Enemies, Allies, and Artificial Intelligence: Developing Rules and Norms for Lethal Autonomous Weapons Systems

Kiseok Michael Kang

With the advent of artificial intelligence in recent years, countries are developing and testing lethal autonomous weapons systems (LAWS), a class of weapon systems that leverage AI to select and engage targets with little to no human intervention. For decades, the international community has regulated the conduct of hostilities through the laws of armed conflict (LOAC) and other relevant norms. Despite the significant impacts LAWS will have on the battlefields, there are currently no agreed rules or norms for the responsible development and use of LAWS. This paper explores the potential challenges that the lack of common standards in the area of LAWS poses to the United States. It finds that the absence of norms can trigger inadvertent escalation and undermine geopolitical stability in the context of renewed great power competition. Furthermore, the U.S. alliances will face serious interoperability problems should the United States and its allies fail to establish common grounds on LAWS. To address these challenges, the U.S. government should pursue a minilateral, not multilateral, and differentiated approach to building rules and norms on LAWS with its adversaries and allies.

I. Introduction

Artificial Intelligence is becoming “the new fire” that transforms various sectors in our society, including international security.¹ Ben Buchanan and Andrew Imbrie argue that AI, like fire, can play both a constructive and destructive role for human civilization depending on how humanity handles it. AI enables computers and machines to perform tasks that normally require human intelligence.² Many believe that AI, playing a constructive role in international security, will revolutionize warfare as it can make military operations more agile, efficient, and lethal with less manpower. Leading AI expert Kai-Fu Lee, calls LAWS “the third revolution in warfare” following gunpowder and nuclear weapons.³ Intelligence analysis, logistics management, and command and control are among many military functions that can benefit from AI. However, perhaps the most watched and debated development pertaining to military applications of AI is the development and use of lethal autonomous weapons systems (LAWS). Like AI itself, there is no agreed definition of LAWS, but the U.S. Department of Defense (DOD) defines it as “weapon system[s] that, once activated, can select and engage targets without further intervention by a human operator.”⁴ This paper will use the DOD’s definition when referring to LAWS from this point on.

From crossbows to drones, emerging technologies time and again disrupted and shaped the law of armed conflict (LOAC), a set of international rules and norms regulating armed conflict.⁵ LOAC has helped humanity to curb “the evils of war”⁶ and make the conduct of hostilities more humane since the time of ancient Greece and Rome.⁷ Agreeing on common rules is important for both enemies and allies alike because the lack of the rules of the game can increase uncertainty on the battlefields, lead to unnecessary escalation between adversaries, and hamper coalition operations. Emerging and disruptive technologies can shape how militaries fight one another and necessitate new rules of engagement. For instance, governments are trying to establish new norms for cyberspace and outer space, the two newest warfighting domains.⁸ Given the
far-reaching impacts AI will have across the military spectrum and the rapid pace of its development, building rules and norms for LAWS is imperative for states. However, when compared to other emerging technologies, international norms regarding the use of AI-enabled weapons remain still at a very nascent stage. The Group of Governmental Experts (GGE), a United Nations working group that addresses emerging technologies in the area of LAWS, has been the main international vehicle for multilateral norm-building for LAWS since 2014. Although the GGE process has facilitated international discussion on the subject and produced some valuable outcomes such as the 11 guiding principles, it has been slow and stalled in recent years.

This paper examines the potential challenges that the ruleless development and use of LAWS will pose for the United States on the battlefield. The next section reviews the existing laws of war and the current status of international efforts applying these rules to LAWS. Section III discusses the ways in which autonomous weapons systems can spark inadvertent escalation and threaten strategic stability in the context of great power competition. Section IV analyzes the implications of LAWS for alliances: the absence of common rules on LAWS may disrupt the interoperability and weaken the reliability of the U.S. alliance systems such as NATO. Finally, Section V recommends that the U.S. government take a minilateral and differentiated approach to developing basic rules and norms for LAWS. Rather than trying to create one-size-fits-all norms through large multilateral settings such as the GGE, the United States should cooperate with its allies to devise a clear and viable code of conduct, while simultaneously seeking high-level, foundational agreements on norms with adversaries that can minimize potential AI-related risks and facilitate confidence-building measures.

