Chapter 23

Are We Ready: Regulating Military Uses of AI

By James E. Baker

I. INTRODUCTION

One way to maximize the promise of AI and to mitigate the challenges is through the meaningful application of policy, law, and ethics. In U.S. national security practice, policy to date has largely been driven by the Department of Defense (DOD). At least DOD has most transparently shaped U.S. government AI policy. Section I of this chapter briefly considers the DOD policy response to AI, largely in the context of what is known as the Third-Offset Strategy and DOD regulations. Section II then considers the role of law in regulating military uses of AI. In brief, national security law serves three purposes: (1) It provides the authority to act and the boundaries of that action; (2) essential process; and, (3) security and legal values, which should inform the evaluation of lawful policy choices.

There is a tendency to think of the military use of AI from an operational and weapons point of view. But regulation of weapons and technology starts with the research and development of AI applications, not deployment. And, as the Third-Offset Strategy illustrates, the use of AI is contemplated across the defense disciplines and applications. Therefore, Section III of this chapter considers existing authorities to regulate the defense industrial and technological base. In particular, the Section considers the Defense Production Act (DPA) and the Inventions Secrecy Act (ISA), two legislative tools that could be used to harness industry and academic R&D to military needs.

Section IV then considers how the law of armed conflict applies, or might apply to the military use of AI within defense systems and weapons. It does so by exploring six questions and principles.

1. The Duty to Test New Weapons for compliance with Law.
2. The Duty to Train personnel in the Law.
5. Prohibitions and Redlines applicable to Weapons Use.
6. The distinction between lawful Ruses and Unlawful Perfidy.

II. DOD POLICY RESPONSE

"U.S. military deterrence plans have historically been driven by offset strategies. An 'offset strategy' does not try to match strength for strength, but instead seeks ways to offset competitor’s advantages." 1 The United States adopted two offset strategies during the Cold War to address real and perceived Soviet advantages in personnel and battlefield hardware as well as eventual Soviet parity in the number of ICBMs. Each was based on the strength and advantages of the U.S. defense industrial base. As described by the Department of Defense:

The first offset strategy occurred in the 1950’s as the Soviet Union reached parity with the United States on conventional weapons. The United States turned to tactical nuclear weapons for conventional deterrence. The second offset occurred in the 70’s and 80’s as strategic nuclear parity was reached, and the United States turned to [sic] its focus to building an advantage in conventional guided munitions. 2

As indicated, these offset strategies were directed at the Soviet Union in Cold War context. The U.S. response came in the form of technological advantages in developing and producing weapons, specifically artillery shells, mines, and missiles in the 1950s, and precision guided munitions in the 1970s and 1980s.

In 2016, the United States adopted a Third-Offset Strategy. It is different than the previous strategies in at least six ways, each of which will test the scope of the government’s legal authority to regulate and access technology.

First, the Third-Offset Strategy relies on technology to do more than enhance weapons and munitions. It is intended to augment human decision, intelligence capacity, and communications.

Second, it relies on a commercial sector that reaches beyond traditional defense contractors and is global in form. Thus, not all the same assumptions about mission alignment and mission compatibility necessarily apply between the Department of Defense and Silicon Valley as once applied to traditional defense contractors.

Third, it is intended to offset the strengths of a variety of potential adversaries, both state and non-state, as opposed to a single adversary.

---

Fourth, it is directed not just to the perceived numeric advantages of potential opponents, like Soviet conventional forces, but to potential opponents’ advantages in asymmetric warfare.

Fifth, AI will play a critical role in the third offset strategy. The Department of Defense has said so.

The initial third offset technology vector is to exploit advances in artificial intelligence and autonomy and insert them into DoD’s battle networks to achieve a step increase in performance. Other areas include assisted human operations (wearables), human-machine combat teaming (unmanned and manned equipment working together), and network enabled, cyber hardened weapons (cyber security).³

Sixth, and finally, this offset will spill over into many aspects of national security beyond traditional defense areas, including homeland security, grand strategy, and national security decision-making. It already has.

To achieve this third strategic offset, the U.S. government will rely on four national security assets: (1) an unmatched national technical intelligence capacity; (2) a robust university educational and research system; (3) the presence of much of the advanced information technology industry in the United States; and, (4) an established defense industrial base. Heretofore, to meld and utilize these assets, the U.S. government has principally relied on the legislative tools Congress has provided with respect to the Defense Industrial Base (DIB). However, these tools do not provide a cohesive legal framework. Neither were they drafted with AI and Silicon Valley in mind. Therefore, it is not a foregone conclusion that the U.S. will sustain these advantages or use them wisely and effectively. A more coherent strategy to regulate national security AI should start with an understanding of current law, in particular, the President’s most important authority, the authority to regulate the DIB, the meaningful and purposeful application of the law of armed conflict, and customary and treaty based norms.

III. THE THREE PURPOSES OF NATIONAL SECURITY LAW

National security law serves three purposes, or rather it can serve three purposes if crafted well and wielded wisely. First, it provides authority to act as well as the left and right boundaries of that action. In the AI sphere, authority to act is found in the Constitution. Constitutional authority for Congress to regulate the national technology and industrial base is found in the Commerce Clause, the Rules Clause, and the Necessary and Proper Clause, among other places. As Youngstown explored, the President’s constitutional authority derives from his role as Commander in Chief, Chief Executive, and the take care clause, among others. When it comes to the

DIB, the primary law the President is taking care to faithfully execute is the DPA. Chapter 148, of Title 10, is also material. As is well known from the *Youngstown* paradigm, where the President acts pursuant to both his constitutional authority and the authority provided in Title 10 and the DPA, he will be acting at the zenith of his authority. The boundaries of action are defined by these same statutes as well as the First, Fourth, and Fifth Amendments to the Constitution, and as discussed below the LOAC, much of which is applicable through domestic implementation as well as international application. Left and right boundaries might also be established through process rather than substance, requiring for example, a certain type of approval before an action can be taken, rather than prohibiting that action altogether.

Second, law can provide essential process. Process is “essential” when it addresses needs like unity of command and purpose, or establishes legal and democratic accountability. Process is also essential when it mitigates pathologies of national security decision-making, like too little time, too much secrecy, incomplete facts, and cognitive bias. To be sure, process can be both good and bad. Good process is timely, contextual, and meaningful. With regard to military applications of AI, process is essential to determine when and in what manner human control will be asserted and in what manner responsibility will be identified and accountability upheld. To some, this is known as the “man-in-the-loop” question, or more correctly, “the person in the loop,” as we should not assume the designer, operator, or user of a system will, in fact be a man. Here, of course, the question is not whether, but where and in what manner, for so long as there is software or hardware involved there is a human making willful, and we hope, purposeful choices about how technology will be used and subject to what controls. The question is will the person be in, on, or out of the loop when key operational decisions are actually made.

Third, the law provides national security as well as legal values. Here too, the source of many of these values is the Constitution, which is designed to provide for both our physical security and preserve and protect our liberty. When and whether to give greater weight to one or the other of these values, or perhaps to find balance, is often a question of choice and of process. The essential points here are, first, that legal values often present security virtues and benefits as well. For example, the concept of checks and balances found in the separation of powers expresses a legal value about constitutional government and power, but it also bears a security virtue. Where the President acts pursuant to his own authority as well as that provided by or delegated by the Congress, he acts at the zenith of his power, and if one is confronting a significant national security challenge that is a good thing. Likewise, the doctrine of command responsibility is a legal concept intended to fix responsibility and provide, if necessary, criminal accountability for the conduct of hostilities. However, a clear and accountable chain of command is also a military necessity. To effectively maneuver and fight a commander must wield effective control over his subordinates and subordinate units.

Second, where there is choice as to which values are weighted and how, policymakers and lawyers should make conscious, purposeful, and honest decisions about those value choices as well as their short and long term consequences.
Third, when, whether, and how process is purposefully applied will depend in part on what legal and national security values are privileged. Process is substance when it dictates outcome.

**IV. AUTHORITY OVER THE DEFENSE INDUSTRIAL & TECHNOLOGICAL BASE**

**A. Title 10, Chapter 148**

The phrase Defense Industrial Base (DIB) is derivative of the Cold War and connotes images of military hardware, like tanks and aircraft, as well as material resources like steel. However, the term is, in fact and law, broadly defined. Congress has amended the enabling statutes to include “national technology” and “critical technology” within the scope of the DIB. Chapter 148, of Title 10, the United States Code section covering the Department of Defense and the Armed Forces, states:

> [T]he term “national technology and industrial base” means the persons and organizations that are engaged in research, development, production, integration, services, or information technology activities conducted within the United States and Canada. [10 U.S.C. 2500(1).]

Chapter 148 helps to set congressional policy regarding the DIB and reinforces authority already found in the DPA. Chapter 148, for example, states unequivocally that

The Secretary of Defense shall develop a national security strategy for the national technology and industrial base. Such strategy shall be based on a prioritized assessment of risks and challenges to the defense supply chain and shall ensure that the national technology and industrial base is capable of achieving the following national security objectives:

1. Supplying, equipping, and supporting the force structure of the armed forces that is necessary to achieve –
   - The objectives set forth in the national strategy report submitted to Congress by the President pursuant to section 108 of the National Security Act of 1947 (50 U.S.C. 404a);
   - …
   - …
2. …
3. Maintaining advanced research and development activities to provide the armed forces with systems capable of ensuring technological superiority over potential adversaries.
4. …
(5) Providing for the development, manufacture, and supply of items and technologies critical to the production and sustainment of advanced military weapons systems within the national technology and industrial base.

(6) …

(7) Providing for the development, production, and integration of information technology within the national technology and industrial base.

(8) …

From an Executive and legal perspective, such statements of congressional policy intent are usually viewed as persuasive, but not as positive enabling authority. However, this statement of congressional policy is clear: The Secretary shall make purposeful and strategic decisions to integrate information technology, including AI, into the defense industrial base and U.S. national security policy.

Chapter 148 provides positive enabling authority to do so. Section 2507, for example, authorizes the President to collect data to administer and enforce Chapter 148. Specifically:

The President shall be entitled, by regulation, subpoena, or otherwise, to obtain such information from, require such reports and keeping of such records by, make such inspection of books, records, and other writings, premises or property of, and take sworn testimony of, and administer oaths and affirmations to, any person as may be necessary or appropriate, in the President's discretion, to the enforcement or the administration of this chapter and the regulations issued under this chapter.

While expansive, this section is subject to scope and exhaustion limitations. First, the President is required to issue regulations “insuring that the authority of this section will be used only after the scope and purpose of the investigation, inspection, or inquiry to be made have been defined by competent authority.” Second, the regulations shall further provide that the data collection authority will only be used if it is “assured that no adequate and authoritative data are available from any Federal or other responsible agency.” Title VII of the DPA provides similar authority. This section is noteworthy because it duplicates and reinforces the Congressional grant of authority in the DPA and because it provides limitations to that authority.

B. Defense Production Act

The Defense Production Act, 50 U.S.C. 2061-2171, is the principal statutory authority available to the executive branch to “shape defense preparedness programs and to take appropriate steps to maintain and enhance the domestic industrial base.” [2062(a)(4).] It provides seven core authorities:
(1) Authority to prioritize contracts and allocate resources;
(2) Authority to incentivize research, development, and production capacity in support of national defense and homeland security; and,
(3) Authority to gather information to uphold and administer the DPA and to assess the U.S. industrial base to support national defense.
(4) Enabling authority and operating procedures for the Committee on Foreign Investment in the United States.
(5) Anti-trust protections for certain defense industry programs and practices.
(6) Personnel authorities to include waiver authorities and enabling authority for a National Defense Executive Reserve.
(7) Judicial jurisdiction and authority to compel compliance with the DPA and to regulate its proper use consistent with constitutional protections.

The DPA was passed in 1950 during the Korean War (and Cold War) with the purpose of providing a mechanism to ensure the “supply [of] materials and services for the national defense and to prepare for and respond to military conflicts.” This declaration of policy was amended to include within its statement of security purpose, “natural or man-caused disasters, or acts of terrorism within the United States.” The Act was also amended to reach energy supplies, emergency preparedness, and critical infrastructure, in other words, to reach beyond traditional Cold War industrial base disciplines.

