Interpreting the Signs

The Prospects of QR Coding the Battlespace

Dr. Manabrata Guha

Dr. Jai C. Galliott

Abstract

Despite the best of efforts, the number of attacks on "protected sites" (hospitals, schools, and other civilian infrastructure of importance) are increasingly with alarming frequency. This article considers this problem primarily from an operational point of view and proposes a specific and implementable "concept-technology" solution involving the use of QR codes/ coding to mark "protected sites" and blockchain technologies to address it. In the process, this article highlights the critical importance of considering seriously the targeting process used by modern militaries in the context of the problem at hand. It also critically examines two recent proposals that have been made to address this problem. The article describes in some detail the architecture, process-flow, and advantages of the solution that it offers vis-à-vis the other currently available options and the ways and means by which emergent combat systems manned and unmanned can, as a default state, incorporate measures by which they can "attend to protected symbols" in complex battlespaces, thereby augmenting and strengthening the United Nations' "deconfliction" mechanism.

Introduction

The Clausewitzian observation regarding the "fog and friction of war" is a well-known, albeit often misconstrued, truism. While, in common parlance, it is indicative of the inevitable turbulence that marks the battlespace, for those in the thick of battle "the fog and friction of war" presents some rather intractable problems. These problems arise not only when contending with the adversaries that they face in battle but also as the modern uniformed soldier strives to wage war in accordance with a code of conduct that is enframed by international laws and conventions.[1] Among other things, this code of conduct involves limiting the potential for damage that may be caused to civilians, noncombatants, and infrastructure and facilities that are not directly and/or indirectly involved in the conflict during the high-intensity operations that characterize the current and emergent conditions of what some have referred to as "accelerated warfare."[2]

Given the complexity that marks the modern battlespace, contending with this responsibility to avoid and/or to prevent damage and destruction to civilians and civilian infrastructure/facilities presents the military with an ethical problem, which not only directly impinges on its operational capability and efficiency but which also, according to some, puts the obligation of incurring higher risks on military personnel.[3] This is because as the battlespace expands to encompass urban and populated areas there is a concomitant increase in the blurring of the distinction between "the civilian" and "the military." Thus, as modern militaries increasingly strive to enhance their speed, agility, lethality, and precision-strike capabilities in a bid to shrink their own Observation, Orientation, Decision, Action (OODA) cycle vis-à-vis that of their adversaries, this blurring of the distinction between civilian and military puts an inordinate amount of pressure and responsibility on how military leaders plan and execute operations and, specifically, on how they design and operationalize targeting capabilities.[4] Thus, increasingly, what is at stake for modern militaries is the need to achieve their operational aims while simultaneously upholding their commitment to respect and act in accordance with the international laws and conventions that guide the prosecution of war.

Aside from contexts involving nuclear warfare strategy where discussions continue on the subject of "force" (i.e., military) and "value" (i.e., population center) targeting, we appear to have moved on particularly in the conventional warfare context from an age in which we saw the use of indiscriminate wide-area bombing campaigns, which intentionally attacked civilian targets as a strategic end. In light of this, while it may be plausibly argued that our collective sensibilities about mass attacks, particularly against civilians and civilian infrastructure, have improved considerably and that laws, conventions, and procedures have been created, changed, and strengthened accordingly, there is, however, evidence that such kinds of attacks continue to persist in the modern battlespace.[5] Thus, for example, *The New York Times* reported that in May 2020, within a span of "12 Hours, 4 Syrian Hospitals [were] Bombed." The article further reported that "Physicians for Human Rights, an advocacy group that tracks attacks on medical workers in Syria, has documented at least 583 such attacks since 2011, 266 of them since Russia intervened in September 2015. At least 916 medical workers have been killed since 2011."[6] Such incidents led the UN to constitute a board of inquiry (BOI) in April 2019 to investigate these and related occurrences, though it warrants mentioning that the BOI's report was in itself disturbing as it failed "to identify the role of the UN in facilitating attacks that it intended to prevent and [did not show] how the UN can avoid doing so in the future."[7]

What is perverse about these and similar incidents is that they have taken place (and continue to take place) even though the UN has a specific process *deconfliction* that is intended to prevent such attacks. Deconfliction, which is essentially, an information-exchange mechanism, has been specifically designed to protect facilities such as schools and hospitals, which enjoy protected status under international humanitarian law (IHL). The mechanism involves requiring those operating such facilities to share their coordinates with the UN Office for Civil and Humanitarian Affairs (OCHA), which then shares this information with the warring parties such that they can put them on a no-hit list, thereby protecting these sites from unintentional and accidental attacks. There remains, however, a critical flaw in the design of this process, namely, that adherence to the deconfliction list is voluntary. Moreover, the warring parties could just as easily use the deconfliction list to specifically target the listed facilities for various ends. This was evident in the report of the April 2019 BOI that the UN had constituted.[8] For our purposes, it is important to pay attention to three significant points that emerged from the UN's BOI report. First, the report clearly states that given the safety and security concerns of its UN OCHA personnel, the presence of UN officials was limited to nonexistent at the targeted sites.[9] Second, the UN OCHA is organized in a stovepiped manner, thus inhibiting the sharing of information within the organization, which led to the reports of the attacks not being recorded, verified, and thus investigated.[10] And third, more concerningly, the report observed that though the purpose of the deconfliction mechanism was to identify and protect humanitarian sites, the mechanism was not intended to be a "protection tool."[11] What this suggests is that regardless of the good intentions of the UN to protect sites of humanitarian consequence, the currently available mechanisms are woefully inadequate. The problem is not simply confined to the inadequacy of the UN mechanism in question. As we segue into an age marked by the proliferation of autonomous weapon systems (AWS) and other technologically advanced machines of war, which, it is often speculated, will acquire the capability to make targeting decisions independent of human involvement and to act on them, we can expect this problem to be further exacerbated. What is necessary, therefore, is a solution or, at the very least, a pathway to a solution that can address this problem.

Given the sensitive nature of the problem and the stakes involved, it is not surprising that there have been various attempts to address this issue. Yet, these attempts have some significant conceptual, methodological, and technical drawbacks. Without undermining the intent underwriting these attempts, we cannot help but observe that, for the most part, while they focus on the humanitarian and ethical aspects of the problem, which certainly warrant attention as they are the

principal concern, they either overtly or implicitly undervalue the military-operational stakes involved, which is surprising given the context in which the solutions are offered. This is perplexing because the problem of accidental (or even intentional) attacks on humanitarian sites is one of targeting, which is a key military-operational capability. Thus, any solution that does not account for the very real complexities involved in the targeting process that militaries must contend with under accelerated warfare conditions and which does not find ways and means to address the problem at hand without compromising a key military capability—targeting—will fall short on two counts. First, it is unlikely that militaries will be amenable to a solution that materially hinders their operational capability, particularly one that is, quite literally, foundational to military operations; and second, a solution that does not take into account the targeting process and/or which underestimates it will likely suffer the same fate as the moribund option provided by the UN deconfliction mechanism—thus, rendering it, like the UN mechanism, more a theoretical model rather than as a "protection tool" of consequence.

This article offers an alternative in the form of a concept-technology solution, which is distinguished by a number of specific features: (1) it remains cognizant of and sympathetic to the military-operational needs—with specific reference to targeting—under modern combat conditions; (2) it seeks to provide a solution that is not only applicable to manned platforms but also, importantly, to the growing number of AWS that are and may be expected to populate the emergent battlespaces of the twenty-first century; (3) it employs a concept-technology pairing that is not abstract in nature (meaning, the "concept" is well-known and not obtuse or controversial, and the technologies involved are readily available and may require, at the most, only minimal reengineering prior to being deployed to achieve the requisite ends); and (4) it provides a means by which the inadequacies of the UN's deconfliction mechanism may be directly addressed in terms of preventing attacks on "protected sites," thereby upholding the core tenets of the IHL and, thus, serving as a viable protection tool. One additional benefit of the solution offered by this article is that it also showcases a way by which ethically grounded concept-technology pairings may be imagined and designed in and for the strategic-military context.

To this end, this article will first as a context-setting exercise briefly discuss the nature and character of the targeting process in the current and emergent military-operational environment. It is necessary to examine this because kinetic effects as experienced and witnessed in the battlespace are a direct outcome of the targeting process and capability. Thus, a clearer understanding of how the targeting process works, its implications (both operational and ethical), and where an intervention

may be possible (and/or warranted) is critically important. Second, considering this context-setting exercise, this article will critically examine two recent proposals that have sought to address the problem at hand. The aim of this examination is to point out how, despite their undoubtedly laudable intent to advance the humanitarian and ethical cause related to attacks on protected sites, these proposals exhibit some important conceptual shortcomings, which includes eliding the critical matter of engaging with military-operational considerations. When considered holistically, these shortcomings and omissions undermine the potential for adoption and implementation of these proposals. Third, having critically examined these two current and representative proposals, the article will then outline in some detail the concept-technology alternative that it offers. It will do so by (1) briefly discussing the technologies it seeks to use and the justifications for doing so; (2) describing the concept-technology pairing, which will include identifying how and in what ways it influences the targeting process to comply with the requirements of the IHL provisions related to protected sites without compromising this critical military-operational capability; and (3) indicating how and in what ways the suggested concept-technology pairing improves the UN deconfliction mechanism. The article will conclude by reiterating how the solution that it proposes serves not only as an example of how the problem of preventing attacks—accidental and otherwise—on protected sites may be addressed but also, from a wider strategic point of view, how such concept-technology pairings serve as examples by means of which the ethical design of militarily oriented solutions may be promoted, which contribute to the development of "trusted military systems"—autonomous and otherwise.

