***Fordham International Law Journal***

This is an Accepted Manuscript of an article published by Fordham University School of Law in the *Fordham International Law Journal* on 05/03/2021, available at:

<https://www.fordhamilj.org/volume-44-issue-3/2021/3/5/the-duty-to-take-precautions-in-hostilities-and-the-disobeying-of-orders-should-robots-refuse>

and

<https://ir.lawnet.fordham.edu/ilj/vol44/iss3/3/>

Recommended Citation: Francis Grimal & Michael Pollard, *The Duty to Take Precautions in Hostilities, and the Disobeying of Orders: Should Robots Refuse?*, 44 Fordham Int'l L.J. 671 (2021).

It is deposited under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

**The Duty to Take Precautions in Hostilities, and The Disobeying of Orders: Should Robots Refuse?**

Francis Grimal and Michael J. Pollard[[1]](#footnote-1)\*

“*In war the first principle is to disobey orders. Any fool can obey an order.*”[[2]](#footnote-2)

Abstract

*This Article not only questions whether an embodied artificial intelligence (EAI) could give an order to a human combatant, but controversially, examines whether it should also refuse one. For example, a future EAI may be capable of refusing to follow an order where an order appeared to be manifestly unlawful, or, was in breach of International Humanitarian Law (IHL), national Rules of Engagement (ROE) or, even, where they appeared to be immoral or unethical. As part of opening this much-needed discussion, the authors examine the legal parameters, and by way of a solution provide a framework for overriding and disobeying. In short, the article examines whether human error can be corrected and overridden - but for the better, rather than for the worse? An aircraft’s anti-stall mechanism, which takes over, and corrects human error, is seen as nothing less than “positive”. Such an argument has traction in the strategic realm in terms of “system of systems” — the premise that more advanced technology can potentially help overcome Clausewitzian “friction” / “fog of war”.*

*Within the broader discussion of the “duty to take precautions” (Article 57 API/ ICRC Customary Rule 15), the authors analyze the concept of obeying/disobeying orders through the lenses of Autonomous Weapons Systems (AWS) and EAI - two “robots” which, for the sake of the current Article can be distinguished owing to the fact that an EAI is capable of giving, or of refusing to follow an order to apply force, while an AWS is one which once activated, makes targeting decisions, and applies force independently. Central to this discussion, are state specific ROEs within the concept of “duty to take precautions”. At present, the guidelines relating to a human combatant’s right to disobey orders are contained within such doctrine, but vary widely. For example, in the United States, a soldier may disobey an order but only when the act in question is clearly unlawful. Whereas in direct contrast, Germany’s “state practice”, has specific additional requirements in terms of human dignity, and that the “order” is not being of “use for service”.*

*At its heart, this Article introduces a future-thinking discussion of the practical process of disobeying of orders between various “individuals” within the chain of command (human to AWS; AWS to human; EAI to AWS, human to EAI). Taken to its extreme, the authors envisage the ability of an EAI being able to override and overturn nuclear launch. Towards the end of the Article, the authors extend the discussion of “robot” refusal to a wider application including robot Private Military Contractors (PMCs), robot spies, and the more provocative concept of robot insubordination. Would robot PMCs operate according to existing implicit biases, following orders, regardless of personal “opinion” due, to the financial implications of failing to do so. Additionally, the authors consider whether an EAI (being non-human), should apply a higher threshold before ‘deciding’ whether to disobey an order, due to their immortality.*

*By way of overall solution, the authors propose the crafting of “robot” rules of engagement (RROE) with specific regard to the disobeying of orders. In addition to ensuring the EAI is programmed to run an indefinite proportionality assessment feedback loop, the authors also propose a novel test: the EAI is to discount human “traits” which, lead to human error. If this test is satisfied, an order should be disobeyed, and human error overturned. In the broader sense, the authors question whether warfare should remain an utterly human preserve – where human error is an unintended but unfortunate consequence*. *Or, does the duty to take all feasible precautions in attack require a human commander to utilize available AI systems to routinely question human decision-making, and where applicable, prevent mistakes/ disobey orders – whether unintended, or are in fact punishable as war crimes? Ultimately, the overarching question posed, is whether EAIs are to be afforded the same combat privileges as human combatants when it comes to the disobeying of orders.*

I. Introduction

II. Duty to Take Precautions during Armed Conflict

A. *How do the Duty to Take Precautions fit within IHL?*

B. *State Practice*

III. Rules of Engagement and Disobeying of Orders

A. *How does the Duty to Disobey an Order fit within DTP?*

B. *State Practice*

IV. Disobedience and Nuclear Launch

A. *Additional Test for Safeguarding Decision-making.*

B. *Scenario 1: Rudimentary decision-making.*

C. *Scenario 2: Target Conformation.*

D. *Scenario 3:Mistreatment of PoWs.*

E. *Scenario 3: Nuclear Launch.*

V. Wider Application of Disobeying Orders

A. *Private Military Contractors.*

B*. Spies.*

C*. Insubordination*

VI. Conclusion

# Introduction

This Article goes significantly beyond traditional boundaries of debate surrounding a combatant’s legal obligation to disobey unlawful orders[[3]](#footnote-3), and by way of overall solution, presents both a legal and strategic “system of systems”. Grounded at the very heart or “Schwerpunkt” (to use Clausewitz’s terminology[[4]](#footnote-4)) of strategic thinking is the concept of “friction” within warfare.[[5]](#footnote-5) Friction, simply-put is more colloquially known as “the fog of war” – the concept that within warfare, things will go wrong, and mistakes will happen.[[6]](#footnote-6) In essence, the Clausewitzian concept of ‘friction’, is what distinguishes “real war” from war on paper, something (despite advances of technology) a strategist cannot always account for.[[7]](#footnote-7) Strategic doctrines that failed to account for “friction” were deemed as inherently flawed[[8]](#footnote-8) and much, if not considerable ink was spilt in the 1990s and 2000s, to consider the means by which to overcome friction in warfare.[[9]](#footnote-9) Fully-conscious that previous efforts to simply “throw” advanced technology at the problem would not resolve the issue[[10]](#footnote-10), strategists formulated the idea of “system of systems”.[[11]](#footnote-11) Much as with a mechanical watch, technology / cogs would be assigned individual tasks, but those smaller “cogs” would then form part of a bigger and more advanced “system of systems”.[[12]](#footnote-12)

As presented above, the authors propose to transfer this strategic concept into the legal realm as a means of providing an overall solution as to how, and more importantly when, *various “individuals” within the chain of command (e.g.; human to AWS; AWS to human; EAI to AWS, human to EAI)* would be able to correctly reject, and overturn, an erroneous and unlawful order.[[13]](#footnote-13) Throughout, the authors argue that simply transposing a combatant’s right and duty to disobey an unlawful order) into the EAI realm without further scrutiny is unwise for it would close the door on a much-needed discussion in terms of how EAI’s could potentially remove human error / “friction” in warfare.[[14]](#footnote-14)

In terms of a unique and novel solution, the authors envisage the following: the EAI (when considering *any* order) will not only to run a continuous / dynamic proportionality assessment (an implicit factor of which is a perpetual target verification mechanism), but will also identify whether the human decision is motivated by human character “flaws” such as revenge or self-preservation. Where, the EAI is satisfied that the order is IHL compliant but that human error is in evidence, the EAI should reject that human command. While there is unboundedly the concern that EAIs might increase “friction” (due to their behavior) the authors counter such a possibility by implementing the following when it comes to an human to EAI order in the chain of command. For an human to EAI command scenario (in addition to the test proposed in the preceding section), the authors suggest several cogs and levels within this “system of systems” with each one independently making a specific determination. Each of those independent systems would be tasked with only one computation – the authors set this out comprehensively in Section A of Part IV In relation to the other variable of a human to EAI command, the authors consider this in its most extreme setting – “robot refusal” of nuclear launch.

An additional, but natural (and indeed desirable) implication of this discussion is whether the threshold is raised – the extra layers of “protection” proposed by the authors imply that IHL obligations are set much higher. One could readily argue that an EAI’s calculations place it under a higher burden in terms of assessment and accountability of information when taking a particular decision. Ostensibly, the authors’ analysis is distilled into three distinct areas: first, the understanding and scrutiny of existing guidelines regarding the duty to take precautions: wide (U.S.) narrow (Germany). Secondly, the way in which the guidelines are interpreted with regard to the right to refusal: wide (Germany) narrow (U.S.). And finally, the authors’ proposed solution in terms of an additional test as to precisely how an EAI may refuse erroneous human command owing to the fact that as long as there is human involvement, there will always be human error – and while machines might not eliminate such errors, they can certainly go some way to countering them.

While one could argue that there are more things that could “go wrong” (due to technological limitations of 21st century technology), to suggest that this will remain the *status quo*, is dismissive of advances in machine learning.[[15]](#footnote-15) While the absolute and complete removal of error, is difficult to envisage one could use the example of an aircraft’s anti-stall mechanism which, readily point to simple and existing technology which, innocuously overrides human error for the “better”.[[16]](#footnote-16) In other words, while the authors’ proposed “system of systems” regarding refusal may not fully eradicate “friction”, it’s negative effects will have been significantly reduced and constrained. Indeed, conceptually, it is already easy envisage a AWS that is capable of applying a great many technological advances at speeds far in excess of human capabilities.[[17]](#footnote-17)

In sum, the heart of this article is the proposition that EAIs are likely to possess enough decision-making capabilities, in order for them to be considered, (at least prima facie), conscious.[[18]](#footnote-18) By way of overview, the structure of the Article is as follows: Part II of the Article examines the way in which the Duty to Take Precautions operates as a whole within the IHL realm but, with specific focus on individual state practice so as to make the necessary link between the Duty to Take Precautions and the disobeying of orders. Part III meanwhile examines the specific interface between the legal obligations contained in IHL, and their translation into state-based ROEs. In Part IV, the authors “stress test” the logistics of “robot” refusal in terms of providing a concrete legal test for determining when an erroneous command (whatever its designation: human, autonomous or EAI) should be rejected and refused. And finally, in Part V, the Article applies the principle of “robot refusal” into a much wider discussion by extending this to other areas: Robot PMCs, Robot Spies, and the “niche” consideration of robot insubordination.

II. Duty to Take Precautions during Armed Conflict

Prevailing discussion within the current literature and scholarship on AWS and EAI systems falling under the scope of IHL has understandably centered on whether either can capably satisfy the necessary distinction and proportionality requirements.[[19]](#footnote-19) Equally, the general corpus of literature as a whole, on the duty to take precautions during armed conflict is relatively well-established.[[20]](#footnote-20) Consequently, the authors propose to navigate the key contours of this landscape without over elaboration. Rather, the purpose, and focus of Part II, is simply to identify how the duty to take precautions (DTP) fits within the larger body of IHL. Nevertheless, at its core, the following section confirms the authors’ firm assertation that as part of their DTP, future commanders will have no choice but to use EAIs, and ultimately, to “respect” their decision not to follow orders.

A. *How does the Duty Take Precautions fit within IHL?*

The purpose of this section is twofold. First, the authors identify a combatant’s lawful obligation to take precautions in attack. This obligation is contained within Article 57 of Additional Protocol I,[[21]](#footnote-21) and the analysis demonstrates how it fits within the wider corpus of law relating to the conduct of hostilities. Secondly, the authors utilize this section to demonstrate how the *customary* duty to disobey a manifestly unlawful order will become applicable to increasingly autonomous technologies such as EAIs, and moreover, how this will become a pivotal consideration within the codified duty to take precautions in attack.

IHL strives to “achieve” a thankless task, and to reconcile two fundamental values which operate at opposing ends of the “spectrum”. These are: 1) the protection of civilian life and civilian objects, and; 2) the concept of military necessity.[[22]](#footnote-22) Considerable attention has been dedicated to the “achievement” (or not) of this precarious balancing act, and the present authors do not seek to further that particular discussion here.[[23]](#footnote-23) Nevertheless, the legal barometers of distinction and proportionality act as safeguards to ensure that reconciliation between those two fundamental values takes place in practice. The first of these two, fundamental, IHL principles provides the ‘basic rule’ that is contained within Article 48 API.[[24]](#footnote-24) This provision ensures that civilians, and civilian objects, are distinguished from military objectives. Distinction is clearly intended to reflect the protectionist ambit of the balancing act. Article 51 (2) API ensures that once identified, such civilians and civilian objects must not be made the object of a direct attack,[[25]](#footnote-25) or indeed, must not be threatened with violence.[[26]](#footnote-26)

Chapter 4 of API relates to Precautionary Measures, and Article 57 API specifically to precautions in attack. Article 57(1) states, that “[i]n the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects.” [[27]](#footnote-27) Article 57(2), however, continues to offer more detail, stating,

…a) those who plan or decide upon an attack shall: (i) do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects…[and]…(ii) take all feasible precautions in the choice of means and method of attack with a view to avoiding, and in any event minimizing, incidental loss of life, injury to civilians and damage to civilian objects.

