental and operational T&E. UxSs that are also semi-autonomous or autonomous weapon system must satisfy both the requirements of DoDD 3000.09 and system safety. Use of safety precepts presented in this document support the design safety of the system and assist in the system safety engineering assessment of the UxS and the characterization of potential mishap risk. This requires more testing to capture low probability outcomes and determine these outcomes as safe. In practice, the effort required to perform verification may become prohibitive as the number of possible outcomes increases.

The V&V challenge is exacerbated when implementing autonomous functions, especially when machines learn from experience how to execute tasks. For such functions there is not necessarily a software specification to verify. Testing cannot necessarily ensure that all possible outcomes are seen in any finite amount of testing. Further, if machine learning is allowed to continue after testing, then outcomes may change and invalidate the prior testing.

Autonomous capabilities in weapon systems may provide additional concerns. DoDD 3000.09 assigns responsibilities for the development and use of autonomous and semi-autonomous functions in weapon systems and establishes guidelines to minimize the probability and consequences of failures that could lead to unintended engagements. It also provides specific

guidance on how to conduct TEV&V of autonomous and semi-autonomous weapon systems. DoD policy regarding autonomous behaviors in weapons systems is established in DoDD 3000.09. Policy provides flexibility in allowing some autonomous functionalities in certain UxS weapons systems when V&V have been performed while limiting the autonomy granted all other weapons functions. UxSs must satisfy both policy and safety. Use of the safety precepts presented in this document supports the design safety of the system and assists in the system safety engineering assessment of the UxS and the characterization of potential mishap risk.

Each function of a UxS can be graphed as a point in Figure 3-1. The V&V and T&E challenge for each function of a system is higher per the arrow shown in Figure 3-1. The importance of meeting the V&V challenge, and later the T&E challenge for each function, from a safety perspective, depends on the severity of risk associated with the machine's failure to safely perform. This creates safety concerns regarding the potential for the machine's actions to diverge from the human's intended actions. While providing an accepted V&V process for AI functionality is outside the scope of this document, this document does provide guidance for safety processes that can help identify which functions are particularly V&V challenged. It also provides mechanisms for safe use of deployed autonomous systems, which can also be used to reduce risks during V&V.

3.2 Ethical Principles for AI and Autonomy

The safety precepts presented within this document provide a solid foundation of actionable steps and procedures which serve as concrete instantiations of the DoD's AI Ethical Principles. Projects which follow these precepts will provide evidence of operationalizing these principles.

As AI and autonomous systems continue to develop at an accelerating pace, ensuring the responsible design, development, deployment, and use of these systems is critical. This is particularly true based on autonomy sophistication and from the transition of decision-supporting to decision-making systems. These systems defy traditional classification and raise and require attention to ethical ambiguities and risks.

In order to address the unique challenges posed by AI, the DoD has adopted a set of ethical principles for the design, development and deployment, and use of AI-enabled capabilities. These principles are interrelated and many grounded in and implemented through good AI engineering and safety practices:

1. Responsible – The Department's personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.

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

3. 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.

4. 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 lifestyles.

5. 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.

3.3 Analyzing Autonomy

Until recently, software safety analysis using LOR task requirements delineated in MIL-STD-882E, with greater detail provided in the DoD JSSSEH were sufficient to adequately assess software's contribution to safety risk. The LOR tasks specify the depth and breadth of software analysis and verification activities necessary to provide a sufficient level of confidence that a safety significant software function will perform as required. While LOR tasks along with T&E and V&V considered together suffices for now, per Figure 3-1, as programming languages use to invoke autonomous functionalities evolve and the software's volume and complexity increase the software analytical techniques used to accomplish the LOR tasks are likely to become insufficient for the software safety analysis of autonomous functionalities. Since completion of LOR task requirements is a critical step in assessing the safety characteristics associated with software performance, any shortfall in accomplishing LOR task requirements could result in an inability to certify the software and subsequently the system. Unless and until this insufficiency is overcome, it will impede the scope of safety analysis required to provide full assurance in the autonomous operation.

The subsections below provide mitigations to protect against the risks associated with any insufficiency of available system safety analysis methods. The mitigations do not fully prevent occurrence of such events, but when used in accord with the precepts, they can help mitigate the risk and reduce the severity of resulting mishaps.

3.3.1 Managed Machine Learning

Machine learning introduces obvious risk into autonomous system operations. Managed machine learning design methods such as those discussed here could potentially mitigate risks related to use of newly learned behaviors. One risk is the possibility that one or more previously learned behaviors will change and invalidate prior validations. Another risk is that it can lead to behavior difference among machines of the same type. For instance, in the case of a family of identical machines, if one machine continues learning during deployment, then that machine behavior may become unique. In some narrowly-scoped, well defined, and constrained tasks, like games with clear-cut rules of behavior and perfect information, machine learning has shown that it can quickly develop novel approaches to problems and outperform humans. A side effect

of machine learning is the potential to introduce and implement unsafe decisions. This would likely be amplified in more complex systems with multiple autonomous functions or operations.

The use of machine learning is expected to increase. Managed machine learning, from a safety perspective, refers to design and operational concepts for allowing safe use of machine learning. Section 2.2.2 provides a detailed discussion relevant to managed machine learning. The concept of "learning mode," discussed in Section 2.2.2, provides a tool to enable or disable machine learning and a mitigation to associated potential risk. The ability to have a learning mode in a given system depends on design that isolates the learning function, such that it can be managed (switched on and off) without interfering with the use of previously learned and validated behaviors.

When enabled by UxS design, managed machine learning might also include validation methods that would verify and validate newly learned behavior either real-time during the operation of the system or post operations after the system or UxS completes its mission. The V&V should validate new behaviors and consideration should be given resynching those behavior across the machines of the same type.

3.3.2 Bounding Autonomous Functions

No UxS is fully autonomous. It can, however, perform what can be considered "autonomous functions" in the performance of a mission. For those autonomous functions that are considered safety significant, appropriate bounding of the performance of each function should be provided. This bounding should include (as appropriate to the safety risks raised by the autonomous function):

- Operator control of the authorization of function performance (human-in-the-loop);
- Operator oversight of the performance of an authorized function (human-on-the-loop);
- Bounds on the duration of function performance;
- Bounds on the space in which the function can be performed;
- Bounds on dynamic control of the UxS that the function might influence or direct.

For those UxS autonomous functions where human-in-the-loop or human-on-the-loop command and control is not feasible, appropriate "deterministic checkpoints" (see Section 2.2.1) should be considered in the UxS software or system architecture or design to provide run-time assurance that the autonomous function does not exceed defined bounds of duration, space, or UxS dynamic control that could lead to loss of life or other system mishap. This is especially needed where AI or ML is used to implement safety significant autonomous UxS functionality. As mentioned in Section 2.2.1, the run-time assurance architecture documented in ASTM F3269-17 provides an example of this kind of deterministic checkpoint.

3.4 Flexible Autonomy

This short discussion on flexible autonomy has its provenance from the United States Air Force Office of the Chief Scientist in its publication titled "AUTONOMOUS HORIZONS: System

Autonomy in the Air Force." Flexible autonomy refers to the ability to activate or deactivate autonomous behaviors post-deployment and without reprogramming. Flexible autonomy allows rapid safe reconfiguration of the system based on validation results, field experience with the system, changing mission parameters or rules of engagement (ROE), DoD policy, and more. It allows people to rapidly grant the system more autonomy as assurance is developed. It also allows people to rapidly revoke autonomy where confidence has been compromised.

Flexible autonomy control may be provided for individual functions whose mode can be individually switched safely or for a set of functions that can be switched safely only as a group. Command and control of the flexible autonomous mode may be restricted for autonomous functions whose inclusion in the system is specifically to mitigate risk.

Flexible autonomy is another function that can be linked to specific deterministic checkpoints. The checkpoint issues the task to the autonomous function, initiates the bounds check on the resulting plan, and then authorizes the execution of that plan based on the results of the bounds check. If operating in semi-autonomous mode for the particular autonomous function, the checkpoint forwards the plan and the results of the bounds check to the human operator for a go/no-go decision.

3.4.1 Unmanned Systems Command and Control

Humans or machines may be authorized to control one or more UxSs. The commander decides whether the control entity will be a human or a machine. In this guide, the term "operator" is defined as human control entity.

The human commander retains authority to revoke or reassign command authority. It is key to ensure the human commander has adequate situational awareness and response time to execute this authority. With this separation of command from control, humans and machines can collaborate in a hierarchical military structure, yet humans command at all levels of that hierarchy.

If separation of command from control is not enforced, the level of risk and probability of mishaps increase. Likewise, if human command can be usurped, the level of risk and probability of mishaps increases. Leveraging system safety engineering early on in the development process, will help ensure that such risks are mitigated by the UxS design.

