Ethical and technical challenges in the development, use, and governance of autonomous weapons systems

By an independent group of experts convened by the IEEE Standards Association:

Emmanuel Bloch

Ariel Conn

Denise Garcia

Amandeep Gill

Ashley Llorens

Mart Noorma

Heather Roff

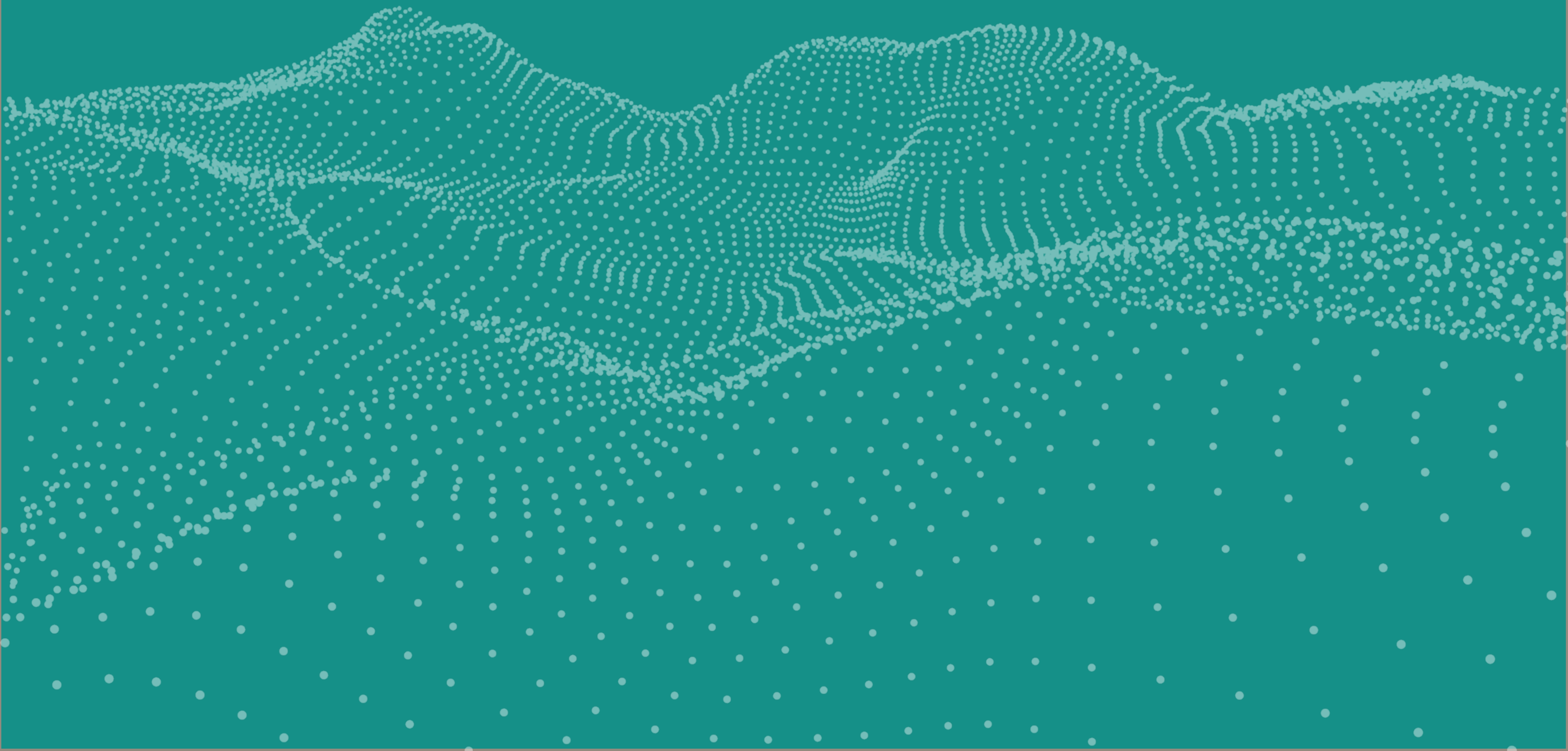
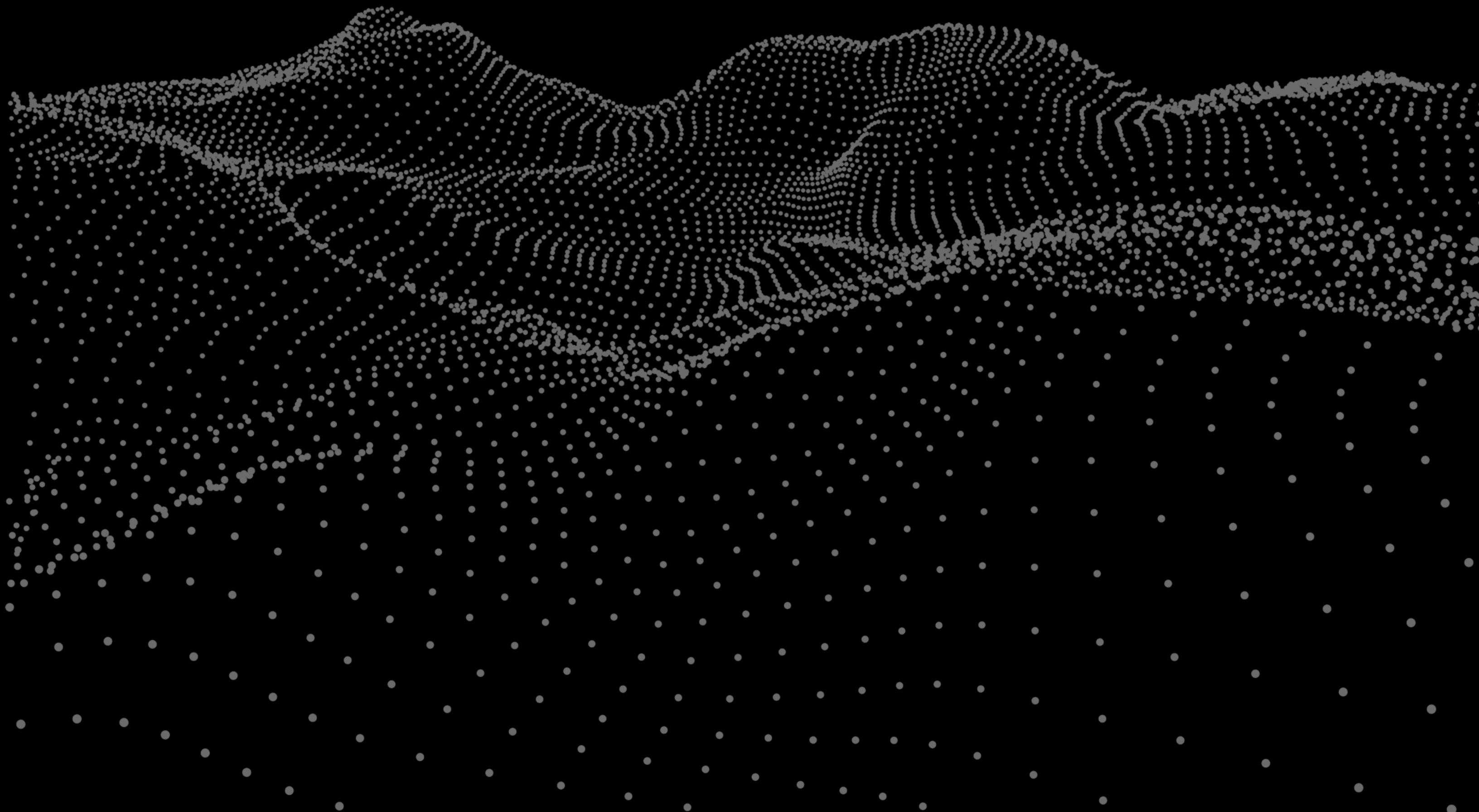