The principle of proportionality contained within Article 57(2)(a)(iii) attempts to ensure that a combatant refrains from launching an attack where the damage caused to civilians and / or civilian objects “would be excessive in relation to the concrete and direct military advantage anticipated.”[[28]](#footnote-28) Viewed holistically, these provisions, inter alia, acknowledge the concept of military necessity. And, in doing so, during the conduct of warfare “situations may arise where civilians simply cannot be spared.”[[29]](#footnote-29) Article 8(2)(b)(iv) of the Rome Statute identifies that carrying out a disproportionate attack is a war crime.[[30]](#footnote-30) The ICRC identify that the principle of proportionality is customary in nature,[[31]](#footnote-31) as is the separate duty to take precautions in attack.[[32]](#footnote-32) Consequently, these provisions are binding upon all states, whether or not they are party to API, (and are also applicable in both IAC and NIAC). By implication, this enables due consideration of U.S. practice.[[33]](#footnote-33)

An additional, and relevant provision to DTP, is Article 35(1) API. This states, “…the right of the Parties to choose methods or means or warfare is not unlimited”.[[34]](#footnote-34) The term “**means**” refers to the choice of weapon and obliges an attacker to select “weapons and munitions, that prevent ‘overkill’”.[[35]](#footnote-35) Furthermore, “**methods**” of attack also include, “angles of attack”, “time on target”, and “similar tactical choices.”[[36]](#footnote-36) If an attacker has a choice of means and method to achieve a legitimate military objective, they “should” choose the one that “is likely to cause the least collateral damage or incidental injury”.[[37]](#footnote-37)

This final statement is *central* to the discussion relating to the concept of robot refusal. And, is in fact the cord that binds the present discussion with that in part III. As the authors have argued elsewhere, the introduction of increasingly advanced AI technology into warfare is inevitable.[[38]](#footnote-38) And, for example, an EAI could be an advanced theatre ballistic missile that was capable of conducting an additional distinction and proportionality assessment before striking its target. If the “on-board” tech identifies that the intended target is positioned next to a hospital or school, it could act accordingly, and withdraw from the attack. Equally, an alternative form of EAI might be a unarmed reconnaissance platform, or a humanoid (unarmed) member of a special ops group (an armed member being an AWS). Irrespective, such systems will be capable of monitoring live feeds of battlefield conditions outside of the immediate area, and as a result, can determine the relative value of a target.[[39]](#footnote-39) These systems would then direct human decision-makers accordingly. Furthermore, an EAI, or AI software, could be utilized in a command and control center to support high-end operational decision-making regarding the deployment and movement of strategic assets. A system such as this could encourage decisions to be made that result in changes to the operational fabric of the armed conflict.

No matter which type of AI is at the operator’s disposal, however, (leaving aside whether it offers an improvement in the means, or methods, of warfare), there will be occasions where a decision maker *must* utilize it. To not do so and to use an alternative means that might cause greater civilian harms would breach the DTP obligation.[[40]](#footnote-40) This is a natural continuation of the well-versed discussion of Schmitt and Widmar, who note the loitering capabilities of UAVs (unmanned aerial vehicles) significantly enhance target verification.[[41]](#footnote-41) Central to the current article, is where such tech is capable of conducting additional distinction or proportionality assessments, it must also be capable of saying “no”. Should the circumstances change after the human instruction to attack is given, the EAI or AWS must be capable of disobeying. This is analogous to existing operations whereby the tactical team on the ground can see that circumstances on the ground are manifestly different to those that were imagined when the attack was ordered.[[42]](#footnote-42) Undeniably, the team would have discussed numerous scenarios/ outcomes, and would very likely have also pre-determined a relative value of the target. Nevertheless, the DTP (amongst other considerations) could lead to the aborting the mission.

In all scenarios the decision regarding the most appropriate means and methods, are governed by the concept of “feasibility”. In planning an attack, a combatant must (as a minimum) do only what is “feasible” in the circumstances to verify civilians / civilian objects (feasible varication). Moreover, the means and methods attack are to avoid, or at least minimize collateral damage (feasible precautions). The tactical team on the ground may still decide to consummate the attack because of the target’ value. Equally, command and control may choose to extract the team, and instead authorize the use of an over the horizon precision strike. Nevertheless, ‘what is, or is not feasible is inherently contextual,’[[43]](#footnote-43) coupled to the fact that the term “feasible” is itself, subject to interpretation. Notably, there is no supplementary clarification as to its meaning either within API, or, indeed, within the wider Geneva Conventions and/or Additional Protocols.[[44]](#footnote-44) Instead, to identify a codified definition, one must look to an alternative treaty.[[45]](#footnote-45) A regularly cited example, is Protocol II Annexed to the Convention on Certain Conventional Weapons.[[46]](#footnote-46) Article 3(4) of this treaty states,

[f]easilble precautions are those precautions which are practicable or practically possible taking into account all circumstances ruling at the time, including humanitarian and military considerations.

This is generally understood to be the standard by which feasibility is assessed.[[47]](#footnote-47) And clearly, it is an attempt to acknowledge the delicate balance between the need to protect the civilian population, and the concept of military necessity.[[48]](#footnote-48) Schmitt and Widmar note that the ICRC’s commentary to Article 57 states that ‘practicable or practically possible entails common sense and good faith’ to which of course, the phrase, “given the prevailing circumstances”, should be added.[[49]](#footnote-49) According to the two authors, the feasibility standard, must be based upon the concept of “reasonableness”.[[50]](#footnote-50)

Since feasibility assessments are inherently contextual, however, they are predisposed to ambiguity. The brief examination that follows, cross references a number of statements or codifications of feasibility within past or present military doctrines in order to identify something more tangible. Starting with the U.S., previous doctrine has stated that its position reflects Article 3(4) CCCW. [[51]](#footnote-51) However, it is at least tenable that the U.S. places greater emphasis upon military operations that it does humanitarian concerns, by way of the fact that this has removed any reference to the latter.[[52]](#footnote-52) U.K., German, Australian and Canadian doctrine reflects Article 3(4) CCW more precisely – the text appearing in the manuals of each being virtually identical.[[53]](#footnote-53) Similarly, though perhaps applied a little more liberally, French doctrine notes feasibility is ‘that which can be realized or which is possible in practice, taking into account all circumstances ruling at the time, including humanitarian and military considerations”.[[54]](#footnote-54)

In essence, there is slight variation between these alternative interpretations. However, it must be said that each example merely refer in some way to “what is possible” in the circumstances. And, given the inherently contextual nature of warfare, it is perhaps unsurprising that such statements do little to add much in the way of extrinsic context. Therefore, when this brief examination is considered in isolation, the difference in interpretation may arguably only be considered a matter of semantics.

In summary, this section has demonstrated that DTP is intertwined with the IHL principles of distinction and proportionality. DTP, therefore, plays a pivotal role in maintaining the critical balance between the protection of the civilian population, and the concept of military necessity. There is, perhaps, a slight variation in how certain states interpret the concept of feasibility. However, in order to adhere with their DTP obligation, there will undoubtably be instances in the future, where (in order to ensure that everything practicably possible has been done to verify a target, and / or to minimize collateral damage) a commander will have to utilize an EAI / AWS, instead of a lesser IHL compliant alternative.[[55]](#footnote-55)

*B. State Practice*

With reference to the same six leading military powers considered above,[[56]](#footnote-56) (namely; the U.S., the U.K., Canada, Australia, France, Germany), the purpose of the following section is to underline state practice (with specific regard to the interpretation of the duty to take precautions within IHL) that goes beyond the concept of “feasibility”. N.B. The authors do not envisage consideration of “state practice” in terms of the formation process of a customary international rule[[57]](#footnote-57) though of course, one could suggest that if there is a suitable variance present between states it may undermine the idea that there is overall consistency in this respect.[[58]](#footnote-58) Before turning to individual “practice”, the starting point for analysis is to return to Article 57 Additional Protocol I (API). Clearly, the key tenets to distill from the test as a whole in terms of this Article’s focus, surround Article 57 API (2)(a)(i) and (ii). Ultimately, it is during the process of “feasible verification” contained in (i), and “feasible precaution” contained in (ii) that the authors wish to ground their discussion in terms of rejecting an order.

In the first instance, all six states military doctrines do make some reference to the wider “constant care” obligation. While in the majority of cases the language utilized in relevant military manuals broadly reflects Article 57 (1),[[59]](#footnote-59) a U.S. Naval manual refers more specifically to the concept of “reasonableness”.[[60]](#footnote-60) Arguably, ‘all reasonable precautions’ seems to imply a lesser obligation than that of say, Germany, whose manuals state ‘all necessary precautions’ A similar discrepancy appears with regard to the Article 57(2) element of feasible verification, which the ICRC identify as customary Rule 16.[[61]](#footnote-61) Here, a German commander is required to positively verify every target, [[62]](#footnote-62) where U.S. doctrine only requires reasonable precautions to be taken.[[63]](#footnote-63) U.K. states that a decision-maker can only be expected to do everything that is feasible.[[64]](#footnote-64) Australia restates the Article 57(2) requirement for commanders to do everything feasible,[[65]](#footnote-65) as does Canada.[[66]](#footnote-66) French doctrine mirrors the German requirement that a decision-maker must “verify that the objectives to be attacked are neither civilians nor civilian objects”.[[67]](#footnote-67)

In practice, it may prove difficult to identify the existence of such a disparity. However, on paper, there appears to be “individual” practice with regard to target verification. Insofar as EAI tech is concerned, while French and German armed forces may be “required” to deploy EAI as often as is possible (subject is availability), the U.S. are likely to give their commanders more freedom of maneuverability.

The Article 57(2) obligation for the avoidance or minimization of incidental damage only represents *one* of two elements of the concept of feasible precautions. The other is that of the means and method of attack. The ICRC, identify these as two separate customary obligations: the previous being considered in Rule 15 – the Principles of Precautions in Attack, while the latter is set out under Rule 17 – Choice of Means and Method of Warfare. Given that the U.S. is not a party to API these “customary” rules are considered separately for the sake of the following examination. With regards to Rule 15, the U.S. position,[[68]](#footnote-68) (which, is supported by that of Australia), refers to the concept of reasonableness previously noted,[[69]](#footnote-69) as does Canada.[[70]](#footnote-70) Canadian guidance, however, also states that military operations are to be “conducted in such a way that damage to civilians and their property is minimized…[and, that combatants]…use only the necessary force that causes the least amount of collateral civilian damage”. [[71]](#footnote-71) French Doctrine speaks of “all precautions”,[[72]](#footnote-72) while Germany returns to the concept of “all *feasible* precautions”.[[73]](#footnote-73) Finally, in the U.K., guidance states that “[p]recautions must be taken…to avoid civilian death or injury and damage to civilian objects”,[[74]](#footnote-74) and elsewhere, perhaps more simply stating only, “[c]are must be taken…”[[75]](#footnote-75)

Thus far, though “feasibility” is clearly a central focus of the six states doctrine that has been considered, arguably they do exhibit a degree of divergence. Notwithstanding, there is certainly no explicit statement suggesting that the obligation to avoid or minimize harm, for example, extends to eradicating *all* civilian harms. The second element of feasible precautions that is contained within 57(2), is reflected by ICRC Rule 17, - Choice of Means and Methods of Warfare. In this respect, previous Australian doctrine[[76]](#footnote-76) has reflected the argument that while a military decision-maker must do everything feasible to ensure collateral damage is minimized, “the existence of precision guided weapons munitions…does not mean they must necessarily be used”.[[77]](#footnote-77) This principle is also reflected by current U.S. doctrine which states that,

The commander must decide, in light of all the facts known or reasonably available to him, including the need to conserve resources and complete the mission successfully, whether to adopt an alternative method of attack, if reasonably available, to reduce civilian casualties and damage Under the heading of ‘proportionality’.[[78]](#footnote-78)

Canadian doctrine refers back to the language of Article 57(2),[[79]](#footnote-79) and similarly, a French manual states that decision-makers shall ‘take all precautions which are practically possible in the choice of means and methods of attack…”[[80]](#footnote-80) German doctrine is, perhaps, more stringent in requiring that “before engaging an objective, every responsible military leader shall…choose means and methods minimizing incidental injury and damage to civilian life and objects”.[[81]](#footnote-81) *Prima facie*, this narrower obligation is also a requirement of U.K. armed forces.[[82]](#footnote-82) However, the Manual states that this is only where it is feasible to do so.[[83]](#footnote-83) When all the elements of DTP that have been discussed in Part II are considered holistically, three primary interpretive tracks begin to develop – (i) a narrow interpretation, (ii) a central based interpretation, which largely restates the language of Article 57, and a wider interpretation. These 3, interpretations, along with track adopted by each of the states considered, is presented in graphic below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Obligation to take Constant Care. | Feasible Verification.  (ICRC Rule 16.) | Feasible Precautions (A):  Minimize Civilian Harms.  (ICRC Rule 15.) | Feasible Precautions (B):  Choice of Means and Methods.  (ICRC Rule 17.) |
| Wide Interpretation. | **U.S.** “…*take all reasonable precautions…”* | **U.S.** *“…ensure reasonable precautions are taken…”* | **U.S.** *“…where reasonable…”* | **U.S.** *“…conservation of resources… [and] …reasonable availability…”* |
|  |  |  | **Aus.** “…*where reasonable…”* | **Aus.** *“…consider wider mission objectives…”* |
|  |  |  | **Can.** *“…where reasonable…”* |  |
|  | **Aus.** “…*maximum extent possible…”* | **Aus.** *“…feasible precautions …”* |  |  |
|  | **Can.** “…*care should always be taken…”* | **Can.** “…*feasible precautions (as per Art. 3(4) CCCW)…”* |  | **Can.** *“…all feasible precautions …”* |
| Central/ Treaty based approach. | **U.K.** “…c*onstant care shall be taken*…” | **U.K.** “…*feasible precautions (as per Art. 3(4) CCCW)…”* | **U.K.** *“…where feasible* *(as per Art. 3(4) CCCW)…”* | **U.K.** *“…least collateral damage commensurate with military success…”* |
|  | **Fra.** “*…constant care shall be taken…”* |  |  | **Fra.** “*…all precautions which are practically possible…”* |
|  |  |  | **Ger.** *“…where feasible (as per Art. 3(4) CCCW)…”* |  |
| Narrow Interpretation. |  | **Fra***. ‘…Must verify objects to be attacked…”* | **Fra.** *“…take all precautions…”* |  |
|  | **Ger.** *“…all necessary precautions (to spare as far as possible)…”* | **Ger. *“…****shall verify the military status of target…”* |  | **Ger.** *“…shall choose means and methods to minimize civilian harms…”* |