On Targeting: Setting the Crosshairs

As Merel Ekelhof cogently puts it, "[t]here seems to be a considerable lack of knowledge and understanding about targeting among individual members of the public, as well as many groups that represent the public in some way, such as lawyers, nongovernmental organizations, political leaders, industry, scientists, and the press."[12] This lack of knowledge, when coupled with the exponential increase in the sophistication of citizen-based media, which often allows for the production of "Insta-News" and leverages "the network effect" to report on battlespace events in near real time, has, in large part, fostered an environment that has brought the *outcomes* of the targeting function of militaries under extremely close scrutiny. While in many instances this close scrutiny is warranted and serves to hold to account the actions of defense and security policy makers and their military counterparts, yet, it remains, for the most part, oblivious to the extraordinarily complex task of targeting.[13] That said, and precisely because targeting is one of the

core functions of the military involving, "essentially[,] . . . practice of destroying enemy forces and equipment,"[14] it is also an absolute necessity to keep in mind that it "is the *sine qua non* of the international law of armed conflict because intrinsic to it are the central tenets of civilized combat: distinction, proportionality, military necessity, and humanity."[15]

Interestingly, targeting as we recognize it today is a relatively new phenomenon, accompanying the advent of aerial warfare. Prior to that targeting though always an important component of warfare was a linear, relatively unsophisticated and tactically oriented function/process. The advent of aerial warfare, however, changed that. Aerial warfare progressively enabled combatants to take the conflict often beyond the immediate battlespace and deep into the enemy heartland. As this ability matured, concurrently, the targeting function found itself becoming increasingly sophisticated, nonlinear, and acquiring a strategic dimension. This was reflected in the nature of the outcomes that the targeting function sought to achieve. While targeting retained its tactical relevance in terms of serving to destroy an adversary's armies and equipment, increasingly, given the expansion of aerial warfare capabilities, it also began to be used to create "strategic effects" that could influence the behavior of an adversary. Thus, for example, the attempts to "blitz" London in a bid to compel the United Kingdom to recognize and accept the futility of continuing the struggle against Nazi Germany though the attempt failed and the sustained Allied bombing campaigns against Nazi Germany's industrial and population centers and against Imperial Japan's major cities in the East Asian theater are cases in point.[16]

The shocking experiences of these strategic bombing campaigns—including the unprecedented use of the two atomic bombs against Imperial Japan—coupled with the mass casualties, military and civilian, sustained over the two world wars and the subsequent wars in Korea, Vietnam, and elsewhere led to a considerable reevaluation of the merits of waging indiscriminate forms of warfare. These concerns and reevaluations found expression in a number of international agreements, such as the Nuremberg Charter, the 1977 Protocol I to the 1949 Geneva Convention, and so forth. This trend continued in the 1990s, as evidenced by the public outcry at the extent of the mass civilian casualties that were sustained, particularly in the wars that followed the break-up of Yugoslavia.[17] Simultaneously, however, there were developments underway in the military-technology domain that aimed to co-opt the use of microelectronics and the then still-nascent information and computational technologies to develop "smart" and "precise" weapons.[18] These advances in military technologies, particularly those that fueled the design and development of precision-guided munitions, were underwritten by at least two considerations. The first was military necessity. Having recognized the futility of

mass bombing campaigns and recognizing the benefits of assessing potential adversaries in systemic terms, strategic-military planners began revisiting the concept of the "extended battlespace" and the possibility of interdicting critical nodes of an adversary's war-waging capabilities in a bid to degrade his fighting potential.[19] The second reason was humanitarian in nature. Militaries also recognized the weight and importance of public opinion and began to sensitize themselves to it and to the need for waging war in a more discriminatory manner in a bid to reduce civilian casualties and to avoid inflicting damage to civilian infrastructure.[20] This, in turn, led to refocusing more closely on targeting.

In an Annex to "the keystone document of . . . joint operations . . . [which] . . . provides the doctrinal foundations and fundamental principles that guide the Armed Forces of the United States,"[21] targeting is explained as "the process of selecting and prioritizing targets and matching the appropriate response to them, taking account of command objectives, operational requirements and capabilities."[22] Taking care to clarify that targeting is a process that is systematic, comprehensive, and continuous, the document goes on to point out that when combined with "a clear understanding of operational requirements, capabilities, and limitations, the targeting process identifies, selects, and exploits critical vulnerabilities of target systems and their associated targets to achieve the commanders' objectives and desired end state."[23] As such, targeting may be understood as being "the deliberate application of capabilities against targets to generate effects in order to achieve specific objectives . . . [thus representing] the bridge between the ends and means of warfare."[24] The processual nature of targeting involving planning, tasking, executing, and assessing is suggestive of the "logical progression that forms the basis of decision-making and ensures consistency with the commander's objectives and the end state."[25] This is represented by the diagram below:

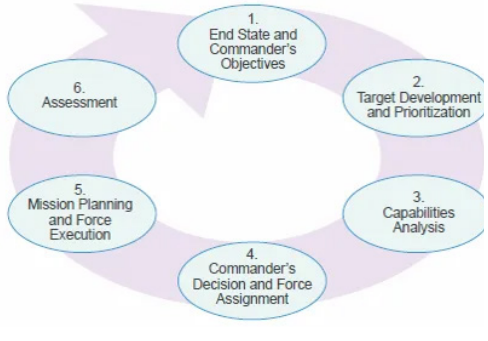

Figure 1. The targeting process. (Source: Annex 3-60, 10.)

In this context, the doctrinal documentation of the US military relating to targeting makes an intriguing point. It suggests that those who are engaged in discharging the targeting function are charged with predicting (or anticipating) and estimating which actions carried out by what means will satisfy the commander's intent. This requires them to fuse inputs received from intelligence, surveillance, and reconnaissance (ISR), strategic assessments, and operational planning exercises to generate a high level of situational awareness, which is a prerequisite for targeting to be effective and is indicative of the integrative nature of the targeting process. Note that the word *effective* in this context is not simply limited to kinetic and lethal outcomes. It also gestures to nonlethal (kinetic or otherwise) outcomes, which may direct or influence an adversary's actions.[26] At the same time, the targeting process is required to give due consideration and weightage to the "sensitivity" of the target, which is a matter of critical concern when dealing with targets whose interdiction may have potentially negative strategic-political (and humanitarian) consequences such as the targeting of the senior leadership of the adversary, attacking stores of dangerous munitions and equipment (for example, weapons of mass destruction), and, in a context most relevant to us, engaging in combat where the risk of collateral damage (particularly involving vulnerable civilians) is high. This emphasis on the sensitivity of the target underwrites the entire targeting process. Thus, the doctrinal documentation insists that the targeting process is not random and ad-hoc. It is "controlled by strategy, law of war, and rules of engagement."[27]

Critical to the targeting process is the understanding of what constitutes a target. The aforementioned Annex explains: "A target is an entity or object considered for possible engagement or other actions,"[28] which may be "facilities, individuals, virtual (nontangible) things, equipment, or organizations."[29] As such, a target is said to possess a set of distinctive characteristics, namely, physical, environmental, functional, and cognitive. While the first two—physical and environmental—relate to the structure, constitution, and location of the target, the third relates to the functions that the target performs within the adversarial system and its relative importance to the adversary's war-waging capability. The fourth characteristic is primarily concerned with the human element of the adversarial war-waging system, which assumes importance especially in the context of effects-based operations, where the aim is to either disrupt an adversary's command and control system or to influence its behavior to achieve a desired outcome.[30] The Annex then provides what is, in our context, a significant clarification. It states that

> a fundamental tenet of targeting [is] that no potential target derives its importance or criticality merely by virtue of the fact that it exists, or even that it is a

> crucial element within a target system and other interdependent target systems. Any potential target derives importance, and thus criticality, only by virtue of the extent to which it enables adversary capabilities and actions that must be affected in order to achieve the commander's objectives.[31]

This clarification is important because it directs our attention to the primacy accorded to the commander's intent and aims, which is reflective of the "control" that a commander exercises in war, thereby underscoring the critical role of the human in the targeting process, which is a subject that has acquired much attention, particularly in the context of the often speculative discussions surrounding the use of artificial intelligence (AI) and autonomous systems in future warfare.

There are, in essence, two types of targeting: deliberate and dynamic. Despite what these terms may superficially indicate, the *Annex to the Joint Targeting Document* cautions us that "[i]t is a mistake to associate deliberate targeting with fixed targets and dynamic targeting with mobile targets."[32] While the former is directed toward targets that are known in advance and/ or have been pre-identified as existing in a definite geophysical space and are known to have specific functions whose interdiction is assessed as being vital for the prosecution of an operation, and thus have been subjected to detailed planning and development, the latter is directed toward those targets which may not have been pre-identified or known in advance or whose identification may have taken place within a compressed timeframe thus preventing them from being subjected to the target planning process. Deliberate targeting is a structured, systematic, and analytical process whose sequence may be illustrated as depicted in figure 2:

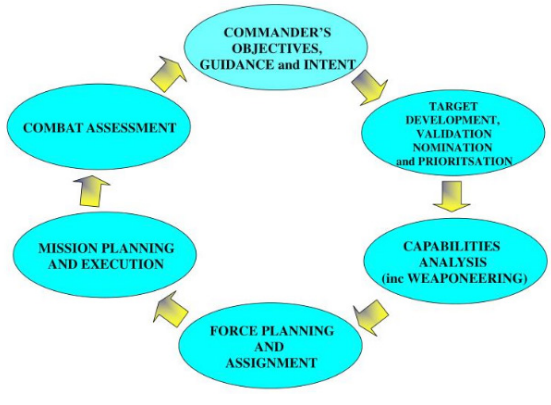

Figure 2. Deliberate targeting. (Source: Bg. H. Walther, "Building Military Corpora for Curricula" [BILC Conference, 2013, Tbilisi, Republic of Georgia].)