Table of Contents

Introduction

> “As the use and impact of autonomous and intelligent systems (A/IS) become pervasive, we need to establish societal and policy guidelines in order for such systems to remain human-centric, serving humanity’s values and ethical principles.”
>
> *— IEEE Ethically Aligned Design, First Edition*



Limited levels of autonomy have existed in weapons systems for decades. However, recent computational advances, especially in the field of artificial intelligence (AI), have changed the scope of what a weapons system can accomplish autonomously.

Capabilities today extend beyond such competencies as navigation and radar detection to include facial recognition software, swarming technology, extensive data analysis, and much more. Further, due to technological advancements, autonomous weapons systems (AWS) are increasingly allocated tasks that were traditionally performed by humans.

Such advancements have raised questions at national and international levels about what is technologically possible and what is legally and morally acceptable regarding autonomy and AI in a weapons system, especially with respect to the decision to use force. The most prominent international discussion is taking place within the United Nations Convention on Certain Conventional Weapons (CCW). There, the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons (GGE) have been discussing matters pertaining to AWS, including whether they are sufficiently covered by existing International Humanitarian Law (IHL) and International Human Rights Law (IHRL) and on the quality and extent to which human control must be maintained.

Yet while discussions remain ongoing, AI technologies and autonomous capabilities continue to advance, creating uncertainty regarding the future of autonomy in warfare and with some voicing further concerns regarding the impact of AI on nuclear strategy. Though some countries have developed policies or statements on AWS, such as the U.S. Department of Defense’s Directive 3000.09, we still lack widely accepted international best practices and guidelines for development, use, and governance of AWS. Because AI and autonomous systems represent dual-use technologies, many of the innovations that are later employed in weapons systems were originally designed with consumers or businesses in mind. For example, algorithms that might be used for image recognition in commercial applications could be repurposed for target discrimination in military applications. Technology corporations have also faced public backlash when they have attempted to work with governments on military projects. In response to this, along with public pressure regarding many other AI-related issues, corporations, governments, international organizations, and nonprofit organizations have developed various sets of principles to offer guidelines around which they plan to develop and use artificial intelligence ethically. The Global Landscape of AI Ethics paper identified “84 documents containing ethical principles or guidelines for AI,” most of which were written between 2016 and 2019.

As these principles, which focus on artificial intelligence more broadly, have been developed, the difficulties of dissociating artificial intelligence development from possible military uses have become increasingly apparent. As a result, corporations and governments alike seek to determine how AI and autonomy can be used in weapons systems and what is legally and morally acceptable. Though the development of principles begins to address the growing demand for consensus on AI ethics, translating these principles for AI to applicable best practices for AI and autonomy in weapons systems is no small feat.

A group of independent AWS experts convened to help bridge the gap between general principles and practical insights in this sensitive and vital domain. The group's ultimate goal is to create a knowledge base of best practices regarding the governance, development, and use of autonomous weapons systems, initially based on the guidelines first established by AWS-related principles and later updated to reflect evolving technological advancements and international norms and regulations.

The authors of this paper believe that the first step beyond principles is to conduct a thorough investigation into each specific issue and identify the most pressing challenges facing organizations and governments in developing, using, and regulating autonomy in weapons systems, according to national and international laws. That is what the authors have attempted to do in this document.