II. Existing Laws of War and Their Applications to LAWS

The modern law of armed conflict, also known as international humanitarian law (IHL) or jus in bello, regulates the conduct of hostilities to protect the vulnerable and alleviate human suffering in war. The Geneva Conventions of 1949 and their Additional Protocols have codified most customary rules and norms of IHL. The core principles of IHL include the prohibition of causing unnecessary suffering, the distinction between combatants and civilians, military necessity, and proportionality. Article 35(2) of the 1977 Additional Protocol I explicitly prohibits weapons that “cause superfluous injury or unnecessary suffering.”

Article 51 of the same Protocol promotes the protection of the civilian population and forbids indiscriminate attacks. Additional Protocol I also requires state parties to conduct a weapons review “in the study, development, acquisition or adoption of a new weapon” to determine whether the weapon in question violates IHL. Lastly, even in the absence of a relevant international law or custom, the principle of humanity and the dictates of public conscience apply to any new weapons systems according to the Martens Clause.

To further cement norms on what kind of weapons should be used in warfare, states adopted the Convention on Certain Conventional Weapons (CCW) in 1980. The purpose of CCW is to prohibit or limit “the use of specific types of weapons that are considered to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately.” Under the framework of CCW, particular weapon types are prohibited, such as projectiles that are not detectable by X-ray, anti-personnel mines, incendiary weapons, and blinding laser weapons. Today, a total of 125 states including the U.S., Russia, and China are party to the CCW. Apart from the CCW, a
separate set of norms and legal frameworks had evolved for weapons of mass destruction. For example, the Chemical and Biological Weapons Conventions have established norms against the use of these weapons in IHL. Similarly, nuclear arms control agreements demonstrate that the major powers have mutual interests when it comes to particularly destructive weapons of war being used. Although many of the Cold War-era nuclear arms control treaties have expired or been terminated, they contributed to the stabilization of strategic competition during the Cold War and led to the conclusion of the New Strategic Arms Reduction Treaty (New START) which is still in force.16

Against this backdrop, the CCW became an appropriate forum to discuss legal and regulatory issues in the area of LAWS. The High Contracting Parties to the CCW decided to establish the aforementioned GGE with a mandate to identify rules and principles applicable to LAWS. Several GGE meetings since 2017 have attempted to achieve a consensus on several key issues pertaining to the use of LAWS. One of the most contentious issues in the GGE has been the question of how to define the technology. Countries differ in how they define LAWS, and the absence of an agreed definition has hindered discussions at the GGE. For instance, the United Kingdom officially opposes the development of LAWS, but its definition of such a system is narrow and futuristic to the extent that it essentially allows the development of “systems with an unacceptably high degree of autonomy.”17 Similarly, China supports a ban on the use of LAWS, but it equivocally defines LAWS as “indiscriminate [and] lethal systems that do not have any human oversight and cannot be terminated.”18

Despite their differences in the definition of LAWS, most countries agree that fully autonomous or “human out of the loop” weapons systems are not compatible with IHL. To understand this debate, it is important to differentiate various levels of human involvement in autonomous systems. “Human in the loop (HITL)” refers to AI-enabled systems that require human intervention during the cycle of observe, orient, decide, and act (OODA). For instance, the U.S. counter-rocket, artillery, and mortar system (C-RAM) identify and track incoming rounds autonomously, but needs humans to make final verification before engagement, thus keeping humans in the loop. “Human on the loop (HOTL)” systems do not require human intervention and can engage targets on their own, but human operators can override the decision made by the algorithm. The Patriot air defense system and Aegis systems are examples of HOTL systems. Both HITL and HOTL systems are considered semi-autonomous systems. On the contrary, “human out of the loop (HOOTL)” systems can search for, detect, decide to engage, and engage targets all on their own without any human interaction.19

The following Table 1 summarizes this.