Since original passage, the DPA has included a five-year sunset clause for a majority, but not all, of its provisions. (At times the reauthorization of the DPA has sunset on a shorter timeline, for example, when the Act was continued to accommodate a later congressional debate and extension.) The Act has been reauthorized over fifty times. Thus, there is—by necessity—ample opportunity to amend or adjust the DPA to address AI applications or ambiguities.

The DPA defines domestic industrial base as:

Domestic sources which are providing, or which would be reasonably expected to provide, materials or services to meet national defense requirements during peacetime, national emergency, or war.

National defense, in turn, is defined as:

Programs for military and energy production or construction, military or critical infrastructure assistance to any foreign nation, homeland security, stockpiling, space, and any directly related activity. Such term includes emergency preparedness activities conducted pursuant to title VI of the Robert T. Stafford Disaster Relief and Emergency Assistance Act [42 U.S.C. 5195 et seq.] and critical infrastructure protection and restoration.

Critical infrastructure, in turn, means:

…any systems and assets, whether physical or cyber-based, so vital to the United States that the degradation or destruction of such systems and assets would
have a debilitating impact on national security, including, but not limited to, national economic security and national public health and safety. (Emphasis supplied throughout)

As this definitional daisy chain indicates, the DPA is broad in scope. The definitional language also clearly reaches AI, if those who wield its authority wish it to reach AI. Further, the Act applies in peacetime as well as periods of national emergency and war.

The Act also has teeth. It carries the potential of civil and criminal sanction (not more than one year of confinement and up to $10,000 fine) for willfully failing to perform a required act or an act prohibited by the statute, or willfully disclosing information the President deems is confidential under the Act. [705(c)(d).] It also provides an enforcement forum, by granting jurisdiction to U.S. district courts over “violations of this Act or any rule, regulation, order, or subpoena thereunder, and of all civil actions under this Act to enforce any liability or duty created by, or to enjoin any violation of, this Act or any rule, regulation, order, or subpoena thereunder.” [706(b)] The Act further provides and contemplates that federal courts will uphold the Act through use of their contempt authority. [705(a)].

1. DPA Process

Application of Title I of the DPA involves five layers of law and regulation: The Constitution, the DPA, Executive order, DPA Implementing Regulations, and departmental and agency internal regulations. The system is implemented using what is known as the Defense Priorities Allocation System (DPAS). In general, DPAS is administered by the Department of Commerce (DOC), Bureau of Industry and Security. However, DOC has delegated many of the DPAS authorities to other agencies. In addition, the DPA in some instances requires certain process, such as presidential determinations with respect to Section 303. And, of course, the President’s own direction takes legal precedence over Agency direction.

The President has delegated certain of his DPA authorities pursuant to Executive Order 13603. First, he has delegated his 101(a) authorities to six cabinet secretaries with respect to six specific areas. However, exercise of this delegated authority “may be used only to support programs that have been determined in writing as necessary or appropriate to promote the national defense” as determined by the Secretaries of Defense, Secretary of Energy, and Secretary of Homeland Security. Specifically:

…the authority delegated by Section 201 of this order may be used only to support programs that have been determined in writing as necessary or appropriate to promote the national defense:

(a) by the Secretary of Defense with respect to military production and construction, military assistance to foreign nations, military use of civil transportation, stockpiles managed by the Department of Defense, space, and directly related activities;

(b) by the Secretary of Energy with respect to energy production and construction, distribution and use, and directly related activities; and
(c) by the Secretary of Homeland Security with respect to all other national defense programs, including civil defense and continuity of government.4

The Department of Commerce has delegated authority to four agencies to place priority-rated contracts and orders: Defense, Energy, DHS, and GSA. The Department of Commerce and these agencies can also place priority contracts for the use of other agencies. While the Department of Defense is the primary user of the system, applying the DO prioritization rating to approximately 300,000 contracts per year, the authority is used by a range of agencies on an episodic basis and by DHS on a regular basis.

The breadth and potential range of this authority is illustrated with reference to the range of uses reported in the 2011 DPAC report.

[The] priorities authority has been used to support, for example, hurricane and flood preparedness and response activities; Homeland Security Technology Programs; emergency preparedness activities related to the 2009 H1N1 flu virus; the Greater New Orleans Hurricane and Storm Damage Risk Reduction System program (by the U.S. Army Corps of Engineers); the International Safeguards, Second Line of Defense, and Nuclear Counterterrorism Incident Response programs (by DOE’s National Nuclear Security Administration); the Geostationary Operational Environmental Satellite, R-Series Program (DOD’s NOAA); and the Terrorist Screening Center program (DOJ/FBI).5

In addition, DHS/FEMA guidance and reporting indicates that DHS primarily uses rated orders to ensure on-time performance of contracts and to address supply chain problems.

The Department of Defense uses two Defense Priority and Allocation System industrial priority ratings—DX ad DO. DO is used for orders “critical to national defense” and requires the approval of the Under Secretary of Defense for Acquisition, Technology, and Logistics. DX rated orders are used for orders of the highest national defense urgency and must be approved by the Secretary or Deputy Secretary of Defense.6 [DoD Regulation 4400.1-M]. DX rated orders have equal priority and take precedence over DO orders. Likewise, DO orders have equal priority and take precedence over regular commercial contracts.

Finally, as required by the DPA itself, E.O. 13603 directs each Agency with delegated authority to “issue regulations to prioritize and allocate resources and establish standards and procedures by which the authority shall be used to promote the national defense, under both emergency and non-emergency conditions…. ” Four

4. 45 C.F.R. § 101.2(b).
6. See, DOD Priorities and Allocations Manual, DOD Regulation 4400.1-M.
agencies have issued such regulations, including the Departments of Agriculture, Health and Human Services, and Transportation. The Department of Commerce, Bureau of Industry and Security, amended existing regulations, which regulations cover the Department of Defense as well as Commerce.

Policy oversight is provided by the Defense Production Act Committee (DPAC), and its subordinate interagency working groups. In fact, a majority of the work of the DPAC is conducted at the working group level on either an interagency basis or an as needed agency basis. The Committee was created pursuant to Section 722(b) of the DPA to advise the president on the effective use of DPA authorities in support of national defense. Seventeen departments and agencies are currently members of the DPAC, which is chaired by the Director of the Federal Emergency Management Agency. Since 2011, the DPAC has issued an annual report on “government contingency planning for events that might require the use of the priorities and allocations authorities, provides recommendations for effective use of priorities and allocation authorities, and provides recommendations for improving information sharing among Federal Departments and agencies relating to the use of priorities and allocations authorities.” The report is currently submitted under the signature of the Administrator of the Federal Emergency Management Agency, as delegated by the Secretary of DHS.

2. Title I: Prioritization and Allocation Authority

Title I of the DPA provides authority to the President to prioritize contracts to promote the national defense as well as distinct authority to allocate materials, services, and facilities to promote the national defense.

(a) Allocation of Materials, Services, and Facilities

The President is authorized (1) to require that performance under contracts or orders (other than contracts of employment) which he deems necessary or appropriate to promote the national defense shall take priority over performance under any other contract or order, and for the purpose of assuring such priority, to require acceptance and performance of such contracts or orders by any persons he finds to be capable of their performance, and (2) to allocate materials, services, and facilities in such manner, upon such conditions, and to such extent as he shall deem necessary or appropriate to promote national defense. (Emphasis supplied).

However, Section (b) of Section 101 further provides that the “this section [allocation authority] shall not be used to control the general distribution of any material in the civilian market unless the President finds” (1) that the material is a scarce and critical

material essential to national defense; and, (2) the requirements for national defense cannot otherwise be met. Section (b) is a limitation on the President's authority, but it also shows that the DPA is potentially expansive. Section 101(c) provides distinct prioritization and allocation authority “in order to maximize domestic energy supplies.” This section, in contrast to the national defense authority in subsection (a), requires three presidential findings antecedent to its use. The law presents several threshold questions.

1. **Scope and Reach**: From first passage, section 101(a) has been viewed as a broad grant of authority, with some referring to it as a “commandeering” authority, granting the government power to take over and run defense industries. However, there is at least one apparent and material question regarding the scope of Title I authority: Could the President direct a company to accept a government contract for AI services, or is the authority limited to prioritization of existing contracts and orders?

2. **The Meaning of Services and other Terms of Art**: In addition, the Act is not as clear as it might be on the meaning of “services.” Absent such clarity, technology firms may choose to litigate the applicability of the DPA to requests for AI services, rather than provide them. Additional regulatory terms, such as “critical technology,” are also subject to ambiguity and thus dispute, especially when applied in new contexts, to new entities, and to new technologies outside of traditional DIB areas, and thus areas of common understanding and existing practice.

3. **Clarifying and Testing Less Traditional Practice**: The authority has primarily been used by the Department of Defense to prioritize defense contracts. There is no question that the law provides for such prioritization and there is a long track record of doing so. However, there is less practice outside of DOD of using Title I authority, and less practice in less-traditional national security fields. Other agencies, especially DHS, are increasing their use of DPA Title I authority for preparedness and disaster response. Based on this practice, some companies may take the view that to the extent the DPA is read to provide a commandeering, i.e., directive authority, beyond existing contract prioritization, such authority has either lapsed or was not intended to apply to them in the first place. Whether such arguments are well founded or not, litigation is likely given the stakes involved.

4. **Academic Institutions**: Finally, it is not clear which sections of the DPA apply to Academic institutions and which do not. A plain reading of the definition of “person” appears to include academic institutions; however, DPA implementing regulations, as of 2018, expressly incorporate academic institutions by reference in Section 705, but not Section 101. It is not clear whether such inclusion and omission is intentional and purposeful. In any event, it arguably creates an ambiguity as to the potential reach of the DPA to academic institutions. The point, or conclusion, is not that it does. But that any source of ambiguity is a potential source of litigation and dispute.
The clearer the law is on such matters, the more likely courts will uphold executive branch interpretations and decisions applying the law and do so in a consistent manner. Conversely, the more ambiguous the law and past practice, the more likely courts will substitute their independent judgment on the scope of the law exercising jurisdiction to resolve disputes. Varied case law results in delay and less certainty about how the law will apply in future and different contexts.

3. **Other DPA Provisions: Title III & VII**

Most commentary on the DPA is addressed to Title I’s prioritization and allocation authority. This is true of the Annual Defense Production Act Committee Report to Congress as well. This makes sense. Title I is a frequently used provision of the DPA. It is also the authority most likely to be used in an emergency. And, it is an authority likely to be subject to litigation because of its potential scope and reach. However, other provisions of the DPA are potentially important as well.

Title III, for example, authorizes the provision of incentives through loan guarantees (301), direct loans (302), purchase commitments and purchases (303(a)), and subsidy payments (303(c)). As stated on the Manufacturing and Industrial Base Policy (MIBP) website, the purpose of the program is to provide “the President broad authority to ensure the timely availability of essential domestic industrial resources to support national defense and homeland security requirements, by authorizing economic incentives to create, expand, and modernize production capacity.” Invocation of the program requires the President to make seven determinations, including:

1. that the resource or technology is essential for national defense;
2. “industry cannot or will not provide needed capacity in a reasonable time without DPA Title III assistance;” and,
3. “Title III incentives are the most cost-effective, expedient, and practical alternative for meeting the need.”

By law, Congress has prohibited the President from delegating these determinations. The program is overseen by the Deputy Assistant Secretary of Defense for Manufacturing and Industrial Base Policy, reporting to the Under Secretary of Defense (Acquisition, Technology & Logistics). The DOD Executive Agent for the Program is the Air Force, specifically, the Title III Office at Wright-Patterson, AFB. According to the MIBP website, there were twenty-nine funded Title III projects in 2011. There were twenty-eight publicly listed projects in 2017. These projects cover such subjects as thermal, solar and lithium batteries, nano-technology, and rocket motors.