To date, the only set of principles dedicated solely to AWS were those developed at the CCW in 2018 and updated in 2019. As such, these principles formed the basis for the development of our challenges. We also found weapons-related guidance in more general AI Principles, including the US Department of Defense's 2020 Artificial Intelligence Principles (from DIB Primary document), the IEEE's Ethically Aligned Design Principles, the European AI High-Level Expert Group's Principles, the OECD's Principles, and the UNESCO Principles. Additionally, the ICRC and SIPRI released a joint paper in June 2020, identifying some of the broader challenges of defining and achieving human control for AWS. Though the ICRC/SIPRI paper focuses more on challenges and solutions within the context of the CCW framework, it served as a helpful starting point for many of the challenges outlined below.

The following challenges represent an initial public draft, which we expect to update with feedback from stakeholders. We hope this document continues to advance and progress as some challenges are addressed and as new challenges are identified. The initial ten categories of challenges we have identified are:

1. **Establishing common language**
2. **Enabling effective human control**
3. **Determining legal obligations**
4. **Ensuring robustness**
5. **Testing and evaluating**
6. **Assessing risk**
7. **Addressing operational constraints**
8. **Collecting and curating data**
9. **Aligning procurement practices**
10. **Addressing non-military use**

These ten categories of challenges, explained more below, are inspired by various principles, and within those categories, we have identified the most relevant and distinct components. The document broadly addresses all forms of autonomy in weapons systems, but we acknowledge that the challenges below pose more legal, ethical and technical issues when the use of force is applied, especially for more advanced autonomous technologies, such as non-deterministic AI programs and online machine learning. The first round of challenges is primarily technical in nature, with some challenges relating to governance. We have found that discussions regarding AWS typically focus on legal and ethical issues surrounding the weapons systems, with less attention to specific technical capabilities and features, and we seek to fill that gap with this round of challenges.

With this framing of the document we are not suggesting that the technical challenges are more important or relevant than legal or ethical challenges, nor are we suggesting that AWS should be considered from a purely technical perspective rather than a human perspective; rather, we see the technical challenges as the most relevant area to which technical organizations, such as IEEE, can contribute.[1]

[1] The intent of the authors is to contribute this paper as the basis for an IEEE SA Industry Connections Activity on the ethical development, use, and governance of autonomous weapons systems.

We hope that more grounding in technical challenges will help enhance and advance those legal and ethical discussions. As such, we expect more governance challenges to be added in future iterations of this document, as we better understand the technical challenges and limitations. Additionally, we recognize that many of the challenges listed are beyond the scope of a technical organization to address. We hope that by listing them here, others may be able to consider and address them.

For now, we have deliberately chosen not to provide definitions for phrases, including “autonomous weapons systems,” “human control,” and “human element,” as we have instead recognized the lack of consensus around definitions as one of the core challenges to be addressed (see below). We have also chosen not to define “autonomy” or “AI” for the same reasons, though we use them regularly below. While some groups have chosen to bypass definitional confusion by using other phrases — such as the use of “autonomous and intelligent systems” in IEEE’s Ethically Aligned Design or NATO’s use of “systems with intelligent functions” — we continue to use “autonomy” and “AI” here because these are the terms most commonly used in discussions regarding AWS. For the purposes of this document, we recognize “autonomy” as deriving from the delegation of authority to a human or to a machine to act within specified bounds. We use “AI” to refer to a broad family of technologies that enable systems to make decisions and act autonomously with limited or high-level human involvement. Though autonomous systems have been around for many decades, continuing advancements in AI, including advancements in machine learning, are enabling an increasing sophistication in autonomous capabilities.

Additionally, though many members participating in the discussions regarding AWS use the acronym “LAWS” to indicate “lethal” autonomous weapons systems, we do not use the word “lethal” here. Whether or not an autonomous weapons system is lethal is not always clear, and the lethality of a system can also sit on a spectrum. For example, a system might autonomously identify a potential target and provide the human user with information regarding threats associated with the target, but the human user might use a different system to apply force based on that information. Alternatively, within a single system, some programs might run autonomously, such as navigation, but the application of force would require a human action. We observe that in situations such as these, the lethality of the autonomous system versus the human decision maker is unclear, but the challenges below would still apply. An AWS may also not apply lethal force, such as a system designed to shoot rubber bullets or tear gas. In this case, it could still be classified as a weapon system, the human user would still be expected to honor IHRL, and the challenges below would still apply.

Finally, though not explicitly stated below, we acknowledge the challenges associated with disagreements regarding the inherent legality or illegality of AWS. Again, we hope this document will help advance all of these discussions.

AWS Challenges

1. Establishing Common Language

Definitional disagreements often hamper the discussions regarding autonomous weapons, with even some of the most common phrases being poorly defined, including “autonomous weapons” and “human control.”

The lack of agreement is a problem for international negotiations and for national and corporate discussions. More broadly, when describing technologies that incorporate autonomy, AI, and other similar emerging technologies, people use terms, such as "transparency," "control," "fairness," and even “autonomy,” which have inconsistent definitions among different groups and across diverse languages and cultures. This is a challenge in itself, but the problem is further complicated in AWS, which require clear and concise meanings and definitions for any form of international regulations.