<table>
<thead>
<tr>
<th>Human in the loop (HITL)</th>
<th>Select targets and deliver force only with a human command</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human on the loop (HOTL)</td>
<td>Select targets and deliver force under the oversight of a human operator who can override the actions</td>
</tr>
<tr>
<td>Human out of the loop (HOOTL)</td>
<td>Capable of selecting targets and delivering force without any human input or interaction</td>
</tr>
</tbody>
</table>

Table 1: Different Levels of Human Involvement20
Drawing a clear line between semi- and the fully autonomous system is challenging since the notion of autonomy is a continuous spectrum and a weapon can consist of multiple components with different levels of autonomy. Hence, many countries highlight the importance of the concept of human-machine interaction—an important factor in deciding whether LAWS can be employed in compliance with IHL. The Campaign to Stop Killer Robots, a coalition of non-governmental organizations led by Human Rights Watch, seeks to ban any autonomous system without “meaningful human control” over the use of force. Germany, France, and many other countries support that human operators must retain control to ensure compliance with the principles of distinction, proportionality, and necessity. On the other hand, the U.S., Israel, and other countries acknowledge the importance of having an appropriate level of human judgment in the use of force but argue that the appropriate level of human-machine interaction should be determined on a case-by-case basis. Israel, in particular, strongly opposes the use of the term “human control” and prefers a more general term like “human-machine interaction.”

Despite vastly differing views of participating countries, the GGE was able to produce 11 guiding principles in 2019. They are not binding norms but can serve as “a useful and valuable starting point to build substance towards a normative and operational framework.” Principle (a) proclaims that IHL continues to apply fully to all weapons systems, including LAWS. Principle (c), which many countries consider to be of primary importance, demands human-machine interaction to ensure that the potential use of LAWS complies with applicable international law, in particular IHL. Although the adoption of 11 guiding principles is a commendable achievement, there have been criticisms of their lack of specificity and enforceability. For example, Principle (a) fails to elaborate on when and how IHL can apply to LAWS in different situations. Will the principle of precautions, which requires a state to take measures to minimize civilian harm, apply only to operators of LAWS, software developers, or data collectors that will enable the algorithm to calculate potential harm to civilians? Similarly, it is unclear how to understand human-machine interaction in Principle (c) given countries’ varying opinions on the appropriate level of human control or judgment over machines. To resolve this problem, the Campaign to Stop Killer Robots and a diverse set of states including Austria, Brazil, China, Mexico, New Zealand, and the Philippines expressed their support for a new legally binding instrument on autonomous weapons systems. Others, particularly the United States and Russia, continue to assert that a new treaty is unnecessary and that existing norms suffice. These countries further argue that LAWS can improve compliance with IHL on the battlefields. For example, AI-enabled platforms can “dramatically improve target identification and discrimination” resulting in fewer unintended civilian casualties. Still, other countries such as Germany and France favor a political declaration over a binding instrument. This “enduring impasse” has paralyzed the GGE process in recent years.

In sum, no international rules currently exist in the areas of LAWS. On the contrary, the multi-year GGE process has demonstrated that there is significant disagreement among countries on how to regulate LAWS. Interestingly, countries are not divided into their usual geopolitical camps on this issue. The United States and Russia advocate a more laissez-faire approach, while some NATO allies such as Germany and France voice concerns over
the unethical development and use of LAWS. In the absence of globally agreed rules, individual countries have sought to write their own rules. The U.S. DOD Directive 3000.09, which lays out U.S. policy on LAWS, is a notable example. Other countries like Russia and China have adopted a more general AI strategy that includes some military implications. In the absence of mediating ground rules, great power competition involving LAWS could lead to dangerous escalation.