Appendix A. Programmatic Safety Precepts Clarification Tables

Table A-1. PSP-1 Clarification Table

Category	Description
PSP-1	• The program office should integrate into their MIL-STD-882E safety program the UxS system safety precepts and the DoD's AI Ethical Principles.
Scope	• The intent of this precept is for all programs to establish and maintain a compliant MIL-STD-882E SSPP or an equivalent standard (which also addresses the DoD's AI Ethical Principles). The program office should establish a common approach to UxS safety in a system-of-systems environment. The OSD UxS precepts, contained herein, provide the foundation for a common approach for UxS safety. UxS system safety programs should demonstrate traceability to these precepts, and assess the potential mishap risk of any non-adherence. Adherence to and any deviation from these precepts are addressed in PSP-3, which requires the program office to review each of the UxS precepts in this document for applicability to their program and incorporate requirements derived from the precepts into program documentation (e.g., contract statement of work, program plans, requirement specifications). This precept implements the requirement to establish UxS safety precepts, while PSP-3 provides for UxS safety precepts tailoring. The precepts, presented in this guide, are provided as a baseline set of precepts for consideration for any UxS safety program. While deviation from these precepts must be justified, it is fully anticipated new precepts will be established for individual safety programs that compliment these precepts.
Rationale	• DoDD 5000.01, "The Defense Acquisition System," requires that safety will be addressed throughout the acquisition process for every acquisition category level program. MIL-STD-882E is the DoD "systems safety standard practice [that] identifies the systems engineering approach to eliminate hazards, where possible, and minimizing risks where those hazards cannot be eliminated." These precepts are intended to establish UxS program management, design, and operation, thereby mitigating potential mishap risk. • These precepts are intended to be applicable regardless of design and technology. • The Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity, DoD (February 2019), directs the Department to lead in military ethics and AI safety. These precepts advance efforts on ethics and AI safety and also act to operationalize many aspects of the DoD's ethical principles for the design, development, deployment, and use of AI capabilities.

Category	Description
Examples	• Many current UxSs have been developed as prototypes, and due to their value to the warfighter, their fielding to theater is accelerated without the benefit of an effective safety program. This is also evident during the recovery and disposal of these UxSs. • The SAE International's Joint Architecture for Unmanned Systems Committee established a standardized architecture that is applicable to UxSs.
Detailed Considerations	• PMs must commit resources to the safety efforts beyond fielding to ensure the appropriate level of safety support is maintained. • The UxS program office should establish a strategy for managing the system's safety to: o Identify safety issues as early as possible, o Assess the risks associated with the system safety issue, o Document, manage, monitor, and mitigate the system safety risks through the UxS life cycle, and o Track the system safety issues through its life cycle or until resolution of the issue. • The UxS program office should ensure that system safety compliance is considered during design, testing, evaluation, demo, deployment, post-deployment, and any other activities throughout the acquisition cycle. • The UxS Program Office should ensure that the DoD's AI Ethical Principles are considered during the design, development, deployment, and use of a UxS system. • The UxS program office's MIL-STD-882E SSPP should be tailored for each program as appropriate by considering all and then adopting/applying applicable UxS programmatic, design, and operational safety precepts. • Service safety centers typically maintain mishap data that should be examined early in the system life cycle in an attempt to incorporate UxS lessons learned. • The UxS Program Office should participate in safety investigations of Class A and B UxS mishaps to provide analysis of hazards that contributed to the mishap and recommendations for materiel risk mitigation measures. • Consider developing joint lessons-learned databases and common processes that can be shared among the UxS community. • Ensure the human system integration is designed appropriately and that all the necessary UxS command and control data requirements are considered in the UxS design. Human system integration analysis and a command and control analysis should be integrated with the SSPP. • Consider common human-machine interfaces for UxSs.

Category	Description
	• With the life of defense systems being extended through block upgrades, technology refresh programs, etc., system designs must consider post-deployment support that will ensure the safety of the system. • Safety issues which arise beyond fielding generally result in the use of procedures to mitigate mishap risk. Procedural updates frequently limit the system's operational utility and effectiveness and should only be considered as the last resort.

Table A-2. PSP-2 Clarification Table

Category	Description
PSP-2	• The program office should ensure the UxS complies with current safety policy, standards, and design requirements and the DoD's AI Ethical Principles.
Scope	• The intent of this precept is to ensure the program office considers appropriate existing policy, military standards, and criteria in the design of the UxSs, consistent with its intended life-cycle use. • The program office should ensure that the environments and scenarios the UxS is designed to operate in are documented and if the UxS is operated in an environment, scenario, or state not consistent with its design that operators are made aware of this inconsistency.
Rationale	• While present system designs are performance driven, design standards specific to potentially hazardous systems such as munitions, weapons, suspension, and release equipment, aviation systems, and laser systems are mandatory. Compliance with these standards is reviewed for adequacy by Service safety organizations or the Joint Weapon Safety Review Board during the Joint Capabilities and Integration Decision System process for joint service programs. • As AI and autonomous systems continue to develop at an accelerating pace, raising new risks and ethical ambiguities, it is critical that UxS systems adhere to the DoD's AI Ethical Principles.
Examples	• None
Detailed Considerations	• Additional references include, but are not limited to, Military Handbook 516C including RTCA DO-178, MIL-STD-2105D, MIL-STD-2088B, MIL-STD-1316F, MIL-STD-1901A, MIL-STD-1472H, Standardization Agreement (STANAG) 4187, STANAG 4586, Allied Ordnance Publication (AOP)-52, DoD JSSSEH, and the DoD's AI Ethical Principles.

Table A-3. PSP-3 Clarification Table

Category	Description
PSP-3	• The program office should ensure adherence to and any deviation from the UxS safety precepts are addressed during program reviews.
Scope	• This precept, along with PSP-2, requires the program office's safety team to review each UxS precept in this guide for applicability to their program; incorporate requirements derived from the precepts into program documentation (e.g., contract statement of work, program plans, requirement specifications); and show adherence to and any deviation from the precepts. Adherence to and any deviation from these precepts should be addressed and approved at the first major program review. Should the program office choose to tailor this guide's precepts, then that tailored set becomes the baseline upon which the subsequent major reviews address continued adherence. Tailored safety precepts should be assessed for system safety.
Rationale	• These precepts were developed by subject matter experts and represent best safety practices intended to influence programmatic, design, and operational activities, but not to dictate specific design solutions.
Example	• None
Detailed Considerations	• The program office should document adherence to and any deviation from each precept, including the associated rationale, as part of the design review technical data package. This information is also critical for continuous improvement to these precepts and expansion of UxS lessons learned. • The UxS SSPP should be reviewed by Systems Safety Working Groups and during system readiness reviews, preliminary design reviews, critical design reviews, and internal program office reviews.

Table A-4. PSP-4 Clarification Table

Category	Description
PSP-4	• The program office should ensure off-the-shelf items, reuse items, original-use items, design changes, technology refresh, and technology upgrades within the system are assessed for safety.
Scope	• This precept applies to every component in the UxS system. All components must be assessed for safety within the context of the overall system. The level of assessment should be commensurate with its safety criticality.
Rationale	• Significant mishaps have occurred related to the reuse of components within a different system. A safety review of off- the-shelf items must provide insight to the hazards and control of these items. • Additionally, as levels of autonomy functionality mature and AI is introduced as potential design improvement attributes, systems must reassess their engineering design functionalities and how changes to methods or mechanisms to invoke new functionality affect system safety and performance. • The complexity of autonomous systems should require system safety reviews as technologies (AI and autonomy) evolve and mature. Such technologies may be introduced into a system during technology upgrades or design improvements.
Examples	• Ariane V Rocket – The June 4, 1996 maiden flight ended in complete destruction 40 seconds into flight. High aerodynamic loads due to a high angle of attack caused the booster to separate from main stage, triggering self-destruct. A high angle of attack was caused by full nozzle deflections commanded by the on-board computer. The on-board computer received a diagnostic bit pattern from the inertial reference system due to a software exception. The software exception was generated by overflow from a 64-bit floating point that was converted to a 16-bit signed integer. The module responsible for the fault was not used during flight, but was used only for alignment of strap-down of the inertial system on the launch pad; it was reused from Ariane IV. • A research fly-by-wire aircraft experienced a failure on the flight line during a group test the day before the flight. That failure caused the flight control computer to crash, resulting in an erroneous response from the flight computer (fortunately the failure occurred while on the flight line rather than during flight). A memory conflict occurred, causing safety significant data to be overwritten by non-safety significant code.
Detailed Considerations	• Ensure that full integration, end-to-end testing is performed on systems containing legacy and/or reuse items. • Correct implementation of software exception handlers is safety significant.

Category	Description
	• Safety concerns from the software system safety analysis must be addressed in software test plans and procedures (boundary value testing, full integration end-to-end testing). • Any off-the-shelf items (e.g., commercial-off-the-shelf, government-off-the-shelf, non-developmental item), reuse items, original-use items, design changes, technology refresh and technology upgrades (software) must be thoroughly assessed and tested for safety within the system into which it is being inserted. • Components, including legacy systems and subsystems, that have been proven safe in an earlier application, cannot be assumed safe in another application. Special attention should be paid to the interfaces.

Table A-5. PSP-5 Clarification Table

Category	Description
PSP-5	• The program office should ensure that the UxS, by design and operation, does not allow subversion of an authorized human command or control of the UxS.
Scope	• Autonomous UxSs cannot have the capability to subvert human UxS command and control. There may be periods of authorized autonomous UxS control; however, there should be no circumstance where the UxS overrides human control or a human request to regain control.
Rationale	• Human command and control of a UxS must have ultimate control authority over the UxS. The autonomous UxS will release control to the human controller when requested.
Examples	• UxSs that deactivate human command and control would disallow the human operator of the UxS to intervene in the case of unexpected UxS behavior, potentially resulting in a mishap that the operator could have prevented. o A UxS learning how to navigate obstructions deactivates human control, resulting in a collision that the operator could have prevented had the operator been in control of the system. o A UxS has less situational awareness than the human operator. The human operator's knowledge, being greater than the autonomous or AI UxS, can be employed to operate the UxS optimally.
Detailed Considerations	• Biases, unwanted outcomes (as described in the "detailed considerations" of DSP-13), and cognitive autonomy are three examples of design aspects that can create the possibility for a system in which the UxS's mission goals and metrics may diverge from the human mission goals and metrics. It is completely unknown whether design can achieve full autonomy while preventing such divergence. Thus, it is up to the program to ensure the human retains command and control if divergence occurs. • The autonomous UxS may complement human command and control, but in accord with human goals and metrics, the boundary between the two must be clearly defined by the program so that it can be enforced in the design. • The program office should consider appropriate human system integration to mitigate mishaps driven by human-machine interactions.