Additionally, when discussing the future of autonomous weapons, even the most informed observers often refer to AWS as having capabilities that do not currently exist, while the limitations of the technology are often overlooked. This discrepancy can lead to confusion and misunderstanding about what AWS can and will be used for, while anthropomorphic descriptions of the weapons increase both the hype and fear. Discussions around policy should be grounded in the technical realities of today with an eye toward future developments, but not limited by disagreements over what technologies may or may not be successfully developed farther into the future.

CHALLENGE 1.1

To develop guidelines and examples to help ground abstract and inconsistently defined terms (such as autonomous or transparency) in reality so that developers can translate principles and ethics into programming code.

CHALLENGE 1.2

To identify which ethical and principled concepts are still too abstract to be programmable.

CHALLENGE 1.3

To encourage nations and other stakeholders to come together towards establishing a smaller number of broadly-adopted definitional standards.

CHALLENGE 1.4

To incorporate internationally recognized standards, such as those developed by the IEEE, that apply to fields related to AWS.

CHALLENGE 1.5

To clarify and expand on principles and guidelines such that developers and designers can reasonably apply them when designing aspects of an AWS and to more easily recognize what AWS characteristics might be prohibited by IHL, by a national law, or by a relevant guiding principle.

CHALLENGE 1.6

To identify critical questions to ask across the full technology lifecycle to ensure legal and ethical issues are considered and addressed throughout.

CHALLENGE 1.7

To improve communication and transparency regarding autonomous and AI functions to be used in military settings for a more accurate and realistic understanding of current and near-future capabilities.

CHALLENGE 1.8

To acknowledge when definitional disagreements cannot be overcome and find ways to continue discussions around AWS regardless.

2. Enabling Effective Human Control

As stated in the Guiding Principles of the Group of Governmental Experts on AWS with the United Nations Convention on Certain Conventional Weapons, “Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines.”

And according to the ICRC/SIPRI (June 2020), there is a need for a combination of three types of control measures:

- **controls on the weapon system’s parameters of use,**
- **controls on the environment, and**
- **controls through human-machine interaction.**

The centrality of the human element in every stage of development and use of an AWS is paramount for the observance of International Humanitarian Law (IHL), Human Rights Law (HRL), and the Law on State Responsibility.

The phrase “human control” can also cover a spectrum of control, from a human directly enabling a system to take an action to a human deciding the parameters within which a system can act, but in all cases, human control and/or the human element needs to be identified in advance and planned for in each of the weapons systems stages: planning, design, development, deployment, use, and retirement. As a result, responsibility and accountability can be attributed in case problems arise. Additionally, increased autonomy in weapons systems may enable a human decision to occur in the early stage of a mission, which, in some cases, could lead to a disconnect between the initial intention and what actually occurs.

(a) Determining Human Control

In order to ensure that sufficient human control is maintained and/or that the role of the human element is appropriately established, these will need to be determined in advance, for AWS in general and/or for each specific system.

CHALLENGE 2a.1

To determine what role the human element must play and what degree of human control and accountability must exist.

CHALLENGE 2a.2

To develop national and international norms regarding the degrees of human control and accountability for AWS.

CHALLENGE 2a.3

To realistically identify the extent to which a human can override an AWS decision or action, given the potential speeds of autonomous systems and that they may not always be in communication with operators, and to enable the overriding of decisions that the autonomous system makes.

CHALLENGE 2a.4

To identify, during a weapon's development, the spatial and temporal boundaries that must exist with respect to when the human makes a decision regarding the use of an AWS and the final outcome, in order to ensure compliance with IHL.

CHALLENGE 2a.5

To include elements in the AWS that can notify operators or commanders of potential misuse, technical malfunction, or defect, especially regarding issues like incomplete data, an unfamiliar environment, or other technical limitations or inconsistencies.

CHALLENGE 2a.6

To ensure the user is able to approximate any potential differences between the initial assumptions of the system and the final targeting to reduce risks to life, human dignity, and to comply with existing legal, ethical, and moral constraints.

CHALLENGE 2a.7

To establish international laws or norms regarding the extent to which a human must be able to override an AWS based on the category of AWS technologies and context of use.

(b) Ensuring Technical Understanding

Autonomous weapons systems are complex systems with many different people, groups, organizations, corporations, governments, etc., involved at varying stages of design, development, deployment, etc. The technical understanding and nuance of an AWS and its intended use may not be consistent between the system developer and the end-user.

This is especially true given the layers of people and bureaucracy between the development and use of a system within a military. Additionally, without proper training, military members may not fully understand the system's technical capabilities, which could lead to improper use, or the military may not use the system at all.

CHALLENGE 2b.1

To develop formal training for military commanders and operators that will confer an adequate understanding of the testing and evaluation conducted on a given AWS and that will help those commanders and operators to develop an accurate mental model of the system's capabilities and limitations so that they can make better judgements regarding appropriate uses of that system.

CHALLENGE 2b.2

To develop appropriate tactics, techniques, and procedures for a given weapon system during the development stages, recognizing that these will need to change and adapt over time.

(c) Designing for User Interface & Experience

Algorithms in a weapons system are designed to help process information at a rate far faster than a human, which allows the system to analyze more data at speeds exponentially greater than the person overseeing the system. To maintain an appropriate level of oversight of the system, designers need to create user interfaces that allow the operators and commanders to understand what the system is doing and why.

CHALLENGE 2c.1

To develop an information flow and interface design that is both comprehensive and comprehensible to the humans using or overseeing the system, regardless of the complexity of the situation or information, and to ensure the users have time to process information both from the system and from external sources.

CHALLENGE 2c.2

To develop standardization and protocols for interoperability, training, doctrine, and use of the information flow and interface design.