While this may convey the impression that the process flow of deliberate targeting is linear and sequential, the reality is that "the targeting process is bi-directional, iterative, multi-dimensional, sometimes executed in parallel, and [as] part of a larger set of processes."[33] Moreover, each stage is not an act in isolation; rather, the stages are closely interrelated and require close coordination. Given that deliberate targeting is directed toward targets whose identity, capability, and importance are known in advance, it generally involves the activation of plans and attack schedules that have been prepared ahead of time in addition to creating "on-call packages or missions that deal with the targets through predetermined CONOPS [concept of operations]."[34] As such, deliberate targeting is generally employed in the opening stages of a battle, where the aim is to neutralize at the earliest possible instance an adversary's offensive and defensive military systems, which may be fixed and/or mobile. One of the stark examples of this in operation was during the US military campaign in Iraq in 2003. The lightning speed with which US (and Allied) air assets neutralized the Iraqi military systems suggests that prior to the initiation of hostilities the US military had conducted a deep and thorough analysis of potential Iraqi targets, which were subjected to the deliberate targeting process. This allowed the US forces to "shock and awe" their Iraqi opponents, which resulted in the US forces acquiring, retaining, and exploiting the initiative in the battlespace.

When considered in abstract terms, while the dynamic targeting process does not differ from the general logic underwriting the deliberate targeting process, there are, however, some significant differences. This is because as we noted above unlike the deliberate targeting process, which is directed toward pre-identified targets, the dynamic targeting process seeks to address targets that were either not pre-identified or were identified too late to be subjected to the deliberate targeting process. The dynamic targeting process involves six specific steps: find, fix, track, target, engage, and assess (F2T2EA), and the process sequence is represented as depicted in figure 3:

While the headings assigned to each of the steps are self-explanatory and descriptive of the activities that take place at each point of the F2T2EA cycle, a brief examination highlights the operative logic that underwrites the process. The first step—*find*—involves detection of a target. This requires what the doctrinal documentation refers to as "clearly designated guidance from commanders,"[35] which implies that there is a prior intelligence input (albeit, perhaps diffused) and a consequent prioritization. This leads to the allocation of ISR resources to detect such targets. On finding the target, a determination is made as to its relevance and the time-sensitivity that may be accorded to it in keeping with the commander's intent. The key point to note here is that despite the moniker—*dynamic*—as-

signed to such kinds of targeting, the find action is not unfocused. The doctrine, thus, specifically states, "Commanders should not task sensors without an idea of what they may collect."[36]

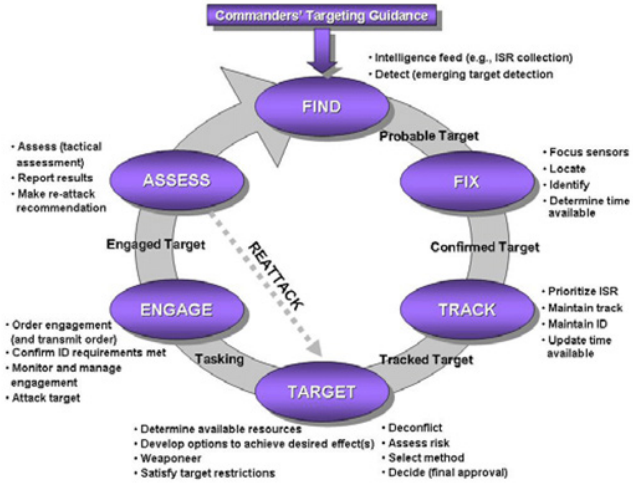

Figure 3. Dynamic targeting. (Source: AFDD 3-60, 48.)

The second step—*fix*—is the concrete determination that a target found by the above method is worthy of serious consideration. This requires the deliberate focusing of various kinds of sensors and other ISR assets to confirm the target's profile and the timeframe within which an active engagement with it may be possible.

The aim of the third step—*track*—is to enhance the situational awareness needed to viably engage with the target. Like the step prior to this, tracking or following the target also necessitates the direction of sensor packages and other ISR platforms toward the target in a sustained manner. This allows for further refinements in the identification of the target, its capabilities, and a continual updating of the situational awareness relative to the target.

The fourth step—*target*—may be considered to be the prelude to the actual engagement of the target. This involves determining the weapons package that will deployed against it (also known as weaponeering) and devising the appropriate targeting solutions that may be required to effectively strike it. This step is also the point at which detailed assessments are made regarding the possible effects of striking the target in terms of potential for collateral damage, determining whether the target is on a no-strike list, and whether the target has been priorly designated

as a "restricted." In effect, at this stage, all measures relating to target validation are undertaken and/or revised.

The fifth stage—*engage*—is the point at which the hostile profile of the target is reconfirmed and the orders to strike the target are issued, which a designated weapon-platform executes.

The sixth and last stage—*assess*—involves assessing the outcomes of the actions taken and the resultant effects on the target by deploying ISR assets to provide immediate feedback. This is critically important because—depending on the circumstances in which the strike is made—there is always the possibility that the strike may not have achieved the desired outcome in full or in part. If the assessment—based on the feedback received through the ISR assets—reveals that the outcomes have not been achieved, a reattack order is generated and executed.

What this brief discussion about the targeting process—involving deliberate and dynamic targeting—reaffirms is that targeting is not an ad-hoc and random activity; rather, it is a systematic and analytical decision-making exercise, which requires a myriad of increasingly granular levels of coordinated actions—each of which is critical to the process and none of which may be considered in isolation—to achieve the commander's intent.

Yet, there is one specific element that plays a central role in targeting, which though being implicit in our discussion, has thus far remained unaddressed. It is also a matter of crucial importance in the specific context of this article. While we noted that regardless of whether the targeting process is oriented toward deliberate or dynamic targeting, the realization of the commander's intent is contingent on—after cycling through the due process—a kill vehicle (attack platform) executing a specific tasking order. Similarly, the triggering of the targeting process and the intermediate steps of validation that co-constitute it is dependent on "the sensor" (or multiples thereof). Put differently, it could be said that the effectiveness of the targeting process is contingent on a sensor-to-shooter link given that, while on the one hand, the triggering of the targeting process is dependent on the sensor, on the other, the achievement of the desired outcome of the targeting process is contingent on the shooter. Alternately, it can be argued that while it is the sensor-to-shooter link that empowers (and validates) the targeting process, equally, it is the targeting process that bears the responsibility to align the shooter to the sensor to achieve the desired outcomes. The diagram in figure 4 is a representation of the sensor-to-shooter link and situates the targeting process in relation to it.

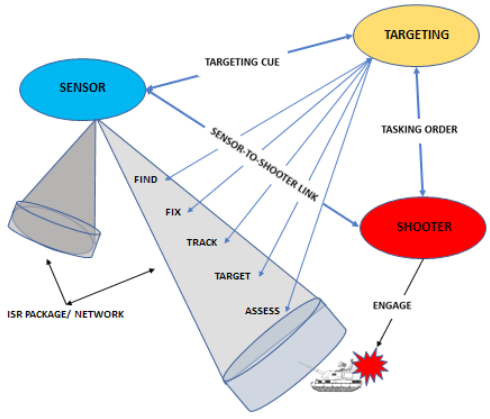

Figure 4. The Sensor-to-shooter link and the targeting process. (Source: Author)

In the context of a multi-domain battlespace, where sensor platforms and shooter platforms may be widely distributed over and across multiple domains, the robustness and efficiency of the sensor-to-shooter link is the key determinant of the effectiveness of the targeting process and, by extension, of a military's combat capability. As such, it is also one of the primary targets for adversarial disruption/interdiction. Recognizing this potential vulnerability, efforts are underway to progressively collapse this sensor-to-shooter link by, among other things, locating both the sensing capability and the shooting capability on the same platform. The advantages of doing so are self-evident. In the first instance, it allows for a near instantaneous reaction. This is particularly true in the case of the deliberate targeting process involving predetermined fixed targets for which preplanned strike packages may have been prepared. Secondly, it allows for the creation of a shock-and-awe effect, which could, potentially, overwhelm an adversary and lead to a rapid degradation of his war-waging capability. And third, it allows for increasing the efficiency of the dynamic targeting process by engaging with targets of opportunity under conditions of compressed timeframes. While the military-operational benefits of collapsing the sensor-to-shooter link may be undeniable, it is important, however, to appreciate the fact that in the context of a collapsed sensor-to-shooter link, the F2T2EA process as discussed above will also, therefore, be compressed. This most certainly will give rise to operational and ethical concerns. The operational concerns would arise because the targeting process would be compressed to a high degree, leading to a potential loss of control for the human operators, while the ethical concerns would be heightened because the

prospect of near instantaneous reaction times may result in unwanted and unwarranted actions that may severely violate the conditions of the IHL and the laws of armed conflict (LOAC).

Given this article's strategic intent, it is necessary at this point to take a step back and assess the salient points that have emerged as a consequence of our discussion of the targeting process, which may be enumerated as follows: (1) with the advent of aerial warfare, the targeting process has transited from being a purely tactical exercise to one that is spread across the strategic, operational, and tactical spaces of war, and is often employed to generate effects as much as to execute kill functions; (2) the targeting process is a structured, systematic, and analytical exercise, which unfolds in a cyclical and iterative manner, and while the targeting process may be rendered in discrete terms, in reality it is a complex process involving continuous interactions between its constitutive parts; (3) the trigger for initiating the targeting process begins with the statement of the commander's intent, which is then broken down into its constituent elements with increasing granularity, reinforcing the fact that targeting is a patently human activity, since it is a means by which a commander's intent (which is premised on perceptions of threats and/or benefits) is realized; and (4) deliberate and dynamic targeting processes are very similar in nature though the latter unfolds at a faster pace than the former and involves six distinct phases (F2T2EA). Further, we noted that implicit in the discussion about targeting is the question concerning the sensor-to-shooter link, which lends a distinct materiality to the targeting process by, on the one hand, providing crucial inputs which informs the commander's intent and, on the other, executing the tasking order to achieve the specific aims and objectives of the commander.