III. LAWS and the Strategic Competition

The absence of common norms not only creates legal and ethical challenges but also can lead to practical problems on the battlefield. Great powers, especially, need to agree on basic rules of the road to avert inadvertent escalation and maintain strategic stability. LAWS is no longer science fiction. The 2017 SIPRI report identified 49 military systems from different nations that could engage targets autonomously “without the direct involvement of a human operator,” meaning HOTL or HOOTL. Some systems such as Israel’s Harpy, a loitering munition that can self-target radar emitters, and U.S. Tomahawk Anti-Ship Missile (TASM) have been fielded for years. As the SIPRI report points out, some of these systems are “automated” rather than “autonomous” as they carry out tasks through predefined procedures. However, with the advance of AI technology, more systems can perform sophisticated decision-making processes and cope with uncertain and complex environments. In this context, militaries around the world are pouring billions of dollars to build more capable LAWS that are not automated weapons but truly autonomous. Both the U.S. Third-Offset Strategy and the 2022 National Defense Strategy underscore the importance of leveraging AI and robotics in fighting and winning the wars of the future. The U.S. military and its industry partners are experimenting and developing a wide range of autonomous weapons systems that include AI-powered flying, underwater, surface, and ground vehicles with lethal capabilities. Some believe the widespread deployment of such systems is still decades away, while others believe that there is a real possibility of deployment within 10-15 years. Russia already deployed and test-drove its robotic tank known as the Uran-9 on the actual battlefield of the Syrian war. Moscow is eager to contend for superiority in AI, which can make Vladimir Putin “the ruler of the world” according to his own words. Besides the Uran-9, Russia is working on other autonomous systems such as self-destructing kamikaze drones. China is likewise aggressively developing a variety of AI-enabled weapon systems in order to counterbalance “technologically advanced adversaries” and achieve “battlefield singularity.”

There will be significant strategic risks should the great powers begin to field LAWS without agreeing to ground rules for the responsible development and use of such systems. The “wait-and-see” approach advocated by the United States and Russia is particularly dangerous. First, LAWS can spark unintended escalation due to their inherent fallibility. Intensified competition can incentivize militaries to deploy LAWS that are fallible and more prone to inadvertently violating LOAC. The fear that the enemies are “operating at machine speed and we’re operating at human speeds” can also make militaries more risk-acceptant. Although the U.S. military promised a responsible, reliable, and governable approach to AI-powered weapons systems and a thorough review process, there is suspicion within the United States that its adversaries will not abide by the same moral codes. But this distrust could be mutual. Since there are no detailed rules regarding the appropriate level of human control, any government could delegate authority to machines.
largely unchecked. A 2020 RAND wargame concluded that in this situation, an autonomous system might employ force without human approval and intention.\textsuperscript{46} Given that LAWS will be trained on data from a controlled environment, the probability of these technologies behaving in an unexpected situation is high.

Second, the lawless development of LAWS could threaten geopolitical stability. This is because LAWS enable governments to pursue more risky operations without worrying about the cost of human lives.\textsuperscript{47} The frequent use of drone strikes today shows how governments might use more capable autonomous weapons to remove high-value targets in the future.\textsuperscript{48} China has already been employing unmanned and autonomous assets to probe contested areas in the South and the East China Sea.\textsuperscript{49} Even inadvertent non-lethal actions by poorly-designed LAWS such as crossing the border could lead to unintended escalation. Widespread use of LAWS will also lower the cost of war in general. The LAWS might in fact weaken deterrence and stability because adversaries would know that striking the opponent’s LAWS, typically unmanned, is less likely to trigger a wider conflict than attacking manned assets.\textsuperscript{50}

Lastly, great powers may seek to incorporate AI into the strategic command, control, and communication systems to improve the speed and precision of their nuclear arsenal. This will shorten the required response time and increase pressure on decision-making in a crisis to the point where human commanders and operators can be sidelined. AI-powered Intelligence, Surveillance, and Reconnaissance (ISR) capabilities coupled with autonomous missile defense systems can also seriously undermine strategic equilibrium. Although this looks like a bridge too far, with the development of hypersonic missiles, an AI-enabled missile defense system is becoming a necessity, not a luxury.\textsuperscript{51} This danger of AI potentially undermining strategic stability is a direct consequence of an absence of regulations limiting its use in warfare.