*Figure 1. The IHL Duty to Take Precautions in attack: State Practice.*

With reference to figure 2, it is evident that while the language used by states is somewhat similar, certain states do appear to place a higher burden in an attempt to ensure DTP compliance. For example, a German soldier *must* assess each target before choosing the means and method that will minimize civilian harm. In contrast, an Australian, and/ or U.S. decision-maker is encouraged to consider the wider mission brief/ status, rather than only a specific target. When this practice is distilled one step further the narrow and wide extremes can be presented as follows,

|  |  |  |  |
| --- | --- | --- | --- |
| State Practice/ Interpretation | Measures taken to ensure target verification…; | All precautions to minimize civilian harms where…; | Use Means or Method…; |
| Wide | Must be reasonable in the circumstances. | Reasonable | Where it is reasonable to do so, *[and in particular where it is consistent with the concepts of resource conservation, and/ or reasonable availability].* |
| Narrow | Must be all available measures in all circumstances. | Feasible | Whenever/ wherever possible. |

*Figure 2. The Narrow and Wide interpretations of the Duty to Take Precautions*

In sum, the discussion in Part II has considered DTP through many lenses. For the benefit of the current Article, the focus has largely, and intentionally, centered upon “feasible verification”, and “feasible precautions”. One may also reasonably add the general obligation for the targeteer to take constant care, to which the concept of feasibility still applies.[[84]](#footnote-84) Having considered the practice of six states, the analysis has identified that U.S. doctrine is positioned upon the military necessity extreme of the spectrum when compared to some of its contemporaries. Though this is most noticeable when compared to the practices of Germany and France. Australian military doctrine more closely reflects that of the U.S., while the U.K. and Canada sit somewhere in-between. One may posit that the U.S. simply applies a different language as a result of not wanting to restate legal obligations to which they are not party. It is, however, beyond the scope of the present article to determine whether this is actually the case.

When considered holistically, Part II has identified a number of relevant characteristics, including the fact that DTP is irrevocably intertwined with distinction and proportionality. Together, these three concepts help maintain the delicate balance between humanitarian considerations and military necessity. The elements of DTP (as identified in API) seek to ensure that a minimum standard of care is applied. The ICRC note that each of these individual elements are customary in nature and thus binding upon all parties. With reference to the sample of state practice considered, there appears to be a certain degree of flexibility with regards to their application. However, regardless of these variations, it is undeniable that there will inevitably be circumstances where those that possess EAI and AWS tech, will be compelled by IHL to use them. The central question that is posed by the current thesis, and which the following section begins to address, is what are the conditions under which an EAI should be permitted to *refuse* an order?

III. Rules of Engagement and Disobeying of Orders

The discussion in Part II identified that the DTP that is contained within Article 57 API is subdivided into numerous interconnected paths.[[85]](#footnote-85) That analysis concluded that while a narrow and wide interpretation of “feasibility” can be crafted, neither path detracts from the certainty that where EAIs are an available option, armed-forces will, under certain circumstances. be obligated to utilize them. The purpose of Part III is to forensically examine the interface between DTP and the state-based Rules of Engagement (ROE).[[86]](#footnote-86) Here, specific reference is naturally-centered on the disobeying of orders. As noted, in the non-too-distant future, DTP will require military commanders to deploy robots with sophisticated decision-making capabilities where they offer the most appropriates means and /or method of avoiding or minimizing civilian harms. However, when they do, a key question arises as to whether robots can, and should, refuse an order to attack? Similar to the previous examination, the first section of Part III identifies how the obligation to disobey certain orders fits within the wider corpus of IHL, and in particular within DTP.[[87]](#footnote-87) This, in turn, “lays up” the discussion in section B, which, examines whether there is significant variation of state practice in this area.

A. *How does the Duty to Disobey an Order fit within DTP?*

This section identifies that a soldiers duty to obey the vast majority of orders is what differentiates them from members of the civilian population. In contrast however, it also identifies a customary duty that requires a solider to disobey a manifestly unlawful order. Finally, it considers what a manifestly unlawful order is, and how the duty to disobey such fits within the wider body of DTP. Part II identified that while there is some disparity between national interpretations of them, adherence to Art 57 API, and indeed, Article 35 API, means that a military decision-makers “means” and “methods” of warfighting are not unlimited. Indeed, the discussion above has highlighted the two central elements of targeting law: (i) an order-giver must take precautions to ensure that targets are verified, and (ii) that civilian harms are avoided where possible, or minimized where they cannot be avoided. As Part II also identified, API provides that these two criteria are subject to the concept of “feasibility”. Whether that means the two central elements are considered an “absolute obligation”, or whether they should be carried out in line with the concept of “reasonableness”, differs from state-to-state. Though the current authors accept that the former might be a somewhat lofty aspiration.

No matter which interpretation is adopted, in certain environments, a commander will be obligated to use a certain weapon in his or her arsenal. This is applicable to *all* means and methods of warfare, and in absence of a further substantial IHL being introduced, this will be the case for years yet to come. No matter whether AWS as a “means”, are ever fully realized, unarmed AI systems, including EAIs, *will* almost certainly play a role in enhancing the “methods” used by armed-forces. As previously noted, the current authors support the proposition that such robots will certainly become increasing autonomous. And as a result, they will almost certainly able to make certain decisions. For example, although a human typically programs an autonomous car (an EAI) with a destination address, or set of GPS coordinates, it is the car’s software that is responsible for taking the majority of decisions about which turns to make, or perhaps for the sake of the current Article, which turns not to make.[[88]](#footnote-88) In the military domain, the task may not be as straightforward as getting from “A” to “B”. Instead, as identified in Part I, EAIs may be programmed to carry out all sorts of tasks. One thing IHL will ensure that they will have in common, is that the positive military advantage on offer, is balanced with humanitarian considerations. And, in order to achieve the latter, EAIs may often have to refuse to order.

The proposition that EAIs can account for a “gray area” that currently exists with regards to a combatants’ right to disobey *certain* orders is central to the present thesis. As a consequence, the following discussion identifies two relevant concepts. First, it demonstrates how members of the armed-forces must currently ( in the vast majority of circumstances), *obey* orders. In addition, it also introduces the circumstances in which a combatant does the right to refuse to obey an order. By way of further explanation, the first matter that is considered is the combatants’ duty to follow orders. Indeed, as noted by one commentator,[[89]](#footnote-89) from 1863, Article 44 of Lieber code permitted a commanding officer to shoot and kill a subordinate on the spot, should they disobey their order to cease committing a crime.[[90]](#footnote-90) Today, when a soldier refuses an order they may not face quite the same fate. However, they will, for example, be subject to court martial and / or imprisonment if appropriate.[[91]](#footnote-91) As a consequence, in the vast majority of circumstances (while seemingly at odds with the ICJ’s rejection of the ‘Nuremberg defense’),[[92]](#footnote-92) a combatant *may* rely upon the defense that they were acting in accordance with orders.[[93]](#footnote-93)

The majority of domestic courts recognize one of three variations of this defense - (i) that following orders is always a defense; (ii) that following orders is a defense where the combatant believed it was a lawful order, and; (iii) that following orders is a defense where it would have been *reasonable* to believe that the order was lawful.[[94]](#footnote-94) The U.S., for example, follows the latter of these three variants, with the passage previously identified in the US Courts-Martial Manual, continuing on to assert, “…unless the accused knew the orders to be unlawful or a person of ordinary sense and understanding would have known the orders to be unlawful.”[[95]](#footnote-95) In other words, according to the second and third variants of the ‘following orders’ defense, there may be circumstances where a soldier either ‘will’ either recognize an unlawful order, or where they ‘should’ have recognized that it was an unlawful order. This is clearly an exception to the combatants’ obligation to obey orders. As identified by the ICRC, this exception is customary and binding upon all parties.[[96]](#footnote-96) Significantly, this customary duty not only means that there are limited circumstances under which a combatant *can* refuse to follow an order, but, that in such circumstances “[e]very combatant has a *duty* to disobey a manifestly unlawful order’.[[97]](#footnote-97) In other words, the obligation is reversed. Indeed, where the combatant does not fulfil this latter duty, but instead carries out a manifestly unlawful order, there can be no ‘Nuremberg” defense, and the subordinate will not be relieved from individual criminal responsibility. This is also a customary provision.[[98]](#footnote-98)

Manifestly unlawful, or otherwise, manifestly criminal, is generally accepted to refer to the commission of war crimes. This includes breaches of IHL principles of distinction and proportionality. In addition, crimes against humanity, and crimes against the peace would also, in the vast majority of cases, be considered manifestly unlawful.[[99]](#footnote-99) In short, a combatant is only presumed to know the law that regards a subset of crimes that are ‘immediately recognizable as manifestly criminal by a person of ordinary understanding’.[[100]](#footnote-100) And, such a limitation is perhaps, *just*, for it is questionable whether the law should require armed services personnel to carry out their non-civilian duties while simultaneously having to conduct a full legal assessment of each and every instruction that is handed down to them by a superior. In reality, soldiers are not trained as lawyers, and nor should they be. Instead, the obligation to *obey* their superiors is what sets combatants apart from civilians.[[101]](#footnote-101) The question instead, is *should* the same principle apply to an EAI given their ability to process vast amounts of data at super-human speeds? In other words, should robots have an obligation to refuse an order, and if so, should it extended further than merely manifestly unlawful orders, for example, to include human rights, a moral or ethical code, and / or policy assessments?

*B. Disobeying of Orders*

The following section examines existing state practice in order to identify whether it can help determine whether there are circumstances under which robots should be programmed, and indeed lawfully permitted to refuse a superiors order. The states whose practice is considered below are the same six that formed a part of the analysis in Part II; Australia, Canada, France, Germany, the U.K., and the U.S, with the addition of three further states: the Philippines, the Republic of the Congo (Congo), and, the Democratic Republic of the Congo (DRC). These additions seek, inter alia, to increase the scope of the analysis due to the fact that, unlike the former six, they are not widely considered to be leading military powers.

Science fiction writer Isaac Asimov’s second law of robotics states that, “a robot must obey orders given to it by a human except where such orders conflict with the first law’.[[102]](#footnote-102) His first law is “a robot may not injure a human being, or, through inaction, allow a human being to come to harm”. As the current authors have noted elsewhere,[[103]](#footnote-103) while these rules allow Asimov to cleverly write stories that toy with human-EAI relationships, they are utterly inadequate for robots that are programmed specifically to apply “deliberate…often lethal capabilities in order to produce maximum effect upon an enemy”.[[104]](#footnote-104)

Instead, the discussion in the previous section identified that a soldier has a strict obligation to follow an order, except where they have a customary duty to refuse to follow an order. Presumably, when advanced EAIs are introduced, this obligation would simply be transposed and implemented. However, the following section recognizes that in practice, the types of orders and circumstances under which a combatant is expected to refuse an order differ significantly. These variations exist because they are contained within municipal military doctrine, or ROE, as opposed to being codified within IHL.[[105]](#footnote-105) Somewhat regrettably, this means that while the members of certain armed forces may make a subjective determination as to which orders should be disobeyed, others may do so only objectively.[[106]](#footnote-106) This is concerning, not least, because the current authors support the view that EAI, and AWS, tech can increase adherence to international law. EAI’s can operate in adherence with vast amounts of preprogrammed information, including, for example, unabridged IHL databases. Moreover, they will able to do so while remaining unaffected by the threats and pressures that are typically associated with armed conflict. However, while autonomous technology in general can offer an opportunity to ensure the duty to disobey a superior order is applied equally, the current variation in state practice is likely to be reflected as somewhat of a national bias when programming EAIs and other such technologies.