Title VII of the DPA includes several provisions addressed to the operation of the DPA, such as Section 702 providing definitions of key terms. Section 706 grants federal district courts jurisdiction over civil and criminal proceedings for violations of the Act or its regulations. Included within this authority is the power to enjoin or enforce provisions of the DPA. This provision reflects Congress’s anticipation that use of the DPA might be contested and thus necessitate a judicial forum for dispute resolution. It also suggests that Congress anticipated the need for an enforcement
mechanism, including one with potential criminal sanctions. Title VII also provides additional positive authorities to study, shape, and influence the defense industrial base. Section 722 is important because it provides the statutory underpinning of the Defense Industrial Base System, which in turn is the statutory foundation for the DPAS, the DPAC, and the DIB information system. Title VII, Section 721, also serves as the statutory enabling authority for the Committee for Foreign Investment (CFIUS), which is addressed to the integrity of the industrial base and related supply chains. (CFIUS, as amended by the Foreign Investment Risk Review Modernization Act, known as FIRRMA, is an important supply chain authority, but one covered in detail elsewhere.)

Section 708 provides for voluntary agreements between the government and industry, where the work of industrial partnerships might otherwise raise antitrust and other fair business practice concerns. “Before a voluntary agreement may take effect, the Attorney General is required to make a written finding that conditions exist which may pose a threat to the national defense or its preparedness programs and that voluntary agreements are necessary to help provide for the national defense.” According to the DPAC Reports for 2013, 2014, and 2015: “There are two active voluntary agreements, both sponsored by the Maritime Administration in DOT: (1) the Voluntary Intermodal Sealift Agreement (VISA); and (2) the Voluntary Tanker Agreement (VTA). The purpose of the VISA is to provide DOD with assured access to commercial, dry cargo sealift capacity and intermodal equipment and systems to support emergency deployment and sustainment of U.S. military forces. The purpose of the VTA is to provide DOD with assured access to commercial tanker capacity in support of DOD contingency requirements.”

Like many areas associated with cyber security and operations, AI will test the government’s capacity to attract and retain the necessary personnel to run, advance, and indeed, regulate government AI policy and activities. The opportunity to perform public service and work on governmental applications may not be enough to attract the right people to the right positions. Moreover, background investigation requirements may prove an impediment. Additional thought and conscious risk management decisions about the need for waivers—for drug use, hacking, and foreign travel and residency—may be warranted. Of course, the Snowden and Manning cases are only the most manifest examples of the inside threat to information security, which threat AI tools will certainly compound. What is needed in the AI personnel context is thoughtful, deliberate, conscious, and accountable choices. The DPA provides two authorities that may bear on the question of recruiting critical expertise for government AI work and response.

Section 703 provides agency heads waiver authority to hire persons outside the competitive civil service system and without regard to the general schedule pay scale. However, salaries are still capped at the GS-18 rate (now calculated at the Senior Level, which replaced the grades of GS-16-18), and thus are not competitive with the sort of NFL salaries many AI innovators and engineers are receiving in the private sector. For example, a first year PhD in computer science can expect to make as
much as $500,000 as a software engineer at a leading commercial technology firm.\textsuperscript{8} In contrast, a similarly situated PhD joining a FFRDC could expect to make one-third or less of that amount. Whereas, using this special authority and the GS-18 rate, a government employee could not make more than 207,000 the highest possible level of ES-I. Of course, an entry level employee would not likely receive the benefit of a 703 waiver for this purpose. If AI is an integral ingredient in national security, as is quantum computing and data analytics, the government needs to prepare now for a future where large numbers of AI specialists, or skills sets, are required in government.

Section 710(e) provides authority to establish a National Defense Executive Reserve, which in theory, would serve as a cadre of executives from critical defense industries and sectors to serve in the federal government during periods of emergency. The original concept appears to have intended use of the NDER as a traditional reserve to backfill government positions vacated by military deployments. However, the law and implementing regulations clearly permit use of this authority in other ways, such as to augment government capacity with civilian expertise in the event of a national defense emergency.

This authority dates to the original DPA but has been used only intermittently since the 1950s. There are only a few readily available benchmarks along the Internet way. In 1957, President Eisenhower addressed a full auditorium at the Department of Commerce at the National Defense Executive Reserve Conference. He thanked the Reserve with the following words:

You are part of the whole machine that will keep that cutting edge sharp and efficient. You not only help keep the economy and the government moving itself; you are part of the hinge between the whole great 172 million of us and those that are in the Armed Forces.

These things I feel very deeply, and so it is difficult indeed for me to find the proper words in which to say to such a group as this “Thank You.” Possibly there are no better words than just those two, because coming here at your own expense you are setting an example for everybody in your community that may know you. You come here to prepare yourselves to be ready, to plan, so that if this country should ever, unhappily, have to face the ultimate in threat, there will be a strong body marching in, mobilizing in and behind the government, to make certain that we overcome whatever enemy may attack us, and to restore life to a free system as quickly as it is humanly possible to do so. [Remarks at the National Defense Executive Reserve Conference, November 14, 1957]

In the 1980s, there are GAO Reports addressing the decline and dysfunction of the NDER, although it is evident from the few reports that there were NDER units at

several agencies. However, eventually the NDER slides into disuse and virtually disappears from reporting citation.

President Obama's Executive order 13603 “re-established” the NDER. Section 501 of the Order states:

[T]here is established in the executive branch a National Defense Executive Reserve (NDER) composed of persons of recognized expertise from various segments of the private sector and from Government (except full-time Federal employees) for training for employment in executive positions in the Federal Government in the event of a national emergency.

This was among the provisions that drew public attention and partisan political commentary at the time the order was issued. Little attempt is evident in this commentary to trace the origins of the NDER and place it in historical or practical context. In any event, the Order further delegates authority over the program to the Secretary of Homeland Security, including issuance of necessary guidance. That in turn was delegated to FEMA, which in 2007 issued interim guidance on the NDER. However, as of June 2018 the FEMA interim guidance appears to remain in effect and there is no easily accessible listing of NDER units or members within the government, if any. (A Google search reveals a 2016 CIA NDER Unit Statement.) The threshold substantive predicate in the DPA to call up members of the NDER is a determination of a “national defense emergency.” Executive Order 13603 delegates this authority to the Secretary of Homeland Security and otherwise precludes the determination from redelegation. The heads of agencies activating an NDER unit or personnel therefore must first obtain a written determination from the Secretary of DHS and notify the Assistant to the President for Homeland Security and Counterterrorism in writing. It is evident from this procedure that President Obama, at least, perceived the NDER as most relevant in a homeland security context, but the law does not establish such limitations. As in other contexts, one can designate an emergency in one context, but not another. In other words, logical or statutory consistency is not required. If it were, the President's constant resort to the “unusual and extraordinary” powers of IEEPA would trigger a cascade of additional statutes, which it does not. One benefit that would accrue from an actual functional and operational NDER would be the opportunity to provide appropriate security clearances to industry officers. The NDER might also help to bridge or diffuse the “Us v. Them” culture that seems to arise between some parts of the government and technology companies.

4. **Section 705: Authority to Inquire and Survey**

In AI context, Section 705 may be as important as Title I or the CFIUS provisions of the DPA. The AI field has been described by some commentators as ungovernable. One reason it is at least hard to govern is that it covers so many different fields and capabilities. It also is spread over a breadth of private and academic actors, many acting in secrecy to protect intellectual property and profit potentials. Section 705 is a tool that could be used to create a mosaic from disparate AI tiles, as well as to
forecast key milestones and pending breakout moments. But its use in this manner will not be without challenge and controversy.

Section 705, as written, is broad and flexible in scope and can be used in support of any provision of the DPA. Thus,

The President shall be entitled... by regulation, subpoena, or otherwise, to obtain such information from, require such reports and the keeping of such records by, make such inspection of books, records, or other writings, premises or property of, and take the sworn testimony of, and administer oaths and affirmations to, any person as may be necessary or appropriate, in his discretion, to the enforcement or administration of this Act and the regulations or orders issued thereunder.

The authority can also be used independent of other sections of the DPA to “obtain information in order to perform industry studies to assess the capabilities of the United States industrial base to support national defense.”

This authority is used frequently by the Department of Commerce and would seem a logical and important authority for determining and assessing AI developments relevant to national defense in both industry and academia. However, there are scope and exhaustion requirements before the Executive branch can resort to such authority. It may only be used after the scope and purpose of an investigation is defined by a component authority and “it is assured that no adequate or authoritative data are available from any Federal or other responsible agency.” The section can be enforced with reference to criminal sanctions contained in Sections 706 and 705(c), as well as through the judicial power of contempt. Significantly, the section protects trade secrets, under threat of criminal sanction. Section 705(d) states:

Information obtained under this section which the President deems confidential or with reference to which a request for confidential treatment is made by the person furnishing such information shall not be published or disclosed unless the President determines that the withholding thereof is contrary to the interest of the national defense, and any person willfully violating this provision shall, upon conviction, be fined not more than $10,000, or imprisoned for not more than one year, or both.

Finally, reminiscent of debates in the context of PATRIOT Act enforcement, there is an express right to consult counsel in response to a subpoena issued under this section.

The 705 survey authority is well established, but not well established regarding AI, and especially AI outside traditional disciplines associated with national defense. To put a sharper point on the pencil, Silicon Valley is not part of the Cold War military industrial complex. Google’s primary commercial interest is not national defense contracting. AI is not a defense function or weapon, it is a universal capacity.

Two issues are especially pressing and should be debated and better resolved—now. One is technical, the other philosophical. The technical question is: If the government is going to collect information from private companies, as it already does,
are there additional safeguards and limitations on its use that should be put in place? The philosophical question is: To what extent, if at all, should the government use its authority to inquire into the activities of private companies engaged in AI research?

5. Conclusion

The DPA provides surprisingly broad authority to the President and Executive branch to prioritize, allocate, incentivize, and survey critical technology related to national defense. The authority is “surprising,” because the language is expansive in its scope and reach, especially in a context where private rights and property are at issue. But in some regards the scope of this authority is untested, by the government, which has not pushed as far as the statutory language might allow. And, by private industry, which has not litigated where it might. This is especially true with respect to critical technologies, including AI, in the commercial sector.

Efforts to use the DPA to reach beyond traditional DIB actors or use the DPA in new ways would likely result in litigation. Indeed, the DPA contemplates litigation and enforcement in its express grant of jurisdiction to federal district courts. Six challenges are likely:

(1) The authority is being used beyond the predicate necessity of “national defense.”
(2) The authority being used does not track with the statutory or regulatory requirements of the DPA.
(3) Use of the authority violates the individual’s or institution’s First Amendment,
(4) Fourth Amendment, or
(5) Fifth Amendment rights.
(6) The authority has fallen into disuse, if it was ever used, and is no longer extent under the doctrine of desuetude.

If litigated, courts will have to weigh the plain language of the statute with apparent practice and legislative history.

C. Inventions Secrecy Act (ISA)

The ISA, found at Title 35 U.S.C. 181-188, is a Cold War era statute (1952) that authorizes the government to impose a secrecy order on certain inventions in the interest of national security. Specifically, “if in the opinion of the Atomic Energy Commission, the Secretary of a Defense Department, or the chief officer of another department or agency so designated, the publication or disclosure of the invention by granting of a patent thereof would be detrimental to national security” that officer shall notify the Commissioner of Patents and the “Commissioner shall order that the invention be kept secret and shall withhold the grant of a patent for such period as the national interest requires.” Secrecy orders may be issued for up to one year and
are subject to renewal on an annual basis upon notification by the requesting agency, or for shorter periods, if requested.

The ISA comes with several enforcement mechanisms. In the event information about a covered invention is published or disclosed, or an application for a patent for a covered invention filed in a foreign country, the invention may be treated as abandoned by the PTO. In addition, violation of the Act with knowledge of a secrecy order carries a potential criminal penalty of a $10,000 fine and imprisonment of not more than two years. Appeals of secrecy orders may be taken to the Secretary of Commerce and from there to the Court of Federal Claims or the U.S. District Court of the district where the claimant resides. In both cases, the courts have authority to issue protective orders and to hear appeals of secrecy orders in a secure manner.

The law is implemented through the Patent and Trademark Office (PTO) within the Department of Commerce and is triggered by an application for a patent. According to media reports, the principal agencies involved in ISA reviews are the Defense Threat Reduction Agency, NSA, DOJ, and DHS. These same reports indicate that by 2012 there were approximately 5,300 such orders, with a high order mark of just under 100 issued in 1998. At least one commentator has suggested that the way for academics to circumvent the law is to preempt it by publishing research before filing for a patent or by seeking a foreign patent first.