IV. LAWS and Alliances

Russia’s massive cyberattacks on Estonia in 2007 perplexed NATO leaders as there was no code of conduct for the alliance on the conduct of cyber warfare. The alliance lacked consensus on how to respond to non-kinetic cyberattacks, let alone whether a cyberattack constitutes “an armed attack” that can trigger Article 5. In the 2012 Wales Summit, NATO drafted the \emph{Tallinn Manual on the International Law Applicable to Cyber Warfare}, affirming that cyberattacks such as the one against Estonia could lead to the invocation of Article 5.\textsuperscript{52} This case illustrates how military alliances can lag behind military innovations—and LAWS are no exception.

Without common standards across the alliance, NATO will face considerable interoperability challenges concerning LAWS. NATO recently adopted an AI strategy in 2021, but the strategy only prescribes broad principles and lacks substance.\textsuperscript{53} As discussed above, NATO allies differ in their views on the ethical use of LAWS. For example, Spain maintains that HITL is only acceptable for strictly defensive systems such as the Iron Dome, while offensive weapons should always include HITL.\textsuperscript{54} Portugal suggests time and space restrictions on LAWS to ensure compliance with IHL.\textsuperscript{55} When allies disagree on \emph{jus in bello}, coalition operations can falter. During the Kosovo War, the United States and France had a considerable disagreement over the legality of bombing a civilian radio station that disseminates the regime’s propaganda.\textsuperscript{56} The discord delayed the operation for days. Similarly, the Trump administration’s reversal of the landmine ban policy troubled NATO operations as all other NATO members banned the use of anti-personnel mines under the Ottawa
The United States and other NATO member states, thus, need to devise a common policy position on the legal and ethical aspects of LAWS to avoid any interoperability problems on the battlefield.

Putting the legal and ethical issues of LAWS aside, NATO still has other pending issues for the alliance’s interoperability with LAWS. First, the free flow of data among the allies will be critical for the success of NATO’s AI strategy, but it is unclear how much data national governments will readily share, as there are strategic as well as privacy concerns associated with data sharing. While NATO established the Data Exploitation Framework Policy to facilitate data sharing among the allies, it still needs to set more detailed rules. Second, NATO should address the technology gap between the advanced and the less advanced. If the technologically advanced allies start to field LAWS on the battlefield while the less advanced continue to operate legacy systems and large infantry, a range of practical and ethical problems will arise. Third, NATO’s former Warsaw Pact states still operate the Soviet-era equipment which can be problematic for automated target recognition whether it uses visual or heat signature to distinguish targets. Fratricide can be a real danger to the alliance’s use of LAWS. For example, during the Iraq War, the U.S. Patriot air defense system shot down a British aircraft when the automated system misidentified the aircraft as an anti-radiation missile and the human operator accepted the system’s incorrect identification.

The introduction of LAWS into the alliance will also complicate the reliability of U.S. defense commitment to its allies. Historically, U.S. soldiers in Europe and East Asia have acted as “tripwires” to deter adversaries. However, maintaining a large military presence has become politically and economically costly for the U.S. government. By enhancing the alliance’s military capability and reducing personnel costs at the same time, the deployment of LAWS provides an attractive option for Washington. The allies, nonetheless, could interpret that as a sign of reduced security commitment by the United States. The United States also needs to reassure its allies that the deployment of LAWS does not necessarily reduce the threshold of war. Furthermore, should more U.S. strategic assets become AI-integrated and machine-controlled, the allies will demand to know more about what algorithms and data are used for these systems. In other words, the development of LAWS might lead U.S. allies to question the credibility of the American nuclear umbrella.