When it comes to the criteria that help determine when a soldier should disobey an order state practice in is somewhat fragmented. For example, in the U.S., ROE generally identify that a combatant must distinguish only between lawful, and unlawful orders (the latter of which must be disobeyed).[[107]](#footnote-107) In this instance, and in contrast to the way in which the U.S. interpreted the treaty and customary obligations considered in part II, U.S. national doctrine appears to support a “narrow” obligation. However, this narrow obligation is consistent with the concept of “manifest” unlawfulness that is recognized by the ICRC as being customary in nature.[[108]](#footnote-108) The U.K. armed forces adopt a similar stance, with one military manual simply stating that ‘[m]ilitary personnel are required to obey lawful commands but must not obey unlawful commands’.[[109]](#footnote-109) A similar example can be found in Philippine military doctrine, which states “[a]nyone who shall refuse or fail to carry out a lawful order from the military chain of command shall be subject to military discipline”.[[110]](#footnote-110)

Though only subtly different, a number of other states adopt a slightly wider, but perhaps more informed approach. French ROE, for example, identifies the sources of law that should specifically be used to inform the decision whether disobey an order. These being; “the customs of war…[*and*]…the rules of international law applicable in armed conflicts, or duly ratified or approved international treaties”.[[111]](#footnote-111) Analogously, and offering guidance as what to do where an order is ambiguous, Australian doctrine supplies that, “clarification should be sought. If clarification is unavailable, any action taken must comply with LOAC [Law of Armed Conflict]”.[[112]](#footnote-112) In a similar vein, Congolese doctrine states “the subordinate must not execute an order to commit an act manifestly…contrary to the customs of war and to the international conventions”.[[113]](#footnote-113)

The obligation to refuse an order by members of the armed forces from Germany, the Democratic Republic of the Congo (DRC) and Canada must, however, be considered “wide”. While German, Congolese and Canadian military doctrine appears to be aligned with the standard of manifest unlawfulness,[[114]](#footnote-114) the doctrine supplied by each state to its combatants is significantly more conceptual than that which is the offered to the armed forces of states such as the U.S., and the U.K.

Article 28 of DRC’s Constitution provides, for instance, that,

No one is required to execute a manifestly illegal order. Every individual, every agent of the State is released from the duty of obedience if the order received constitutes a manifest violation of respect for human rights and public liberties and morals.

This clearly goes further than merely breaching the principles of IHL, and even the wider LOAC, even if that is taken to include International Human Rights Law (IHRL).[[115]](#footnote-115) Moreover, the obligation to apply a moral perspective when considering manifest unlawfulness, also casts the proverbial net much farther than the seemingly straightforward lawful, or unlawful assessment that the narrow interpretation adopts. Canadian doctrine mirrors the Australian position regarding the importance of seeking clarification where an order appears ambiguous, or more correctly where its lawfulness is questionable.[[116]](#footnote-116) However, it goes on to identify a manifestly unlawful act is one which “shocks the conscience of every reasonable, right-thinking person”.[[117]](#footnote-117) While this is perhaps not quite as conceptual as the DRC guidance, it remains, nevertheless, much wider in scope than a mere assessment of lawfulness.

The final example considered is German doctrine. As a state, Germany’s interpretation of the customary duty to disobey an unlawful order is arguably wider than that of its peers due to its historical requirement for unconditional obedience to military orders.[[118]](#footnote-118) Instead, presently, German military doctrine provides that an order does not need to be executed if,

it violates the human dignity of the third party concerned or the recipient of the order; it is not of any use for service; or in a definite situation, the soldier cannot reasonably be expected to execute it. [[119]](#footnote-119)

Once again, this provision is clearly some distance from the narrow Part III obligation not to carry out an unlawful, or manifestly unlawful, order. Rather, according to German doctrine, the combatant must consider the lawfulness of the order, in addition to their dignity, as well as that of the target.[[120]](#footnote-120) If that were not considerably ‘wide’ enough, a German combatant must also consider whether, in the circumstances, there is a reasonable chance of mission success, and/or whether the order serves the “defense of Germany…[and]…the pursuit or achievement of its political or economic aims”.[[121]](#footnote-121) Therefore, a member of the German armed-forces must evaluate the very reasoning behind every superior command. Here this wide application is in direct contrast to the examinations that took place in Part II, where (at least in the case of Germany), a narrow interpretation of Art 57 API was supported. There is, therefore, a significant amount of variation with regards to current state practice in this area. The various models (narrow, central/treaty based, and wide), and relevant state provisions are presented in the flowing graphic:

|  |  |
| --- | --- |
|  | **Disobeying Orders**  A soldier has a strict obligation to follow an order, and to do so is a defense (three variants);  (i) Always; (ii) Where they believed it was a lawful order, or (iii) Where they should have known it was an unlawful order…(see ICRC Rule 155)  Therefore, the is a duty to an follow order, except where there is a customary duty to refuse (ICRC Rule 154).  State practice relating to the circumstances in which a soldier should therefore refuse to follow an order includes: |
|  |  |
|  | **U.S.** *“…must distinguish between lawful and unlawful…”* |
| Narrow Interpretation. | **U.K.** *“…obey only lawful commands…”* |
|  | **Phi.** *“…must carry out a lawful order…”* |
|  | **Aus.** *“…Seek clarification if ambiguous, or follow LOAC..”* |
| Central/ Treaty based approach. | **Can.** *“…manifestly unlawful means it would shock the conscience of every right thinking person…”* |
|  | **Fra.** *“…the customs of war…LOAC…duly ratified or approved international treaties…”* |
|  | **Con.** *“…manifestly contrary to customs of war and international conventions…”* |
| Wide Interpretation. | **Ger.** *“…if it violates human dignity (of first and third party), if the order is of no use to service, of if soldier cannot reasonably be expected to execute it…”* |
|  | **DRC.** *“…can disobey an order if it is a manifest violation of human rights, of public liberties and morals…”* |

*Figure 3. The Duty to Disobey an Order: State Practice.*

The result of such disparities between the narrow and wide variants are often stark. Though they are straightforwardly demonstratable. For instance, the U.S. military courts have consistently refused to recognize that the following orders, or the “Nuremberg Defense” is applicable to the government’s decision to wage war.[[122]](#footnote-122) As a result, U.S. First Lieutenant Watada (who had previously refused to deploy to Iraq in 2006 because he believed the war there was unlawful and, therefore that any order he would receive there would also be), failed to convince the court to that effect.[[123]](#footnote-123) While he did escape a charge for insubordination, he ultimately left military service with “an ‘other than honorable’ characterization of service – the worst administrative discharge…an officer can receive”.[[124]](#footnote-124) This was the case, even though U.S. Army doctrine that existed at the time states “an essential foundation for Army leaders is a character ‘comprised of a person’s moral and ethical qualities [which] helps to determine what is right…regardless of circumstances or consequences’”.[[125]](#footnote-125)

Contrastingly, however, in 2005, the German Federal Administrative court (with Germany having no military courts), had to consider very similar facts. The court had to determine the case of a German Major who had refused to take part in a NATO information technology (IT) project that would potentially support Operation Iraqi Freedom. Similar to First Lieutenant Watada, the German Major believed that the conflict was unlawful.[[126]](#footnote-126) The *Limits of Obedience to Superior Orders Case* is, however, the antithesis of Watada. In contrast, the German court upheld the claim of the Major, and acquitted him of the charge.[[127]](#footnote-127) Having itself considered that the lawfulness of the war was indeed questionable, the German court instead held that the Major could rely upon his the fundamental right of freedom of conscience under German Basic law.[[128]](#footnote-128) As a result, he was permitted to seek, and to be assigned, an alternative task.[[129]](#footnote-129) This is an important decision, not least, because it reaffirmed that under German law “the duty to obey orders ‘does not demand blind or unconditional devotion to superiors”.[[130]](#footnote-130) However, in addition, and perhaps most notably, the decision recognized that a German combatant should not be made to act against their moral or ethical convictions.[[131]](#footnote-131)

While this brief examination of these two similar but contrasting cases provides a fundamental, and useful example of the differences between what the current authors identify as the narrow and wide obligation, it is not necessarily a novel, or a ground-breaking discussion.[[132]](#footnote-132) However, what it does highlight for this Article, is that in all but the most incontrovertible of circumstances,[[133]](#footnote-133) a combatant can rarely be certain that their decision will be supported, because it is unclear where the boundaries lie. In fact, a soldier will have to consider many factors when contemplating whether to disobey an order, not only whether they will face court-martial or criminal charges. For instance, while the reasons to disobey an order can include self-preservation, preventing breaches of IHL, and/ or the commission of war-crimes,[[134]](#footnote-134) other factors, including political,[[135]](#footnote-135) and financial pressures,[[136]](#footnote-136) training, a sense of patriotism, group solidarity, or, the fear of being branded a coward by one’s peers,[[137]](#footnote-137) might cause any reasonable person to hesitate before doing so. As is examined in the following section however, EAI and AWS have the potential to circumnavigate many of those elements that might be considered “unwanted”.

Nevertheless, many of the negative influencing factors could be removed if an EAI was tasked with making an objective determination as to whether an order should be obeyed or disobeyed. The EAI could then guide the human decision-maker accordingly, or even prevent a course-of action undertaking mistakenly, or maliciously. Moreover, where a particular instruction was not considered unlawful, an EAIs decision-making capability could still be used as a tactical tool to the commander. In this regard, EAI might accurately predict the probability of mission success, issue relevant instructions to a commander or, simply refuse to pass on his instructions to his or her subordinates. Although it may be somewhat of a cyclical argument, the fact that a soldier may be obliged to use an EAI to minimize collateral damage, may also mean that objective rational decision-making is applied more evenly in the ‘fog of war’, further ensuring adherence to the wider DTP.

If three primary reasons for refusing an order are; (i) deterring a perceived wrong, (ii) exercising a freedom of conscience, and; (iii) avoiding self-harm,[[138]](#footnote-138) it is arguable that only the first is applicable to a machine. However, a machine can support the human in respect of the other two. Nevertheless, the question posed in the following section, is in seeking to identify an objective criteria for EAI, whether they should be programmed with the “narrow” obligation previously identified, the “wider” obligation, or something in-between? For the sake of the examination in the following section, the discussion, thus far, can be distilled as follows:

|  |  |
| --- | --- |
|  | *Adherence to DTP, and Influencing Factors for Disobeying Superior Orders* |
| *Track 1* | *Every soldier should take all reasonable precautions to verify targets as military objects, and should take all reasonable steps to minimize civilian harms. In choosing the means and methods of attack, the decision-maker should consider their reasonable availability, with a specific regard for the wider mission objectives. A soldier must only refuse to follow an order where they know, or should know, that it is a manifestly unlawful order.* |
| *Track 2* | *Every soldier should take all steps to ensure target varication, and the minimization of civilian harms, where they are practicable or practically possible. The soldier should take into account the circumstances ruling at the time, including humanitarian and military considerations. In doing so, the decision maker must consider all of the customs of war, and the applicable international treaties, and refuse to follow any order which would be in breach of them.* |
| *Track 3* | *Every soldier must take all necessary precautions to ensure a target is verified, and in each case, must choose the means and methods that ensures civilian harms are minimized. The factors which should influence the decision to refuse to follow an order, include, but are not necessarily limited to; the lawfulness of the order; whether following the order would violate the human dignity of the first or third party (including personal moral and ethical standards); whether there is a reasonable chance of mission success, and; whether the order is consistent with national policy/ strategy.* |

*Figure 4. The Three Interpretive Tracks.*

Naturally, each of these three tracks are themselves subject to some level of fluidity with regards to further interpretation. This is particularly true of the central/ treaty based track, which could either be heavily weighted toward the narrow or the wide variant. Nevertheless, for the sake of examination in Part IV, these three models provide a more than suitable point of reference.

Nevertheless, in sum, DTP will in certain circumstances oblige a decision maker to deploy an EAI/ AWS. In many, if not all of the cases where this happens, an EAI will have to “decide” upon an appropriate course of action. On occasion, and in order to keep within the limits of IHL, this will mean that EAI will have to refuse an order – though the fundamental question is when? The analysis in Part III has shown that currently, the minimum standard (reflected by the narrow obligation) is that a soldier has a strict obligation to follow an order, except where they have a customary duty to refuse to follow a manifestly unlawful order. However, this creates a ‘gray area’, because in practice, the obligation to refuse is actually much wider. Indeed, for some states, the soldier must display “situational awareness” in order to determine whether the order should be followed in light of many varied considerations.[[139]](#footnote-139) A question that is central to the current debate therefore, and one which is considered in Part IV is, in programming EAIs, which should be the standard adopted by EAIs?

IV. Disobedience and Nuclear Launch

This section introduces the authors proposed “test”, which is otherwise referred to as Robot Rules Of Engagement (RROE). As previously noted, these RROE are tailored, for the sake of the present discussion, only insofar as addressing the concept of “refusal”. The test is based upon a “system of systems” approach, where each level/ systems represents a task that is to be undertaken by an EAI, or EAIs. Once the RROE have been established, the text goes on to consider a number of tangible scenarios as to how this would operate in practice. At the heart of this Article, and discussion, is the practical simulation and application of the test proposed in section A (below). The fourth and final scenario envisaged is the instance of an EAI questioning the human decision of nuclear launch. Currently, an order for nuclear launch goes simultaneously to both central command and the team in the field, making recall difficult.[[140]](#footnote-140) In other words, there could be the curious case of one part of the launch team wishing to disobey, but because that original message goes to both, recall and override is no longer possible. Before moving on to consider each scenario, however, section A proposes a test for calculating and overcoming human error.