The ISA would seem directly on point in the event a U.S. corporation or academic entity created an algorithm or AI application with security importance the public disclosure of which could undermine national security. This might be the case, for example, if disclosure would alert adversaries to an emergent U.S. capability or lead to the possible theft or copycat emulation by an adversary. The question is not whether the law would apply to AI. It clearly does. The questions are whether, when, and why the government might invoke its authority with respect to AI and subject to what litigation risks. Another policy question is whether the government should hold additional authority, beyond the patent process, to protect technology and know-how in the research and development stage before or outside the context of patent applications. These are at root policy questions. However, they also present process and legal questions to the extent they implicate the government’s ability and authority to determine what is occurring within industry and academia in the first place. The government cannot very well invoke the ISA if it does not know whether an invention is in the research pipeline. The process questions include whether key agencies and offices, like the Patent Office, understand the national security applications and implications of AI well enough, and whether they have received sufficient policy guidance, to timely trigger a process of ISA review.

---

The law provides for a right to compensation for use of the invention by the government. Thus, another string of legal questions addresses the extent to which an inventor is entitled to compensation where an invention is not used by the government but where the inventor is otherwise barred by a secrecy order from taking an invention to market. Embedded in this issue is the question of valuation, which can be highly speculative before an invention is deployed and thus before it realizes its market potential. There are also questions of secrecy, First Amendment rights, and court room procedure.

A second string of legal questions considers the extent to which the government can look within academia and industry to determine what is occurring to better determine when and whether to invoke the ISA or other similar authorities. As earlier noted, this is one of the central AI questions posed by the Defense Production Act (DPA).

V. DEPLOYMENT AND EMPLOYMENT

A. Introduction

As the DPA and ISA are addressed to the research and development of weapons and defense systems, the law of armed conflict is addressed to the deployment and employment of weapons. It is said that the difference between an armed mob and a professional army is discipline. Discipline is rooted in leadership and law. Where warfare is concerned, the most relevant law is the law of armed conflict, also known as international humanitarian law and the law of war. The national laws, and in particular military law and the military codes of individual nations are also important, in part, because they are ultimately controlling where there is a real or perceived conflict between international and domestic law. Further, international tribunals like the ICC and ICTY notwithstanding, it is predominantly through national laws that individual service members are held to account for their adherence to the law of armed conflict.

The law of armed conflict is based on treaty as well as customary international law. The leading treaties are generally grouped into two lines of law. The Geneva line is largely addressed to the treatment of persons engaged in or subject to warfare as embodied by the 1949 Geneva Conventions. The Hague line is addressed to the conduct of warfare, also known as the means and methods of warfare. A number of leading treaties were in fact concluded at The Hague and in Geneva, hence the monikers; however, there are any of a number of additional treaties, such as the Biological Weapons Convention and Chemical Weapons Convention that were concluded elsewhere.

Customary international law is based on the conduct of states practiced out of a sense of legal obligation. The fundamental principles of military targeting—necessity, proportionality, distinction, military objective, and minimization of suffering—for example, are generally accepted as customary international law, although they also find expression in treaty law. (U.S. military doctrine also includes honor as one of the
core targeting principles.) Some principles are found in both treaty and customary international law, but only accepted by certain states as binding under customary law because they are not parties to the treaties in question. Article 36 of the Additional Protocol I, discussed below, is an example. The United States is not a party to the Protocol, but recognizes Article 36 as customary international law.

Adherence to these and other IHL norms occurs, if it occurs at all, because states have a reciprocal interest in their application; states care how they are perceived especially if they seek alliance; the participation of some states in military operations is contingent on compliance with international law; and, because there is an effective, or possible, means of domestic or international enforcement. A disciplined armed force is also more effective in accomplishing the mission than one that is not.

In U.S. military practice, the law of armed conflict is implemented through military doctrine and manuals. It is also based on leadership, which is the backbone of good order and discipline. Military law and regulation buttress leadership and doctrine by establishing a code of conduct and means of enforcement. Thus, the Uniform Code of Military Justice (UCMJ) serves two core national security interests, justice and good order and discipline. Additional criminal sanctions for grave breaches of the law of armed conflict are found in Title 18, Section 2441. The question posed in this subsection is whether some of these principles should apply, in relevant manner, to national security uses of AI, and if so, in what manner.

**B. Outline**

This section first considers three requirements from the law of armed conflict for the deployment and use of weapons, and thus also AI enabled weapons and weapons systems. Those requirements are:

1. the *duty to test new weapons for compliance with the law before deployment*;
2. the *duty to instruct* personnel on the law of armed conflict; and,
3. *Command responsibility* for the actions of one's subordinates, including with respect to (1) and (2).

The section then considers three questions derived from the law of armed conflict.

1. Should the principles of targeting identified above apply differently in the case of AI enabled weapons and systems?
2. Should additional principles apply in the case of AI? For example, should the law of armed conflict prohibit certain uses of AI, such as the use of AI to enable nuclear weapons? Certain forms of autonomous systems? Or, the deployment of AI in certain domains, like space?
3. Finally, in light of AI’s potential to mimic and to fake voices and images, should the distinction between lawful efforts to deceive one's opponent, known as ruses, be further distinguished and clarified from unlawful efforts
to deceive the enemy known as perfidy? Restated, where AI is concerned, should special rules apply to the distinction between ruse and perfidy?

C. New Weapons Review

Article 36 of Additional Protocol 1 to the Geneva Conventions states:

In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to a High Contracting Party.

“The aim of Article 36 is to prevent the use of weapons that would violate international law in all circumstances and to impose restrictions on the use of weapons that would violate international law in some circumstances, by determining their lawfulness before they are developed, acquired or otherwise incorporated into a State’s arsenal.” Biological and chemical weapons, for example, are weapons that would fall into the first category. They would violate international law in all circumstances, not just because they are expressly prohibited by Treaty and customary international law, but also because they are inherently indiscriminate in their effect. The use of munitions whose fragments are not-detectable by x-ray also fall into this category, because they inherently cause undue suffering. In contrast, whether the use of white phosphorous or lasers is lawful, or unlawful, depends on how they are used. If, for example, they are used as anti-personnel weapons, to burn or to blind, then they would be unlawful for causing undue suffering. However, if they are used to mark targets, for example, to help pilots or artillery spotters distinguish between civilian and military objects, their use would be lawful if such use otherwise satisfied the principles of targeting. Indeed, proponents of discriminate warfare should encourage their appropriate and lawful use.

Scholars and states debate where nuclear weapons fall on this continuum. On the one hand, opponents argue nuclear weapons are inherently indiscriminate and disproportionate to any lawful military objective, and they cause undue suffering. On the other hand, proponents conceive of the possibility that they might be used in a tactical manner that is necessary, discriminate, and proportionate to the military objective and need at hand, for example, to penetrate hardened targets. They further posit that they are necessary for stability as a belligerent reprisal deterrent to the use of WMD.

What about AI? Interestingly, Article 36, which was included in the Protocol opened for signature in June 1977, appears to have contemplated some form of future AI and AWS. This is reflected in the 1987 ICRC commentaries, which state:

The use of long distance, remote control weapons, or weapons connected to sensors positioned in the field, leads to automation of the battlefield in which the soldier plays an increasingly less important role. The countermeasures developed as a result of this evolution, in particular electronic jamming (or interference) exacerbates the indiscriminate character of combat. [ICRC, 427-428].

Although the United States is not a Party to Protocol I, it recognizes Article 36 as reflecting customary international law. Indeed, the United States has conducted weapons reviews since well before the adoption of Article 36, as a matter of sound policy and consistent with Article 36 and pursuant to customary international law.

The United States has implemented the requirements of this provision as well as the general duty to follow U.S. law in among other places Department of Defense Directives. DoD Directive 5000.1, The Defense Acquisition System, states

The acquisition and procurement of DoD weapons and weapons systems shall be consistent with all applicable domestic law and treaties and international agreements...An attorney authorized to conduct such legal reviews in the Department shall conduct the legal review of the intended acquisition of weapons or weapons systems.

The Directive further establishes a process for conducting such reviews as do Department Regulations for each of the military services with respect to the procurement of new weapons systems.

The DoD Law of War Manual (2015 and electronically updated since) states that the review "should consider three questions to determine whether the weapon's acquisition or procurement is prohibited:

Whether the weapon's intended use is calculated to cause superfluous injury;
Whether the weapon is inherently indiscriminate; and
Whether the weapon falls within a class of weapons that has been specifically prohibited."

In addition, in 2012, DOD issued a Directive addressed specifically to Autonomy in Weapons Systems.11 The directive:

a. Establishes DoD policy and assigns responsibilities for the development and use of autonomous and semi-autonomous functions in weapons systems, including manned and unmanned platforms.

b. Establishes guidelines designed to minimize the probability and consequences of failures in autonomous and semi-autonomous weapons systems that could lead to unintended engagements.

Further, the Directive establishes DOD policy regarding AI in weapons systems, including the following:

a. Autonomous and semi-autonomous weapons systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.

(1) Systems will go through rigorous hardware and software verification and validation and realistic system development and operation test and evaluation...

(2) ...

(3) In order for operators to make informed and appropriate decisions in engaging targets, the interface between people and machines for autonomous and semi-autonomous weapons shall:

(a) Be readily understandable to trained operators.

(b) Provide traceable feedback on system status.

(c) Provide clear procedures for trained operators to activate and deactivate system functions.

b. ...

c. ...

(1) ...be designed such that, in the event of degraded or lost communications, the system does not autonomously select and engage individual targets or specific target groups that have not been previously selected by an authorized human operator.

(2) ...

(3) Autonomous weapons systems intended to be used to apply non-lethal, non-kinetic force, such as some forms of electronic attack, against material targets in accordance with DoD Directive 3000.3 [Policy for Non-Lethal Weapons].

The Directive, however, also states that systems intended to be used in a manner that falls outside the policies in c.(1) and (3) must be approved by the Under Secretary of Defense for Policy, the Under Secretary of Defense for Acquisition, and the Chairman of the Joint Chiefs of Staff before formal development and again before fielding. In addition, the Directive assigns responsibilities, and thus accountability in the
review of autonomy in weapons systems, to include the General Counsel of the DoD for legal review in accordance with Article 36 and Directive 5000.1.

Thus, the Directive establishes important substantive technology redlines and thresholds.

- Commanders and operators shall exercise appropriate levels of human judgment over the use of force.
- The interface between people and machines shall be “readily understandable,” provide traceable feedback, and provide clear procedures.
- In the event of lost communications, the systems will not operate autonomously against targets not previously selected by an authorized human operator.

The Directive also establishes a process of responsibility and accountability by making specific senior officials accountable for adhering to the Directive. Of course, there is always room for wiggle where words like “appropriate” are used, essentially deferring to a later and unseen decision as to how much human judgment is appropriate.

Two immediate legal policy questions arise:

1. Does the U.S. take the view that these redlines are necessary in order to comply with Article 36 or other law, or are they a matter of policy discretion?
2. If not already viewed as part of customary international law, should the U.S. (and other actors) seek to embed these principles in customary international law and military doctrine?

There are also several interpretive questions regarding the application of Article 36 and its customary norm.

First, what is the meaning of new weapon and how should it be read in conjunction with the clause that follows, “methods and means of warfare.” The ICRC commentaries take the view that when read together, “the words ‘methods and means’ include weapons in the widest sense, as well as the way in which they are used.” Thus, most transparent state applications of the article, such as Australia’s, define weapons broadly to include “an offensive or defensive instrument of combat used to destroy, injure, defeat or threaten. It includes weapon systems, munitions, sub-munitions, ammunition, targeting devices, and other damaging or injuring mechanisms.”12 The ICRC Guide to the Legal Review of Weapons indicates that review should extend to new uses for weapons or modifications to existing weapons, or weapons that are new in the inventory of the country involved.

A second question regarding new weapons review is what is required? There is general agreement that the requirement covers the normal anticipated use(s) at the time of evaluation. States are not required to foresee and evaluate how a weapon might be misused or used for unintended purposes. The review is also contextual.