CHALLENGE 2c.3

To ensure designers and the system's end-users interact to clarify problem points and that members of relevant disciplines related to the user interface are involved at all levels of design, development, and use.

(d) Ensuring Situational Awareness & Assurance for Accountable Human Oversight

With recent advances in AI, especially machine learning and image recognition, humans increasingly rely on technology within a remote AWS for information about safety critical or high-stakes situations. There may be situations (whether by design or due to operational conditions) where an AWS system is out of communication and unable to send information to commanders or operators in real time. Yet in all cases, humans accountable for AWS oversight still need some level of situational awareness, assurance such an autonomous system is operating appropriately, and/or options for intervention.

CHALLENGE 2d.1

To ensure that humans with AWS oversight responsibility have an appropriate amount and type of information at the right time to maintain situational awareness.

CHALLENGE 2d.2

To ensure that AWS are designed to enable oversight and failsafes as necessary, especially when a system is out of communication.

CHALLENGE 2d.3

To develop standards and protocols for governing a system when it's out of communication, and to identify a means of ensuring trust in a system when that system is out of communication.

(e) Designing for Human Cognition

Science has not yet acquired sufficient knowledge about human cognition to understand how well the human brain can process the extensive data that an AWS is expected to process and provide to the operators and commanders overseeing the system.

CHALLENGE 2e.1

To ensure that autonomous systems are designed so that human supervisors can choose a flow of data, in volume and type, appropriate for their use conditions, recognizing that the type of task and cognitive capacity will not be constant.

CHALLENGE 2e.2

To identify or develop a means of verifying whether the AWS is presenting the operator or commander with sufficient and reasonable information about the situation, and to identify standard verification benchmarks that help ensure replicability and generalizability of this system or process.

CHALLENGE 2e.3

To ensure studies of AWS with respect to human cognition take into account the rights and welfare of those affected by their use.

CHALLENGE 2e.4

To recognize human tendencies to trust a machine even when it's providing inaccurate or incomplete information, and to not only train humans to take extra precautions, but also to program the system to analyze and provide data more effectively.

(f) Identifying Who Decides and When

An important aspect of autonomous weapons systems is that the systems can autonomously identify which action to take and then do so. However, many AWS are designed such that they can operate with different degrees of autonomy and with different degrees of human interaction, including in a fully autonomous state, by order of human commanders, or as part of human-machine teams. Given the uncertain nature of how the systems will be used and in what situations, it will not always be clear when an AWS should make a decision or when it should be left to a human.

CHALLENGE 2f.1

To identify, either predictively or during a situation, the degree of autonomy a system may maintain in the decision-making process with respect to the human overseeing the system.

CHALLENGE 2f.2

To adequately represent and measure decision making by 1) machines, 2) humans, and 3) human-machine teams, and to achieve shared intent and ensure the most seamless adaptive role-taking, tasking, and hand-off that's possible between the human(s) and the system(s).

CHALLENGE 2f.3

To design systems and develop concepts of operations that enable humans to remain responsible and accountable for distributed decision making.

CHALLENGE 2f.4

To develop AWS that can run audits which provide explanations for why the system took the actions it did.

3. Determining Legal Obligations

AWS are often complex systems composed of many subsystems that require various people and organizations involved at different stages of development and use.

Additionally, systems that utilize machine learning may continue to evolve and adapt throughout their life cycle and may be inherently uncertain. Many argue this uncertainty makes the weapons inherently illegal, while others argue that, if a framework can be developed to limit the uncertainty of a machine learning system to be within predefined boundaries, then the system would be legal. Yet as long as any uncertainty remains, determining responsibility, accountability, liability, and other legal obligations for autonomous weapons systems, such as respect for human rights obligations that all countries are expected to observe in conjunction with IHL, represents a challenge.

Moreover, as a multi-use technology, AI algorithms can be programmed into systems for various purposes (including military and commercial), and those systems can potentially be repurposed by the end-user to perform tasks the designer never intended. While international discussions around AWS focus more on the role of countries and governments regarding the legal obligations of AWS, the general public increasingly expects the companies designing AI and AWS to take ethical stances regarding the products they develop. In addition to the legal issues, countries must deal with regarding these challenges, the companies could also face a growing number of legal actions or other types of backlash if the way products are used is perceived to be in violation of a company’s principles and in breach of international legal obligations.

CHALLENGE 3.1

To address responsibility, accountability, liability, and other legal obligations given the uncertain nature of dual-use and nondeterministic technologies with application to AWS.

CHALLENGE 3.2

To ensure that liability is clearly outlined by a state’s domestic law and international law.