Table A-6. PSP-6 Clarification Table

Category	Description
PSP-6	• The program office should ensure that the UxS safety significant functions and components are not compromised when utilizing flexible autonomy.
Scope	• This precept requires the program's careful consideration of safety significant functions and components when designing a flexible autonomy to ensure that these functions and components are not subverted in any way during the process of modifying, adding, or subtracting capabilities or functions. UxSs with flexible autonomy allow customization of a UxS to meet mission needs.
Rationale	• While the advances in flexible autonomy provide more versatility, adverse impacts to safety significant functions or components might be overlooked. This precept reinforces the importance of assessing safety significant functions and components after adding, removing, enabling, or disabling capabilities or functions.
Examples	• Removing the remote-control function of an unmanned aerial system (UAS) also removes verification of all communication to the UAS. • A corrupted message to the UAS is not detected and UAS software becomes unstable, resulting in UAS failure.
Detailed Considerations	• The program office should maintain documentation of safety significant functions and components. The program must ensure adequate assessment of these safety significant functions and components for every combination of functionality and capability of the UxS.

Table A-7. PSP-7 Clarification Table

Category	Description
PSP-7	• The program office should prioritize personnel safety in UxS intended to team with or operate alongside manned systems.
Scope	• This precept is intended to ensure personnel safety is prioritized where there are UxSs and manned systems operating in proximity and/or together (e.g., swarms or unmanned/manned teaming).
Rationale	• In situations where there is an alternative between loss of a UxS or endangering personnel, the safety of personnel should be prioritized.
Examples	• During missions where UAVs are in a swarm with a manned aircraft, any misbehavior or failure of a UAV could result in a collision with the manned aircraft. • Under hostile fire situations, the UxS should prioritize defense of personnel even if it could mean catastrophic damage to itself.
Detailed Considerations	• None

Table A-8. PSP-8 Clarification Table

Category	Description
PSP-8	• The program office should ensure authorized and secure control (integrity) between platform and controller to minimize potential UxS mishaps and unauthorized command and control.
Scope	• The UxS design should ensure the human command entity and the human or machine control entity have decision-making authority and are designated to command the UxS. The UxS communication link will ensure robust, secure, and safe operation with positive feedback status. If a command or control entity communication link loss or corruption occurs, the UxS system should transition to a predetermined safe (or acceptable risk) state and mode while re-establishing command and control links.
Rationale	• Secure command and control, communication integrity, and human command with human or machine control of the UxS is pivotal to ensuring safe execution of UxS mission requirements and safe operations within mission parameters.
Examples	• An unmanned ground vehicle (UGV) continues to roll down a slope when communications are lost, resulting in collision with personnel or other vehicles. • An unmanned surface vessel is taken over and commanded to collide with a host ship. • Video information from an unmanned surface vessel is compromised and used to provide false information to ship personnel on potential threats. • UxS communication loss results in a system in an unknown mode that crashes, returns armed, or performs unsafe maneuvers that create a top-level mishap with the UxS controller or nearby friendly troops. • UAV loses communication link and returns to a known area to reestablish, or returns to a known mode/state/way point until reestablished, to minimize collateral damage from total mishap.
Detailed Considerations	• A command entity, controlled by a human in a command position and paired with a UxS, can authorize a control entity. A command entity is the entity that exercises immediate control over the UxS. • As UxSs evolve and increase in autonomy, a system operator or human controller (control entity) may no longer be a valid assumption. • A UxS control entity can be a human or software, authorized by the paired command entity, and capable of controlling the UxS.

Table A-9. PSP-9 Clarification Table

Category	Description
PSP-9	• The program office should ensure that software systems exhibiting unpredictable, nondeterministic behavior are employed safely and comply with current policy.
Scope	• Systems with AI could perform logical deductions or inferences and make decisions based on acquired information that is insufficient or conflicting. If these decisions are then used in safety significant functions, there must be an assessment of the mishap risk of using such technology. This is especially true if the new deductions, inferences, and decisions have not been validated and verified to ensure they could not result in a mishap.
Rationale	• All safety significant functions in current weapon systems should be properly verified and validated prior to exposing personnel, equipment, and the environment to system operations. • From a safety perspective, autonomy and AI present unique challenges where there is the potential for unverified decision-making to occur. A detailed assessment should be conducted that provides better insight or understanding on the behavior and boundaries of the technology so that a proper assessment of mishap risk can be developed (inadequate clarity will result in elevated mishap risk).
Examples	• Deterministic vehicle stability checks on a UGV ensure autonomous path generation does not cause a vehicle rollover.
Detailed Considerations	• Programs with software systems exhibiting unpredictable, nondeterministic behavior should work to assess risk levels of unsafe operation of nondeterministic systems, based on best efforts at understanding the range/statistics of nondeterminism in a given environment and operational mission. • Safe operation resulting from software commands is instantiated in Software System Safety Engineering practices in MIL-STD 882E, the DoD JSSSEH, and AOP-52.

Appendix B. Operational Safety Precepts Clarification Tables

Table B-1. OSP-1 Clarification Table

Category	Description
OSP-1	• The control entity of the UxS should have adequate mission information to support safe operations.
Scope	• The intent of this precept is to ensure safe operation of the UxS, given adequate mission information is provided. Adequate mission information includes, but is not limited to, specific data requirements for all operational phases influenced by mission; mission objectives; concept of operations (CONOPS); ROE; tactics, training, and procedures (TTP); available intelligence; environmental and meteorological conditions; geographic position; spatial separation from other systems/users; data from vehicle indicating operation within defined operational criteria; weapon status; sensor status; remaining "life"; known anomalies; and UxS status and health. Fundamental to this precept is the clear identification of the control entity applicable to all phases of operation.
Rationale	• The availability of adequate mission information is critical for safe UxS operation. This precept is dependent upon a thorough job of defining and processing adequate mission information and thorough TTPs.
Examples	• Loading appropriate maps. • Having an appropriate predefined target. • Having accurate situational awareness. • Ensuring adequate time to regain UGV control after a civilian incursion of a robotic convoy.
Detailed Considerations	• None

Table B-2. OSP-2 Clarification Table

Category	Description
OSP-2	• The UxS should be considered unsafe until a safe state can be verified.
Scope	• Positive determination of state must be verified by any control entity. Verification of state includes, but is not limited to, UxS mobility, weapons, and hazardous system appendages, items retrieved by the UxS during operations, and the system following exposures to hazardous environments such as chemical, biological, or radiological (CBR) hazards.
Rationale	• Safe transitioning between operational states must be addressed to ensure safe human and UxS interface during any operational mode. • Requirements also encompass UxSs that have been out of sight or out of communication from the control entity.
Examples	• Visual confirmation with safe and arm status. • Health status from the UxS. • Positive identification of state.
Detailed Considerations	• Careful consideration of defined techniques and training, standard operating procedures (SOP), TTPs addressed through control entity training, and user guidelines with respect to this precept must be made to avoid conflicts with either programmatic or operational design of the UxS.

Table B-3. OSP-3 Clarification Table

Category	Description
OSP-3	• The control entity should verify the state of the UxS to ensure it is a known and intended state prior to performing operations or tasks.
Scope	• The operator must be aware of the state of the UxS. The system must be in a safe state prior to performing operations. This OSP implies the operator has the requisite competencies to ensure the state of the UxS prior to performing operations. Tactics, techniques, SOPs, training, and user guidelines should appropriately address the relationship of the state of the UxS or its weapons, hazardous system appendages, and items retrieved by the UxS during operations and human interface with the UxS.
Rationale	• While identifiable safe transitioning between operational states must be addressed through design to ensure safe human and UxS interfaces during any operational mode, the same state requirements must be enforced through TTPs. These TTPs should also enforce the identifiable state requirements for UxSs that have been out of sight or communication from the control entity.
Examples	• Verify power-on self-test and built-in test. • Verify health status and communications.
Detailed Considerations	• Consider intrusion prevention/detection.

Table B-4. OSP-4 Clarification Table

Category	Description
OSP-4	• The UxS weapons should be loaded and/or energized as late as possible in the operational sequence.
Scope	• This OSP addresses weapons, lasers, and other hazardous devices such as emitters.
Rationale	• This OSP limits the exposure of personnel to a potentially high-risk condition and is in keeping with the standard procedure for manned systems.
Examples	• None
Detailed Considerations	• None

Table B-5. OSP-5 Clarification Table

Category	Description
OSP-5	• Only authorized, qualified, and trained personnel using approved procedures should operate or maintain the UxS.
Scope	• This OSP addresses requisite operator competencies in operating and maintaining UxSs.
Rationale	• Appropriate skills and expertise are consistent with DoD policy and reduce the potential for mishaps caused by human error. Appropriate authorizations prevent unnecessary exposure of unqualified personnel.
Examples	• None
Detailed Considerations	• Consider who needs to be in the developmental process to determine, document, and train operators. • Identify needs for training aids or training modules. • Use engineering mitigations instead of administrative mitigations to condense training needs.