---

12. IRRC, p. 937.
Modelling can be used or actual firing, in the case of kinetic weapons. According to the ICRC, the review is supposed to be multidisciplinary taking into account “military, legal, environmental and health-related considerations.” Where necessary the review is intended to result in modification or cessation of the weapon’s development, or appropriate training and guidance, for example, in the case of a munitions, like white phosphorous, or riot control agents that might have both lawful and unlawful uses. Commentary also points to the importance of testing weapons at each stage of the development process and not just prior to deployment. This is the better course because of the difficult budget and bureaucratic challenges of cancelling or modifying a weapons program after it is fully developed. Finally, there is general agreement that a state is not required to make its review findings public for reasons of security.

1. Outstanding Questions

   a. How should Article 36 apply to AI-enabled systems?

Although it is clear that the Article 36 applies to new weapons or new methods and means of warfare, there is ample room for interpretation if one is talking about AI-enabled systems that are not directly used as weapons, but may nonetheless impact military operations, such as systems that process, sort, and identify targets; queue targets; or warn and alert watch officers and radars linked to weapons response, and thus are inexorably linked to how existing kinetic or cyber weapons are used.

   b. Should military AI applications be subject to some form of “new technology” review?

The new weapon requirement presents a potential template for AI applications outside the construct of weapons systems. Six subordinate questions immediately arise.

- Should states be required to test AI for compliance with the law prior to development and deployment?
- If so, should entities other than states, such as corporations and academic labs also be required to engage in new AI review? And, if so, at what stages of research and development?
- If so, with what law or regulations should AI comply? In the absence of agreed law, should states be required to meet certain safety and security standards?
- Should such a requirement apply to AI generally, or just national security applications, or just to AI associated with weapons? The question is like that posed by the meaning of new weapon. Line drawing with AI is more complex than with weapons, but here, one might ask whether one could reasonable distinguish between categories of AI. One might also distinguish

13. IRRC, p. 935.
between systems that are attached to critical infrastructures, like the Internet or the energy grid, and those that are not.

- Should states be required to publish their regulations providing for such testing?
- Should an international organization, like the OPCW or IAEA, or a non-profit, like IANA/ICANN in Internet context, oversee the conduct of such safety and security testing?

**c. If so, or in any event, who should be held responsible and accountable for new AI technology and applications in the context presented?**

As a matter of internal U.S. process, the DoD Directive on Autonomous Weapons also suggests the question: Who, or what institutions, should approve AI research and development? In the context of autonomous and semi-autonomous weapons systems, the key decisional actors are the Under Secretaries of Policy and Acquisition as well as the Chairman of the Joint Chiefs of Staff. Should additional decision-makers outside the Department in question also be involved? And, in private and academic context, which actors should be involved, responsible, and accountable? Should university Presidents and General Counsel be required to review AI R&D and be held to account for the ethical outcomes of the results? In industry, the responsible actor is presumably the CEO. But should law or regulation expressly require General Counsel approval or a decision of the Board of Directors?

**d. Should critical actors, including the U.S., press for greater transparency?**

The long march starts with a single step. As a first step, states might consider publishing regulations that apply to safety and security testing of AI, or at least AI-enabled weapons and systems linked to weapons use. As noted, there is no requirement to publish the results of a new weapons review. This makes sense. If there were a requirement, states would either decline to test weapons or lie about the results out of concern that they would reveal national security capabilities to potential adversaries. However, the law might usefully require states to publish their regulations regarding new weapons review, as a first step toward verifying compliance with the overall requirement. Few states do so, which leaves one to wonder whether states generally adhere to this IHL requirement or not. Moreover, without publication of state regulations it is hard to determine whether there is agreement on what the law requires, how it applies, or determine where the gaps and fissures are. It will also help identify ambiguities in law and application.

In any event, the Department of Defense should further refine and make public, if possible, an express statement of how it will comply with Article 36 in the context of the five scenarios DOD appears to contemplate for autonomous weapons systems. In doing so, the United States can help to establish customary international law on
the matter as well as pressure other states to make public their procedures. Former Deputy Secretary of Defense Bob Work has described the five building blocks or components to the Third-Offset Strategy. They are:

1. Autonomous deep learning systems;
2. Human-machine collaboration;
3. Assisted human operations;
4. Human-machine combat teaming; and
5. Autonomous weapons.

The specific question is: How does the new weapons requirement apply to each of these areas and is the United States prepared to state so publicly to help develop and assert customary international law in this area? Recall that the goal here, as elsewhere, is for decision-makers to make purposeful, thoughtful, and informed decisions and not default to the lowest common denominators established by states racing for national security advantage.

D. Training

Most LOAC treaties include requirements to implement those treaties through training specific to the treaty. This practice dates to the 1906 Geneva Convention. These provisions are sometimes general in nature, as in the case of Common Article 1 to the four 1949 Geneva Conventions, which states:

The High Contracting Parties undertake to respect and to ensure respect for the present Convention in all circumstances.

This Article is viewed as creating a binding obligation to implement the Convention through training and enforcement. (Contemporary disputes in interpretation revolve around the extent to which this obligation imposes on States Parties an obligation to ensure their allies, with whom they might share intelligence and weapons, are adhering to the law as well in the conduct operations.)

Other requirements to train, are more specific, like the obligation found in Article 144, of the 1949 Geneva Convention IV, “Relative to the Protection of Civilian Persons in Time of War.”

The High Contracting Parties undertake, in time of peace as in time of war, to disseminate the text of the present Convention as widely as possible in their respective countries, and, in particular, to include the study thereof in

their programmes of military and, if possible, civil instruction, so that the principles thereof may become known to the entire population.

Interestingly, in contemplating how this might apply in AI context, this provision seems to require training for the civilian populace at large, as well as military members. This is an extraordinary undertaking, creating some skepticism that such an obligation is in fact binding law rather than aspirational thought. If it is operational law, it is not in fact put into operation in U.S. practice. But where AI is concerned it may well be a good policy idea. Shouldn't we want our senior civilian officials trained in AI enabled operations, given the speeds, complexities, and impacts involved?

Perhaps the best baseline statement of the duty to train personnel in the law is found in Article 83 of AP 1, which also makes it clear that any obligation to train on the law should be tailored to one's duties and responsibilities.

In order to prevent and suppress breaches, High Contracting Parties and Parties to the conflict shall require that, commensurate with their level of responsibility, commanders ensure that members of the armed forces under their command are aware of their obligations under the Conventions and this Protocol.

The United States implements its obligations under these and other treaties through the DoD Law of War Program, DoD Directive, 2311.01 E., 2011. Among other things, the Program recognizes that the United States should adhere to these requirements not just because it is obliged to do so, but because it is makes military sense to do so. For example, the legal principle of proportionality, which prohibits the use of force which is excessive in relation to the direct and concrete military advantage anticipated dovetails with the military principle of economy of force. Likewise, principles of law requiring distinction between combatants and noncombatants and the humane treatment of detainees also reflect national security values. Adherence to these values can protect one's own forces when the law is applied in reciprocal manner. Adherence is also more likely to encourage detainees to share intelligence and information. And, in the context of counterinsurgency or counterterrorism operations, adherence will help keep the local population friendly, or at least neutral rather than hostile. Adherence to the law generally also leads to wider domestic support for the military at home. Consider how Abu Ghraib and My Lai served as turning points in public perceptions about the Iraq and Vietnam conflicts. Adherence to the law and training on the law also garner and sustain allied support for U.S. operations.

Of course, Parties to Treaties may go beyond any required obligation. The DoD Law of War Program does so. For example, the Secretary of Defense has directed that:

The Heads of the DoD components shall... Make qualified legal advisers at all levels of command available to provide advice about law of war compliance during planning and execution of exercises and operations[.]. [Dir. 2311.01E, Sec. 5.7.3].
The Commanders of the Combatant Commands shall: ... Ensure all plans, policies, directives, and rules of engagement issued by the command and its subordinate commands and components are reviewed by legal advisers to ensure their consistency with this Directive and the law of war. [Dir. 2311.01E, Par. 5.11.8].

In addition, as is generally known, the military provides case specific Rules of Engagement to supplement the general principles of the law of armed conflict, which always apply.

Although the obligations to train are textually clear, several genuine questions arise. First, when is training required? The better view, based on both the text of Protocol I and state practice is that training is required in peacetime as well as during wartime, and on an ongoing basis. But treaty text could be more express on this point.

Second, how specific is the required training? Again, the text could be clearer. Protocol I uses the language “commensurate with their level of responsibility.” However, it is not clear, nor expressly required, that training occur on the record or be conducted by lawyers or ethicists.

Third, whatever the obligation, it is also clear that in conflict, training is only effective if violations that deviate from training are investigated and addressed and, where appropriate, warfighters held accountable.

Finally, however clear or unclear the requirement to train is with respect to States, it is far less clear with respect to non-state actors and parties. According to the ICRC, “Article 19 of Additional Protocol II, [applicable to Non-International Armed Conflicts] states that the Protocol 'shall be disseminated as widely as possible,' and this provision binds armed opposition groups.” To address this potential loophole, there are cases where parties in conflict have agreed on a bilateral basis and as a matter of legal obligation to disseminate and adhere to the law of armed conflict. See, Memorandum of Understanding on the Application of IHL between Croatia and the Socialist Federal Republic of Yugoslavia; and, Agreement on the Application of IHL between the Parties to the Conflict in Bosnia and Herzegovina, cited in ICRC, CIL, Rule 142. In other words, the fact that treaty law or CIL can be read to exclude non-state actors (substitute here—certain AI applications) does not mean that states can't otherwise choose to establish and adhere to or apply norms on a unilateral, bilateral, or multilateral basis.

1. Outstanding Questions

Beyond the application of AI in contexts where the law of armed conflict applies, should international and domestic law require training on AI or the law of AI? Such training might encompass safety, security, law, and include an express acknowledgment of responsibility and accountability, in effect, a chain of custody for the proper use of the AI in question.

Taking Article 144 of Geneva Convention IV as a model, international law might also require the training of civilians as well as military personnel in AI applications to warfare. Under current U.S. practice, civilians receive virtually no formal or
required training in the law of armed conflict. Rather, the training that is received, is received informally and on-the-job in the form of advice and memos. As a matter of domestic law and regulation, if not international law, this could and should be changed. Why not require something similar for AI? States might consider a requirement for legal or ethical training for persons engaged in military AI research and development in industry and academia based on the IHL model.

States might also agree to require publication of any national requirements to provide training, thus creating pressure on other states to do the same while at the same establishing customary international law norms.

Finally, proponents of AI for military purposes should consider the lessons learned from the DOD law of war program. It is not only implemented out of a sense of legal obligation. It is implemented because adherence to the law results in greater public support and sustainment for military operations, underpins allied support and assistance to U.S. military operations, and is the foundation upon which the reciprocal application and enforcement of the law depends. It also leads to better national security results.

E. Command Responsibility

As noted above, Protocol I and customary international law hold commanders responsible for “ensur[ing] that members of the armed forces under their command are aware of their obligations under the Conventions and this Protocol.” Among those obligations is the principle of command responsibility. Command responsibility includes five elements.

**Individual Responsibility**: All members of an armed force are responsible and accountable for upholding the law of armed conflict; not just commanders. This means that following superior orders is not a defense to violating the law of armed conflict. In U.S. practice and law, an order is presumed to be lawful, unless a reasonable person would know it was unlawful, in which case the recipient has a duty not to follow the order.

**Commanders are responsible for War Crimes they knew of or had reason to know of and did nothing to stop**. The most important case in this line of law is the *Yamashita* case. General Tomoyuki Yamashita was the commander of Japanese forces in the Philippines. He was also one of Japan’s most successful and revered commanders, known as the “Tiger of Malaya,” for his successful conduct of the Malayan campaign and capture of Singapore. He also was in charge of holding and defending the Philippines from General MacArthur’s allied forces. All of which is relevant to his case.

Yamashita was charged and tried before a U.S. military court-martial in Manila for the war crimes of troops under his command committed in the Philippines, including mass murder, torture, and rape. Yamashita was not charged with directly committing, directing, or ordering the offenses himself. Nevertheless, the Court concluded that Yamashita was culpable for the offenses of his subordinates because he “…unlawfully disregarded and failed to discharge his duty as a commander, to control the operations of the members of his command, permitting them to commit brutal
atrocities... and he... thereby violated the laws of war.” Yamashita was convicted and hanged.