What is equally striking—though perhaps underappreciated—about the targeting process is that throughout the various stages that constitute it, the process remains mindful of the ethical dimension of combat. The evidence of this lies in the fact that, in the first instance, the commander's intent is always (at least, in theory) guided by the dictates of the IHL and the LOAC. Further, to ensure that the IHL and the LOAC are adhered to, continual assessments are made at the various stages of the targeting planning and development process. Thus, for example, this mindfulness of the conditions imposed by the IHL and the LOAC is particularly evident at the weaponeering stage where the appropriate strike packages are created keeping in mind the concerns regarding proportionality and appropriateness. Taken together, this reiterates a point that the doctrinal documentation strives to emphasize repeatedly, namely, that targeting is not a random and ad-hoc process; rather, that it is a carefully considered analytical exercise.

The above analysis notwithstanding, it also cannot be denied—as we noted at the very outset—that incidents involving attacks on protected sites continue to occur with alarming regularity. Thus, two critical questions stand: (1) given our brief exegesis of the targeting process, which revealed its deliberate and analytical nature, why do attacks on protected sites persist; and (2) having examined in some detail the mechanics and dynamics of the targeting process, can an *effective* intervention be made to ensure that such attacks can be prevented? While the first of the two questions may have (possibly nefarious) political reasons and implications, the second presumes that despite the apparently rigorous methodology informing the targeting process, there may exist the possibility of some form of intervention that can augment the targeting process, thereby addressing the problem on hand *more effectively* than is currently the case. This presumption underwrites two recent proposals that seek to employ technological means to intervene in the targeting process and thus warrants our brief critical attention.

Seeking Solutions: A Brief Review of Two Recent Proposals

The preceding discussion about how the targeting process unfolds highlights the systematic, analytical, and detailed steps that are involved in the identification, development, assessment, and engagement of a target. When cast within the context of the perpetual presence of the "fog and friction" of war, targeting emerges as one of the most complex of tasks that a military performs. This is particularly true when we account for the imperative of war fighters to maintain the operational or battle tempo. This is important because warfare, as Clausewitz cogently explained, is a duel between at least two entities who are aiming to outdo the other—both in terms of the capabilities that they bring to bear on each other *and* the speed with which they can act—in battle. Targeting, as we have seen, is also one of the most critical functions of a military, since it involves directly interdicting and degrading an adversary's war-waging abilities. Thus, the speed at which the targeting process unfolds is also of critical concern and is one of the key metrics by which the effectiveness of a military force is gauged. Equally, as we have seen above, it is also precisely for this reason that the targeting process of the military is scrutinized so carefully, since the effects that it generates in the battlespace have real and tangible humanitarian consequences. Thus, any attempt to address humanitarian concerns—aside from measures to reinforce and/or expand the *jus ad bellum* framework—will have to focus on the military's targeting process. This, for the reasons mentioned above, is a sensitive matter for it directly impinges on the effectiveness of a military force.

Labeling itself as Protected Assurance Understanding Situation Entities (PAUSE), the first of the two solutions that we will examine involves "the inte-

gration of two key technologies: blockchain and artificial intelligence (AI)."[37] The solution is proposed in the context of the recognition that "[i]nformation flows in conflict and disaster zones continue to be marked by intermittent communications, poor situation awareness, mistrust and human errors,"[38] and the claim is that deploying blockchain and AI technologies—particularly "Protective AI"—will help "catch human mistakes and complement human decision-making."[39] The conceptual premise of the solution appears to rest on the assumption that "[m]uch of the error of war could be reduced if decision-makers knew more,"[40] which, as the proponents point out, leads their solution to focus "primarily on the challenges of awareness [while acknowledging] that awareness absent humanitarian intent or capability is ineffective and leads to a lack of trust."[41] Contending that "[a]wareness of ignorance is a virtue associated with intellectual humility," the proponents of PAUSE argue that given the conditions of radical uncertainty that marks the battlespace, military decision makers who are subjected to such uncertain conditions are justified "to make decisions when a certain threshold for evidence is met and the perceived risk of inaction is greater than the risk of action."[42] Of course, they also observe that with the increasing sophistication of ISR technologies, militaries are expected to "hold fire under uncertainty . . . [given that the] higher the humanitarian risk, the greater the evidential expectations in accordance with just war principles of discrimination and proportionality."[43]

Essentially, the PAUSE architecture consists of two technological layers. The first involves what has been identified as *Whiteflag*, which is a "digital communications protocol based on blockchain technology that provides a reliable means for both combatant and neutral parties in armed conflicts to digitally communicate."[44] The need for this protocol is justified on the grounds that (1) the profusion of "digital technologies . . . has changed information availability in conflicts . . . [and is] driving a new requirement to share [presumably information] among disparate groups"[45]; (2) there appears to be "very little uptake of message data" since "real-time messaging data being contributed by bystanders and those affected by a disaster has been deemed as unverifiable and untrustworthy, and has not been incorporated into established mechanisms for organizational decision-making"[46]; (3) since "smart phones and social media are readily used for many purposes [including by state/non-state actors, humanitarian groups, local population engaging in citizen-journalism] . . . there is an opportunity for these groups to exchange these new sources of information to better meet humanitarian goals."[47] Noting in passing that the effectiveness of such exchanges is contingent on them being neutral, secure, and providing undeniable proof of receipt, the Whiteflag protocol is promoted as being "a reliable means for both combatant and neutral parties in armed conflicts to digitally communicate."[48] The reliability

of the Whiteflag protocol is assured given that the messaging system is underwritten by "the underlying blockchain."[49] In effect, "Whiteflag operates by sending messages as transactions on a blockchain . . . [thus securing] the messages [and consequently ensuring that] . . . Whiteflag is neutral and cannot be controlled or manipulated."[50] In this way, Whiteflag, it is claimed, provides information assurance and establishes the trustworthiness of the information by providing "instant verification of the originator, authentication of reliable sources, cross-checking facts with persistent information on the blockchain to evaluate reliability of sources, confirmation by multiple sources, duress functionality, and implementation-specific measures such as filtering, blacklisting, and other sources."[51]

The second technological layer of the PAUSE solution is categorized by its proponents under the wide rubric of Protective AI. While the proposal is unclear about what specifically constitutes Protective AI and which particular aspect of Protective AI it wishes to leverage, its proponents gesture to two recent approaches to prospective designs of AI technologies that seek to ensure that such technologies meet ethical and legal standards, namely, MaxAI, which is a maximally-just ethical machine, and MinAI, which is a minimally-just ethical machine. A review of the literature provided by the proponents of the PAUSE solution suggests that they are aware of the speculative nature of the MaxAI solution since it requires a level of "reasoning" that is beyond our current technological capabilities. This accounts for the proponents of the PAUSE solution directing their attention to the MinAI solution. Noting in passing that we will examine the MinAI solution in some detail below, for our present purposes, it is interesting to note that the proponents of the PAUSE solution while, for the most part, remaining cognizant of some of the weaknesses of the MinAI solution, appear to consider it (or something approximate to it) as being representative of Protective AI.[52]

Based on this, the proponents of the PAUSE solution "propose a trusted human-AI network based on the Whiteflag protocol and Protective AI . . . [whose] architecture mirrors trust relationships between military and civil authorities to increase efficiency and timeliness of information processing and exchange."[53] According to them, their architecture also "makes use of AI and automation to extract, clarify, identify, categorize, locate, assess, and most importantly fuse information from eye-witness sources (with variable trustworthiness) to improve the accuracy and accountability of decision-makers."[54]

Aside from noting the paucity of details about exactly how the PAUSE solution would operate under real-life conditions, there are a few observations that we can make that are pertinent to the strategic objectives of this article:

1. The PAUSE solution, while paying lip service to "decision makers," does not identify who they may be and where they may be located. Given that at

least one of the areas where the PAUSE solution may be expected to be deployed is in a high-intensity combat situation, surprisingly, there is no attention paid to the targeting process. Thus, the following questions stand: Where will the PAUSE solution be implemented? How and in what way will the PAUSE solution impact the targeting process? Will the PAUSE solution provide a refined product, which the targeting process can incorporate easily into itself, or will it require further processing after it is received by the targeteers?

2. While the use of citizen-based inputs may be helpful, the timelines involved in culling authentic information from such inputs remain unclear. This is especially critical since such inputs—as per the PAUSE solution—will co-constitute the data that the military will use during the target development process. The risks are simply too high.[55]

3. From the documentation provided by the proponents of the PAUSE solution, it is unclear whether the Whiteflag protocol is open-source or proprietary. If it is the latter, then invariably the question will arise whether such a closed protocol will be advisable to use in matters relating to sensitive contexts such as targeting and the protection of sites of humanitarian importance. Moreover, it appears that the inclusion of the Whiteflag protocol is made to suit the PAUSE solution. In other words, there appears to be no overriding necessity to specifically use the Whiteflag protocol. If this is indeed the option to be taken, then there are other similar solutions available or, indeed, a tailor-made solution may be constructed under the watchful aegis of an internationally recognized body such as the UN OCHA or the International Committee of the Red Cross (ICRC).

4. As we will see when we review the MinAI solution, a key component of the PAUSE solution is the use of convolution neural networks (CNN) as applied to computer vision. The PAUSE solution gestures toward the more recent developments in the field of region-based convolution neural network (R-CNN) technologies but gives no indication as to how and where such technologies and their related processing systems will reside. This is a critical consideration, since operationalizing such a solution will require integrating it within the battlespace. The proponents of the PAUSE solution leave unaddressed precisely how this can be done.