Table B-6. OSP-6 Clarification Table

Category	Description
OSP-6	• The operator should be aware during all phases of the mission when autonomous behaviors are utilized.
Scope	• To help the operator avoid inadvertent functioning of potentially hazardous operations, the operator should be aware of controls that involve or are influenced by AI or unpredictable, nondeterministic autonomous functions. Initiating a function that results in AI or unpredictable, nondeterministic autonomous behavior should be well known and understood.
Rationale	• Initiating AI functionality should be restricted to that which is necessary for mission operations so unintended hazards are not introduced.
Examples	• AI functions/controls for search and identify are separated from other controls. • Functional control for engage/fire are separated from other controls. These functions should not be AI controlled. • Mission phases would include initialization, launch, search, identify, and target engagement.
Detailed Considerations	• Operator awareness may take the form of labeling on the controller, block flow diagrams – or other easy to read diagrams – in TMs, or training the operator. • The controller functions should be separated by functionality for the various phases of the mission and be clearly different from other user controls implementing other types of operations. • AI-based modes of operation utilized for specific purposes should be independent and partitioned as much as is feasible from other AI-based modes of operation utilized for other purposes. • Segregating or partitioning AI functions allows other non-AI functions to be verified and validated independently.

Table B-7. OSP-7 Clarification Table

Category	Description
OSP-7	• The operator should be able to establish alternative recovery points prior to or during mission operations.
Scope	• Missions may include target points beyond a safe point of return to home. Mission planning should include alternate recovery points along the mission path that can be designated fail-safe zones for recovery, self-destruct, sterilization, or other disposal. • As the mission progresses, the operator may need to change recovery points as operational awareness dictates.
Rationale	• This limits unintended loss of vehicle, enemy recovery of assets, property damage, loss of life, and reduces unexploded ordnance.
Examples	• Designation of an emergency landing zone for UAV recovery. • Defining criteria for initiating UGV emergency stop functionality.
Detailed Considerations	• The consideration of an alternate recovery point should be based on available fuel or energy to return to the original (initial) recovery point. • Potential for collisions or other mishaps must be considered during execution of recovery, emergency, or retrograde mobility operations.

Table B-8. OSP-8 Clarification Table

Category	Description
OSP-8	• Weapons should only be fired or released with human consent, or with control entity consent in conjunction with preconfigured criteria established or verified by the operator.
Scope	• Weapon firing/release criteria must be in accordance with Paragraph 4.c of DoDD 3000.09 or by a system that has undergone senior review and approval pursuant to paragraph 4.d of DoDD 3000.09. • Initiation of a UxS mission must be done with human consent as the control entity for the system.
Rationale	• If the control entity is other than a human, then preconfigured criteria (location, time, event, successful state transition) must be defined or verified by the operator before the firing/releasing event. • Human command and control is preferred in the decision-making process for firing/releasing of weapons as opposed to autonomous control of weapon releases.
Examples	• The launch sequence event can be considered a checkpoint and thus authorization to engage/fire. • UxS is launched but human consent to fire is given at a later time.
Detailed Considerations	• The operator must set mission-specific preconfigured criteria/data prior to weapon launch/release. This includes possible starting conditions for any AI or autonomous behavior for which the system is designed.

Table B-9. OSP-9 Clarification Table

Category	Description
OSP-9	• The operator should have the ability to take control of the UxS, as appropriate and feasible.
Scope	• The operator must assume manual control of autonomous UxS functions or operation in the event of compromised operation (security), degraded functionality, or physical damage. The operator needs the ability to take control of the UxS (including, for example, remote or teleoperation) to the greatest extent possible.
Rationale	• In order to maintain safe operation of the UxS, the operator may need to take over manual control. It is recognized there may be situations or environments where this action may not be achieved.
Examples	• Manual override switch. • Regaining control of a UxS that is behaving abnormally. • A UxS that requires manual control of launch and recovery.
Detailed Considerations	• The look, feel, and timeliness of this control is subject to human factors and other DoD guidance. • Strong connection to OSP-12. • Where there is another control entity controlling an autonomous system, the operator should have the ability to command the control entity • Consideration must be given to cyber security of the autonomous system and prevention of commands being given by other than the authorized entity

Table B-10. OSP-10 Clarification Table

Category	Description
OSP-10	• The operator should have the ability to abort the mission of the UxS, if appropriate and feasible.
Scope	• The system should have a unique override function apart from a manual operation, where appropriate and feasible. This feature should be separated from the AI software so that it provides the operator the ability to terminate the system.
Rationale	• The system should have the ability to abort, self-destruct, sterilize, disable, render safe, or other similar action upon command. • This feature provides a bypass mechanism (back door) to override any AI functionality that would otherwise prevent the abortion and termination of the system.
Examples	• Mission abort (kill) switch.
Detailed Considerations	• This feature may require a separate command pathway to prevent common-cause failure. • The look, feel, and timeliness of this control is subject to human factors and other DoD guidance. • Strong connection to OSP-12. • Where there is another control entity controlling an autonomous system, the operator should have the ability to command the control entity to abort the mission of the autonomous system. • Consideration must be given to cyber security of the autonomous system and prevention of commands being given by other than the authorized entity.

Table B-11. OSP-11 Clarification Table

Category	Description
OSP-11	• The operator should be able to disable learning mode.
Scope	• The operator should be able to control the transition out of learning mode. • This includes the ability of the AI system to operate in learning mode.
Rationale	• Enabling and disabling learning mode is a necessary feature that is required in order for an AI system to train and be validated for use in the field. • This allows the AI system to limit its new knowledge in executing the mission.
Examples	• Learning mode switch.
Detailed Considerations	• The look and feel of disabling learning mode AI control is subject to human factors and other DoD guidance. • Re-enabling learning mode introduces safety risks; it may result in unknown behavior changes in the system. • Strong connection to OSP-12.

Table B-12. OSP-12 Clarification Table

Category	Description
OSP-12	• One operator should maintain positive and active control of the UxS during transfer of control.
Scope	• When transferring command and control of a UxS from one operator to another control entity, the UxS should remain in positive control of the losing operator until confirmation that all data necessary to operate the UxS has been received by the gaining control entity/operator. Once confirmed, the losing control entity/operator should send a message to the gaining control entity/operator to take control of the UxS. The entire operation of transferring control should be seamless to the operation of the UxS.
Rationale	• The process of transferring control of a UxS from one control station to another must be completely deterministic and include confirmation between the controllers such that safe operation of the UxS is maintained at all times. • Operators must be trained on transfer of control procedures with both the losing and gaining controllers.
Examples	• Emergencies where eminent threat of the operator or the position of control may require the operator to transfer control to an operator with a control station in an alternate location that is not under threat. • Transfer of control may be needed in order to extend the communications range of control of the UxS. • Transfer of control may be necessary due to mission requirements or operational workload of the operator/controlling crew.
Detailed Considerations	• The operator should plan for contingencies that would require transferring control of the UxS between control stations. • If losing and gaining operators are remote from each other, two-way communications between the operators should be established before transfer of control begins. • If the control station is experiencing technical issues, transfer of control may not be possible. In such cases an operator may need to take control of the UxS if possible.

Appendix C. Design Safety Precepts Clarification Tables

Table C-1. DSP-1 Clarification Table

Category	Description
DSP-1	• The UxS should be designed to minimize the mishap risk during all life-cycle phases.
Scope	• The intent of this DSP is to ensure the safety order of precedence, as prescribed in MIL-STD-882E, is applied.
Rationale	• In accordance with MIL-STD-882E, mishap risk mitigation is an iterative process that culminates when the residual mishap risk has been reduced to a level acceptable to the appropriate authority. The system safety design order of precedence, discussed in detail in MIL-STD-882E, identifies alternative mitigation approaches and lists them in order of decreasing effectiveness. These approaches, in their order of effectiveness, are: o Eliminate hazards through design selection; o Reduce risk through design alteration; o Incorporate engineered features or devices; o Provide warning devices; and o Incorporate signage, procedures, training, and personnel protective equipment.
Examples	• None
Detailed Considerations	• The UxS should be designed to safely operate across life-cycle environments, as defined in the system CONOPS, including storage, transportation, maintenance/servicing, deployment/launch, states and modes operation (that includes machine-learning mode), and reuse/recovery. • The UxS design should provide adequate protection against hazardous scenarios such as: o Uncommanded control (e.g., weapons release and firing, navigation and stability of platform, movement of weapon system, radiation). o Unintended loss of UxS (e.g., denied global positioning system, loss of communications link, loss of status, latency issues). o Unintended target engagement (e.g., enabling and disabling fire, independently selecting and discriminating of targets/re-targeting, keep out of exclusion zones, proper control of anti-tamper measures). o Loss of platform situational awareness (e.g., sensor failure or degradation, detection and reporting of system damage, power loss). o Cyber threats (e.g., jamming, take-over control, inadvertent control, denial of service).

Category	Description
	○ Revert to a fail-safe condition for abnormal control (e.g., emergency recovery procedure to include E-stops, re-authentication of controller). ○ Prevention of false display status: accurate safe-status information should be provided to the controller entity (e.g., weapon safety information, platform status information). ○ Accountability of the individual platform or collection of platforms' status and their states over the duration of an employment beginning at component emplacement and continuing through mission completion (e.g., if employing five UAVs, those five must always display required information throughout mission end). ○ Loss of power source (e.g., platform response to loss of system power or degraded system power, control of power recharging).

Table C-2. DSP-2 Clarification Table

Category	Description
DSP-2	• The UxS should be designed to fulfill valid commands only from the control entity.
Scope	• The UxS should be designed to fulfill or execute commands through a process that includes accepting commands only from control entities, determining whether the command is valid, and performing only valid commands. In addition, the UxS should detect when the commands are not valid, provide and alert/report message failure, and not perform the command/action and fail-safe.
Rationale	• The intent of this precept is to address the validity of commands and the hierarchy of the controlling entity.
Examples	• Prevent dual control. • Prevent inadvertent or unauthorized control of the UxS.