In the Queenfish case, the commander of a U.S. submarine was court-martialed for torpedoing a Japanese ship, the Awa Maru. The ship, a passenger cargo vessel, had been marked as a hospital ship. It had also been used to transport essential supplies and medicines to Allied POWs held in Japanese camps. In exchange, the vessel received safe passage from the U.S. Pacific Command, notwithstanding the possibility that the vessel would also be used to transport war material and personnel to and from Japan during the outbound and return voyages. The Pacific Command sent multiple encrypted and clear messages to fleet submarines. One message stated: “Let pass safely the Awa Maru carrying prisoner of war supplies. She will be passing through your areas between March 30 and April 4. She is lighted at night and plastered with white crosses.” The Queenfish received the message, but the message was not specific as to location or route. A more detailed message was not read by the Commander, although it was read by his communications officer. Later, the Queenfish was advised by other submarines that there were targets in her area near the Taiwan Strait and she proceeded to engage a vessel in a dense fog, which the commander believed was a Japanese destroyer. It was the Awa Maru.

There was one survivor. He was picked up by the Queenfish, at which point Commander E. Elliott Loughlin learned he had torpedoed the Maru. He reported what had occurred. The Queenfish was ordered to port and CNO Admiral King eventually ordered, “Detach Loughlin from his command and have him tried by general court-martial.” A subsequent court-martial acquitted Loughlin of culpable inefficiency in the performance of duty, and disobeying orders, but found him guilty of the third and least serious offense of dereliction of duty “for negligence in obeying orders.” Loughlin was sentenced to receive a Letter of Admonition from the Secretary of Navy. Admiral King ordered that Loughlin not command submarines again. However, in due course, Loughlin did command a squadron of submarines and eventually served as and retired from the Navy in the grade of Rear Admiral. Two Navy Crosses for wartime service as well as successful post-war service as the Naval Academy’s Athletic Director no doubt contributed to his rehabilitation.

The Queenfish case, like the Yamashita case, is cited for the proposition that commanders are responsible for the actions of their commands, whether they hold the specific intent to commit the underlying delict or not. In Loughlin’s case, he was convicted, in part, for having an ineffective process in place to ensure compliance with orders and law. No doubt, wartime pressures to keep a unique line of supply open to Allied POWs, even if the Japanese were also using it to transport war material and personnel, contributed to the decision to court-martial Loughlin.16

There were other problems and lessons to learn from the incident, some of them relevant to current consideration of AI. The Pacific Command did not repeat the safe passage warning to the officers of the Queenfish during its operational briefing but relied on message traffic alone to communicate the safe passage. Lacking visibility in the fog, the officers of the Queenfish relied on radar to profile the Awa Maru along with its speed and direction. The Awa Maru was running fast and low in the water like a destroyer, rather than zig-zagging like a merchant vessel. This dovetailed with what the officers were predisposed to expect (confirmation bias). And, of course, there was the fog of war, both literally and in the sense that Clausewitz meant.

**Commanders must take reasonable measures to prevent violations of the law of armed conflict:** The requirement to take “reasonable measures” is found in the case law and statutes of the International Criminal Tribunals for the former Yugoslavia (ICTY) and Rwanda (ICTR) as well as the Rome Treaty establishing the International Criminal Court. However, Article 86 of Protocol I, purports to require commanders to take “all feasible measures within their power,” rather than “reasonable measures,” a higher standard, and one with an additional burden of proof. The United States has adopted the “reasonable measures” standard and has done so since the 1956 U.S. Army Field Manual on the Law of Land Warfare. The 2015 DoD Law of War Manual applies this standard as well, stating:

> Commanders have duties to take necessary and reasonable measures to ensure their subordinates do not commit violations of the law of war. Failures by commanders … can result in criminal responsibility. [18.23.3]

In the Queenfish case there was no evidence the commander had seen the communications granting the Maru safe passage, but the communications were received. The commander was found responsible for his failure to have established proper procedures on his vessel for the handling of such communications. Thus, this case can be viewed as a “reasonable measures” case as well as one upholding the commander’s duty to know of the actions of his subordinates.

**Civilian Command Responsibility:** Although the principle of command responsibility is most often articulated with reference to military command and commanders, recent case law in the context of the ICTY and ICTR make clear that the concept can extend to civilian command responsibility for the actions of military subordinates. The ICRC Commentaries, for example, state “Not only military personnel but also civilians can be liable for war crimes based on command responsibility. The International Criminal Tribunal for Rwanda, in the Akayesu case in 1998 and in the Kayishema and Ruzindana cases in 1999, and the International Criminal Tribunal for the Former Yugoslavia, in the Delalic case in 1998, have adopted this principle.” [ICRC, Rule 153]. These principles are reinforced at the Head of State level by the arrest and subsequent conviction of former Serbian President Slobodan Milosevic for war crimes associated with the Balkans Wars of the 1990s.

---

17. Dingman & ICRC Lesson Plan generally.
Investigation and Prosecution: Finally, lawyers generally agree that the doctrine of command responsibility includes the duty of the command to investigate credible assertions of war crimes and to hold accountable those responsible for war crimes. The issues that are usually debated here are whether the predicate for “investigation” is met, and then whether the country in question engaged in a credible investigation and effort to hold the responsible parties accountable. There is often a tension, of course, between victims and outside observers, who view “conviction” as synonymous with “investigation,” in contrast to military authorities, who tend to view the obligation as requiring good faith investigation, and where appropriate, prosecution in an open and fair trial. In U.S. practice, this of course requires proof beyond a reasonable doubt of each element of an offense.

Finally, the Uniform Code of Military Justice, the criminal code applicable to members of the U.S. Armed Forces, does not have a provision on command responsibility. However, it accomplishes the objectives embodied in the principle through punitive Article 92, which covers among other things, failure to follow orders and dereliction of duty, which orders and duty include adherence to the law.

1. Outstanding Questions

When it comes to military uses of AI in armed conflict the command responsibility question is clear. Lawyers and technologists can argue about who should be held responsible when software does not work as intended. But what is clear, is that in conflict, the commander could, and likely should be held responsible, if he knew of, or had reason to know of violations of law, including that an AI-enabled weapon or system could not, or likely could not, be used in a manner consistent with the law of war. Ignorance, willful or otherwise, is not a defense as the *Yamashita* and *Queenfish* cases demonstrate.

One question is whether the United States should seek to refine the law regarding how command responsibility does, or should, apply to AWS specifically, or AI generally. This can be done by changing domestic law, or by asserting as a matter of legal obligation certain requirements in customary international law. Policymakers may want to do so not because they believe in good order and discipline, but because they want better security outcomes and greater confidence that in the race to obtain AI for military purposes shortcuts in safety and speed will be accompanied by a corresponding increase in the level of responsibility and accountability.

Third, in what manner, if at all should the United States, other countries, or other actors like NGOs, seek to develop and expand the concept of command responsibility, including civilian responsibility, to better address AI applications in warfare, grey zones, and peacetime. AI may be a field warranting development of a doctrine of civilian command responsibility.

F. The Principles of Targeting

Since the advent of the digital age, lawyers and operators have been debating the application of the law of armed conflict to the cyber domain and cyber operations.
The discussion over when and whether a cyber-event rises to the level of “armed attack,” and thus gives rise to a right to kinetic or other self-defense seems to attract inordinate attention. U.S. public policy regarding armed attack in cyberspace appears to be “we will know it when we see it” and we will continue to reserve the right to respond with a full array of policy options, cyber, kinetic, and other. The Tallinn Manuals (1.0 and 2.0) are an effort by scholars to come to grips with some of these questions from the perspective of international law. Scholars in public, and operators in less public settings, also debate how and whether the core principles of targeting should apply in cyberspace in peacetime, or in that peculiar grey-space within the cyber domain which is neither at war nor at peace, but marked by hand-to-hand cyber engagement.

One question for military law and doctrine presented by AI is whether the targeting principles should apply in a different manner with AI-enabled tools than they already do to kinetic and cyber operations. For sure, one of the strengths of the targeting principles, is that they are principles, and thus adaptable to different contexts. But while the definitions may be the same, and must be the same, principles often apply differently in different contexts. This is not unusual with legal principles. The Fourth Amendment is a case in point. The text of the Amendment is static in protecting persons from unreasonable searches and seizures. However, what is “reasonable” will vary depending on the facts and circumstances.

So how might the principles apply differently with AI? Let’s look at proportionality, which is defined in the DOD law of war manual as: “the principle that even where one is justified in acting, one must not act in a way that is unreasonable or excessive.” In applying this principle to kinetic operations, military lawyers and commanders will weigh the military advantage against the potential unintended consequences. If there is time to do so, as in the case of planned or static targets, then sophisticated modeling may be used, which incorporates attack azimuths, angles of attack, plume analysis, fuse type, and so on. And, where there is less time, as in the case of targets of opportunity or infantry pop-up targets, military operators rely on practice and the experience of having used their weapons before.

However, Stuxnet illustrates that it is a lot harder to make these judgments when dealing with technological weapons that persist and for which the only real model of effect will occur when they are used. In the Stuxnet case, for example, it was possible to determine the scope of the kinetic effect the Stuxnet malware might have on the centrifuges themselves and it was possible to make predictions as to whether the air-locked introduction of the malware might jump the rails. But it was much harder to predict what would happen if the malware jumped the rails and attacked unintended civilian targets. It did, which is one reason the virus was discovered and one reason the weapon might be used again by others. Simply put, cyber modeling is not as accurate and reliable as plume analysis for kinetic weapons. There are more variables, including the unpredictable role of humans and the persistent nature of some cyber weapons.

AI weapons, or AI-enabled weapons may present some of the same challenges. Especially if one is talking about weapons that are persistent or that can act on their own, to escape and evade with human level intelligence or better than human level
intelligence. That is one reason that the Defense Department Directive states that in the event of a loss of communications the systems will not operate autonomously against targets not previously selected by an authorized human operator. Where a weapon has autonomy and autonomous features, is it really good enough to weigh the proportional effect of a weapon's use based on modeling and prediction? Or, should there be some additional burden of proof to demonstrate the safe or proportionate use of the weapon. This, of course, is one of the questions an Article 36 review might consider. The question posed here is whether the doctrine of proportionality should evolve to address AI-enabled systems as well. A further question is whether AI, which is viewed as "brittle" now, will ever be subtle enough to make the sort of intuition, situational, and judgment based decisions many proportionality assessments require.

The principle of military objective may present challenges as well. The DOD Manual defines military objective as: "persons or objects that may be made the object of attack. Certain classes of persons and objects are categorically recognized as military objectives...other objects are assessed as to whether they meet the definition of ‘military objective.’" This principle is easy to apply in the case of traditional World War II types of targets, like tanks, aircraft, and ball bearing factories dedicated to weapons and munitions production (although cover and concealment may make it harder to discern the true nature of a target). But the line is less clear when one is talking about less traditional objects. This is illustrated with respect to the one time debate about whether civilian oil tankers transporting fuel to market from ISIS occupied Syria and Iraq were lawful military objects subject to attack. On the one hand, scholars and the government pointed out that there was a direct link between the selling of the fuel in Turkey and on the black market and the financing of ISIS's terrorist and military operations. The tankers, in effect, had been appropriated by ISIS. On the other hand, other scholars pointed out that the fuel was being transported in civilian tankers driven by civilian drivers who had been engaged in this activity before ISIS occupied the region and likely would be again after ISIS was gone, if they or the vehicles survived to drive another day.

If AI is the transformative military technology of the 21st Century, one searches for the right and trite metaphor—can of worms, Pandora's Box, slippery slope. What then would constitute a military object where AI is concerned? And, how far down the supply chain is the law, and are commanders, prepared to go? If one contemplates a potential target continuum one might put an autonomous weapons system on one end and on the other end an R&D lab at a federally funded research and development center. But where along the continuum might one put the privately owned cloud in which the AI-enabled weapon stores essential data? And, if the majority of AI R&D is being conducted in private industry and at universities when do those facilities become more like ball bearing factories and less like schools?