The highly institutionalized NATO is better positioned to address some of these challenges than other U.S. allies. As discussed above, NATO has already begun to discuss AI-related issues internally and made some progress such as the adoption of the NATO AI strategy and Data Exploitation Framework. In contrast, the United States lacks similar arrangements with its bilateral allies such as Japan, South Korea, and Australia, although their position on the legality of LAWS is more aligned with the United States than their European counterparts. South Korea, for instance, has already deployed autonomous sentry guns along the demilitarized zone and Australia explicitly opposed any premature ban on LAWS. Since these American allies do not have formal alliance relations with one another, data sharing for the Indo-Pacific theater could be particularly challenging. Also, several important actors in the region like India and Vietnam do not belong to either block. The multiplicity of actors will increase the complexity of the operating environment for LAWS and thus render unanticipated behaviors more likely.
V. Policy Recommendations

With the ongoing war in Ukraine and the atrocities committed there, the United States has a strong incentive to develop rules and norms regulating LAWS vis-à-vis enemies and allies. Moscow allegedly fielded “killer robots” in Ukraine, while Kyiv also employed Turkish kamikaze drones that have some autonomous capabilities. Future wars, in which the belligerents might seek to employ LAWS to win, could be even more catastrophic and inhumane if LAWS continues to be unregulated.

It took two devastating world wars for humanity to universally adopt the laws of war. After the atomic bombs were dropped on two Japanese cities and the Cuban Missile Crisis brought the world to the precipice of a nuclear war, the superpowers agreed to establish bilateral nuclear arms control and global non-proliferation regimes. The international community should learn from history and it still has a chance to prevent an AI-led world war. With the most advanced military and global interests, the U.S. has an imperative not only to develop bleeding-edge AI technologies but also to build international rules and norms that can create more predictable and secure environments for these systems to operate. As such, I recommend the following two approaches to the U.S. government.

1. Minilateral Approach

The United States should take a minilateral, not multilateral, approach, which could expedite the process of drafting rules and norms pertaining to LAWS. While technology is advancing at a rapid pace, international efforts to build common standards for technology have been lagging. The GGE, the main international forum on the subject, struggles to move forward because of the diverging interests of participants. This is in part unavoidable since it includes a very broad group of stakeholders and adopts the final reports on a consensus basis. The most recent GGE on LAWS meeting was plagued by the conflict between Russia and Ukraine.

Given the arduous and protracted nature of international negotiation on regulating LAWS, it is wise to pursue minilateralism, a more focused approach with a small number of countries. Such an approach is more efficient since like-minded states can initially agree to basic rules that are easy to negotiate and build upon. The Tallinn Manual, produced by NATO allies, is a good example. The international community failed to come up with binding rules on cyber warfare even after decades of negotiations, but the NATO-led Tallinn Manual, although non-binding, has provided a basis for global norms on cyber war. Similarly, the World Trade Organization (WTO), which champions consensus-based decision-making, failed to produce any meaningful trade agreements since 2000 due to the profound disagreement among its member states. Another valuable example comes from the “plurilateral trade agreements,” which allow a subset of WTO member states to negotiate and conclude a mini trade agreement as a solution to inefficient and inoperative multilateral negotiations.

2. Differentiated Approach

Since Washington has different sets of concerns for its adversaries and allies when it comes to deploying LAWS on the battlefield, it should build different sets of rules. As for its near-peer competitors such as Russia and China, the United States should focus on finding the lowest common denominator: the rules for risk mitigation and management. Although great powers wish to realize the full potential of LAWS, they nonetheless can agree on basic safeguard rules to avoid inadvertent escalation and geopolitical instability. Some skeptics may argue that it is infeasible to produce any meaningful and satisfactory agreement on LAWS. They will further point out that, “the more
useful a technology is at providing armies
with an edge, the harder it is to effectively
regulate.\textsuperscript{66,7} The international community
was mainly successful in restraining the
use of weapons that do not decide the
outcome of war, such as blinding lasers.
Nonetheless, states had successfully
negotiated many important and complex
issues in the past when they shared
a sufficient common interest. To avoid
nuclear catastrophe, superpowers
voluntarily limited the number of
warheads and missiles each country could
possess.