Detailed Considerations	• Valid commands/input are commands that the system is allowed or capable to perform in the current mode (a system cannot violate predefined rule sets).
	• In response to invalid commands, the system should
	  o Recognize the commands are invalid,
	  o Not perform the invalid commands,
	  o Alert/report of the invalid command, and
	  o Revert to a predefined action (e.g., fail-safe, do nothing).
	• Control entities should be validated and should allow for many-to-one and one-to-many relationships between platforms and controllers. Provisions should be made to ensure platforms are receiving valid commands from a single authorized entity at a time. It may be necessary to specify a hierarchy of control of the platforms to the control entity.
	• Delegation of authority of the AI should always be revocable by the human operator.
	• Consider assigning a unique identifier to each weapon and/or platform as mitigation to selecting the incorrect weapon/platform.

Table C-3. DSP-3 Clarification Table

Category	Description
DSP-3	• The UxS should be designed to provide command and control for safe operations.
Scope	• This precept addresses operational situational awareness and controls feedback of the system to make decisions and execute functions for all states and modes of operation.
Rationale	• The intent of this precept is to ensure that appropriate, sufficient, and timely resources are provided to process safety significant functionality.
Examples	• Communication modes of operation such as radio frequency, signals, voice recognition technologies, etc. • Prioritization of communication modes if multiple forms of communication are used. • Proper control feedback.
Detailed Considerations	• Communication reliability, network availability/quality of service, and data/information assurance should be commensurate with the safety criticality of the functions supported by the communication. • The integrity of the network availability/quality of service and data/information assurance should be commensurate with the safety criticality of the functions supported by the communication. • The level of on-board information processing capability should be adequate and commensurate with the intended method of control. • Delivery of the information to the controlling entity includes, but is not limited to, selection of data to be collected, the means of conveyance, ordering of importance, and reliability and timeliness of data of situational awareness, target selection, and target engagement. • The human-machine interface should be designed using a defined set of symbols and terms to include warnings and alerts that are common to platforms and operational services. • UxS information processing capabilities and constraints are appropriate and compatible for the operation being performed. AI will not impact safety information processing or safety significant information display status. • With an increasing number of UxS platforms for the human operator to manage, the human operator actions should be prioritized and minimized to ensure critical tasks are performed first. When AI functionality is utilized to assist the human operator for critical tasks, AI will not impact safety significant functionality. • UxSs should be designed to optimize the proficiency of the controlling entity in all operations, training configurations, and environments.

Category	Description
	• The system should be designed to detect degraded performance of the controlling entity and provide notifications to the command entity/operator. • The system should be designed to provide positive identification of the UxS including its weapon system and its existing configuration, modes, and states to control and command entities. This should include confirming preconfigured entered mission parameters, settings, and operator actions. • The UxS should provide "real-time" system status when requested, in addition to the commanded status, to the controlling entity. • The UxS should provide control and informational feedback necessary to support safe movement, navigation of the system(s) and accountability of all systems under control. UxSs require safe movement assurance in order to discriminate between potential obstacles and humans. • The human-machine interface should be designed to minimize the use of complex operational procedures to ensure safe operations. Operational procedures should not be used to replace safe design practices. • System design should consider separation of weapon system locations and sensor locations to preclude interference that could result in degradation of situational awareness. For example, the design should minimize auditory or visual degradation as the result of weapons fire.

Table C-4. DSP-4 Clarification Table

Category	Description
DSP-4	• The UxS should be designed to prevent unintended release and/or initiation and firing of lethal and nonlethal weapon systems or any other form of hazardous energy.
Scope	• This precept applies to systems and subsystems utilizing ordnance, rocket motor initiation circuits, bomb release racks, energetic materials, explosives, propellant, directed energy equipment, harmful radio frequency, radiation, lasers, etc., and the preparations for the release of energy. These systems and subsystems cannot be controlled by AI functionality and should be designed in accordance with applicable design standards such as MIL-STD-1901A, MIL-STD-1316F, and MIL-STD-1911A.
Rationale	• The intent of this precept is to preclude the inadvertent release of hazardous energy.
Examples	• Isolating power for firing and/or releasing a munition/missile from a UAV as late as possible prevents the unintentional initiation from an electrical short. • For a defensive system, inadvertent release of energy could occur if a false radar image was detected and the power source had not been isolated. • Smoke grenades on UGVs require the control entity to take multiple actions to fire the weapon so inadvertent contact with the fire command or an inadvertent action by the control entity does not fire the weapon.
Detailed Considerations	• AI or autonomous functionality should not be capable of subverting safety features for arming in accordance with design standards such as MIL-STD-1901A, MIL-STD-1316F, and MIL-STD-1911A. • AI should not subvert or compromise the firing of the weapon. • A UxS design should prohibit premature fuze arming or functioning if any or all electrical safety or energy control features fail in any given state or credible mode. • The UxS should be designed to provide verifiable safety design measures to isolate platform power from weapons or ordnance initiation circuits and other forms of energy release until intent to initiate. • The on-board weapons systems for UxSs should be designed to minimize tampering with or unauthorized physical reconfiguring of the weapon. • New UxS designs should comply with appropriate design safety requirements of pertinent STANAGs 4586 and 4737 and MIL-STDs 1901A, 1316F, and 1911A for initiation systems and hand emplaced munitions.

Category	Description
	• UxS should provide weapon safety status to the controller entity. • Each weapon/platform should have a unique identifier and should respond only to arming and firing commands with its identifier in the command. • The firing of weapon systems should require a minimum of two independent and unique validated messages in the proper sequence from the controller entity, each of which should be generated as a consequence of separate controller entity action. • The arm and fire commands should each require a unique verified message generated as a consequence of a distinct controller entity action; both messages should not originate within the UxS launching platform. • The arm and fire command messages should be sent in the proper sequence and acted upon only after being recognized as being in the proper sequence. • The UxS should be capable of determining the order in which the arming and firing messages were issued. • The design should consider platforms that may require reusable platforms and reloading of expended weapon systems. • For reusable platforms, the UxS should alert the human operator of system faults prior to reloading of expended weapon/munition system.

Table C-5. DSP-5 Clarification Table

Category	Description
DSP-5	• The UxS should be designed to prevent release and/or firing of weapons into the UxS structure itself or other friendly UxS/weapons.
Scope	• This precept addresses potential damage by firing a weapon into the UxS structure itself or weapon(s) and addresses potential damage by one friendly UxS firing into another friendly UxS, to include those UxSs within the same mission command.
Rationale	• The intent of this precept is to prevent damage by the UxS to itself or to other friendly UxSs.
Examples	• The UxS used the incorrect no-point/no-fire area and fired a weapon into the UxS platform. • During swarming of UAVs, one UAV loses navigation control and collides with other UAVs. • A self-defense UxS weapon opened fire immediately after launch of an offensive system, resulting in a weapon-to-weapon collision. Timing and definition of the target cut-out area must consider the dynamics of the operational environment.
Detailed Considerations	• The design should identify and define the specific no-point/no-fire requirements for the weapons systems. • Design of hard and/or soft stops should preclude entering the designated target cut-out area (no-point/no-fire areas). • System design should consider separation of weapon systems and sensor locations to preclude interference that could result in degradation of weapon targeting. • The design should consider dynamic no-point/no-fire zones created by the weapon timing sequences. • The design should consider unintended collision. • UxS should be integrated/de-conflicted into the overall battle management system for common operational picture perspective for management resources.

Table C-6. DSP-6 Clarification Table

Category	Description
DSP-6	• The UxS should be designed to minimize the potential for releasing or firing of a weapon on a friendly or unintended target group selection.
Scope	• The intent of this precept is to have the UxS designed in order to prevent inadvertent firing or release of weapons onto an unintended target.
Rationale	• The intent of this precept is to ensure accurate UxS target engagement. When human operator overwatch is not available, design features must be incorporated that prevent firing upon an unintended target group selection.
Examples	• Control entity mistakenly requests permission to fire upon friendly forces, but is denied by command entity, who correctly identifies forces as friendly. • During maintenance procedures, the UxS's maintenance state has anti-tamper detection and response functionality turned off to prevent inadvertent weapon release. • The UxS design should alert the operator prior to executing a self-destruct event. • During system training mode, the UxS weapon fire control communications are disabled to prevent inadvertent message transfer. • The UxS design should as part of target selection include time stamp as part of sensor and situational data.
Detailed Considerations	• The control entity must validate target groups and obtain permission by Command Authority prior to firing and/or weapon(s) release due to TTP and ROE. • The control entity must validate the target selection prior to firing and/or weapon(s) release. Identifying the target prior to, or as early as possible (e.g., during the planning phase, reactions to detection of tamper events). • Final target engagement must follow the delegation of authority. Note: AI or autonomous functionality should always be revocable by the human operator. • Methods of weapon neutralization or self-destruct should complete without causing collateral damage to other UxSs and without creating unexploded ordnance. • The UxS should support execution of preprogrammed instructions (engagement modes and tactics), based on situational awareness information received and ROE. • The UxS should allow the operator to change tactics or override commands before committing to fire or self-destructing/self-deactivating. • UxS designs should distinguish among tactical, system training, and machine learning modes to prevent inadvertent firing.