In sum, therefore, while the intentionality underwriting the PAUSE solution is undeniably positive, as an implementable solution, there are many operational-level questions that remain unaddressed. Equally, at the conceptual level, the solution appears to be lacking a thorough appreciation of the critical importance that targeting plays in the context of modern warfare. As we have noted earlier, any

proposal that seeks to either intervene or influence the targeting process (and prowess) of a modern military force must necessarily account for how its integration will impact strategic and operational-level competencies. This is not a matter that any military organization will (or should) take lightly, since it directly impinges on its ability to discharge its mandated duties.

The second proposal that we will consider is the MinAI solution, which proponents of the PAUSE solution also invoked. Compared to the former, the MinAI solution is a more elegant and conceptually robust proposal. In effect, it identifies what it refers to as "the ethical machine spectrum," at one end of which lies "maximally-just autonomy using artificial intelligence (MaxAI) guided by acceptable and nonacceptable actions [which] has the benefit of ensuring ethically obligatory action . . . [while at the other] a constraint-driven system . . . [allowing] what is ethically impermissible . . . [based] on the need to identify and avoid protected objects and behaviors."[56] What is interesting about the MinAI proposal is that not only does it resist falling into the trap of invoking the need for what Ronald Arkin has referred to as an "ethical governor"[57], thus avoiding the pitfalls of some of the more speculative constructs that plague proposals like MaxAI, it also unabashedly promotes its simplicity. Claiming—not without reason—that there is a "general disdain for simple technological solutions aimed at a better state of peace," the proponents of the MinAI solution assert, "It does not seem unreasonable to ask why weapons with advanced seekers could not embed AI to identify a symbol of the Red Cross and abort an ordered strike. Additionally, the location of protected sites of religious significance, schools, and hospitals could be programmed into weapons to constrain their actions."[58] It is in keeping with this sentiment that the MinAI solution is proposed.

The proponents of the MinAI solution also argue that "to meet fundamental moral obligations to humanity, [they] are ethically justified to develop MinAI systems. The ethical agency embedded in the machine and, thus, technologically mediated by the design, engineering, and operational environment, is less removed from the human moral agency than it is in a MaxAI system."[59] They also attempt to defend themselves from the charge that when considered from a long-term perspective, it may be more productive to seek a MaxAI solution than to expend energies on MinAI. They argue (1) that a realistic assessment suggests that Artificial General Intelligence (AGI) remains elusive and will likely remain so in the foreseeable future; (2) that there "are currently irresolvable problems with the complex neural networks on which the successes in AI are based,"[60] which, most likely, will escalate with the emergence of AGI; (3) that there is the unavoidable problem of the black-box phenomenon in the context of deep-learning systems, which may become even more acute as the operative algorithms mature and be-

come increasingly sophisticated[61]; and (4) that Moore's Law may not be immediately available to potential AGI technologies, which would likely make the cost of deploying such technologies on combat systems prohibitive.

Keeping in mind the need to modify existing weapon platforms to integrate the MinAI solution, its proponents, unlike the proponents of the PAUSE solution, address the requirements of the Commentary of 1987 to Article 36 of the IHL, which requires that a state must review not only new weapons but also any existing weapon that is modified in a way that alters its function—or a weapon that has already passed a legal review that is subsequently modified.[62] Observing that while this may require a further review of Article 36, they draw attention to the older Saint Petersburg Declaration, which served as a precursor to Article 36 of the IHL, which states that "The Contracting or Acceding Parties reserve to themselves to come hereafter to an understanding whenever a precise proposition shall be drawn up in view of future improvements which science may effect in the armament of troops, in order to maintain the principles which they have established, and to conciliate the necessities of war with the laws of humanity."[63]

But these considerations, which are well-intentioned and undoubtedly aimed toward strengthening the case for improving "humanitarian outcomes through embedded weapon capability to identify and prevent attack on protected objects,"[64] do not give us any deeper insight into the precise mechanism by which such an objective would be achieved. Aside from identifying some of the more common countermeasures that may be launched against the proposed system and the counter-countermeasures that may work in addressing such adversarial actions, the proposal remains bereft of any design or operational detail. In a manner similar to the PAUSE solution, we are only left with some indications of the employment of R-CNN technologies applied to computer vision and the integration of the outcomes to a weapon platform. Precisely how this would happen and how the necessary failsafe mechanisms would work remain unexplained. Also unexplained is how such inputs would feed into the targeting process and how they would or could materially impact the tempo of battle that the targeting process is charged to maintain and augment under active combat conditions.

Given the above, our assessment suggests that the proponents of the PAUSE and MinAI solutions, while well-intentioned, may not have given due consideration to the problem on hand—protecting sites of humanitarian importance—in the context of the hard military problem of targeting. In the section that follows, we will provide a solution that aims to address the problem of protected sites, but in a manner that pays due consideration of the hard military problem of targeting under active combat conditions, thus, potentially, serving as a functional protec-

tive tool that can, conceivably, radically augment the UN-mandated deconfliction mechanism.

Interpreting the Signs: A Protection Tool for Protected Sites

In addition to noting the paucity of detailed information— in terms of technologies and processes—our review of the two proposed solutions also revealed one other curious fact. Both solutions invoked a set of technologies, namely, blockchain and AI (principally, R-CNN). While the former affords the possibility to create a "trustless" system (a paradoxical point in the context of wanting to create a "trusted Human-AI network" that the proponents of the PAUSE solution appear to have either missed or elided), the latter, in the form of R-CNN (which both the PAUSE and the MinAI solutions invoke), is at the cutting edge of work being done in the computer vision field which, arguably, has great potential in the military domain. Nevertheless, there remains the question as to the need for the deployment of these technologies. While we will see how the blockchain technology may be integrated into a possible solution (though not in the manner and for the reason proposed by the PAUSE solution), the need for employing emergent computer vision technology may not be necessary at all. Indeed, it may be both a case of overkill and of a needless complexification in what is already an overly complex operational space.

The solution offered by this article is grounded on two specific technologies: (1) QR (Quick Response) Codes and (2) blockchain. However, it also involves a number of other technologies that, though integral in the context of the proposed solution, are also ones that are almost always integrated within existing military platforms. These include the following: (1) all-weather sensor technologies; (2) encrypted information and messaging datalinks; and (3) dynamic machine- and human-readable battlespace mapping systems (including human-machine interface). While the justification for invoking the two primary technologies—QR Codes and blockchain—will be discussed in short order, it is first necessary to explicitly state the objective of the solution and then outline its fundamental nature and character.

As we noted above, the proponents of the MinAI solution—asserting that "simple technological solutions" can achieve "a better state of peace"—sought to make a case for embedding "weapons with advanced seekers" with AI that would, presumably, be underwritten by their MinAI design—thus enabling such weapon platforms with the capability to "identify a symbol of the Red Cross and abort an ordered strike" and to explore the possibility whereby "the location of protected sites of religious significance, schools, and hospitals could be programmed into weapons to constrain their actions."[65] Yet, their appeal to seek simple technologi-

cal solutions aimed at a better state of peace by embedding AI modules, particularly those involving R-CNN and related computer vision technologies, appears contradictory. Among other things, the science and technology of convolutional and deep neural networks continues to be a complex field of study and research and the state of the technology is still relatively nascent to be employed, particularly in real-life combat conditions where "protected sites of religious significance, schools, and hospitals" are at stake.

Contrasted against this, the solution offered by this article, by invoking a set of much simpler technologies, seeks to radically augment the UN deconfliction mechanism that, in its present form, has proven to be ineffectual as a protection tool for protected sites. Unlike the PAUSE and MinAI solutions, the proposal set forth in this article aims to serve as a viable, effective, and immediately implementable protection tool—not by insisting on an "AI solution" but by leveraging the QR Code and blockchain technologies, which are, by many magnitudes, simpler than any AI module. It warrants reiterating that the singular focus of the proposed solution is to protect infrastructure and not humans, though, as the examples of human casualties in the context of attacks on protected sites show, humans will be indirect beneficiaries of the proposed solution. As such, the proposed solution may be considered to be a "concept-technology" paring, which brings together the concept of a protection mechanism with a set of technologies (QR Codes, blockchain, coupled with sensors, data and messaging links, and battle mapping systems). The design principle of the proposal is deliberately oriented to seamlessly integrate with the targeting process—involving deliberate and dynamic targeting—that militaries employ under combat conditions and may be integrated relatively effortlessly with manned and unmanned systems.[66] Additionally, when considered outside and beyond the operational-tactical sphere, the proposed solution may also serve specific strategic-political aims, which we will have occasion to briefly examine below. And, lastly, the proposed solution eschews the tendency to "moralize machines"; instead, it serves as an example of how a value-sensitive design orientation may be adopted when thinking through the design of military systems.

The proposed solution exhibits a number of distinctive characteristics. Thus, for example, the two core technologies that it invokes are cheap, secure, and easy to deploy (as is the case with QR Codes/Coding) and leverage the benefits of distributed ledger systems, which guarantees a trustless context and immutability of records, timestamps, and so forth (as is the case with blockchains). The solution is designed to ensure that the most critical function of QR Code generation is entrusted to an impartial/neutral agency (the UN OCHA and/or the ICRC). The solution is integrable within the military targeting process and is particularly sen-

sitive to the needs of dynamic targeting. As such, it does not disrupt or needlessly extend the sensor-to-shooter link; rather, it augments the targeting cycle by providing "detargeting cues" and, therefore, contributes to the updating of active and passive battle maps in near real time—thus, enabling "collective engagement capability" and "shared awareness" (particularly in battle swarm contexts), which are in keeping with the basic tenets of network-centric warfare. Lastly, and importantly, the proposed solution avoids the trap of the black-box problem that often afflicts AI solutions (particularly those involving convolutional and deep neural networks) by enabling humans to be involved both in the QR Code generation process and in reviewing, monitoring, and analyzing all the activities that take place, which are recorded on the blockchain. This has the added benefit of relatively "hardening" the proposed solution against potentially malicious actions in which bad actors may engage.