It is doubtful that governments will spend, or have spent, much time contemplating these questions in advance of conflict or crisis and were they to do so even more doubtful that they would choose to reveal their thinking on this point. They should. At the very least the question warrants internal review and clarity at a time free from the pressures and consequences of conflict. It is at times of relative stability, when there is time and space, that policymakers might more deliberately consider
the intermediate and long-term consequences of any doctrinal shift. Whereas crisis tends to focus decision-making on immediate impacts and needs.

G. Prohibitions and Redlines

In addition to providing core targeting principles the LOAC also includes absolute, or black letter rules. These include prohibitions on attacking hospitals, cultural and historic sites, and religious sites, among others. However, like most law, most black letter rules have exceptions. For example, the misuse of a protected target, such as the use of a school or hospital to store arms or command military operations, does not immunize the site from potential attack. Rather, it can make it a lawful military object of attack; however, it does change the application of the targeting principles, including the necessity of targeting the site and determination as to whether the direct and concrete military advantage in attacking the target are proportional to the anticipated collateral consequences. In other words, the value of attacking the target and its necessity must be considered in light of the presence of civilians and civilian objects. This is one of the strengths of the principles, they adopt to context. This makes sense, otherwise, bad actors would use protected sites with impunity. Instead, they are exposed to both appropriate attack and investigation and potential prosecution for having used a protected site in such a manner.

The first question presented in this subsection is whether States and non-state actors should consider specific black letter rules, including prohibitions, regarding AI-enabled systems, before AI is fully integrated into weapons and weapons systems and thus before states have too large a stake in the outcome of their own research and development to restrict or prohibit any real and perceived AI advantages.

As already discussed, one possible redline might incorporate the principles contained in DOD Directive 5000.1, including the statement that “systems will not operate autonomously against targets not previously selected by an authorized human operator.” Defense has further stated that “DoD does not currently have an autonomous weapon system that can search for, identify, track, select, and engage targets independent of a human operator’s input.” This too, might provide the basis of an international legal redline, albeit one that is hard to verify when and where software code can be easily hidden and modified. Other redlines might be addressed to:

- Linking AI to certain weapons, including nuclear weapons.
- Deploying AI enabled systems, or autonomous systems into certain domains, like space.
- Requiring traceable feedback from and to any weapon relying on AI generated target data or queuing.
- Designating additional sites and objects as inherently non-military in nature, provided they are not used for military purposes.

18. Roadmap, p. 22.
A further question is whether states and non-state actors who are not engaged in AI weapons development, or the AI arms race, might nonetheless seek to regulate the field or impose specific prohibitions. This is the subject of various expert group meetings convened under the auspices of the UN and the States Parties to the Convention on Certain Conventional Weapons (CCW).

The CCW was concluded in 1980 and entered into force in 1983. As the full title of the Convention indicates (“Convention on the Prohibitions and Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects as Amended on 21 December 2001”), the purpose of the Treaty is to create a framework for banning or restricting the use of weapons that cause undue suffering or are indiscriminate in nature. Protocol I to the Treaty, for example, adopted with the Treaty along with Protocols II and III consists of one sentence: “It is prohibited to use any weapon the primary effect of which is to injure by fragments which in the human body escape detection by X-rays.” Protocol II, and Amended Protocol II (May 3, 1996), are addressed to minimizing the impact of mines and booby-traps on civilian populations by prohibiting among other things the deployment of booby-traps that look like toys, or are attached to dead or wounded persons. The Protocol also requires the recording of the placement of mines and booby-traps. Protocol III prohibits the use of incendiary weapons against civilians and civilian objects. Protocol IV (July 30, 1998) prohibits the use of blinding laser weapons prohibits the use of laser weapons that are designed as their sole combat function or one of their combat functions to permanently blind. Protocol V is addressed to the explosive remnants of war and the duties of combatants to protect the civilian population and humanitarian missions following the conclusion of conflict (November 28, 2003). The Treaty entered into force in 1983 and has 125 State Parties, including China, Russia, the United States, and Israel. The first three states are parties to all amendments and protocols to the Treaty; Israel is not.

Based on a 2013 decision of the State Parties an informal group of experts has met since 2014 to consider a potential ban on Lethal Autonomous Weapons (LAWS). In 2017 a more formal Group of Governmental Experts was established to consider the topic, indicating an intent, at least in diplomatic circles to stick with the topic. The talks have been popularly characterized as an effort to ban killer robots.

So far, the GGE has met approximately twice a year. Four positions have emerged. First, certain states generally associated with the one-time, non-aligned movement, but notably including China, have advocated on behalf of a binding legal restriction on LAWS. At the September 2018 GGE meeting, twenty-six states favored this position. Second, other states, largely associated with the EU favor a political declaration regarding LAWS, but not an outright legal prohibition. A third group of states, favors requirements for positive human control over critical functions of weapons systems. And, a fourth group of states, is opposed to a prohibition, legal or political, and generally favor at least in diplomatic context, some form of annual weapons review akin to but not necessarily required by Article 36. Of course, there is a fair bit of fluidity between the positions as well as within the positions. Russia and the United States, for example, might both be characterized as being against a LAWS ban, but their positions, outlook, and interests are hardly aligned. In addition
to divisions on fundamental positions, the talks have also revealed differences and difficulties in defining key terms, like “man in the loop” and LAWS itself. Nor have the experts begun the arduous task, and thus resolved, how in any event such a ban would be verified.

In short, efforts at regulating AWS in the context of the CCW have been evasive. Those states and entities that wish to regulate AWS might do well to consider some of the lessons learned from earlier efforts to regulate land mines. The Ottawa Treaty prohibits the use, transfer, and stockpiling of anti-personnel land mines (APLs). Although the Treaty is not without challenges and limitations, it is among the most successful discrete efforts to limit a means of warfare. From the standpoint of AI, it is perhaps most apt because of the way it was negotiated and concluded, and the speed with which it was concluded. In brief, the treaty was the product of a grassroots campaign that synergized the energy of NGOs, prominent private actors, the public, and like-minded states willing to take the negotiations outside normal multilateral channels to bypass or overcome roadblocks. One lesson from Ottawa—while helpful, the participation of the big three—the United States, China, and Russia—is not always necessary to accomplish significant arms regulation.

“The Convention on the Prohibition on the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and Their Destruction” (1997), more widely known by its popular name, the Ottawa Treaty, as its title suggests, prohibits States Parties from “using, developing, producing, otherwise acquiring, stockpiling, retaining, or transferring to anyone, directly or indirectly, anti-personnel landmines, or to assist, encourage, or induce, in any way, anyone to engage in any activity prohibited to a State Party under the Convention.” (Article 1). In addition, states are required to identify, remove, and dispose of land mines on their territory within ten years of becoming Parties, with the assistance of other State Parties if requested. An APL under the Convention is defined as

\[A\] mine designed to be exploded by the presence, proximity or contact of a person…. Mines designed to be detonated by the presence, proximity or contact of a vehicle as opposed to a person, that are equipped with anti-handling devices are not considered anti-personnel mines…. [Art. 2.1].

As of December 2017, there were 164 States Parties to the Convention, including many of the United States’ closest allies, including Australia, the United Kingdom, and Canada. Thirty-six UN members are not Party to the Treaty, including the United States, Russia, China, Iran, Israel, and India.

The Treaty is credited with helping to reduce the number of APLs that maim or kill civilians each year in conflict zones and former conflict zones. It has spurred twenty-nine states to eliminate APLs from their weapons inventory and dispose of all APLs on their territory. The Treaty has also established a norm against the use of APLs, whether that norm is accepted as customary international law or not. However, the Treaty has shortcomings as well.

First, critical actors are not parties to the treaty. For the United States this reflects a national security determination that it could not meet its defense commitments on
the Korean peninsula without the presence of APLs in the DMZ and in contingency reserve. In the DMZ, for example, it is estimated that the two Koreas have deployed over one million land mines, which is one estimate of North Korean land mines alone. Thus, even if the United States were a party to the Treaty, it could not meet its obligations in the context of its alliance with South Korea and joint command structure. Because of this U.S. practice and that of other states, the U.S. takes the view that the Ottawa Treaty does not reflect customary international law binding on non-States Parties.

Second, the treaty is limited by its terms to APLs and not to mixed munitions, which are used against both vehicles and personnel, like cluster munitions, or anti-vehicle mines. These munitions kill or maim thousands of civilians notwithstanding their designation as mixed or anti-tank/vehicle.

Third, the treaty lacks a compliance mechanism, like the OPCW in the case of chemical weapons. This means that in the case of the thirty-four states that have fallen behind on their ability to identify, collect, and destroy APLs on their territory there is no central authority or mechanism to obtain technical or financial assistance.

Fourth, this also means there is no standing enforcement and verification mechanism. Some States Parties continue to use APLs, while others stockpile APLs or discuss doing so in the event of a national security need. Turkey, for example, continues to maintain APLs along its border with Syria, a border over which refugees as well as combatants in Syria cross or attempt to cross. Parties to Libya’s conflict have also deployed APLs. And, it is likely that other states may do the same where there are real or perceived threats. For example, Russian aggression in Crimea and Eastern Ukraine, as well as against Georgia and the Baltic states, may yet place additional pressure on the Baltic State Parties to reconsider their commitment to the treaty, or at least the prospect of reserving a right to deviate in the case of national security exigency. Indeed, this experience may impact the degree to which they are willing to sign on to absolute prohibitions in other contexts as well.

Finally, the Ottawa Treaty applies to States Parties and not to non-state actors. Thus, it has had no bearing on the widespread use of IEDs, unquestionably a form of APL, by non-state actors, in Syria, Iraq, Afghanistan, and Yemen among other places. It would be naïve to think that many, if not all, of these non-state actors would follow treaty or customary international law even if it applied. The point is to illustrate one significant limitation of such treaty-based mechanisms—their application to private actors. Of course, the use of IEDs by non-state actors might violate other law of armed conflict norms subject to criminal sanction, such as murder and indiscriminate killing.

What makes the Ottawa Treaty potentially of greatest interest in an AI context, beyond the general lessons about arms control it offers, are its origins and its effect on U.S. practice. The original efforts to negotiate a land mine treaty occurred within the structure of the Conference on Disarmament (CD) in Geneva. Thus, the negotiations consisted of the usual U.N. parties as well as the accompanying dynamics of U.N. structure. The CD, one might say, was not a coalition of the willing, or like-minded states; but rather a U.N. conference with all the endemic disincentives that break consensus, and drive U.N. efforts to lowest common denominators and division.
Certain states sought to avoid this dynamic by taking the negotiations over an APL Treaty outside of the U.N. to what was essentially a conference of the willing. Canada agreed to host the process, which was spurred by the personal interest and commitment of Canadian Foreign Secretary Lloyd Axworthy. This process coincided and was in part a product of two parallel events. The first, and most important, was the grassroots effort starting in 1991 on the part of individual actors and NGOs to pressure states, and then the international community, to ban land mines. The most visible, but not only, NGO in this effort was the International Committee to Ban Land Mines (ICBL), led by Jody Williams, for which the ICBL would receive the Nobel Peace Prize in 1997. The campaign also had the support of a letter-writing campaign and the visible support of popular figures such as Diana, Princess of Wales.

At the same time, and in parallel, several states were on their own banning APLs under domestic law. This started with Belgium in 1995, which banned the use of APLs, but not their stockpiling in the event of subsequent need. Austria was the first state in 1996 to prohibit APLs outright, including their possession, and not just use. It was also an Austrian diplomat, Dr. Werner Ehrlich, who proved instrumental in drafting and providing treaty text at decisive diplomatic moments, when ideas needed to move from concept to text to move forward.

Ottawa is also noteworthy that while the United States did not join the treaty for its own (and South Korea’s) national security reasons, Ottawa nonetheless influenced United States behavior. Among other things, in 1997 the United States adopted a policy of only using APLs that automatically self-destruct and to no longer use APLs outside of Korea. Indeed, most APLs in U.S. inventory disarm or self-destruct within four hours, and thus do not remain a persistent threat to either civilians or combatants. The exception to U.S. policy is on the Korean peninsula where South Korea uses persistent mine fields to deter invasion from North Korea, and if need be, channel any attacking forces.