Table C-7. DSP-7 Clarification Table

Category	Description
DSP-7	• The UxS should be designed to safely initialize in the intended state, safely and verifiably change modes and states, and prevent hazardous system mode combinations or transitions.
Scope	• This precept applies to all states and modes related to every phase of UxS CONOPS including storage, transport, maintenance/servicing, deployment/launch, transit, operation (that includes machine-learning mode, user-training mode, and tactical/engagement modes), and reuse/recovery. Both initialization and re-initialization must establish a known, confirmed safe state. Contingency or other alternative plans should be prepared for foreseeable occurrences of degraded or hazardous modes and states.
Rationale	• This precept ensures the system modes and states and their transitions and different combinations are designed for safe operation of the UxS while operating under the command and control of an operator. The UxSs, while operating autonomously, should be able to detect failure conditions that may lead to a hazard upon allowable state/mode transitions and fail safe/alert the operator.
Examples	• System power-up mode should ensure mission data is valid. • System power-up mode and safe states (not armed state) should ensure propulsion is deactivated and weapons system is disabled to prevent uncommanded movement and firing. • Machine-learning mode should be enabled/disabled and confirmed by human operator (or by a pre-allowed state for autonomous mode) to prevent hazardous states/modes combinations. • User-training mode is different from operational mode, which necessitates restrictions of communication (e.g., broadcasting of messages). • Predetermined states/modes without ROE and human operator authorization should prevent inadvertent access to restricted/not-allowed functions (e.g., target engagement) for the intended mission. • Maintain remote control over the UxS states and modes and control over inadvertent state transitions or functions. • Ensure ability of system to fail safe when unsafe system faults/errors are detected. • Prevention of hazardous conditions due to malfunctions/anomalies or lost communications.
Detailed Considerations	• The design should include specific measures to test and verify safe mode and state transitions.

Category	Description
	• Upon detection of a degraded safe condition, the UxS should alert the human operator of the degraded condition and should provide positive confirmation on the response (e.g., revert to a fail-safe state or continue). • During initialization the UxS should start in a safe state; if the UxS reports an unsafe state, the system should revert to a safe state and the UxS should alert the operator of the failure. • Unexpected loss of communication for a predefined time duration should result in the UxS reverting to a safe state. In addition, the controller/display unit should immediately notify the operator of the unexpected loss of communications. • The system should have the ability to operate as intended in a mode where various AI functionality has been enabled. This includes the ability of the system to operate in learning mode or its ability to utilize new knowledge of what it has learned. • The system should reject a command not valid for the state and should prevent sending or acting on commands not valid for the state/mode the UxSs are currently operating in. In addition, the controller/display unit should immediately notify the operator of the rejected command. • Any reconfiguration capability (any hardware or software and machine-learned changes to the system configuration) should ensure that the UxS remains in the intended safe states and modes of operation. • The UxS should ensure that priority message processing cannot cause transition to, or remain in, an unsafe mode, state, or combination thereof. • Ensure latency conditions of mode and/or state transitions do not adversely affect safety. • The system should be in a verifiable safe state before transitioning between modes. Mode transitions may occur without verification of the safe state if the resulting mode is "safer" in the operational context. There may be various safe states, depending upon the system operational and physical environments (e.g., training, test, underwater, airborne). • The system may require information on last known states and/or configurations to recover from an unintended shutdown or abort. • The system should be designed to include reset capabilities, such as warm boot of individual functions or subsystems, which support safe transitions between states and modes. • The UxS should ensure command messages are prioritized and processed in the correct sequence within the intended state and/or mode. • System initialization and re-initialization should not result in motion, weapons loss, or unwanted energy transfer that may harm servicing personnel or operators. • Consideration should be given for a time of validity for all safety-significant commands.

Table C-8. DSP-8 Clarification Table

Category	Description
DSP-8	• The UxS should be designed to include an abort function that transitions the system to a safe state.
Scope	• The primary intent of this precept is to provide an ability to abort an operation (e.g., mobility, machine learning, autonomous behaviors, weapon fire sequences). The secondary intent is for the UxS to automatically transition to a safe state upon abort and alert the control entity of such an operation.
Rationale	• The dynamics of the operational environment require all systems to provide an ability to immediately cease the current function and make safe the system, if possible.
Examples	• An abort switch can be used to immediately cease weapons fire or vehicle movement. • The UxS should provide a capability to abort specific commanded actions, to include autonomous behaviors. If a UxS is about to engage a friendly target, an abort capability is necessary. • If a UAV is flying a fixed pattern and encounters an aircraft in the area, an abort is needed to stop that flight pattern and go to another waypoint. • If a UxS that is part of a group of UxSs fails to act upon a valid safe state command which was intended for the entire group of UxSs from the controlling entity, it must be capable of an individual system abort.
Detailed Considerations	• Delegation of control/command entity of the AI is always revocable by the human operator. • Autonomous operational modes are always revocable by the human operator. • The design should include specific measures to test and verify abort transitions. • The set of functions that can be aborted should be identified and should be included in the abort command. • For range testing or training, consider the need for multiple operators to have the ability to cease fire or abort weapon fire. • UxS should have provisions to detect external tampering of its critical components that are required to execute its mission plan (e.g., engine, battery housing, weapons/munitions) and should be capable of system abort. • UxS should have provisions to detect internal system failure faults and should be capable of system abort.

Table C-9. DSP-9 Clarification Table

Category	Description
DSP-9	• Safety significant software should be appropriately physically and functionally partitioned.
Scope	• The intent of this precept is to ensure the integrity of safety significant software and avoid impacts from nondeterministic functionality.
Rationale	• Any safety significant software functionality should be appropriately isolated within the system's overall functionality. This should be both to protect the safety significant software from detrimental interaction with non-safety significant functionality and to safely bound any non-deterministic autonomous behavior that might increase mishap risk.
Examples	• Isolation of safety significant AI software function to bind its behavior and to reduce V&V for the system. • Isolation of safety significant functionality from the rest of the system to reduce the possibility of interference or data corruption from non-safety significant functionality. • Safety significant vehicle stability checks on a UGV executed in separate software running on a different processor from nondeterministic autonomous navigation functions.
Detailed Considerations	• Appropriate physical and functional partitioning of safety significant software from non-safety significant software is necessary to ensure the integrity of safety significant processing. • Physical and functional partitioning of safety significant AI functionality through the implementation of an appropriate run-time assurance/verification framework to ensure the integrity of AI processing.

Table C-10. DSP-10 Clarification Table

Category	Description
DSP-10	• The UxS should be designed to minimize single-point, common-mode, or common-cause failures that result in high and/or serious risks.
Scope	• This precept is intended to mitigate UxS failure modes of either a single function or multiple functions that could fail from the same causal factor. Integral safety design functions are greatly preferred over procedures.
Rationale	• MIL-STD-882E requires failure analysis to include component failure modes and human errors, single-point and common-mode failures that occur in subsystem components, and functional relationships between components and equipment comprising each subsystem for reducing risk. For UxSs, system design must address redundant safety mitigations for failure modes that, in a manned system, would be prevented by the human operator. The amount of autonomy that the UxS contains should be commensurate with fail-safe/recovery mechanisms.
Examples	• A weaponized UxS requires the control entity to perform independent actions to fire weapons. • A weaponized UxS requires the control entity to acknowledge and perform independent actions to select and engage target threats. • A swarm of UAVs requires unique and redundant navigation controls per individual UAV to prevent loss of guidance. • UxS platform should have independent safety features preventing inadvertent release. • UxS display should have an independent safety display status indicating location information, responsible control entity, state and mode, weapon status, etc. • A steering actuator for a convoy UGV designed with redundant position sensing and control functionality.
Detailed Considerations	• The design should incorporate a minimum of two independent safety features, each of which will prevent subsequent commanded or uncommanded launch/release/firing/arm enable of the weapon. • Safety significant data should be stored in more than one area of memory and compared for consistency prior to use. • The safety significant commands/functions should require a minimum of two independent and unique validated messages in the proper sequence.

Category	Description
	• The individual UxS design should detect when a safety significant failure occurs and revert back to a predefined, fail safe response. Failure controls must address both individual and multiple systems loss and revert back to a predefined, fail-safe response and/or human operator control. • The design should incorporate redundant preconfigured failure responses (e.g., unexpected loss of communications timeout, location waypoints).

Table C-11. DSP-11 Clarification Table

Category	Description
DSP-11	• The UxS should be designed to transition to a preconfigured safe state and mode in the event of safety significant failure.
Scope	• This precept addresses the overall UxS design management in the event of any safety significant failure such as unexpected loss or corruption of the communications link. Safety significant failures must result in safe and graceful degradation of the system.
Rationale	• The intent of this precept is to compel analysis of failure modes to anticipate their safety criticality and develop necessary contingency actions, which may include graceful degradations. Preconfigured safe states can be provided by the combination of the UxS design or by the mission parameters that are entered by the human operator. Preconfigured safe state and mode status should be based on CONOPS, mission profiles, threat hazard assessments, and other parameters.
Examples	• A UAV would continue to fly out of range upon loss of command link if no return data points were programmed into the navigational software. • A UAV has been directed upon loss of link to return to base. It currently has mission parameters loaded, weapons have been energized, and it has been commanded to fire, all just before the communications link was lost. If communications are reestablished, the UAV and weapons should default to the planned or expected state. • When a UAV operating in autonomous mode lost propulsion, it attempted to glide to a preplanned safe waypoint for recovery. • When a UGV lost its command signal, it defaulted to preplanned navigation in autonomous mode to a preset egress waypoint. This UGV has a rear obstacle avoidance system and egresses at a reduced speed. • Upon detection of a safety significant failure (a snapped steering cable), an unmanned sea-craft transitioned to a preset safe state, resulting in "spinning" until it ran out of gas. An alternate design could have provided a remote shut-off switch.
Detailed Considerations	• The design solutions might include selecting contingency plans for UxS system or subsystem failures or recognition of hazardous operational or environmental conditions that would require a modification to the mission parameters to ensure safe operations. • The UxS should provide design features that include awareness of battle damage of safety significant functions. • The system should have the capability to alert the operator and automatically transition into an anticipated, preconfigured safe state (e.g., recovery, sterilization, maintenance).