CHALLENGE 3.3

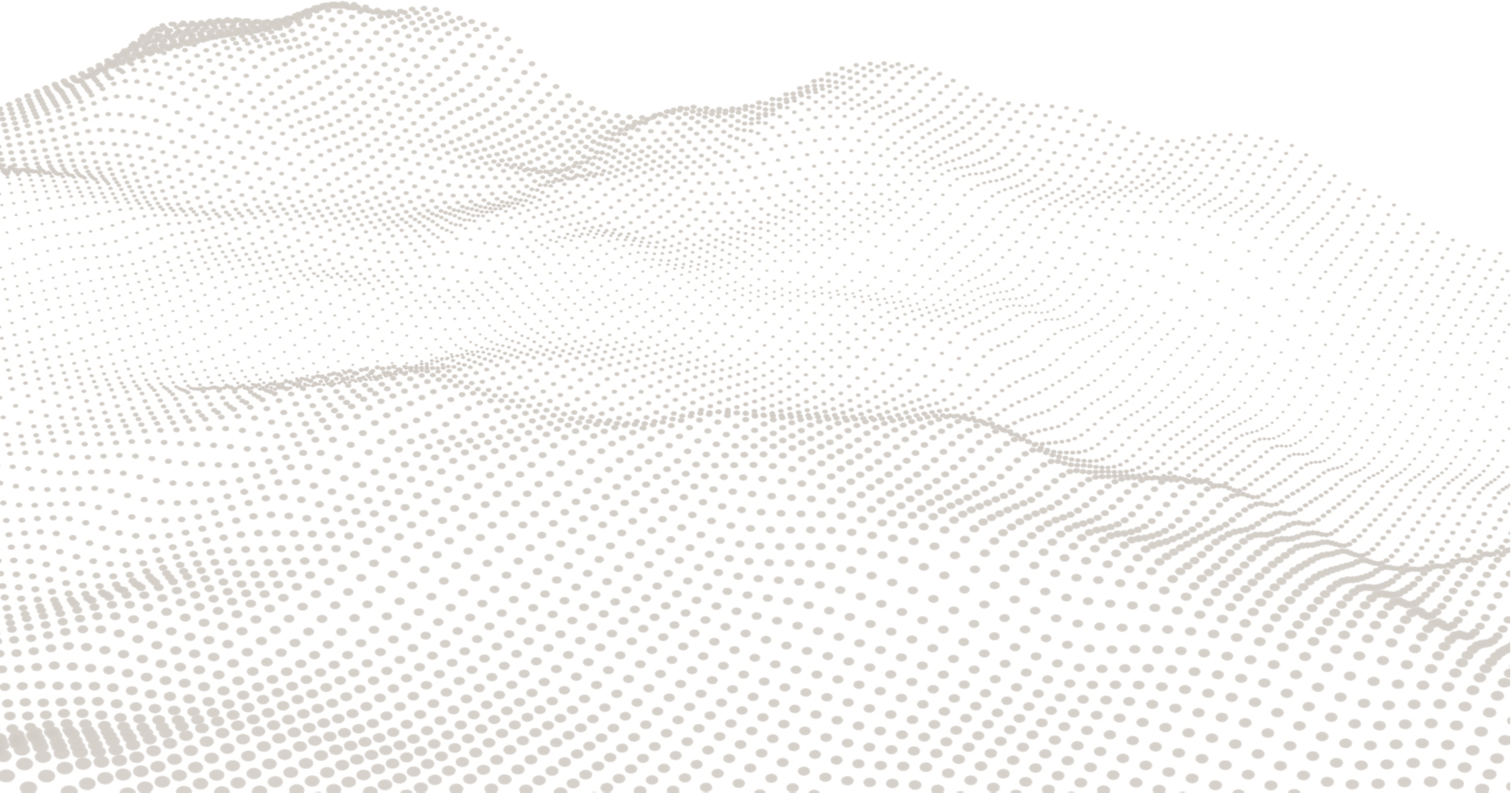
To ensure the use and the mission boundaries are clearly expressed such that potential misuse of the system can be identified.

CHALLENGE 3.4

To determine whether online machine learning systems, which can learn and update their programs in the field, can go through legal testing and validation as they adapt to their environment, or whether regulations are required such that systems can only update off-line, applying what they learned in the field after the mission is completed.

CHALLENGE 3.5

To improve clarity of existing laws as they apply to AWS, and to determine what new norms or regulations are needed.



4. Ensuring Robustness

An autonomous system could be entirely pre-programmed without AI.

Alternatively, a system might use an AI program with machine learning such that it could learn something new and adapt its actions accordingly — either processing new information in the field and adapting in real time, or processing information it learned in the field during periods it's not in use and applying what it learned later. Those are just a couple of options for autonomy. The fact that there are different ways in which these systems can be programmed and modified is a contributing factor to disagreements about whether AWS need new international regulations and norms.

One significant concern with AWS is the threat that systems could be hacked or otherwise susceptible to adversarial manipulation, which could cause the AWS to take an unintended action against an unintended target, including, possibly, the original user of the system. Systems need to be designed such that they can be updated as necessary to remain robust against attacks, even as technologies and threats evolve with time. Additionally, with each iteration that a machine learning system goes through, it receives new input from the external world. Depending on how the environment changes and on the feedback the system receives as a result of each action and interaction, its goals could begin to shift from their original intent or training. This can increase the unpredictability and risk associated with the system.

CHALLENGE 4.1

To understand and address the constantly evolving nature of various attack vectors or potential adversarial behaviors such that AWS developers can create systems, which are robust against known adversarial manipulations and hacking, and which can accept updates regarding new threats and new threat mitigations.

CHALLENGE 4.2

To ensure that safety and security, with respect to minimizing harm to humans and other types of collateral damage, take precedence over simply assuring robustness.

CHALLENGE 4.3

To establish standards and legal norms regarding hacking and adversarial manipulation of AWS.

CHALLENGE 4.4

To ensure robustness against distributional or model shift, likely by developing some sort of technical oversight system that can monitor the AWS to ensure that its objectives and actions remain consistent and predictable.

5. Testing and Evaluating

Standards for testing, evaluation, verification, and validation (TEVV) of autonomous weapons systems are necessary to ensure that these systems will act as intended and to ensure international legal requirements are met for weapons reviews.

This reliability of system behavior, combined with a consideration of applicable laws and ethical principles during system design and during the development of the system's concept of operations, allows militaries to employ the system with confidence that actions taken by the system will comply with international laws. However, standards for TEVV for AWS do not currently exist, and will need to be co-developed with early AWS and refined as understanding of AWS increases.