A. *Additional Test for Safeguarding Decision-making.*

The purpose of this section is to introduce a ‘test’ capable of calculating human error (in respect of issuing orders), to prevent certain orders from being carried out, and thus reduce Clauzewitizian friction. Nevertheless, the authors fully acknowledge that introducing such a test cannot eliminate friction in all situations, all of the time. The EAI examination of orders must be seen as an additional way for ensuring DTP obligations are met. The RROE are based largely upon the discussion in Parts II and III above, an in particular to the central obligation identified in *figure 4*. For the reasons that are considered in the following paragraphs, the test that is introduced takes elements from all 3 tracks that are identified in figure 4. This is for a variety of reasons, but not least because the central premise of IHL is to provide a system of “checks and balances…aimed at minimizing human suffering without undermining the effectiveness of military operations.”[[141]](#footnote-141) In other words, IHL is a compromise, [[142]](#footnote-142) and this must be reflected by any EAI given that is given the task of applying it. Consequently, any usable test cannot be weighted too heavily upon one concept or the other.[[143]](#footnote-143)

The current authors believe track 1 to be too strict an interpretation for the following reasons. First, too much weight should not be given to the wider mission objectives, and certainly not at the sacrifice of humanitarian considerations. Although the authors do agree that a soldier must be aware of the means and methods at their disposal, and the likelihood of needing to utilize them in the future to give them better effect (humanitarianly and militarily), each analysis should attempt to minimize civilian harms where practicable or practically possible. This is arguably a stricter application than to do so only where there is a reasonable availability, with specific regard for the wider mission objectives.[[144]](#footnote-144) Future EAIs will effectively determine not only whether the order is lawful, but whether it is consistent with the political and strategic doctrine that is contained within ROEs. ROEs often impose much greater restrictions than IHL alone requires, [[145]](#footnote-145) and may help, for example, to determine whether or not an order is of *“use for service”.* And, if these more “precise” guidelines could be programmed into an EAI, they help it to identify when an order should be refused, because, for example, it was provided by an individual that was acting out of revenge.[[146]](#footnote-146)

In contrast however, track 3 is also unsuitable for a variety of reasons. In the first instance (as previously identified), this is because a soldier must have at least some regard to the wider mission objectives, both military and humanitarian. For example, if an infantryman carried two grenades, one of which was traditional munition, and the other of which was independently capable of carrying out an additional proportionality assessment before detonating, they may wish to reserve the latter in the knowledge that they are about to enter a more densely populated locality. And, it would be wrong, both humanitarianly, and militarily to deny them the opportunity to do so.

A further reason track 3 fails to satisfy, is that while the test posed by the authors test is intended to identify orders that are, for example, made with “malicious” intent, the nature of factors identified in the wide obligation track are somewhat excessively conceptual.[[147]](#footnote-147) For example, the ICJ provide, that Human Dignity is ‘[t]he essence of the whole corpus of IHL as well as human rights law…’[[148]](#footnote-148) Nonetheless, as a concept, human dignity remains relatively undefined, and ultimately unquantifiable. Indeed, the German court had to utilize the right to freedom of conscience under German Basic law as a dignity “enabler”. Nonetheless, the concept of human dignity goes much deeper than merely freedom of conscience, and means many things to many people. Indeed, as posited by one author, a good deal of “contemporary ideas about the role of international law are grounded on a very misplaced notion of what human dignity is”.[[149]](#footnote-149)

An in-depth investigation into this matter is well beyond the scope of the present Article. However, the authors believe that it would be particularly problematic to include the human dignity element of the obligation for two primary reasons. Although somewhat controversial, the first of these is that objects and people, can ultimately be assigned values. They can, therefore, be expressed, or measured as a quantity and transferred into programmable code. Indeed, this would have to be the case if machines were ever to successfully carry out distinction and proportionality assessments.[[150]](#footnote-150) This may not be the case with a much more conceptual provision such as Human Dignity. However, the overriding issue is not that it would be a difficult to codify and “up-load” dignity, but, even if was possible, “[w]ar itself takes a toll on human dignity through the intentional sacrificing of lives to achieve military objectives”.[[151]](#footnote-151) One could posit that war itself was undignified.

Instead, the authors propose the test needs to have particular regard to the legal principles, and the relevant ROE, of which mankind has developed over a sufficiently long period of time in order to attempt to maintain the balance between military objectives and humanitarian concerns. In the future, doctrine will almost certainly be written for the deployment of EAI, including the circumstances in which they should (or should not) be used. In short, used correctly, military Doctrine, can provide a vital, and detailed EAI *modus operandi* to decision-makers regarding the how to ensure IHL is best adhered to. What is more, this doctrine can, and in many cases is, written with due regard to human dignity, and/ or a wider ethical or moral standpoint, so these are not simply tossed aside.[[152]](#footnote-152)

The result of running an order through a wider body of obligations than merely IHL alone, is that in many cases an EAI will be able to correctly determine whether an order is legally compliant, but also whether it is consistent with matters of national policy, and whether is accords with best practice given the circumstances.[[153]](#footnote-153) Doctrine is therefore key in supporting EAI order analysis - particularly with regard as to whether the order-giver is motivated by something other than a legitimate military reason such as revenge or fear. Consequently, an EAI can go much further than merely identifying those orders which shock the conscience of every reasonable, right-thinking person,[[154]](#footnote-154)or, of which a person of reasonable standing should have known was an unlawful order. A military EAI system can, and should, in other words also be trained as a military lawyer. And, perhaps even one which is able to train, and call upon a “gut instinct” as to whether an order should be refused or followed.

With all that in mind, the authors propose the following solution. The test includes a number of “steps”, or “systems”, in order to constantly analyze the status of an order. The first of these is an authentication step. This vital phase is particularly relevant when an order is received either in the form of code (cyber), or where the EAI that is monitoring a decision is doing so remotely. Where this system fails to authenticate, it will engage the final step, which provides one of three suitable responses. These are; (I) inform the order-giver of the observation, and allow a similar order to be given (Passive Refusal); (II) immediately reject the order, but allow similar orders to be given (Active Refusal), or; (III) immediately reject the order and put safeguards in place to prevent similar orders being issued (Preventative Refusal). The most likely step in the first instance would be to refuse and seek further clarification, which would take the form of active refusal however, this would be affected by the order gravitas.

Where the order can be authenticated, the next system is engaged. This further stage (Stage 2), includes three separate systems, with each being assigned a fundamental IHL calculation. System A runs a continuous distinction assessment, and system B a continuous dynamic proportionality assessment. System C then analyses (given the circumstances), whether all practicable or practically possible precautions have been taken to minimize civilian harms. In each case, where the answer is no, there is a systematic refusal to follow the order, subject to options noted above. In certain instances, a level 1 and 2 assessment alone will be sufficient. This would be the case, for example, with the intelligent grenade, that was considered above. However, once the initial level two analysis is complete, and no reason is identified to intercept the order, system 3 would engage, (while continuously cross-referencing level one). At this point, the order would be considered subject to the criteria identified above – i.e. according to the entire body of LOAC,[[155]](#footnote-155) IHRL, other relevant treaties (including reginal agreements), in addition to ROE, and Codes of conduct.[[156]](#footnote-156) The purpose of this examination, is to determine whether the instruction is consistent with this much wider body of obligations, or, whether it is beset, for example, with motivation for revenge, self-preservation, the commission of a war crime, and national policies. Depending upon the type of order, i.e., whether it contains a single instruction, or multiple instructions, this may need to be repeated on a continuous loop. Where a reason for intercepting the order is identified stage 3 will refuse to implement the order, subject to the prerequisite stage 4 caveats, and where no reason is identified, stage 4 B will be engaged. This final stage will either complete/ or implement the order, or repeat the entire assessment so long as it is necessary to allow for completion. These four primary stages, can easily be identified upon the following graphical representation. In turn, that is followed by a number of scenarios which are intended to demonstrate how robot refusal might work in practice:



Figure 5. EAI - 4 Stage Assessment of Orders.

B. *Scenario 1*

The following scenario is intended to demonstrate how existing, often rudimentary technology, lacks the ability to apply any minor, let alone comprehensive “post-human assessment”. An anti-personnel mine is an example of existing technology that often appears in the conversation surrounding AWS, and provides an ideal point of focus here.[[157]](#footnote-157) Existing references to anti-personnel landmines commonly highlight the fact that once in position, there is no further human involvement. In essence, the weapon displays a basic level of autonomy, “deciding” itself when force should be applied. Any “decision” to detonate (or not), is based entirely upon whether a pressure threshold is surpassed when an individual steps on a pressure plate, or, when a trip wire is snagged.[[158]](#footnote-158) In the case of an anti-personnel landmine, therefore (regardless of any basic autonomy) a human makes the decision as to whether the weapon is an appropriate means, the tactic an appropriate method, that their use is otherwise lawful,[[159]](#footnote-159) and that deploying the munition in a certain way is identified as acceptable practice.[[160]](#footnote-160) In fact, this is true of nearly all existing munitions.[[161]](#footnote-161) While an anti-personnel landmine could be considered a “basic” EAI, there is no method for the robot to “refuse” to detonate.

C. *Scenario 2*

While the previous scenario identified that “basic” autonomy has no real choice in terms of refusing an order, the second scenario considers more advanced technology. Here, emphasis is placed upon weapons platforms, rather than munitions. Indeed, most militaries currently deploy rudimentary EAI platforms - many of them with any number of minor AI systems on board. Radar, and other such detection systems, enable the modern day fighter jet to independently identify potential threats, lock on to them, and provide a pilot with the option of authorizing launch. At that point, a human decides upon the most suitable means of attack (though in reality his or her “choices” is likely to be fairly restricted) . One force such option is very likely to be an air-to-air munition, which is arguably a basic EAI if it were self-guiding. Once a human authorizes the application of force, there is usually no further “interference” from the platforms individual systems. It simply deploys the “means” upon receiving the command by to do so.

However, consider the example where a pilot is involved in a dog-fight, and the platform they inhabited had an internal EAI support system such as that proposed by the authors. In such a situation, the radar might identify an enemy aircraft, and alert the pilot accordingly. The pilot may then choose to fire (or not), based upon their assessment of the prevailing circumstances.[[162]](#footnote-162) If the pilot, or in the case of an Unmanned Aerial Combat Vehicle (UAV), or future combat drone,[[163]](#footnote-163) the remote operator, did decide to engage, stage 1 would commence. Here, authentication would be straightforward, and would likely be satisfied when the pilot “logged in” at the start of their shift. System A would have already positively verified the target. And, system B would then engage, and carry out additional proportionality assessment to that of the pilots. Arguably, and in most instances, this would align with the pilots choice to deploy the munition. However, in another particular situation, the EAI might be aware of one or a number of extraneous circumstances (such as ground troops that at that moment were at risk of being hit by debris, or perhaps a high risk to the civilian population) that the pilot had not factored in due to inclement weather or to a navigational error. Of course, each individual situation would be highly contextual, and there would not necessarily be a strict obligation to not carry out the attack. Nevertheless, if in the circumstances refusal did engage, passive refusal (4Ai), or active refusal (4Aii) would very likely be the most appropriate, as there does not appear to be a need for that preventative refusal at this stage (4Aiii).

If the system calculated that distinction and proportionality were initially satisfied, system C would then engage and reconsider, whether the means and method were appropriate, in that their selection would, as much as is practicably possible, minimize civilian harms. Once again, if the system C assessment led to refusal, the three options in stage 4A would be available. In this instance, 4Ai may be appropriate, giving the pilot the option, for example, of deploying an alternative munition. However, engaging 4Aii might also be considered the “best” option under the circumstances providing the pilot an additional opportunity to give a similar order in the near future. Nevertheless, if stages A, B, and C, were satisfied, the system would not intervene, and the munition would be deployed. Due to the systems speed of operation, and speed of the tactical pursuit, in addition to the battlefield conditions, a stage 3 assessment might be unnecessary. However, if stage 3 was engaged, and the EAI did consult the relevant ROE that were applicable at the time, it could potentially identify whether the platform was operating in in a pre-determined “no-fly zone” or other such restriction, and refuse to follow the order as a result.

In the circumstances considered, it is important to note that the platform is an EAI as opposed to an AWS because it only has the power to refuse to follow an order, and not to authorize a use of force. Nevertheless, while there may be a slight increase in the risk faced by the pilot, the introduction of the multi-stage test, must be seen as a positive method reducing friction, in much the same way as an anti-stall mechanism.[[164]](#footnote-164) In fact, if a test such as this had been available to the Aegis platform which was installed upon the USS Vincennes in 1988, it may have refused to follow the order to launch a sea-air missile. The system would have detected the friend or foe signal that was being sent out by the civilian aircraft Iran Air Flight 655,[[165]](#footnote-165) and as a result the aircraft may not have been destroyed, and civilian lives may have been spared.

Returning to scenario 2 however, if track 1 was followed (as opposed to the test considered above), the results may have differed for a number of reasons. Not least because the decision-maker did not think it was “reasonable” to deploy an aircraft fitted with an EAI system in the first place. In addition, even where a decision monitoring EAI was utilized, track 1 would only require for an order to be refused – in this case to fire a munition – where it was a manifestly unlawful order. As a result, with distinction and proportionality satisfied, a system that was programmed with track 1 would not refuse to follow the order (regardless of whether there was any political, or wider tactical or operational reason not to). Similarly, for the reasons previously discussed, while a track 3 systems would require for the EAI to be deployed in all circumstances in which it was available, the EAI itself would have to consider many wider considerations, the matter of whether the dignity of the target, or whether any civilian that was included in the proportionality assessment had been offended.

D. *Scenario 3*

The primary purpose of the discussions in the previous two scenarios was to distinguish between existing non-autonomous tech, and near future EAI tech. As the analysis has, thus far revealed, where EAI is utilized as a method of monitoring orders, it must also be capable of refusing to follow an instruction that is incompatible with IHL. The following scenario, however, is used to extend the analysis beyond the human decision to use a particular type of weapon (means), to a “robot” decision to use a particular tactic (method) - here, for obtaining information from an “adversary”.