Category	Description
	• System faults should mandate the UxS transition to an alternate safe mode of operation. • The system should be designed to allow for safe and graceful degradation upon system-level or subsystem-level failures throughout the mission. • UxSs should be able to distinguish between a planned and an unexpected loss of communications occurrence. • The design should define state and mode transitions, including a desired and/or preconfigured course of action (such as move physically to a safe zone or crash in a safe zone) in the event of loss of command link or safety significant failure. The criteria for preconfigured states and modes, and the courses of action include: o The latest policy/ TTPs /ROEs/human operator takes precedence, o Autonomous mode and selection of control entity, o The operating environment (e.g., training, test, underwater, airborne), o Location of the UxS, and o Performance degradation (e.g., battle damage, power supply source). • The UxS design should consider retention of pertinent mission information (such as last known state and configuration, etc.) for the UxS and the controlling entity to recover from safety significant failure. • The UxS design should consider limiting the count and duration for which undelivered messages are considered valid within the communication system. • The UxS should ensure command messages are prioritized and processed in the correct sequence and in the intended state and mode. • A UxS implemented as an applique kit to an existing vehicle should consider failures of the base vehicle that may be difficult to detect through the applique kit.

Table C-12. DSP-12 Clarification Table

Category	Description
DSP-12	• The UxS should be designed for safe recovery if recovery is intended.
Scope	• This precept only applies when safe recovery is required. This precept covers three main points: the design supports a recovery process that is adequately safe in normal operations; the design supports a recovery process in degraded or damaged mode; and the system is designed to be "safed." This precept addresses the recovery of a UxS when the state of the UxS and/or weapons may not be known, which includes the return of UxSs with weapons unexpended and possibly armed, and platform and equipment configuration upon landing and/or recovery. Methods for recovery may include back-up systems or procedural controls.
Rationale	• UxSs typically are valuable assets. Their value can be represented by such things as cost, sensitive information, and if captured, their reuse by unauthorized entities. Therefore, design features should be included to ensure safe recovery of the UxS, ancillary equipment, and unexpended weapons stores.
Examples	• A UAV with a jettisonable weapon attempts to release a weapon unsuccessfully, creating a hang-fire situation. Upon UAV recovery, the design should allow for safing of the jettison rack and weapon stores.
Detailed Considerations	• The UxS design should consider that the recovery site may have assets that need to be protected from injury, death, system damage, or environmental damage. • The UxS design should allow the weapons to be identifiably safed upon recovery. • The UxS design should provide a means for inhibiting subsystem movement during recovery. • UxS CBR contamination or unexploded ordnance threats could exist in association with recovery.

Table C-13. DSP-13 Clarification Table

Category	Description
DSP-13	• Use of the UxS newly learned behavior should not impact the UxS's safety functionality until the newly learned behavior has been validated.
Scope	• For learned behavior, including machine learning, AI, etc., the newly learned behavior cannot be acted upon until the appropriate level of V&V is completed if it influences the safety significant functionality. The operationally learned behavior influencing safety functionality can be stored for future consideration/analysis, but the impact of changes and evidence of safety must be adequately assessed before the learned behavior is accepted and integrated.
Rationale	• Ensures "operationally learned behavior" is not acted upon if the learned behavior has the potential to influence safety significant functionality until that learned behavior related to UxS functional hazards has been properly assessed. Each new or revised learned behavior that influences safety significant functions will undergo appropriate LOR activities and V&V to characterize the system's new behavior.
Examples	• A UAV learns a new environmental operational profile from reading temperature and air density during routine mission operations. • A perception system is mounted on a fielded manned ground vehicle to gather data for a pedestrian detection algorithm to be used on a future UGV.
Detailed Considerations	• Any methods or techniques used to validate machine learned changes to safety significant software should satisfy software safety LOR analysis and testing requirements based on the software criticality index (SwCI) assigned (see MIL-STD-882E for SwCI and LOR requirements). • While incorporation of AI into a system does not lessen the LOR required for software V&V under MIL-STD-882E, learned behavior presents difficulties when attempting to verify and validate within a system. It is recommended that detailed functional hazard analysis be performed prior to implementing AI in a system in order to bound learned behaviors to known safe conditions, if practical. The proposed boundary conditions must be defined, deterministic, and verifiable for this condition to be true. Embedding learned behaviors within known safe constraints should allow for more streamlined V&V of the resulting software and should allow for increased learning options during operation of the system without violating safety requirements.

Category	Description
	• Biases can be introduced by the humans developing the machine learning software code. Risk of unwanted outcomes should include an assessment of the associated machine learning and any biases that such learning may include. Machine learning biases may be unknown to the user. Testing may not identify such biases because the user and validation communities are unaware of the machine learning biases. Current techniques such as extreme gradient boosting and deep neural networks may exploit these biases for improved performance, but they may fail in making safe predictions due to unknown shifts in the data domain or inferring incorrect patterns or harmful rules. Learning models are complex and it will be difficult to understand how they will react to such shifts and whether they will produce harmful outcomes as a result unless there is a deliberate effort to identify and document such bias. • Unwanted outcomes, such as those human biases introduced during machine learning software development, can be avoided by defining the confidence boundaries of the learning model. When any model prediction is outside these boundaries, then that option is rejected. When an option is rejected because of a confidence boundary issue, the system can identify the need for human intervention. This does not eliminate the option but moves the decision back to the human-in-the-loop. • Data used to baseline the level of a machine learning system (e.g., target set parameters, operational environmental criteria) should be assessed, validated, and verified for its safety impact.

Table C-14. DSP-14 Clarification Table

Category	Description
DSP-14	• Autonomy should only select and engage targets that have been predefined by an authorized human.
Scope	• The intent of this precept is to comply with the existing DoD policy to allow the autonomy selection and engagement of valid targets by an authorized human operator to mitigate unintended engagements. • Weapon firing/release criteria must be in accordance with Paragraph 4.c. of DoDD 3000.09 or by a system that has undergone senior review and approval pursuant to paragraph 4.d of DoDD 3000.09. • Initiation of a UxS mission must be done with human consent as the control entity for the system.
Rationale	• Selection and engagement of individual targets or specific target groups will be in accordance with the latest DoD policy, TTPs, and ROEs. The human control entity for selection and engagement takes precedence over autonomy selection and engagement. Any autonomy used for selection and engagement of targets incorporated in the weapon system also falls under DoD policy and cannot be used to replace the human control entity to select a human target for engagement without senior-level review.
Examples	• An autonomous classification algorithm can recommend human targets to an authorized human operator of a weapon system, but the human operator must make the final decision of any of the recommended targets before the UxS can engage the target. • The human operator must approve the target engagement of any of the recommended selected target engagements provided by an autonomous classification.
Detailed Considerations	• The human as a target selection is not allowed for lethal and nonlethal autonomous weapon systems in accordance with DoDD 3000.09, without senior-level review. • Target selection derived by autonomy should be validated by the human operator prior to target engagement. • Final target engagement must follow the delegation of authority. Note: autonomy should always be revocable by the human operator. • In the event of unexpected communications issues, the UxS weapon system should not autonomously select and engage individual targets or specific target groups that have not been previously selected and predetermined by a human operator. In addition, any previously selected targets or target groups that were confirmed by the human operator should have an expiration time on their authorization that is appropriate for the context of the mission (e.g., ROE).

Category	Description
	• In the event of a critical system fault (e.g., identified faults which could impact target engagement safety), the UxS weapon system should be designed to provide adequate fail-safe responses and should not autonomously select and engage individual targets or specific target groups. • In the event of a detected fault of the autonomous target classification; the UxS weapon system should alert the human operator entity of the detected fault and should not autonomously select or engage individual targets or specific preconfigured target groups.

Table C-15. DSP-15 Clarification Table

Category	Description
DSP-15	• Common user controls and display status should be utilized for similar functions such as manual override, terminate mission, and learning mode.
Scope	• The intent of this precept is to address the standardization and efficiency of common user controls and display status information to mitigate for human operator error across all UxSs.
Rationale	• A common set of standardized user interface functions (e.g., warnings, alerts) helps to reduce the probability of human error. User interface efficiency for safety significant functions directly reduces mishaps. Interface commonality for all functions helps improve efficiency.
Examples	• Enabling a kill/abort switch requires two independent actions that can only be operated in a specific sequence to prevent inadvertent activation. A common and efficient interface helps enable timely execution of the two actions before a mishap can occur. • Timely execution of the abort command from the command entity requires efficient interface so that abort function is completed in the required timeframe. • Common location of kill/abort switches should be utilized across platforms.
Detailed Considerations	• Mission and safety information that is displayed for human view and decision making should be organized in an intuitive, clear, and efficient manner in accordance with applicable human factors engineering guidelines. Operations that occur most often or have the greatest impact on safety should be the easiest to perform. Where functions that require higher cognitive memory are required, help tools should be easily accessible. • Platform controls and user interfaces should be designed using recognized human system interface and engineering factors with carefully prepared system design data in order to design platform control and information systems that are cost effective, efficient (ergonomically and cognitively), and intuitive. • The controlling entity should have the capability to view/access the latest safety significant information of all UxSs being controlled at all times. • User controls accessed for autonomous functionality should be separated and clearly different from user controls implementing other types of functions. • The human operator should have the ability to switch intuitively and efficiently from any autonomous mode of operation to a manual mode of operation. • Operator control soft keys (if utilized) should be strategically positioned to minimize erroneous operator actions that could cause inadvertent commands to the UxS.

Category	Description
	• In the event of compromised operation (security/physical tampering), degraded functionality, or physical damage, the operator needs the ability to take control quickly of the UxS to the greatest extent possible. • The operator should have the ability to override operations and abort the mission. • Display information should include relevant UxS's communications and health monitor status to include, but not be limited to, network connectivity/communications, location, critical platform battery/power levels, message processing status, tamper status, weapon status. • The operator should have the ability to efficiently terminate a command to the greatest extent possible and to receive acknowledgment of the cancellation status. • Strong connection to OSP-9, OSP-10, and OSP-11.