At this juncture, a brief overview of the two core technologies—QR Codes/coding and blockchain—invoked by the proposed solution is warranted. First released in 1994 in Japan by Denso Wave, Inc., a QR Code is a machine-readable, two-dimensional optical label that contains information about the item to which it is attached. QR Codes—in which 7,089 characters can be encoded in one symbol—are capable of handling all types of data, such as numeric and alphabetic characters, symbols, binary, control codes, and so forth. As such, QR codes are able to overcome the informational restrictions imposed by the previous barcoding system and found their first applications in the auto industry for use in electronic Kanban systems.[67] One significant development that boosted the adoption and widespread use of QR codes was Denso Wave's decision to make the specifications of the QR Code publicly available so that anyone could use it freely. Although Denso Wave retains the patent rights to the QR Code, it declared that it would not exercise them, which enabled QR Codes to be used at no cost and to become a "public code" used by people all over the world.[68] In 1997, QR codes were approved as an AIM standard[69]; in 1999, they were approved as a standard 2D code by the Japan Industrial Standards and made a standard 2D symbol on the Japan Automobile Manufacturers Association's EDI standard transaction forms; and in 2000, they were approved by the International Organization for Standardization as one of its international standards.[70] By 2002, QR codes achieved mass popularity (beginning in Japan) with the release of mobile phone with integrated QR code readers. While there are five basic types of QR codes (QR Code Model 1&2, Micro QR Code, iQR Code, SQRC, and Frame QR), for our purposes, the most relevant variants are iQR Code (because of its storage capacity of 40,000 numerals), the SQRC (because it can carry public and private data), and the Frame QR (because it includes both design flexibility and security).

It is also important to note that QR Codes do not necessarily have to be small in size. Indeed, as is directly relevant for our purposes, QR Codes, scaled to size, can be affixed to large structures like buildings and so forth, as depicted in figure 5.

Figure 5. A huge billboard advertisement in Shibuya (Tokyo, Japan). A pedestrian may use a cellphone to read the QR code (barcode) with the phone's camera. The phone will open a web browser to the URL encoded in the QR code[71])

Blockchain technology is, in effect, a "distributed database containing records of transactions that are shared among participating members. Each transaction is confirmed by the consensus of a majority of the members, making fraudulent transactions unable to pass collective confirmation. Once a record is created and accepted by the blockchain, it can never be altered and disappear."[72] Blockchains have three distinguishing properties, namely, decentralization, transparency, and immutability, which make the technology suitable for our purposes. Further, blockchains have the following capabilities that are of particular interest to us:

1. Shared governance and operation, which addresses "the scenario in which a collection of entities . . . want to participate in a communal system but do not trust each other or any third party to operate the system single-handedly. By deciding on the system details (governance) and then deploying networked devices . . . to run the system, each entity can be assured of correct operation."[73] There are two basic governance models—open governance (or permissionless blockchain systems) where "any party that is willing to par-

ticipate in the consensus protocol is allowed to do so, regardless of their identity" and consortium blockchain systems (or permissioned systems) wherein "participation in the consensus protocol is limited to . . . [actors] approved on a whitelist defined at system initialization."[74] Given the nature of our specific requirements, the type of governance model that will suit our purposes would be the latter.

2. Verifiable state, which reinforces the trust users have in the system (i.e., "that the current state of the system accurately reflects the transactions that the consensus protocol allowed to execute in the past"[75]). To this end, all transactions are written to a cryptographically verified append-only ledger, providing full-system provenance, thereby allowing for the system's current and past states (operations) to be audited.

3. Resilience to data loss, which points to the fact that in the event of a data loss, the content of the ledger is retrievable given that the data is replicated among all the actors/users on the blockchain. The diagram below illustrates outlines of these capabilities of the blockchain.

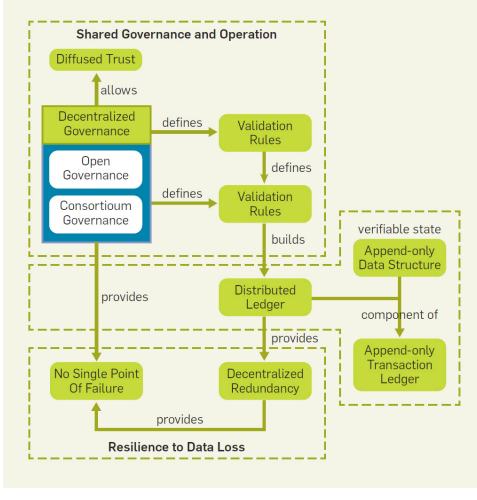

Figure 6. Blockchain capabilities. (Source: Ruoti et al, 2020.)

The above being a brief outline of the core technologies involved, the proposed solution makes four working assumptions:

1. The current UN deconfliction mechanism requires protected sites located in and around potential conflict sites to update the UN OCHA—within a specific timeframe (normally 72 hours)—with details regarding their geolocation, nature and function of the site and the activities that are conducted at

the site, which is then distributed to the belligerents via a so-called no-strike list. The proposed solution assumes that this procedure will continue, but it adds a number of additional steps to the process. The UN OCHA will, in addition to distributing this information to the belligerents, also populate a dedicated permissioned blockchain with the information it receives, including records of it having distributed the information to the belligerents. The UN OCHA will control the permissioned blockchain, and all members of the UN will be on a whitelist allowing them access to it. Moreover, as the blockchain is updated, a push notification would be sent to each member on the access list, informing them about the update.

2. Each member on the access whitelist will be able to retrieve the updated records. More specifically, since it is assumed that the belligerents will be state actors, they will have special access to these records, which will enable them to retrieve the information in a format that will allow them to populate their military targeting systems.

3. All combat platforms will be equipped with all-weather scanners and sensors that will be able to read QR codes in the battlespace. There is nothing extraordinary about this requirement as combat elements during the intelligence preparation of the battlespace (IPB), and subsequently during combat, are continually scanning the battlespace and its environs. Thus, their scanning and sensor systems that, it is assumed, are already scanning the battlespace will be able to scan and read the QR Codes. If modifications are required to be implemented to the combat platforms to accommodate this specific feature, then the same may be accomplished under the provisions of the St. Petersburg Declaration or by invoking the Commentary of 1987 to Article 36 of the IHL. For newer platforms, the inclusion of this capability would be mandatory prior to approval being given for their use in battle.

4. In addition to the modifications involving the installation of the appropriate scanners/sensors, one specific and critical modification is warranted. This involves a locking mechanism that comes into play when an attack platform either scans the UN-issued QR code and/or when its accompanying sensor package cues it with that information. It is necessary to emphasize the point that this locking mechanism should be a "limited feature." It should be *limited* in the sense that it should only prevent an attack platform's weapon system from firing at a specific geolocation (after adjusting for a circular error probability [CEP] factor to account for localized blast-radii) as identified by the scanning of the QR Code.

Given these assumptions, the solution's workflow, which is indicative of the sequencing of its practical implementation, may be listed as follows:

1. The UN OCHA receives inputs as per the methodology employed for its current deconfliction mechanism for protected sites.
2. The UN OCHA generates and issues QR Codes to the protected sites while simultaneously updating a permissioned blockchain that, in turn, would trigger high-priority push notifications to all UN members, but also in a specific format to the belligerents, which can be incorporated into their individual tactical battle planning and management systems.
3. Prior to the initiation of hostilities and during the inevitable IPB phase during which the belligerents (but also any other authorized user of the UN OCHA blockchain) are engaged in actively scanning the battlespace, the QR Codes will be automatically scanned by their sensors. The information resulting from the scans will be transmitted—by dedicated communication and battle-networks, which are already in use by the military—to tactical and operational centers. This allows for a continuous updating of the targeting solutions that are prepared for deliberate and dynamic targeting processes that are estimated to be underway at this initial phase of the battle. Additionally, the activity of each scan and reading of a QR Code is directly updated to the UN-operated blockchain, which allows for a near-real-time updating of the record.
4. Similarly, the protected site, which has been scanned, will also record the scanning activity and will transmit that information to a localized database and/or uplink the information to the permissioned blockchain operated by the UN. This creates an additional record in the blockchain, which is also shared with all entities who are enumerated on the access whitelist. As a backup mechanism, a dedicated UN satellite platform may also be employed to facilitate the transacting of information to and from the blockchain.
5. With the commencement of active hostilities, the role of dynamic targeting takes on a greater importance. While dynamic targeting is not ad-hoc, nevertheless, it usually takes place within highly compressed timeframes. During such targeting actions, all attack platforms and their accompanying sensor packages continue their active scanning activities to support their military functions. Thus, during such active scanning activities, any protected site that has not yet been scanned will fall within the scanning envelope of either the attack platforms and/or their supporting sensor packages. As and when a scan takes place, the actions outlined in steps 3 and 4 will be executed. Diagrammatically, the process may be depicted as shown in figure 7.

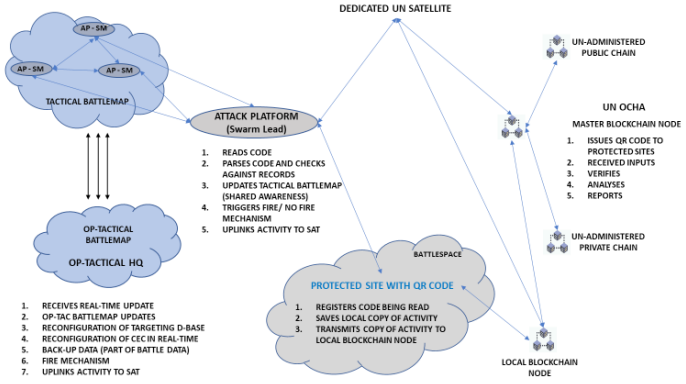

Figure 7. The Concept-Technology Solution. (Source: Author.)