The rules of the road can include
limiting the incorporation of AI into
nuclear control and command systems and
enhancing transparency about the
development and deployment of LAWS.
Great powers can also agree on “off-limit”
geographical areas for LAWS that can
include contested areas around the world.\textsuperscript{68}
Alternatively, the great powers can jointly
prohibit the use of LAWS in densely
populated areas. Russia’s flagrant violation
of IHL in the Russia-Ukraine war casts
doubts on whether powerful countries will
abide by these rules and norms. Indeed, the
presence of norms does not translate into
compliance.\textsuperscript{69} However, the existence of
rules and norms makes it possible to
distinguish acceptable behavior from
unacceptable and provides a basis for
collective action against one who crosses a
line. Thus, agreed rules and norms can
have a deterrent effect on others.\textsuperscript{70}

At the same time, the United
States should work with allies to build
more granular rules on LAWS. It should
facilitate intra-alliance dialogue on the
development and use of LAWS. The
United States and its allies can bring
experts, scholars, and practitioners
together to explore possible rules and
norms applicable to LAWS as they did for
the Tallinn Manual. Building consensus
among like-minded countries can be less
laborious. The agreed inter-alliance
standards could be then scaled up. The
United States employed a similar approach
in the cyber and space domain. Relatedly,
in the latest GGE meeting, the United
States and its like-minded allies circulated
“Principles and Good Practices on
Emerging Technologies in the Area of
Lethal Autonomous Weapons Systems.”\textsuperscript{71}
In addition, the United States and its allies
should start discussing both the tactical
and strategic implications of LAWS. At
the tactical level, the allies should identify
and explore potential interoperability
problems. At the strategic level, the allies
should engage in high-level political
dialogue on the impacts of LAWS on
extended deterrence and regional stability.

As the Biden administration’s
National Security Strategy rightly asserts,
the United States has led the creation of
global institutions, norms, and standards
that “advanced America’s economic and
geopolitical aims” as well as benefited
people around the world.\textsuperscript{72} Playing its
traditional role of norm setter, the United
States should take efforts to shape global
norms on LAWS in a way that enhances
the existing LOAC and bolster strategic
stability, so that humanity can avert
another “failure of civilizations.”\textsuperscript{73}

About the Author: Kiseok Michael Kang
is a second-year M.A. student in the
Security Studies Program, concentrating
on technology and security. Before coming
to Georgetown, Michael served as a
foreign service officer at the Ministry of
Foreign Affairs of the Republic of Korea.
APPENDIX I

Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System

(a) International humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems;

(b) Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system;

(c) Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context and the characteristics and capabilities of the weapons system as a whole;

(d) Accountability for developing, deploying, and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control;

(e) In accordance with States’ obligations under international law, in the study, development, acquisition, or adoption of a new weapon, means, or method of warfare, determination must be made whether its employment would, in some or all circumstances, be prohibited by international law;

(f) When developing or acquiring new weapons systems based on emerging technologies in the area of lethal autonomous weapons systems, physical security, appropriate non-physical safeguards (including cyber-security against hacking or data spoofing), the risk of acquisition by terrorist groups, and the risk of proliferation should be considered;

(g) Risk assessments and mitigation measures should be part of the design, development, testing and deployment cycle of emerging technologies in any weapons systems;

(h) Consideration should be given to the use of emerging technologies in the area of lethal autonomous weapons systems in upholding compliance with IHL and other applicable international legal obligations;

(i) In crafting potential policy measures, emerging technologies in the area of lethal autonomous weapons systems should not be anthropomorphized;

(j) Discussions and any potential policy measures taken within the context of the CCW should not hamper progress in or access to peaceful uses of intelligent autonomous technologies;

(k) The CCW offers an appropriate framework for dealing with the issue of emerging technologies in the area of lethal autonomous weapons systems within the context of the objectives and purposes of the Convention, which seeks to strike a balance between military necessity and humanitarian considerations.
APPENDIX II

AI Ethical Guidelines Recommended by Defense Innovation Board

1. Responsible: DoD 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 so that staffers have an appropriate understanding of the technology, development processes, and operational methods that apply to AI. This includes 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.