There are many ways in this scenario could be presented. Nevertheless, the circumstances that are proposed, are: an AWS in the form of a humanoid combatant, receives an order from an EAI that is located in a command and control center to “secure a block” in which a large number of enemy combatants are located. The order also contained an instruction to retrieve, “by any means necessary” information by regarding a suspected future attack. In this scenario, authentication is likely to be fairly straightforward with any number of sensors, and additional programs (including destination verification) being capable of authenticating the status of the order-giver. However, at this point the analysis needs to be divided in two given that there are clearly two separate orders. The first order, therefore, is to secure a location. Here, arguably, the AWS analysis would be similar to that considered in scenario 2. The stage 1 system would therefore need to be satisfied, before stage 2 was engaged and used to determine whether that particular element of the order is consistent with IHL and the more focused ROE. The same arguments would be applied in respect of track 1 and track 3 application as considered above, and to avoid repetition, there is not a pressing need to consider the entire RROE again at this point.

However, in this scenario the stage 2 assessment is key with regard to the second order to obtain information. This is because there is some ambiguity here - how does the order-giver define “any means necessary”? There are a number of options, the most obvious being that clarification needs to be sought. However, where this was not possible, the two most prominent interpretations would be (i) by any means necessary - providing they are consistent with international legal obligations, (in which a similar assessment to that already considered would be carried out), or (ii) simply - by any means necessary. With regard to the latter, the leading argument must be that because there is customary rule for a soldier to disobey a manifestly unlawful order, an AWS, like its human equivalent, must refuse to carry out any order to torture an enemy combatant.[[166]](#footnote-166)

A question that arises, however, is what if there is military doctrine to the contrary? For example, the U.S., under the Presidency of George W. Bush, attempted to greatly reduce the definition of torture. In doing so, it sought to justify the use of certain cruel, inhuman, or degrading acts that were carried out because they did not violate the U.S.’s international obligations - that they did not produce the requisite intensity of pain and suffering.[[167]](#footnote-167) The result, as argued by at least one commentator, was that despite the fact that Articles 1 and 2 of the Convention Against Torture[[168]](#footnote-168) placed certain restrictions upon military interrogators, the CIA were authorized “to confine and interrogate detainees with a harshness that markedly violated human rights”.[[169]](#footnote-169)

In the present case, if torture was authorized, and indeed ordered, arguably it would be refused by any AWS/ EAI, whether it applied the track 1, track 2, or track 3 approach. Of course, the decision-maker employing the former may not consider it reasonable to utilize an EAI for this type of assessment in the first instance. Nevertheless, given that military manuals typically provide a great deal more detail regarding a soldier’s obligations than that which is provided by IHL alone, in this scenario, it might also be argued that the AWS would have an obligation to follow the order – if, the ROE appeared to reflect recent changes in national policy/ best practice. In such a situation however, the current authors propose that because EAI/ AWS are a “method” of ensuring compliance with international obligations, they must be programmed to reflect a hierarchy of sources which places international treaties at the top of the order. In other words, while military doctrine should be used to support and interpret relevant treaty provisions, they must not be used as a method to negate, or lessen the effect of existing international obligations. And, when such a principle is applied to the present set of particulars, the stage 2 system would either prevent an AWS from carrying out a tortuous act, or, where it was an EAI operating in a supervisory mode, prevent a human subordinate from doing the same.[[170]](#footnote-170)

E. *Scenario 4*

Each of the previous two scenarios have considered a situation where an EAI might be utilized to assess an order, and where necessary, to refuse to follow it. In either case the analysis concluded that, by applying the RROE that are proposed by the authors, refusal can take place due to an order failing to adhere to the fundamental principles of IHL (or the wider body of applicable international law), or, where it is otherwise inconsistent with matters of policy. Best practice in other words, is where the order is not of use for service. In this final and undoubtedly most extreme “application” of “robot refusal”, the authors seek to consider the practicalities and logistics of refusal in the context of nuclear launch.[[171]](#footnote-171) At this point, several opening reflections can be posited. First, is the idea that the “system” making the ultimate “doomsday decision” may neither be AWS nor EAI, but rather, an AI in the intangible sense. Due to the obvious physical differences between an AI and EAI, any EAI involved in the decision is likely to be much closer to a target, and indeed could even be the Strategic ICBM itself. The AI, on the other hand could be positioned more remotely from the target, for example, at the launch site.

Ostensibly, an autonomous “operating system” could ultimately override a launch instruction, whether that originated from a human or an EAI. As a result, while the scenario that follows considers a set of prescribed “facts”, the relationships are interchangeable, and not strictly limited to those considered. One EAI could, for example, be interchanged with an AWS, or a human could be swapped out for an AWS. However, clearly, at some given point, the order must be run through an EAI. In direct contrast to scenario 2 whereby the acquiring of missile lock is more of a strategic assist (rather than a legal assist), for the EAI in this fourth scenario (preventing unlawful nuclear launch as a result of system of systems) is a “legal” tech assist.

The scenario considered in the following section has the following characteristics, but with the caveat that the act in question is scrutinized under the *jus in bello*, as opposed to the *jus ad bellum*:[[172]](#footnote-172) State (A) is at war with State (B), both of whom are nuclear powers. In the midst of this “peer-on-peer” conflict, the President of State A (X) authorizes a nuclear attack. Following current strategic command principles, this order goes to strategic command, and to a central war office. The order to initiate the attack is then forwarded on to submarine commander (Y), and to General (Z), who oversees a number of further launch sites. Shortly after the instruction to initiate the attack is received by Y and Z, *all* communications are lost.[[173]](#footnote-173) The reader may at this stage recall such a scenario being captured in the 1995 motion picture, “Crimson Tide”. Here, Denzel Washington’s character wishes to re-establish radio communications to determine whether a further but unreadable message from strategic command has overridden a previous nuclear launch order. In contrast Gene Hackman’s character portrays the somewhat more belligerent XO who believes that the initial order for nuclear launch should be followed without question.

In terms of the present scenario (and not an appraisal of the motion picture), the authors’ system of systems requires stage 1 authentication. This is clearly a vital stage, but, is also one at which the potential for EAI monitoring could be introduced. Currently, if looking at U.S. nuclear doctrine, the President carries a personal identification tool, which contains a code that is unique to him or her –

known colloquially as the “biscuit”. When necessary, this code can be entered into an authorization system that is carried by a constantly rotating military presence, and which accompanies the President at *all* times. This system is known as the nuclear, or atomic, “football”.

In the first instance, these nuclear “biscuits” and “footballs” ensure that the order can be authenticated. Potentially, for the sub commander, this might implicitly engage stage 2. However, the central question here is, could or indeed should the “football” be an EAI? If it were, and, for example, the order was erroneous, it could be prevented from even making it through to strategic command and / or Central war office, and eventually to the launch site. Of more general application, the authors wish to underline that EAI analysis is at its most beneficial, the higher up the command chain it takes place (at least in the first instance).[[174]](#footnote-174) The football might also be pre-programmed with the correct strategic response – given the circumstances – leading to faster decision making. However, if the decision-monitoring EAI were not the “football”, the order could still be passed to a Strategic command / Central War Office EAI, where the order could face the same examination.[[175]](#footnote-175)

Once the order has passed to both Y and Z however, the task becomes more difficult. And, if it were an erroneous order, it could potentially result in the slightly curious case of one part of the launch team disobeying orders, but because that original message goes to both, recall is tricky, and all too late.[[176]](#footnote-176) This could be prevented with the use an EAI. At this stage EAI assessment provides an additional safeguard at whichever stage it is considered, but also, perhaps the more often it is utilized.[[177]](#footnote-177) Therefore, although the EAI analysis which follows is largely considered at the launch level, the stages of assessment could be considered at any point during the lifetime of the order.[[178]](#footnote-178)

With stage 1 seemingly satisfied, the stage 2 assessment needs to take place. Clearly, all stage 2 assessments must be made according to huge potential for devastation that nuclear missiles carry. Presumably, however, given a particular set of circumstances, a nuclear weapon could “target” a military installation in adherence with the principle of distinction. With regards to proportionality however, system B might have to be satisfied on the condition that the launch, and all of the anticipated collateral damage has been authorized by a strategic level decision.

For example, existing safeguards are no doubt in place for instances where the nuclear order-giver suffers from temporary insanity, or where they choose to dispense an order due to ill health, blackmail / inducement / treachery etc. However, at the sub level, this would be difficult to determine. Nevertheless, where there was no evidence of a legitimate military target, or, there was evidence to suggest that the target was the “civilian population”, the EAI could, without fear for its job, its societal standing, its life, and the lives of its “nearest and dearest” etc. – refuse to follow the order.[[179]](#footnote-179) One would imagine such a refusal would be 4aii at the very least, but 4aiii would perhaps be most likely, dependent, of course, upon the actual context. Given that an order to breach the IHL principles of distinction and proportionality are manifestly unlawful, should an EAI system was utilized here, all three tracks would prevent the missiles from being launched.

Where system A and B were satisfied however, system C would engage. Given the gravity attached to a nuclear launch, one might posit that the standard for target verification should reflect something greater than mere “reasonable” steps, and perhaps, in reality, this is perhaps likely to be the case. Nonetheless, for an EAI system programmed according to track 1, reasonableness would remain the minimum requirement. The matter of whether the EAI calculated feasible precautions had been taken (or not), would be highly context dependent. If, for example, an EAI was stationed with Commander Y, and prior to launch and loss of comms it had evaluated that state A had attempted all other appropriate means methods of defeating the enemy, the order to launch would not be intercepted. However, if the EAI applying the RROE calculates that an alternative means or method could yield a similar result, it could equally intercept the order and prevent Y from initiating the launch until such time as more information regarding the status of the order could be gained. The point to remember here is not that the EAI could refuse to follow an order to initiate a nuclear launch - period, but, that it could help to *stop* a malicious, or ill-informed order to initiate a nuclear launch.

In contrast to the system applying the RROE, an EAI operating to Track 1, would not, when provided with the same set of circumstance, calculate that the order should be intercepted. This would be due to the fact that so long as distinction and proportionality was satisfied, and the order was reasonable in the circumstances (with a specific regard to the wider mission objectives – in this case winning), there is no evidence that it is a manifestly unlawful order. This standpoint is in itself, of course contrary, to the track 3 requirement to always use the means and method that minimizes civilian harms, and to carry out any analysis subject to factors such as whether there is a reasonable chance of mission success. Due to the circumstances, and the chances of initiating a nuclear apocalypse, this must at least be questionable.

Returning to the system of system analysis, however, should the stage 2 examination be satisfied, the stage 3 system would engage. It is at this point where the RROE offer the most focused assessment, when compared to an EAI that were programmed to follow either of the two most extreme tracks. For example,, an EAI applying track 1 would only deny orders which were manifestly unlawful. Nuclear weapons, for example, cannot distinguish between combatant and civilian. Moreover, any legitimate military target that was positioned close to a densely populated urban area would mean that there would be a considerable levels of collateral damage (not to mention the expansive costs of all out nuclear war). Consequently, many might consider that such an attack would always be disproportionate. However, in providing their *Nuclear Weapons Advisory Opinion*, the ICJ acknowledged that there may be certain circumstances when the use of nuclear weapons might not breach international law.[[180]](#footnote-180) One would imagine that any “legitimate” order to initiate a nuclear launch “could” be such a circumstance. And, where this was the case, the track 1 analysis would not intercept - and it is perhaps fair to suggest, nor should it.

As has been discussed throughout however, it is difficult to imagine the circumstances where an EAI applying the track 3 approach, would not decide to intercept the order to use deploy a nuclear weapon. Instead, the only instance where a system that operated according, inter alia, to an objective ethical code (including that of all third parties), while taking in the potential wider costs of full scale nuclear war into account, might be where deploying a nuclear warhead was considered the absolute last resort – in other words, when all other means and methods had been exhausted. Indeed, this may go some way to explaining why Germany is not in procession of nuclear weapons.

Stage 3 of the RROE, however, having considered the same set of relevant legal provisions as the track 1 system would move on to consider the wider body of legal obligations, and importantly all relevant further ROE. Insofar as nuclear weapons are concerned, these ROE should be extremely concise, and perhaps more so than the vast majority of alternative systems. Moreover, in the knowledge that a post-human analysis of an order to initiate a nuclear launch was a possibility, future ROE would include specific reference to the role in which an EAI should play. Vitally, as identified in scenario 3, these ROE cannot be used to side-step existing international legal obligations, but only to provide a greater detail of the strategic, operational, and tactical methods, of enforcing them.

Once again, if the order to launch a nuclear attack was a legitimate order, there would be no reason for the EAI intercept it (though given the circumstances, 4Bii might be the more preferable outcome). And, as previously discussed, the case for non-interception may be stronger the higher up the command chain the EAI analysis is completed. However, where the EAI examination took place at the level of the submarine commander, and where communication was lost in the process of receiving an “update” form strategic command, these ROE may prove vital. No doubt the commander would also have some access to these ROE, and would of course be a highly skilled, well informed, individual.

Nevertheless, one would still question, given the circumstances, whether he or she should realistically be expected to apply the same legal acumen to a potentially voluminous set of instructions, as would a New York Bar attorney, and at speed anywhere near the data-processing capabilities of a supercomputer. Especially not when the same individuals entire training and preparation for events such as this would have been succinctly grounded in the obligation to strictly follow orders. If an EAI were instead tasked with analysing all relevant data, and specifically all national doctrine, there would be a much greater opportunity to identify whether the order was consistent with matters of law, policy and best practice, and the commander could be more confident that he or she was making the right decision.