1. **Outstanding Questions**

1. **When, if ever, has a state or states foresworn the development of new weapons or given up weapons already within their arsenals?** It has happened, but not often. The United States, for example, unilaterally renounced the development, retention, and use of biological weapons in 1972. Most of the rest of the world followed in the context of the Biological Weapons Convention. Likewise, some states, but not all, have foregone the prospect of possessing nuclear weapons in the context of the Nuclear Nonproliferation Treaty. Some states that possessed nuclear weapons, or were on the verge of doing so, have also given up those weapons or renounced the capacity to obtain them. Policymakers should ask what conditions might make it more or less likely that one or more states might do so today in the context of AWS or other AI enabled systems. This threshold question leads to following policy questions.

   A. Should the United States and/or the international community—unilaterally, bilaterally, or multilaterally—forego or renounce an AI Competency (e.g., AGI/ASI) use (e.g., AWS/military/defense) or domain (e.g., space/cyber)? And, if
so, what is the most effective vehicle by which to do so, e.g., international law, domestic law, Executive directive, or policy statement? And then, whether such efforts should be pursued on a unilateral, bilateral, and/or multilateral basis?

B. Stepping beyond the CCW, will other states seek to regulate the use of AI as a weapon or national security instrument if the governments of the United States, China, and Russia do not? At present, the race for AI technology is propelled by three states, although other states clearly are engaged in AI research as well. Given the apparent, and perceived, risks associated with LAWS, one should ask whether other states should, or will, band together to regulate AI outside the CD process. There is precedent for doing so as the Ottawa Treaty illustrates. As noted, states have also sought, with less success to use the Conventional Weapons Treaty framework to address LAWS. Each of these instances help to identify the challenges, benefits, and shortfalls of multilateral control efforts where the outliers are the states with greatest impact.

2. Are there arms control lessons applicable to AI? Arms control, including weapons-renunciation, is usually predicated on reciprocal application, which in turn depends on verification. AI presents distinct, even unique control and verification challenges given its ubiquitous, dual-use, and opaque nature; how do you verify software and algorithms? Nonetheless, are there lessons to learn from earlier efforts to control dual-use and difficult to verify capacities? Three capacities seem particularly apt to consider: (1) cyber weapons; (2) biological weapons; and, (3) chemical weapons. Each in ascending order represents a higher degree of successful regulation and verification.

A. Should the United States or international community establish an Agency, domestic, international, or both, to regulate the development of AI and/or ensure its use for specific or delimited governmental purposes?

B. Should the United States develop a doctrine for AI in National Security context, to include principles of command and control?

C. Should the United States, and/or select actors, and/or the international community, develop Rules of the Road and other Confidence Building Measures in the AI field? If so, what form should these endeavor(s) take? And, if not now, when?

H. Perfidy and Ruse

A ruse is lawful trickery. Or, more precisely as defined in the 2015 DoD Law of War Manual, “acts which are intended to mislead an adversary or induce him to act recklessly but which infringe no rule of international law applicable in armed conflict.” The original and classic ruse, in every sense, was the Trojan Horse. D-Day also offers numerous examples of lawful, and successful ruses, designed to fool or confuse the Germans as to the time and location of the Allied cross-channel landings. These included the use of inflatable tanks and aircraft (a ruse still used today); the assignment of George Patton to Scotland to command a fictitious Army Group aimed
at Calais; and, on D-Day itself, the use of metal particles to mimic aircraft formations on radar, and mini-paratroopers with firecrackers to simulate drop zones.

At its core, the prohibition on perfidy seeks to outlaw and make punishable as a war crime, the use of protected symbols to gain an opponent's confidence in order to kill them. Thus, Additional Protocol I, Article 37, defines perfidy as “acts inviting confidence of an adversary to lead him to believe that he is entitled to, or obliged to accord, protection under the rules of international law applicable in armed conflict, with intent to betray that confidence.” The DoD Law of War Manual defines perfidy in similar manner stating, “The key element in perfidy is the false claim to protections under the law of war in order to secure a military advantage over the opponent. The claim must be to legal protections.” Classic examples of unlawful perfidy involve the use of internationally protected symbols to gain advantage over an opponent like use of a false flag of surrender or the wearing of a Red Cross or Red Crescent emblem by combatants.

However, perfidy reaches more than protected symbols, and more than combatants. It is also generally viewed as treacherous to use assassins, assassination being a form of treacherous killing, although there is no general agreement on what exactly assassination means. Further, some definitions of perfidy appear to reach further and protect, “individuals belonging to the hostile nation” and not just members of its army.19 In addition, perfidy bars the wearing of the uniforms or emblems of neutral countries not party to a conflict.

However, there is debate about whether and when an adversary can wear the uniform of one's opponent. The DoD Law of War Manual takes the view that POWs can lawfully wear an adversary's uniforms to effect escape, and “outside of combat.” Article 23 of the 1907 Hague Convention prohibits the “improper” wearing of an enemy combatant's uniform. Wearing an opponent's uniform in combat is viewed as treacherous. But there is ample debate on what is considered “outside of combat” and the degree to which an adversary's uniform can be worn to gain the element of surprise and stealth. At minimum, while it may not be a war crime under the rubric of perfidy; it does expose the perpetrator to the loss of combatant status and potential exposure as a spy. This was the case, for example, with the German soldiers, under the command of Otto Skorzeny at the outset of the Battle of the Bulge who wore American uniforms and generally sought to sow confusion and chaos behind American lines by cutting communications lines and directing traffic in wrong directions. Skorzeny was later charged with a war crime for this conduct but was acquitted for lack of direct evidence that he gave the command for German soldiers to wear U.S. uniforms while in combat rather than for purposes of infiltration and deception. One defense witness was a British SOE member who testified that SOE agents wore German uniforms as disguises behind enemy lines. However, eighteen German soldiers captured during the battle wearing American uniforms were executed after brief military trials.

19. Article 23(b) 1907 Hague Regulations; The Law of Land Warfare, FM27-10, Dept. of the Army Field Manual, par. 31, p. 17 (July 1956).
There remains active debate among experts on whether the wearing of an adversary’s uniform is a war crime, or just risky. It is unresolved, in part, because the major powers that tend to drive the law of armed conflict debate may have a foot in both camps. One does not need to look hard to find media pictures of Special Forces units from multiple countries dressed in local attire to blend in. Local attire may also be the “uniform” of the adversary if, for example, one is fighting the Taliban or ISIS. The circumstance is somewhat like cyber debates where the law remains opaque because the driving states are not quite sure whether they want offense or defense to prevail, and thus they leave the law vague and contextual, permitting situational choice.

Virtually no commentators, or for that matter U.S. policymakers, push back on the prohibition on using protected symbols. Other black letter law applications of perfidy may at times seem a bit oxymoronic. If the overall purpose of international humanitarian law is humanitarian, wouldn’t it minimize civilian suffering to permit the assassination of key actors, which might otherwise help to end a conflict? However, the perfidy prohibition makes more sense in this context if one considers the two root purposes of the prohibition. First, the law is intended to capture, perhaps in dated terms, concepts of fair play and chivalry. In this regard, it is noteworthy that the 2015 DoD Manual includes “honor” as one of the core principles of the law of armed conflict, along with distinction, necessity, and proportionality. Second, the law is intended to protect civilians. Were assassination lawful, civilians might be inherently suspect as potential assassins, and the distinction between combatants and non-combatants harder to distinguish and enforce. The debate thus returns to the meaning of assassination and whether there is a principled difference between targeting an opposing military leader with a precision munition, a sniper’s rifle, or a car bomb.

1. Outstanding Questions

What does any of this have to do with AI?

First, of course, the law of perfidy and ruse will apply to AI-enabled weapons systems and techniques just as it would apply to any other weapon or effort to deceive the enemy. One cannot use AI capacities to mimic Red Cross call signs or mimic the voices of aid workers. However, it may be less obvious how the law might apply in the context of camouflage and concealment. Might a drone be camouflaged to look like an Amazon delivery vehicle or a recreational toy? Is that a ruse, or only up to the point of actual attack, i.e., entry into combat? Wearing an adversary’s uniform may take on new meaning in the context of weapons systems designed to look like the adversaries or that are the adversaries, because they have been captured and are now controlled by the opponent through AI-enabled cyber means. Is this a form of wearing the other guy’s uniform? And, should such conduct be permitted under the law of armed conflict? What about the use of unwitting third party computers? Is that a ruse? Poor cybersecurity? Or, perfidy in the form of wearing a neutral country’s uniform?

Second, should cyberattacks in the form of phishing or other methods be treated like a Trojan Horse, or perfidy? At this time, they are treated like the former, indeed
some cyberattacks are known as “Trojan Horses.” But is that the right rubric going forward where the potential risks of AI-enabled cyberattacks are greater? Should there be legal redlines, for example, that prohibit acts that inherently undermine stability, or the ability of commanders to responsibly command? Should the use of malicious BOTs, i.e., software applications that can rapidly run repetitive scripted tasks like denial of service attacks, or other methods of attack that mask attribution, be treated as one more cyber challenge, or as akin to wearing the uniform of a neutral party, which is perfidious under the law of armed conflict?

Third, what about AI applications that undermine stability, in peace as well as in war, and in between. For example, should it be viewed as perfidious to interfere with or run false flag operations against a state-to-state hotline? What about interfering with nuclear command and control mechanisms or systems that help to predict weather or alert civilians to emergency circumstances? Should these systems be off limits from AI intrusion, and if so, should those restrictions apply in peace as well as in war and everything between? And, in any event, would any of this be detectable or verifiable before it was too late to react, in which case post-facto sanction alone would be the primary deterrent.

For sure, the “law of ruse and perfidy” is narrow and applies solely to international armed conflict. The underlying strategic question posed here is whether states should consider the concepts of perfidy and ruse and apply them to AI now, in peacetime, given the capacity of AI to mimic, before states have sufficiently developed a capacity that would discourage the identification of redlines later. The tactical questions posed here are not rhetorical, with the answer presumed, but to prompt conscious consideration of the new risks that come from AI’s potential to deceive. These are hard questions, but they will not become easier to answer when states develop these capabilities. They will become harder as states seek to protect and preserve real and perceived advantages. They will also become harder to address as some states try to have it both ways by preserving offensive options even as they suffer the consequences of a porous defense.

1. Conclusion

AI comes with great promise and also potential risk. Much of the commentary in this regard has focused on the existential risk AI may present to humanity. This is an interesting debate. It is interesting in substance, for sure, exploring as it does the meaning of intelligence, the concept of humanity, and the functions of the brain. But it is also interesting for its seeming lack of urgency and largely apparent public and policy indifference. Contrast the advent of the nuclear weapons age and the Cold War years that followed where every war college, think tank, and political science department was studying and debating the contours of nuclear war and deterrence. It was the rare school child who was not familiar with terms like “fallout,” “mutually assured destruction,” and “the triad,” even if they were blithely optimistic that a desk could protect them from radiation and thermonuclear wind.

Perhaps the threat to humanity is overwrought. Or, just far enough away to escape our interest and attention. Or, perhaps, a new AI winter will follow the current
AI summer. But that is not likely. There is too much money involved, too many smart people involved, and national security is at stake. AI is here and here to stay. Game on. But if it is too early to join the existential risk camp, it is not too early to consider the near-term impact of AI on military operations. Technology rarely works entirely as intended, at least at the outset. And, scientists in the weapons field, and others, have not demonstrated a long track record of self-regulation when peril and promise converge on the national security road to knowledge.

AI-enabled weapons and trip wires, may increase the risk of mistaken war as well as intended war. This risk already occurs in cyberspace, but heretofore, it has been contained to cyberspace. AI has the potential to combine the risk of Cold War nuclear first strikes, real and perceived, with the immediacy of cyber operations. When AI enables weapons across the spectrum from space to sea, it has the potential to place warfare on a hair trigger, meaning at machine speed.

One way to mitigate this risk and the others identified in these chapters is to engage in a different sort of arms race, a legal and ethical arms race to consider how and when to regulate the pursuit of AI for national security purposes. In U.S. law and practice that mean conscious, informed, and accountable choices and decisions about how the DPA, ISA, and LOAC should apply to AI weapons and systems as currently drafted or through analogy.