As can be seen in figure 7, the solution described is intended to cover contingencies arising from determined and dynamic targeting practices. In this connection, it is important to underscore the point that when an attack platform (either singly or as part of a larger attack package) is updated with the data resulting from the scans of the QR Codes affixed to protected sites, its weapon systems are limited by a mechanism that prevents them from firing at the geophysical coordinates that the scan of the QR Code will reveal. Note that this limitation is specific only to the geolocation identified by the scan of the QR Code and does not result in "locking up" the weapon systems. In other words, as the attack platform reorients itself away from the specified coordinates, it immediately reacquires its firing capability. It also warrants mentioning that this limitation will not (and should not) impact the attack platform's defensive (nonlethal) countermeasure systems.

Considering the above, it is now possible to list some of the advantages of employing the proposed solution. First, the proposed solution applies to manned and unmanned combat systems. In the latter instance, assuming that the unmanned combat system is capable of autonomous action, the "automated" locking mechanism, which is cued as a consequence of the scanning activity, serves as a narrow operational bottleneck. We suggest that this model of "embedding automation within autonomy" paradigm may be, prospectively, a remunerative way to consider problems related to autonomy in the context of emergent combat and combat-related technologies.

Second, the proposed solution does not require complex technologies to achieve its aim. In fact, it takes the basic principles of MinAI (without the AI component,

which is unnecessary in this context) and applies it to a technologically simpler and likely more cost-effective solution.

Third, the proposed solution intervenes in the targeting process, but it does so in a manner that is not disruptive to the maintenance of the operational/battle tempo. This will likely make the solution more palatable to military organizations given that with such a solution they will be able to maintain the delicate balance between discharging their military functions efficiently while at the same time upholding the tenets of the IHL and the LOAC without compromising the tempo of operations/battle.

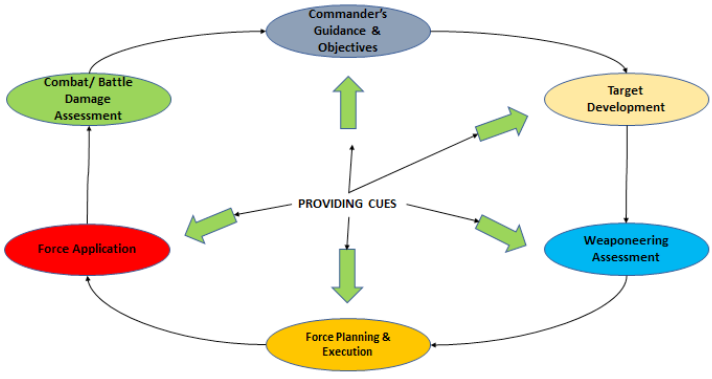

Figure 8. Cueing the targeting process. (Source: Author.)

Fourth, in many ways, the proposed solution not only augments the military's ability to achieve better battlespace knowledge/awareness but also promotes two of the core principles of network-centric warfare—namely, shared awareness and collective engagement capability. This is because, as the figure 9 depicts, when any one attack or sensor platform scans an existent QR code affixed on a protected site, not only are the tactical and operational centers of the military force and the UN-operated blockchain updated, simultaneously, the networked battle map of an attack and sensor package (potentially comprised of multiple attack/sensor platforms) is also updated on a real-time basis.

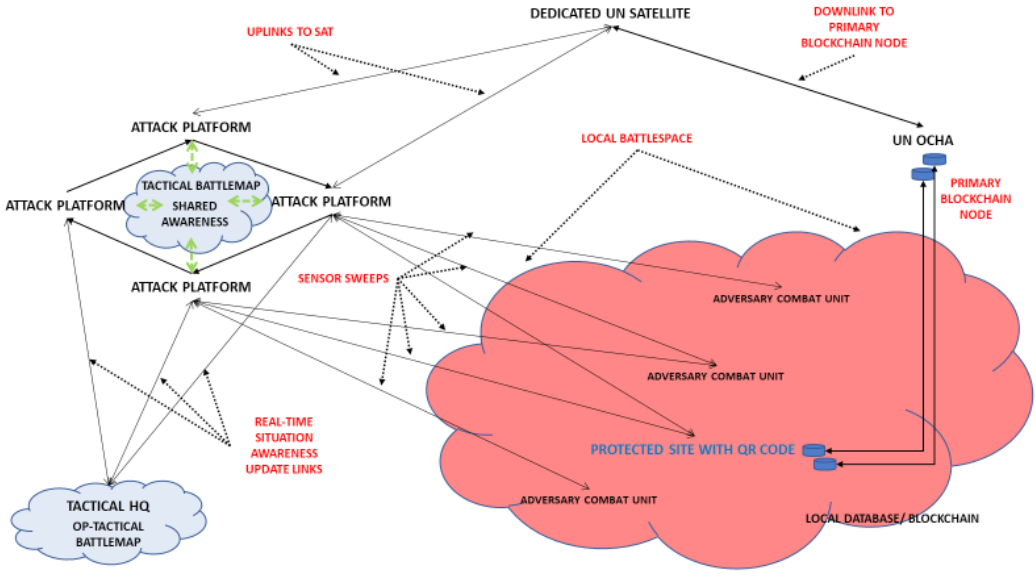

Figure 9. Promoting collective engagement capability and shared awareness. (Source: Author.)

Moreover, since each platform in a given attack package is also actively scanning the battlespace, the initial readings are reconfirmed multiple times and the records—both at the operational and tactical centers and the UN-operated blockchain—are updated accordingly, which only serves to augment the redundancy factor.

Fifth, the recording of the issuance of the QR codes to the protected sites, the recording of the scanning activities (both by the active scanner on the military platforms and by the protected sites) on the UN-operated blockchain creates an immutable and time-stamped record that is undeniable by any entity. Thus, if, despite such records being available, an attack is launched—as was tragically the case in Kunduz in 2015[76]—then the option of claiming plausible deniability is not available, and the transgressor can be immediately taken to task.[77] Moreover, since access to the record is available—as per the proposed solution—to a whitelist, which will optimally include all UN members, the ability of a "cover-up" will also be greatly diminished. In this context, it may also be worth considering allowing specific neutral parties engaged in humanitarian work (for example, the ICRC) to also have access to the UN-operated blockchain in real time.

Conclusion

The problem of attacks on protected sites—despite the UN deconfliction mechanism being in place—continues unabated. The concept-technology solution offered by this article is one way by which this problem may be addressed. As we have seen, previous solutions that have attempted to incorporate technologies, though well-intentioned, have tended to be complex and have, potentially, multiple points of failure. The solution offered here is more simplistic—thus reducing the number of points of failure. The core technologies that the solution invokes, namely, QR Codes and the blockchain, are secure in their respective domains. Moreover, and particularly in the case of the QR code, if a further hardening is necessary, this may be accomplished with minimal effort. In relative terms, the cost of hardening the QR technology would likely be much lower than attempting to refine the currently available computer vision, R-CNN, and related technologies. Additionally, the need for a mechanism that performs better than what the UN can currently offer is immediate, and the other more sophisticated technologies that the PAUSE and MinAI solutions invoke will take time to refine to a degree that an active deployment in a conflict site would warrant.

The proposed solution also does not make unwarranted demands on the military. It respects the fact that targeting is a critical function and the solution's intervention in the targeting process may be considered to be supportive rather than disruptive. Further, it also does not make demands on the military to incorporate additional layers of technologies on its systems and platforms, for it seeks to leverage the very same technologies (scanners, sensors, communications, and data links, etc.) that the military currently uses to create opportunities for it to not only execute its core functions but also to do so in keeping with the dictates of the IHL and the LOAC.

And, last but not the least, the strategic-political climate in which the debate on the research and development work on AI and autonomous systems in the military context is taking place is fraught with contradictions, vested interests, and sometimes extreme ideological positions. Thus, to propose a fresh set of requirements that may in some form appear to restrict the freedom of action that nation-states may insist on would also be problematic. The current solution avoids such pitfalls. Indeed, if configured and packaged appropriately, the solution can be put forth virtually as a *fait accompli* to the global community of nation-states. After all, every nation-state recognizes the humanitarian cost of war. And, even if they transgress the laws governing war—inadvertently or knowingly—they recognize that there is a significant political cost to bear. The solution offered in this article is one low-cost means by which a degree of unanimity may be gained

among nation-states on the question of preventing accidental and deliberate attacks on protected sites. Moreover, in an age where the functional relevance of the UN and other humanitarian organizations like the ICRC is being eroded, their championing of a solution such as the one offered in this article can, if implemented with care, actually serve as a protective tool of consequence (unlike the currently available deconfliction mechanism). In this way, not only would such humanitarian organizations reassert their relevance, but they would also take the lead in addressing a key concern that has plagued the phenomenon of war—namely, how to reduce the human cost of war. ✪

Dr. Manabrata Guha

Dr. Manabrata Guha is a Senior Research Fellow and Senior Lecturer at the School of Engineering & IT, University of New South Wales at Australian Defence Force Academy. He is the author of "Re-Imagining War in the 21st Century: From Clausewitz to Network-centric Warfare" (Routledge, 2011), and has published widely in academic and policy-centric journals.