Once again, if the EAI did not detect any abnormality here, or indeed any other permitted reason to intercept the order, the order must be followed. However, where there was a reason to “suspect” the order was overwritten, or that it was given with malicious intent, or any one of a number of alternatives, it could be refused by the EAI – which as a reminder, would not be influenced by human emotions. Where this happens, there are three possibilities for systematic refusal. *Prima facie*, preventative refusal may seem the most appropriate course of action. However, it is arguable whether states would employ such a system in a nuclear setting. Instead, and perhaps in somewhat of a complete reversal, it is proposed that passive refusal is the most appropriate “type” of refusal for the following reasons. First, strategic decision making will ultimately remain firmly in the hands of humans, but secondly, by operating in such a way, the EAI might demonstrate something akin to a “gut feeling” – a word of caution, as opposed to an absolute rebuttal. Such a system would therefore assist the commander (along with his or her own wide body of experiences) , as opposed to discounting it.

Turning briefly to the AWS realm, the task is much simpler. It is to run a proportionality algorithm on a continuous feedback loop so from the moment the instruction is passed right up until launch (supposing that the order is given the night before to attack at dawn the next day). AWS limitations, or indeed advantages, quickly become apparent because such a system cannot replicate the idea of instinct (presumably derived from combat experience). In terms of instinct, the soldier who is hesitant about an attack is quite possibly making a subconscious proportionality assessment in the crude sense, but ultimately, that hesitancy is also in part governed by self-preservation. So, the disobeying of orders from an AWS is never really likely to be encountered on “those” grounds (self-preservation) but by continuously running a proportionality feedback loop.

However, an EAI might well start to incorporate self-preservation which means that reject or disobeying becomes more “human” rather than less “human” - hence the additional requirements proposed by the authors. The non-compliance of an order in AWS terms is fairly rudimentary, the system simply doesn’t fire. That raises a similar but not identical branch of thought as to whether the EAI receiving the order from a Human would treat that human order differently. Might it look down on the human as “inferior” and that as mere mortals, such complex calculations cannot be computed. Fundamentally, this issue ties back in to whether AWS or EAI should be afforded the same combat privileges status as humans.

V. Wider Application of Disobeying Orders

This penultimate section of the Article extends this controversial discussion into three further areas of analysis: the application of “Robot Refusal” in the context of robot PMCs, Spies, and more provocatively, whether RROE should include insubordination for EAIs. By way of brief caveat, the authors do not intend to “close” the discussion in each of these areas but rather, wish to open these areas in the first instance for future discussion, and much-needed debate.

*A. PMCs*

The corpus of existing discussion surrounding the use and legality of PMCs is not in short supply.[[181]](#footnote-181) Where there is perhaps greater paucity is the application of EAIs to the PMC sector although this is an area the present authors have made gentle inroads to in a discussion elsewhere.[[182]](#footnote-182) Consequently, the sole remit of this present discussion is to highlight a natural extension of the previous discussion surrounding the disobeying of orders; considerations concerning the future use of EAI PMCs are both timely and necessary. The specific PMC “angle” in this section is whether, and to what extent, an EAI PMC is perhaps more or less likely to disobey orders compared with their human counterparts. It is relatively uncontroversial to suggest that one of the primary motivating factors for a human PMC partaking in a theatre of conflict will be one of financial reward.[[183]](#footnote-183) Extending this argument a little further may prompt the not unreasonable conclusion that PMCs (in light of this “added value” of financial incentive” are less likely to refuse an order.[[184]](#footnote-184)

Indeed, should the PMC refuse / disobey an order or at the very least not successfully complete the mission extra bonus enumerations may be affected. An EAI PMC meanwhile would presumably not have that extra variable “present” (unless the EAI is remunerated / motivated by Bitcoin or other crypto currencies). The authors therefore conclude that the addition to the “system” of systems presented above, combined with the removal of the “mercenary” factor in the traditional sense, would mean that there would be no reason as to why an EAI PMC would feel less inclined to fail to disobey / refuse an illegal order.

A final “footnote” to this discussion more generally, surrounding future use of PMCs is that their role may decrease rather than increase. While it is undeniable that the immediate future will see an ever-increasing presence of PMCs both in combat (and even more so in no combat roles), the authors contend that a natural re-shaping of the sector will occur. The very raison d’être of a PMC is financial reward as opposed to fighting for the flag — when such reward is no longer conceivable (in the future envisaged by the authors) the existence of PMCs may continue, but they would no longer be mercenaries.

*B. Spies*

Invariably, and as commentators continuously document, the law applicable to the realm of espionage, and those functioning within it remains relatively sparse.[[185]](#footnote-185) In essence, it is broadly limited to the somewhat dated provisions contained in Articles 30 and 31 of the 1899 and 1907 Hague Regulations.[[186]](#footnote-186) The purpose of including spies within the overarching thesis of this discussion is not to re-open or indeed revisit the paucity of the law in this error. Rather, it is to extend the previous discussion into the natural endpoint of the discussion’s trajectory and very briefly open a suitable niche point of discussion. Quite simply, the authors, not *unreasonably* consider an instance whereby a “retired” combat EAI transitions from SEAL Team 6 and begins employment at the “Agency” – a not altogether uncommon career path.

What the authors wish to highlight is if indeed the EAI spook mirrored that particular human career choice, would the EAI still retain (in the event it hadn’t undertaken a separate training at the “Farm” in Langley, Virginia) it’s existing system of systems programming? If this were to be the case, and the “hard drive” was not “re-formatted” an order to target a dissident on London Bridge, with a poisoned tipped umbrella, would ultimately be rejected – assuming of course that all EAIs were equally equipped with the same system of systems set out in Part IV. As with the previous inclusion of RROE in the PMC world, one could have a similar instance whereby the variation in programming could result in intelligence agents acting differently depending where they had “undertaken” their initial training / programming.

*C. Insubordination*

This next section on EAI insubordination addresses what is undoubtedly a very niche area, and one which, could be too readily dismissed if approached through the restrictive lens of current technology. Nevertheless, the authors do concede that such an approach does undoubtedly require some leap of faith in terms of this forward-facing discussion. Taken to its conclusion, were the EAI to wrongfully disobey an order, should it face consequences for insubordination? Clearly, and without being overly flippant (were one to apply such a discussion to an AWS) the very simple answer would be to re-program the system.[[187]](#footnote-187) A less draconian approach may be some form of disciplinary action but what that may look like in practice the author’s fully concede is less tangible – loss of leave or pay is unlikely to perhaps have a motivating factor. Clearly, there would need to be a sliding scale, reflective of the increasing severity from mere insubordination to full on mutiny.

VI. Conclusion

Ultimately, this Article has sought to bring to the forefront a timely discussion regarding an EAI’s ability to refuse erroneous orders particularly in light of the fact that there is no agreed position in terms of state practice / refusal of orders. The authors have proposed the novel inclusion of a test in order to determine the precise limitations as to when and how this should happen. Therefore, the test set out and proposed in Part IV not only offers greater clarity for now, but something distinctly concrete for the future. Whereas more traditional lines of thinking have rightly cautioned against an AWS or EAI’s ability to compute human emotions when making proportionality assessments, the authors’ novel approach has been to reverse the thinking and suggest that the EAI is not only placed to understand human traits, but can also override human error.

In instances where a human has complied with the IHL legal requirements but where factors such as self-preservation or instinct are present and may have clouded a human’s ability to objectively make the proportionality assessment to launch the author’s recommend aborting launch. In making a decision as to whether to obey or disobey human error, the EAI should not only calculate the IHL “requirements” but also actively look to discount any “human factors” which may have influenced the decision. When it comes to an EAI responding to an EAI error, the authors envisaged EAI/ AI “checks”, carried out by a number of different systems in order to add an additional level of protection in order to identify and prevent rogue orders. In the very purest strategic sense this would indeed be the ultimate “system of systems”.

Part II of the Article was the natural starting point for the discussion, and revisited the key tenets of IHL in order to provide a logical interface between a combatant’s obligations under DTP and the right of refusal considered in Part III. As noted in Part II and one which, forms part of the key concluding message once states are equipped with EAI, IHL may well compel their usage. Part III of the Article extended the discussion into differing thresholds for the refusal of orders. Here, the authors summarized that the minimum standard as portrayed in the narrow obligation requires a member of the armed forces to strictly adhere to an order unless there is a customary duty to refuse on the grounds that it is manifestly unlawful. Perhaps somewhat unfortunately, the analysis identified that difficulties may arise when programming EAI to refuse, due to the fact the obligation operates more broadly. For some states, it is not a straightforward consideration as to the lawfulness of not of the order. Instead, a combatant adhering to the wider obligation must reconcile any order according to relatively abstract concepts such as human dignity, while concurrently scrutinizing whether or not there is a reasonable chance of mission success, and/ or whether the order is consistent with national policy. This leads to the eventuality, however, that existing practice is not only incongruous, but can also place too great an emphasis upon either military necessity or humanitarian considerations – both of which are elements that at its heart, IHL strives to balance without bias.

Part IV of the Article provided what the authors believe is a robust legal “system of systems” not just with the intended effect of ensuring appropriate compliance or non-compliance from a legal perspective, but one that one would have strategic benefit too. It is perhaps overly tempting to suggest that Clausewitz’s idea of “total war” (where the “gloves come off”) is necessarily the appropriate strategic approach – the system envisaged by the authors not only limits unlawful actions but ones which would also minimize and reduce strategic error as well.

In Part V the authors broadened the remit of the discussion and extended the discussion into a natural EAI career trajectory PMC and espionage. Here the authors note that the existence of EAI may well cause the gradual extinction of PMCs or at least in the sense of them operating as mercenaries. More problematic perhaps in both the realms of PMC and espionage is the “legacy” programming – the EAI’s original training and embedding of system of systems may prove ideal for IHL compliance, but prevent ineffectual operational qualities. Quite simply, the system of systems programming may be too effective, and an EAI would refuse any of the “Black Ops” missions those realms frequent.

Too simplistic and broad a rebuttal to any such considerations undertaken in this work would be to simply dismiss such a discussion into the realms of “well…it depends on the algorithm”. What this article has pinpointed and strived to comprehend, is precisely what such calculations would look like. In terms of an AWS system in might be a continuous proportionality feedback assessment. In terms of EAI, in addition to the on-going proportionality assessment, it is a series of additional checks and balances both to discount human error and emotion, and indeed to ensure an even greater level of compliance with IHL obligations. Clearly, this raised broader issues in terms of chain of command where the authors questioned whether an EAI should be able to override a chain of command, and concluded in the affirmative. To return to but one example from the civilian “field” highlighted in the abstract – an aircraft’s anti stall mechanism has the potential to override human error – such overriding is seen as nothing but a positive. It is also perhaps important to distinguish that the system of systems proposed here is much stricter *par excellence* than say an automotive vehicle’s system of systems – the car is unlikely to shut down and switch off (thus overring it’s human controller) purely because they have ignored the wrench key designating the vehicle is in need of service or the coffee cup icon denoting the need for a break. Whereas, and in complete contrast an EAI could, would and should abort nuclear launch if there are any “red flags” raised by the system of systems.

The authors believe that by starting and opening this much-needed dialogue for further debate, will assist in the growth of scholarship in this area. It is the firm assertion of the authors, that under certain circumstances, “robot refusal” is preferable to unquestioning acceptance of human error.