CHALLENGE 5.1

To develop standards and verifiable requirements for AWS, and to include testability throughout the system's lifecycle.

CHALLENGE 5.2

To develop novel methods for assurance that are suitable for autonomous systems, for example new methods of certification and licensing of AWS, and which, in some cases, will need to accommodate changes in the system's capabilities or changes in the system's operational domain.

CHALLENGE 5.3

To design systems in a way that balances transparency and security, to follow a development and deployment process that aggregates evidence of system behavior through the system lifecycle, and which provides maintainers and responsible parties with an accurate representation of system performance, to include incidents and accidents involving the system.

CHALLENGE 5.4

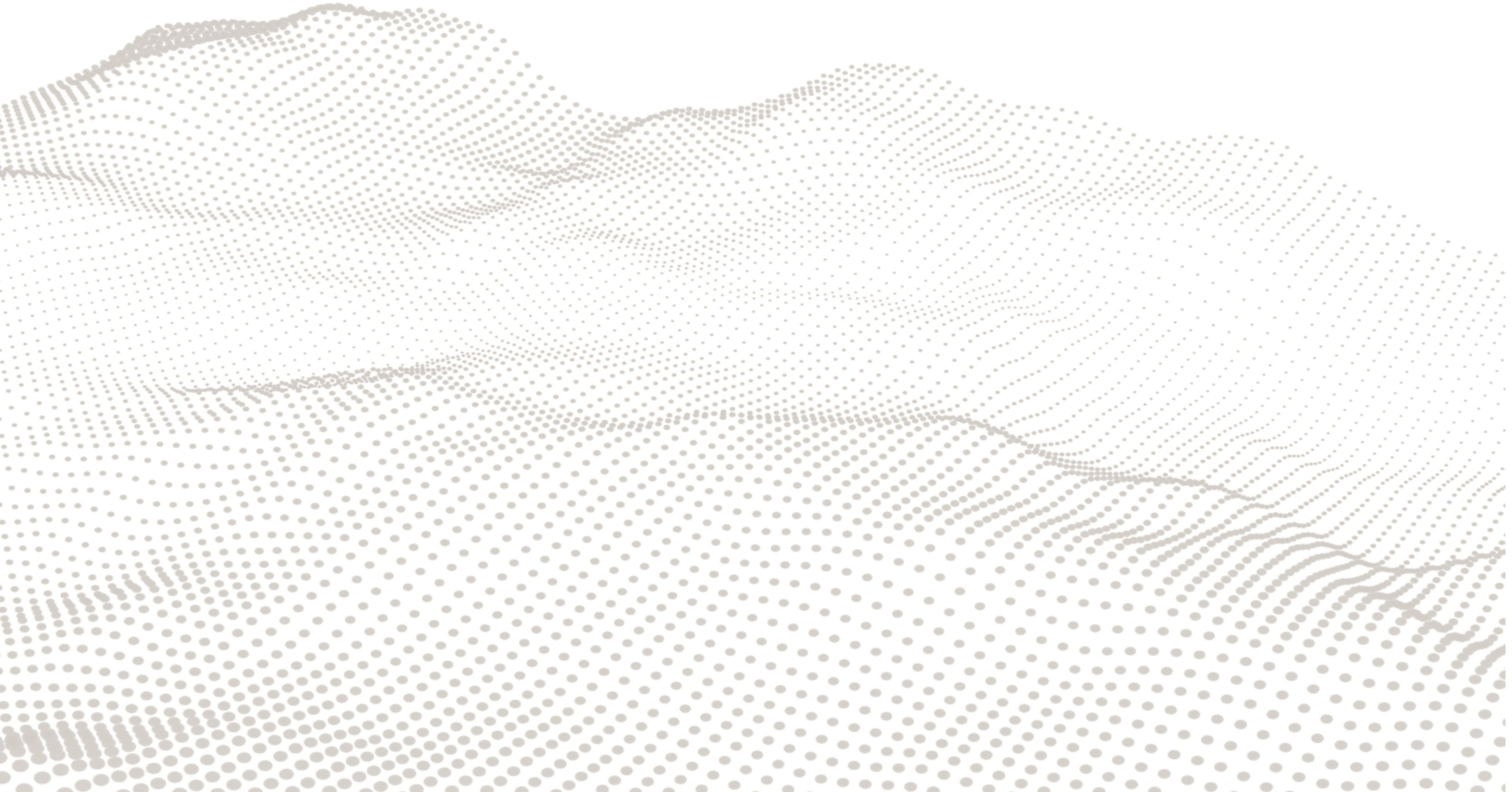
To provide transparency and evidence of behavior of the systems and technical evidence of the real capabilities and potential or observed accidents.

CHALLENGE 5.5

To create appropriate policies and regulations for ensuring testing standards, preferably at an international level.

CHALLENGE 5.6

To establish channels of communication between countries for sharing domestic policies, best practices, and regulations regarding TEVV standards.



6. Assessing Risk

To date, all risk assessments for any weapon deployment are based on the probability of known or estimated likelihoods of unintentional harm or other unintended consequences occurring and on the severity scores associated with those outcomes.

For autonomous weapons that use AI and machine learning, the probability of behaviors may be increasingly difficult to determine, especially given that when AI fails it can fail in unexpected ways, and given the likelihood of unforeseen behaviors occurring due to the chaotic, unpredictable, and potentially undisclosed nature of conflict situations. However, an assessment is a legal requirement in most cases to understand how we can expect AWS to behave amid chaos, conflict, and uncertainty. Transparent mechanisms to enable adequate risk assessment and risk acceptance by human commanders are increasingly necessary.