Dr. Jai C. Galliott

Dr. Jai C. Galliott is Director of the Values in Defence & Security Technology Group, School of Engineering & IT, University of New South Wales at Australian Defence Force Academy. He is the author of "Military Robots: Mapping the Moral Landscape" (Ashgate 2015/Routledge 2016), and has also published numerous books, articles and books chapters on matters related to ethics and autonomous weapon systems.

Notes

1. The reference here, of course, is to International Humanitarian Law (IHL) or what is otherwise known as the Laws of Armed Conflict (LOAC). With specific reference to the legal and ethical considerations that a modern combat soldier has to take into account see Paul A.L. Ducheine et al (Ed.) *Targeting: The Challenges of Modern Warfare*, (The Hague, The Netherlands, 2016), particularly Chap 5-8

2. Accelerated Warfare: Futures Statement for an Army in Motion, *Commander's Statement, Australian Army*, 3 January 2020. Available at https://researchcentre.army.gov.au/sites/default/files/2020-01/futures_statement_accelerated_warfare_a4_u.pdf (alternatively https://bit.ly/371SA7U)

3. Thus, referring to Michael Waltzer's comments on the "double effect" problem, Martin L. Cook notes, "[i]t is insufficient that the combatants do not intend directly to harm the non-combatants and then make a very subjective proportionality calculation. Rather, they should actively engage in actions to protect non-combatants, even if those actions increase their own risk", which is a point of view that Cook (and presumably Waltzer) records as possessing a strong moral force. See Martin L. Cook, "Ethical Issues in Targeting", in Ducheine et al (Ed.) *Targeting: The Challenges of Modern Warfare*, 2016, p153.

4. Col. Jason M. Brown, USAF, "To Bomb or Not to Bomb: Counterinsurgency, Airpower, and Dynamic Targeting", in *Air & Space Journal*, Winter 2007, pp 76-81

5. J. Marshall Beier, "Discriminating Tastes: 'Smart Bombs, Non-Combatants, and Notion of Legitimacy in Warfare" in *Security Dialogue*, Vol 34, No. 4, Dec. 2003, p 413, 422.

6. Evan Hill and Christiaan Triebert, "12 Hours. 4 Syrian Hospitals Bombed. One Culprit: Russia," *New York Times*, 4 May 2020, https://www.nytimes.com/.

7. "UN Fails to Acknowledge Own Failures in Hospital Attacks Inquiry", April 16, 2020, The Syria Justice and Accountability Centre (SJAC). Available at https://syriaaccountability.org/updates/2020/04/16/un-fails-to-acknowledge-own-failures-in-hospital-attacks-inquiry/ (alternatively https://bit.ly/2J2bx28)

8. "Summary by the Secretary-General of the report of the United Nations Headquarters Board of Inquiry

into certain incidents in northwest Syria since 17 September 2018 involving facilities on the United

Nations deconfliction list and United Nations supported facilities", *United Nations Organization*. Available at https://www.un.org/sg/sites/www.un.org.sg/files/atoms/files/NWS_BOI_Summary_06_April_2020.pdf (alternatively https://bit.ly/3nLMors)

9. Ibid., para 82-84

10. Ibid., para 91-92

11. Ibid., para 94

12. Merel A. C. Ekelhof, "Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting", in *Naval War College Review*, Summer 2018, Vol. 71, No. 3, p2

13. Major Gen. Charles J. Dunlap (USAF, Retd.), "Preface", in Ducheine et al (Ed.) *Targeting: The Challenges of Modern Warfare*, 2016, p vii.

14. Ekelhof, 2018, p4

15. Dunlap, 2016, p viii

16. See, for example, Horst Boog, Gerhard Krebs, Detlef Vogel, *Germany and the Second World War: Volume VII: The Strategic Air War in Europe and the War in the West and East Asia, 1943-1944/5*, Illustrated Edition, (Oxford: Oxford University Press, 2006), pp 7-153

17. Beier, pp 419-420

18. See, for example, Paul Dickson, *The Electronic Battlefield: Origins of America's 21st-Century Way of Warfare*, (Bloomington: Indiana University Press, 2012)

19. Donn A. Starry, "Extending the Battlefield", *Military Review*, March 1981, pp. 31–50.

20. David R. Mets., "The Long Search for a Surgical Strike: Precision Munitions and the Revolution in Military Affairs", *Research and Education Paper No.12*, (Maxwell Air Force Base, AL: College of Aerospace Doctrine, 2001), p 39

21. Kevin D. Scott, Vice Admiral (USN), "Preface", in *Joint Publication (JP) 3.0: Joint Operations* (17 January 2017

Incorporating Change 1, 22 October 2018), (Washington, DC: The Joint Staff, 2018)

22. Joint Publication (JP) 3.0: Joint Operations, *Annex 3-60 Targeting*, (Last Updated: 15 March 2019), Curtis E LeMay Center for Doctrine Development and Education, (Montgomery, AL: Maxwell Air Force Base), p 3 (hereafter Annex 3-60 Targeting)

23. Ibid.

24. Ducheine, Schmitt, Osinga, "Introduction" in Ducheine et al (Ed.), *Targeting: The Challenges of Modern Warfare*, 2016, p2.

25. Annex 3-60, p 3

26. *Air Force Doctrine Document 3-60, Targeting* (Incorporating Change 1, 28 July 2011), LeMay Center for Doctrine Development and Education, (Montgomery, AL: 2011), p 2 (Hereafter AFDD 3-60)
27. Annex 3-60, p 40
28. Annex 3-60, p 4
29. Ibid.
30. AFDD 3-60, pp 3-6
31. Ibid.
32. Annex 3-60, p 4
33. Ibid., p 36
34. AFDD 3-60, p 17
35. Ibid., p 45
36. Ibid.
37. S. K. Devitt. J Scholz. T. Schless. L, Lewis, "Developing a Trusted Human-AI Network for Humanitarian Benefit", *unpublished paper*, Jan, 2020, p1.
38. Ibid.
39. Ibid.
40. Ibid., p2
41. Ibid., p3
42. Ibid., p4
43. Ibid.
44. Ibid., p9
45. Ibid., p7
46. Ibid., pp 7-8
47. Ibid., p8
48. Ibid., p9
49. Ibid.
50. Ibid.
51. Ibid., p11
52. We say "something approximate to it" because the proponents of the PAUSE solution do not describe in any detail in their paper as to what the exact nature of the "Protective AI" technology that they seek to employ may be. They only state that "[d]esign of systems to achieve Protective AI based on a starting point of MinAI, [sic] must consider its own weaknesses and errors." (ibid., p 14)
53. Ibid., p15-16
54. Ibid., 16. We also note with concern that despite the use of AI technologies, there have been, in recent times, extensive and effective disinformation campaigns, which have had (and, arguably, continue to have) deleterious strategic effects on, among others, the US government and polity, and whose long-term effects are virtually impossible to ascertain. Thus, the confidence expressed by the proponents of the PAUSE solution in this context is both baffling and, in our opinion, dangerous. For an overview of the problem see, Katarina Kertysova, "Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered", in *Security and Human Rights,* 29, 2018.
55. This is directly related to the observation we make in footnote #55 above, albeit applied to the narrower military-operational context.

56. Jai Galliott and Jason Scholz, "Artificial Intelligence in Weapons: The Moral Imperative for Minimally-Just Autonomy", in *Journal of Indo-Pacific Affairs*, Winter, 2018, pp 58-59

57. Ronald C. Arkin, Patrick Ulam, and Brittany Duncan, "An Ethical Governor for Constraining Lethal Action in

an Autonomous System", *Technical Report*, (Fort Belvoir, VA: Defense Technical Information Center), 1 January 2009)

58. Galliott and Scholtz, 2018, p58

59. Ibid., p60

60. Ibid., p61

61. See Davide Castelvecchi, "Can We Open the Black Box of AI?", in *NATURE*, Oct. 5, 2016 for an overview of the so-called "black-box problem".

62. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, Commentary of 1987, "New Weapons," 421. Available at https://bit.ly/2HwtyVW

63. St. Petersburg Declaration Renouncing the Use, in Time of War, of Certain Explosive Projectiles, Saint Petersburg, 29 November/11 December 1868". Available at https://bit.ly/33aBkMB

64. Galliott and Scholz, 2018, p 61

65. Ibid., p57

66. Note that in this iteration of the proposed solution, the orientation is towards aerial combat systems – manned and unmanned. The fundamental design principle, of course, is applicable in a multi-domain context with minimal modifications.

67. Electronic Kanban: A communication tool used in production management systems.

68. *History of QR Code*, "Section 2: Release of the QR Code and subsequent efforts to spread its use", *QR Code.com*, DENSO WAVE INCORPORATED. Available at https://www.qrcode.com/en/history/

69. AIM: Automatic Identification Manufacturer

70. *History of QR Code*, "Section 2: Release of the QR Code and subsequent efforts to spread its use", *QR Code.com*, DENSO WAVE INCORPORATED. Available at https://www.qrcode.com/en/history/

71. This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. Available at https://commons.wikimedia.org/wiki/File:Japan-qr-code-billboard.jpg

72. D. Efanov and P. Roschin, "The All-Pervasiveness of the Blockchain Technology", in *Procedia Computer Science*, vol. 123, p 116, 2018

73. Scott Ruoti, Ben Kaiser, Arkady Yerukhimovich, Jeremy Clark, Robert Cunningham, "Blockchain Technology: What Is It Good For?", in Communications of the ACM, Vol. 63 No. 1, January 2020, Pages 46-53

74. Ibid.

75. Ibid.

76. Tim Craig, Missy Ryan and Thomas Gibbons-Neff, "By evening, a hospital. By morning, a war zone", in *The Washington Post*, October 10, 2015. Available at https://wapo.st/338X0ce

77. The operative assumption here is that a "bad actor" has somehow been able to either bypass the proposed "limited locking" mechanism and/ or there is a catastrophic system failure of the proposed mechanism.