Acronyms

ACRONYM	MEANING
AI	artificial intelligence
AOP	allied ordnance publication
CBR	chemical, biological, or radiological
CONOPS	concept of operations
DoDD	DoD Directive
DoDI	DoD Instruction
JAIC	Joint Artificial Intelligence Center
DSP	design safety precept
JCS	Joint Chiefs of Staff
LOR	level of rigor
MIL-STD	military standard
ML	Machine Learning
OSP	operational safety precept
PM	program manager
PSP	programmatic safety precept
ROE	rules of engagement
SOP	standard operating procedure
SSPP	system safety program plan
STANAG	standardization agreement
SwCI	software criticality index
TBD	to be determined
T&E	test and evaluation
TEV&V	test, evaluation, verification, and validation
TTP	tactics, training, and procedures
UAS	unmanned aeriel system
UGV	unmanned ground vehicle
OUSD(A&S)	Office of the Under Secretary of Defense for Acquisition and Sustainment
OUSD(R&E)	Office of the Under Secretary of Defense for Research and Engineering
UxS	unmanned system
V&V	verification and validation

Glossary

TERM	DEFINITION
AI	Software and system functionality that exhibits such traits as recognize patterns, capacity to learn, understand, reason, plan, cognate, and problem solve.
authentication	A security method for establishing the validity of a transmission, message, or originator to protect against acceptance of a fraudulent transmission or simulation.
causal factor	One or several mechanisms that trigger the hazard that may result in a mishap.
command entity	A human in a position of authority to command and control the UxS, who can authorize a control entity to execute commands, and who can rescind authorization or transfer it to a different control entity at any time.
control entity	A human or software, when authorized by a command entity, that is capable of controlling the system functions to execute commands as given (by the command entity) and accommodating the state of the UxS.
deterministic checkpoint	A point in the process at which the control entity, whether operator or software, may review the planned behavior chosen by the system's software and decide to intercede or not (for supervised autonomous), or give approval or not (for semi-autonomous) software. This can be provided through the implementation of an appropriate run-time verification or monitoring function or through the provision of human-in-the-loop or human-on-the-loop review and command and control.
DSP	General design guidance intended to facilitate safety of the system and minimize hazards. These safety precepts are intended to influence, but not dictate, specific design solutions.
fail safe	Reverting to a predicted or known safe state.
flexible autonomy	The capability for the real-time, seamless change or reconfiguration in system autonomy.
fully autonomous	A mode of operation whereby the system is permitted to perform all designed functions without further human interaction/intervention necessary to safely execute a specified task.

TERM	DEFINITION
human-machine interface	The means by which the human operator interacts with the UxS system. It includes the software applications, graphics, and hardware that allow the operator to effectively give instructions to or receive data from the UxS.
learning mode	A mode in which a UxS modifies its knowledge base or decision logic to determine its future responses and behaviors.
LOR	A specification of the depth and breadth of software analysis and verification activities necessary to provide a sufficient level of confidence that a safety-critical or safety-related software function will perform as required.
mishap	An event or series of events resulting in unintentional death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment.
ML	A subfield of AI that allows machines to learn from data without being explicitly programmed. The principle branches of ML are supervised learning, unsupervised learning, and reinforcement learning.
operator	Human control entity.
OSP	A safety precept directed specifically at system operation and including operational rules that must be adhered to during system operation. These safety precepts may generate the need for DSPs.
PSP	Program management principles and guidance that should help ensure safety is adequately addressed throughout the life-cycle process.
safe state	A state in which a platform or system can perform no harm to humans, or other entities.
safety critical	A term applied to a condition, event, operation, process, or item whose mishap severity consequence is either catastrophic or critical (e.g., safety-critical function, safety-critical path, and safety-critical component).
safety criticality index	Index or number assigned based on MIL-STD 882 software safety criticality matrices: SwCI is the combination of the severity category and software control category. The SwCI corresponds to the

TERM	DEFINITION
	minimum LOR tasks required to assess the software contributions to the system-level risk.
safety precept	A basic truth, or presumption intended to influence management, operations, and design activities, but not dictate specific solutions. A safety precept is worded as a nonspecific and unrestricted safety objective that provides a focus for addressing potential safety issues that present significant mishap risk.
safety significant	A condition, event, operation, process, or item that is identified as either safety-critical or safety-related.
semi-autonomous	A mode of operation whereby the system is permitted to perform only selected decisional functions without further human interaction/intervention.
subsystem	A grouping of items satisfying a logical group of functions within a particular system.
supervised autonomous	A UxS operating mode that enables a human operator to intervene and terminate selected operations or activities.
system	An integrated composite of people, products, and processes that provide a capability to satisfy a stated need or objective.
UxS	A system comprising the necessary subsystems to operate partially or fully independent of human command and control within specified constraints in any tactical or operational domain.
UxS command and control	The exercise of authority and direction by a properly designated command entity over assigned and attached UxSs in the accomplishment of the mission.
validation	The process of evaluating a system or software component during or at the end of the development process to determine whether it satisfies specified requirements.
verification	The process of determining that a UxS accurately represents the user's conceptual description and specifications.

References

AOP-16/STANAG 4187 Rev 4, "Fuzing Systems – Safety Design Requirements," September 3, 2007

AOP-52/STANAG 4452 Edition 1, "Guidance on Software Safety Design and Assessment of Munition-Related Computing Systems," November 19, 2016

ASTM F3269-17, "Standard Practice for Methods to Safely Bound Flight Behavior of Unmanned Aircraft Systems Containing Complex Functions," September 1, 2017

AUTONOMOUS HORIZONS, System Autonomy in the Air Force – A Path to the Future, Volume 1: Human-Autonomy Teaming, United States Air Force Office of the Chief Scientist June 2015, Section 4.1

DoD Directive 3000.09, "Autonomy in Weapon Systems," November 21, 2012, as amended

DoD Directive 5000.01, "The Defense Acquisition System," September, 9 2020, as amended

DoD Instruction 5000.80, "Operation of the Middle Tier of Acquisition (MTA)," December 30, 2019, as amended

DoD Instruction 5000.87, "Operation of the Software Acquisition Pathway," October 2, 2020, as amended

DoD Instruction 5000.88 "Engineering of Defense Systems," November 18, 2020, as amended

DoD Instruction 5000.89, "Test and Evaluation." November 19, 2020, as amended

DoD Directive 5135.02, "Under Secretary of Defense for Acquisition and Sustainment (USD(A&S))," July 15, 2020, as amended

DoD Directive 5137.02, "Under Secretary of Defense for Research and Engineering (USD(R&E))," July 15, 2020, as amended

DoD Instruction 5105.18, "DoD Intergovernmental and Intragovernmental Committee Management Program," July 10, 2009, as amended

Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence," February 11, 2019

Software System Safety: Implementation Process and Tasks Supporting MIL-STD-882E, Joint Services – Software Safety Authorities (JS-SSA), October 2017

Joint Software Systems Safety Engineering Handbook, Joint Software Systems Safety Engineering Workgroup, August 27, 2010

Military Standard MIL-HDBK-516 Rev C, "Airworthiness Certification Criteria," October 7, 2019, as amended

Military Standard MIL-STD-882 Rev E, "Department of Defense Standard Practice for System Safety," April 23, 2012, as amended

Military Standard MIL-STD-1316 Rev F, "Safety Department of Defense Design Criteria Standard: for Fuze Design, Safety Criteria," August 18, 2017, as amended

Military Standard MIL-STD-1472 Rev H, "Human Engineering," September 15, 2020, as amended

Military Standard MIL-STD-1901 Rev A, "Department of Defense Design Criteria Standard: Safety Criteria for Munition Rocket and Missile Motor Ignition System Design, Safety Criteria," June 6, 2002, as amended

Military Standard MIL-STD-1911 Rev A, "Department of Defense Design Criteria Standard, Safety Criteria for Hand-emplaced Ordnance Design," July 24, 2014, as amended

Military Standard MIL-STD-2088 Rev B "Bomb Rack Unit (BRU), Single Store, Aircraft," September 13, 2019, as amended

Military Standard MIL-STD-2105 Rev D, "Department of Defense Test Method Standard: Hazard Assessment Tests for Non-Nuclear Munitions," April 19, 2011, as amended

NATO Standardization Agreement 4586

NATO Standardization Agreement 4737

Secretary of Defense Memorandum, "Artificial Intelligence Ethical Principles for the Department of Defense", February 21, 2020

RTCA DO-178, "Software Considerations in Airborne Systems and Equipment Certification,"

Stone, Morley O. "DoD Priorities for Autonomy Research and Development," Morley O. Stone, October 21, 2011, www.dtic.mil/get-tr-doc/pdf?AD=ADA554722

Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity, DoD, February 12, 2019

Technology Investment Strategy 2015-2018, Autonomy Community of Interest (COI) Test and Evaluation, Verification and Validation (TEVV) Working Group, May 2015 www.dtic.mil/docs/citations/AD1010194

Unmanned Systems Safety Guide for DoD Acquisition, OUSD (AT&L) Systems and Software Engineering/Developmental Test & Evaluation, June 27, 2007

Unmanned System Safety Engineering Precepts Guide for DoD Acquisition

Office of the Under Secretary of Defense for Research and Engineering
Deputy Director for Engineering
3030 Defense Pentagon
Washington, DC 20301-3030
https://ac.cto.mil/engineering

Office of the Under Secretary of Defense for Acquisition and Sustainment
3010 Defense Pentagon
Washington, DC 20301-3010
https://www.acq.osd.mil/