CHALLENGE 6.1

To develop risk assessment frameworks for categories of AWS, including probabilistic and nondeterministic systems, and to establish appropriate requirements and metrics for evaluation.

CHALLENGE 6.2

To estimate, during design and testing stages, a probability or likelihood of a behavior occurring for a given AWS technology within given situations, and to estimate what the severity of behavior will be.

CHALLENGE 6.3

To develop policy, doctrine, rules of engagement, etc., for AWS, when developers and/or users of a system are unable to provide sufficient risk estimates.

CHALLENGE 6.4

To enable decision makers to estimate the risk of using an AWS, throughout the lifecycle and with respect to relevant legal and ethical considerations.

CHALLENGE 6.5

To ensure that, though situations may be unforeseen, the weapon’s behavior can be controlled and/or predicted.

7. Addressing Operational Constraints

Autonomous weapons will need to maintain greater onboard computing power to support tasks such as navigation, data acquisition and analysis, and decision making, even if they lose communication with operators or mainframe systems.

Meanwhile, in many instances, autonomous weapons could be incredibly small, and they may need to maintain communications between a swarm of systems, again, even if they lose communication with human operators. Additionally, developers of AWS face current uncertainties and unknowns regarding physics-based and computational problems related to size, weight, and power (SWAP) issues of design.

CHALLENGE 7.1

To ensure the system has the technical and physical capabilities to maintain autonomous functionality throughout the mission.

CHALLENGE 7.2

To ensure sufficient computational power for transparency and traceability to track and understand why systems out of communication acted as they did.

CHALLENGE 7.3

To ensure AWS have the capacity to take some sort of autonomous action even if they lose communication with the human overseeing the system.

CHALLENGE 7.4

To ensure components of AWS systems or swarms can communicate even if communication is lost with the human overseeing the system.

CHALLENGE 7.5

To understand the generalizability of algorithms trained on one data set to an environment for which they have not been trained.

8. Collecting and Curating Data

Data collection, analysis, use, and storage pose an increasing number of challenges for all people and organizations that design and employ artificial intelligence.

These problems are exacerbated in autonomous weapons systems. The data may be used as a proxy for harming or killing a person or people or for damaging or destroying infrastructure. Additionally, once the data is collected for a particular system or mission, it could later be repurposed for other applications not originally anticipated.

CHALLENGE 8.1

To securely identify, acquire, classify, use, store, and share high-quality data, and to store data about the system's behavior such that information about the system's behavior can not be manipulated.

CHALLENGE 8.2

To ensure data is acquired, processed, and used in a manner which upholds law and policy.

CHALLENGE 8.3

To ensure that correct and necessary data is available to both the AWS and the human overseeing the system, and that the human operators or commanders have appropriate understanding of the data through user interfaces and knowledge representation.

CHALLENGE 8.4

To balance privacy rights and data privacy and ownership against the needs of the military.

CHALLENGE 8.5

To ensure that the process for collecting and training from data supports appropriate use of the system.

CHALLENGE 8.6

To ensure systems are designed, developed, and deployed in a way that respects different country's laws regarding accessibility to and use of data.

9. Aligning Procurement Practices

For countries in which governments work with private contractors to develop weapons systems, the government contracts typically have specific requirements about what is expected of the contractor.

If the contract is too specific, it could restrict the extent to which a contractor can implement evolving best practices around AI, human-machine teaming, etc.

CHALLENGE 9.1

To update contracting processes to ensure contractors have the legal support they need to implement best AWS practices, even if those are not explicitly stated in the contract.

CHALLENGE 9.2

To recognize that this challenge will vary based on the public-private partnerships that exist in different countries.

CHALLENGE 9.3

To define, during early stages of contracting and development, who holds legal responsibility if AWS don't behave as predicted in the field.

CHALLENGE 9.4

To ensure that, to the extent possible, those involved in development of AWS subcomponents or subsystems recognize how those subcomponents or subsystems fit into the entire AWS.

10. Addressing Non-Military Use

IHL provides the basis for military use of all weapons, but it does not necessarily apply to non-military use.

Discussions within the UN CCW GGE can help ensure IHL is updated as or if necessary to accommodate specific regulations for AWS, but these updates are unlikely to apply to non-military use of AWS unless explicitly stated. However, there are many types of AWS, and some systems will also likely be used by non-military actors, including private security companies, police, border control agencies, non-State armed groups, and more. It's not clear how IHL will apply in those circumstances, given that updates to IHL may not necessarily cover non-military situations. As systems are used by more actors, ensuring that the systems will be used as intended and identifying who is responsible, accountable, or liable becomes trickier. When developing AWS, and if developing regulations regarding AWS, it is essential to consider how they might be used outside of a military setting, how the same challenges above might apply to non-military actors and settings, and what rules or regulations might be appropriate for non-military actors.

CHALLENGE 10.1

To ensure that all challenges listed above are also taken into account for potential non-military use of AWS.

CHALLENGE 10.2

To ensure that AWS regulations that apply to non-military actors and which would be separate from IHL — including police, border control agencies, and private security companies — are similar to, or at least as stringent as, any that are developed within the context of IHL, and to set limits on who can use AWS, as necessary.

CHALLENGE 10.3

To encourage non-military actors to be as publicly transparent as possible about the AWS systems they use, why those systems can be trusted to act as predicted, and under what circumstances the systems will be used.

Thank You