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.
6 The preamble of the Hague Convention of 1907 states that “[t]he provision, the wording of which has been inspired by the desire to diminish the evils of war, as far as military requirements permit, are intended to serve as a general rule of conduct for the belligerents in their mutual relations and in their relations with the inhabitants.”
8 For cyber norms, the U.N. Group of Governmental Experts had been established since 2004. The Convention on Cybercrime also entered into force in 2004. For outer space, the international space law regime had been established since the late 1960s and the U.N. Committee on the Peaceful Uses of Outer Space (COPUOS) has been one of the main international bodies for emerging space issues.
10 See Appendix I for the complete list of 11 guiding principles.
12 Ibid.
13 Ibid.
14 The Martens Clause affirms that “in cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience.”
19 Scharre, *Army of None*, 47.
24 Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico, and New Zealand, “Joint ‘Commentary’ on Guiding Principles A, B, C and D,” September 2020, https://documents.unoda.org/wp-
It is noteworthy to mention that drones are fundamentally different from LAWS because they are remotely piloted and thus do not autonomously select or engage targets. (Santa Monica, CA: RAND Corporation, 2020).  

Lewis, “An Enduring Impasse on Autonomous Weapons.”  


Scharre classifies these weapons systems as “fully autonomous weapon systems.” See Scharre, Army of None, 47. But, Work labels them as “bounded search weapons,” a subset of semi-autonomous weapons. See Work, Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities, 7.  


To illustrate the difference between automation and autonomy, one can compare cruise control and self-driving car. The former helps a driver with a narrow and preprogrammed task such as maintaining speed at a certain level and a driver has to retain control while its vehicle is on cruise control. An autonomous vehicle, on the contrary, can drive itself and respond to various unexpected situations on its own. See Toby McClean, “The Path From Automation To Autonomy Is Swarming With Activity,” Forbes, April 1, 2021. https://www.forbes.com/sites/forbestechcouncil/2021/04/01/the-path-from-automation-to-autonomy-is-swarming-with-activity/  

Governments already spend around $11 billion annually on LAWS according to some estimates. See Buchanan and Imbrie, The New Fire, 137.  

The current ongoing projects include, but are not limited to, Skyborg (autonomous drone), Orca (underwater), Sea Hunter (surface), and Ripsaw M5 (ground).  

Forrest E. Morgan, Benjamin Boudreaux, Andrew J. Lohn, Mark Ashby, Christian Curriden, Kelly Klima, and Derek Grossman, Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World (Santa Monica, CA: RAND Corporation, 2020), xvi.  

The system is known as KUB-BLA and it is claimed that the drones are used in its invasion of Ukraine. See https://www.thenationalnews.com/world/uk-news/2022/03/14/russias-kub-bla-kamikaze-drone-intercepted-in-ukraine/.  

Military Applications of Artificial Intelligence, 60.  


Buchanan and Imbrie, The New Fire, 149.  

Yuna Huh Wong, John Yurchak, Robert W. Button, Aaron B. Frank, Burgess Laird, Osonde A. Osoba, Randall Steeb, Benjamin N. Harris, and Sebastian Joon Bae, Deterrence in the Age of Thinking Machines. (Santa Monica, CA: RAND Corporation, 2020).  

Singer, Wired For War, 319-323.  

It is noteworthy to mention that drones are fundamentally different from LAWS because they are remotely piloted and thus do not autonomously select or engage targets.  

Military Applications of Artificial Intelligence, 65.
Again, the drone warfare can provide a glimpse into what the future LAWS warfare will look like. Amy Zagart argues that the downing of drones would not spark war as Iran’s shoot down of the Global Hawk did not provoke the U.S. military response. See Amy Zagart, “Drones Don’t Spark War, People Do,” *The Atlantic*, June 21, 2019.


“Chairperson’s Summary,” 79.