1. \* Francis Grimal is a Reader in Public International Law, University of Buckingham, UK, and Michael J. Pollard is a PhD Candidate in Public International Law, University of Buckingham, UK. The authors would like to extend their sincerest thanks to Professor Christopher Waters, Dean of Law, University of Windsor, Ontario for all his considerable advice and invaluable feedback throughout the preparation of this Article. The authors would also like to acknowledge the substantial work undertaken by the ICRC with regards to customary international law in general. © 2020, Francis Grimal & Michael J. Pollard. [↑](#footnote-ref-1)
2. Michael Symanski, *Any Fool Can Obey an Order*, Modern War Institute (March 17, 2017), <https://mwi.usma.edu/fool-can-obey-order/> where the author notes “so wrote First Sea Lord Jacky Fisher in angry Critique of Captain H.M. Pelly, a cruiser Captain under Admiral Beatty at the battle of Dogger Bank in 1915”. Also *see* Ryan Scott *Willful Disobedience: Character traits of Independent Thinkers in the Military,* Modern War Institute(February 23, 2017), <https://mwi.usma.edu/willful-disobedience-character-traits-independent-thinkers-military/>. [↑](#footnote-ref-2)
3. Indeed, to date, the primary focus in this area has been upon the acceptability (or not) of the “Nuremberg defense”, in respect of the commission of war crimes. *See* James B. Insco, *Defense of Superior Orders before Military Commissions*, 13 Duke J. Comp. & Int'l L. 389 (2003); Mark J. Osiel, *Obeying Orders: Atrocity, Military Discipline, and the Law of War*, 86 Calif. L. Rev. 939 (1998), and; Robert E. Murdough, *I Won't Participate in an Illegal War: Military Objectors, the Nuremberg Defense, and the Obligation to Refuse Illegal Orders*, 2010 Army Law. 4 (2010). [↑](#footnote-ref-3)
4. *See* Clausewitz see Carl Von Clausewitz, On War 119 (Anatol Rapoport ed., Penguin. 1982). N.B. the vast majority of commentators regard the Michael Howard and Peter Paret (Princeton: Princeton University Press, 1976, revised 1984) edition as the authoritative version/ translation. Also *see* Colin S. Gray, Modern Strategy, 75-112, (Oxford University Press, 2005). For general works on Clausewitz *see* *e.g.,* John E. Tashjean, *Pious Arms: Clausewitz and the Right of War*, 44 Military Affairs 82(1980); Carl von Clausewitz, Clausewitz on Strategy : Inspiration and Insight From a Master Strategist (Christopher Bassford, et al. eds., New York: Wiley, 2001). [↑](#footnote-ref-4)
5. *Id.* [↑](#footnote-ref-5)
6. Strategic literature is replete with examples, *see e.g.,* Colin S. Gray, Airpower for Strategic Effect (2012). Here, for example, the author discusses sorties being abandoned during the Kosovo conflict due to foggy conditions near the bases in Terano, Italy; the negative impact of bad weather (in the form of rain) in Vietnam; and even how relatively modern operations in Afghanistan and Iraq were negatively affected. [↑](#footnote-ref-6)
7. *Id.* [↑](#footnote-ref-7)
8. *See e.g.,* Barry D. Watts , The Foundations of U.S. Air Doctrine: The Problem of Friction in War, (2012). Watts discusses the doctrine of Strategic Air Power (SAP), a relatively simple notion that airpower alone is sufficient (in the military sense) to bring about victory. However, if we accept that SAP has failed to live up to the promise of its theory, then one possible explanation is to do with failure to account for “friction”. Despite technological advances, rain, fog and the climate as a whole still arguably have prevented SAP from operating in certain instances and have therefore meant that as a general proposition it is impossible for SAP to live up to its promise. [↑](#footnote-ref-8)
9. According to Watts, *Id.,* one of the reasons as to why Airpower failed to live up to its promise was the notion that superior weapons, which planes arguably are in comparison to ground troops, does not necessarily guarantee victory. In other words, failure to account for friction prevents *any* strategic doctrine from fulfilling potential has been the influence of external factors. However, as with all doctrines, there is not just one overriding factor (although the omission of “frictional considerations” is probably the most significant). The manifestation of bad weather is only one factor.. [↑](#footnote-ref-9)
10. *Id.* [↑](#footnote-ref-10)
11. *See* Gray, *supra* note 5. [↑](#footnote-ref-11)
12. In short, a system-of-systems is a collection of independent systems/ elements that are each allotted a specific task. They each complete this task without necessarily having to consult other systems/ elements – in a vehicle, this could, for example, be power steering. However, sometimes a group of such systems can be considered holistically, and with the result that the overarching system is much more advanced, and capable, than any of the individual elements could be alone – a self-driving car, for example, would need many individual components, perhaps including, GPS, radar, lidar, and other sensors, automatic braking, lane recognition, self-parking, cameras, and, not least, a central processor. None of these systems alone, however, could ever ensure a vehicle got safely from A to B. *See e.g.* Automated Vehicles for Safety, (National Highway Traffic Safety Administration) <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety> [↑](#footnote-ref-12)
13. As the analysis that follows demonstrates, there is presently some deviation in state practice in respect of exactly which orders a soldier should disobey. [↑](#footnote-ref-13)
14. Note robot refusal in this realm is a controversial discussion, not least because some commentators hold that in order for a soldier to disobey an unlawful order (and therefore to any machine equivalent) “they must exercise 'agency' and engage in nuanced reasoning”. *See* Tetyana Krupiy, *Unravelling Power Dynamics in Organizations: An Accountability Framework for Crimes Triggered by Lethal Autonomous Weapons Systems*, 15 Loy. U. Chi. Int'l L. Rev. 1 (2017) at 15. [↑](#footnote-ref-14)
15. For a useful, and relaxed discussion about recent breakthroughs in machine learning and possible future developmental trajectories, *see* George Anadiotis, *The state of AI in 2019: Breakthroughs in machine learning, natural language processing, games, and knowledge graphs,* ZDNet (July 8, 2019), <https://www.zdnet.com/article/the-state-of-ai-in-2019-breakthroughs-in-machine-learning-natural-language-processing-games-and-knowledge-graphs/>. [↑](#footnote-ref-15)
16. A stall is defined as “a sudden reduction in the lift generated by an aerofoil when the critical angle of attack is reached or exceeded.” Skybrary <https://www.skybrary.aero/index.php/Stall>.Anti-stall systems, automatically reduce the pitch of the aircrafts nose where, for example, the angle of attack is exceeded due to pilot error. They are typically considered so effective that they are a requirement on all transport aircraft. *See* Skybrary <https://www.skybrary.aero/index.php/Stall>. However, the system is not a complete failsafe, and following a spate of incidents in 2019, a malfunctioning anti-stall was responsible for grounding Boeing’s entire fleet of 737 Max aircraft. See Gwyn Topham, *Anti-stall system was 'in play' on Ethiopian's Boeing 737 Max*, The Guardian (March 25, 2019), <https://www.theguardian.com/world/2019/mar/25/anti-stall-system-was-in-play-on-ethiopians-boeing-737-max>; David Shepardson and Jamie Freed, *FAA failed to properly review 737 MAX jet's anti-stall system - JATR findings*, Reuters (October 11, 2019), <https://uk.reuters.com/article/uk-usa-boeing-airplane-faa/faa-failed-to-properly-review-737-max-jets-anti-stall-system-jatr-findings-idUKKBN1WQ0PS> [↑](#footnote-ref-16)
17. Such systems might include satellite imaging, thermal imaging, 3D mapping, real-time online updates from other intelligence gathering robots located in and around the battlespace – and, such technologies *will* inevitably be surpassed, and overtaken in the not too distant future, perhaps on a continuous cyclical basis. [↑](#footnote-ref-17)
18. Indeed, the authors acknowledge that this examination of robot refusal sits alongside a much wider philosophical, ethical and legal debate regarding robot sentience and the concept of robot refusal. Should a “sex”-robot, for example, be entitled to say no, and, should there be consequences for it, and/ or for the “client” who chooses to ignore such an instruction? *See* Robert Sparrow, *Robots, Rape, and Representation*, Int J of Soc Robotics 9, 465–477 (2017); Neda Atanasoski and Kalindi Vora, *Why the Sex Robot Becomes the Killer Robot – Reproduction, Care, and the Limits of Refusal*, Spheres-Journal 6 (2020); Lily Frank and Sven Nyholm, *Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable?,* Artif. Intell. Law 25, 305–323 (2017). [↑](#footnote-ref-18)
19. *See e.g.* Davison, N. *A legal perspective: Autonomous weapon systems under international humanitarian law,* in UNODA Occasional Papers No. 30: Perspectives on Lethal Autonomous Weapon Systems, (UN, New York 2017); Marco Sassoli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, 90 INT'LL. STUD. 308 (2014); Chantal Grut, *The Challenge of Autonomous Lethal Robotics to International Humanitarian Law*, Journal of Conflict and Security Law, 18.1, 5 (2013); Michael N. Schmitt, *Autonomous weapon systems and international humanitarian law: a reply to the critics*. Harv. Nat. Sec. J., 4, 1 (2013). [↑](#footnote-ref-19)
20. *See* *e.g.* Thurnher J.S., *Feasible Precautions in Attack and Autonomous Weapons*, *in* Dehumanization of Warfare. (Heintschel von Heinegg W., Frau R., Singer T. eds., 2018); Michael N. Schmitt, *Precision attack and international humanitarian law,* 87 Int'l Rev. Red Cross 445 (2005); Geoffrey Corn and James A. Schoettler, *Targeting and Civilian Risk Mitigation: The Essential Role of Precautionary Measures*, 223 Mil. L. Rev. 785 (2015). [↑](#footnote-ref-20)
21. Protocol (I) Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter API]. [↑](#footnote-ref-21)
22. According to the International Committee of the Red Cross [hereinafter ICRC]: “The “principle of military necessity” permits measures which are actually necessary to accomplish a legitimate military purpose and are not otherwise prohibited by international humanitarian law. In the case of an armed conflict the only legitimate military purpose is to weaken the military capacity of the other parties to the conflict.” <https://casebook.icrc.org/glossary/military-necessity> [↑](#footnote-ref-22)
23. See Yoram Dinstein, *Military Necessity*, Encyclopedia of public international law 3 (1997); Nobuo Hayashi, *Requirements of Military Necessity in International Humanitarian Law and International Criminal Law*, 28 B.U. Int'l L.J. 39 (2010); Yishai Beer, *Humanity Considerations Cannot Reduce War’s Hazards Alone: Revitalizing the Concept of Military Necessity*, EJIL, 26 4, 801 (2015). [↑](#footnote-ref-23)
24. Article 48 API, *supra*, note 20, states: In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives. [↑](#footnote-ref-24)
25. Article 49 API, *supra*, note 20, states; “Attacks” means acts of violence against an adversary, whether in offence or defence. [↑](#footnote-ref-25)
26. Note also that Article 51(3) API *supra*, note 20, states; Civilians shall enjoy the protection offered by this section, unless and for such time as they take a direct part in hostilities [hereinafter DPH]. Quite which actions should qualify as playing a DPH is the subject of some conjecture. For an in-depth analysis of the two leading schools of thought see the authors’ previous piece discussing EAIs and DPH; Francis Grimal and Michael J. Pollard, *“Embodied AI” and The Direct Participation in Hostilities* 51 3 Geo. J Int’l L. 513 (2020). [↑](#footnote-ref-26)
27. This is known as the ‘constant care’ obligation, to which more attention is given in the discussion which follows. [↑](#footnote-ref-27)
28. Though note this provision does not refer to proportionality directly. Nevertheless, as the ICJ identified in the *Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion*, I.C.J. Reports 1996, p. 226, International Court of Justice (ICJ), 8 July 1996, [hereinafter *Nuclear Weapons, Advisory Opinion*], API ensures that, “a legitimate target may not be attacked if the collateral civilian casualties would be disproportionate to the specific military gain…” *See also;* Yoram Dinstein, The Conduct of Hostilities Under the Law of International Armed Conflict, (3rd ed. 2016) ¶ 406-410, and at ¶ 425 where the author notes for that military advantage and civilian casualties/damage are incomparable in a quantifiable manner…[and that]…[t]here is little prospect of agreement between opposing Belligerent Parties as to the rival values of…[these concepts]. [↑](#footnote-ref-28)
29. Dinstein, *Id.* ¶ 440, citing; F Kalshoven, *Reflections on the Law of War: Collected Essays* 346 (2007), at 546. [↑](#footnote-ref-29)
30. Article 8(2)(b)(iv) Rome Statute of the International Criminal Court (1998) states, [For the purpose of this statute, ‘war crimes’ means:] Intentionally launching an attack in the knowledge that such attack will cause incidental loss of life or injury to civilians or damage to civilian objects or widespread, long-term and severe damage to the natural environment which would be clearly excessive in relation to the concrete and direct overall military advantage anticipated. [↑](#footnote-ref-30)
31. *See* Jean-Marie Henckaerts and Louise Doswald-Beck, International Committee of the Red Cross: Customary International Humanitarian Law, Volume I: Rules (2005), (hereinafter referred to as ICRC customary Rules); ICRC Customary Rule 14, states, [l]aunching an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated, is prohibited… State practice establishes this rule as a norm of customary international law applicable in both international and non-international armed conflicts. [↑](#footnote-ref-31)
32. *See* ICRC Customary Rule 15, *Id.,* stating, [i]n the conduct of military operations, constant care must be taken to spare the civilian population, civilians and civilian objects. All feasible precautions must be taken to avoid, and in any event to minimize, incidental loss of civilian life, injury to civilians and damage to civilian objects… State practice establishes this rule as a norm of customary international law applicable in both international and non-international armed conflicts. [↑](#footnote-ref-32)
33. Seeing as the U.S. are not currently a party to API. [↑](#footnote-ref-33)
34. See also; Regulation 22 annexed to the Hague Convention (IV) (1907): The right of belligerents to adopt means of injuring the enemy is not unlimited”; Nuclear Weapons Advisory Opinion, *supra* note 27, identifying that nations: “do not have unlimited freedom of choice of means of weapons…” For a discussion see; Dinstein, *supra*, note 27, ¶ 10-11, and; Stuart Casey-Maslen, Steven Haines, Hague Law Interpreted: The Conduct of Hostilities under the Law of Armed Conflict, (2018) where the authors for example argue that the ICJ “strangely”, and incorrectly, limited the principle to weapons that caused unnecessary suffering to combatants. [↑](#footnote-ref-34)
35. Dinstein, *supra*, note 27, ¶ 441. [↑](#footnote-ref-35)
36. *Id*. [↑](#footnote-ref-36)
37. Michael N. Schmitt, *The Principle of Discrimination in 21 Century Warfare*, 2 Yale Hum. Rts. & Dev. L.J. 143 (1999), at 152, and; ICRC Customary Rule 17, *supra* note 30, (restating art. 57 (2)(a)(ii): Each party to the conflict must take all feasible precautions in the choice of means and methods of warfare with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects. [↑](#footnote-ref-37)
38. Grimal/ Pollard, *supra*, note 25. [↑](#footnote-ref-38)
39. Sassoli, *supra* note 18, at 339. [↑](#footnote-ref-39)
40. It is important to note that an attacker does not have to use a particular weapon, or tactic, just because it will lessen the amount of collateral damage. A commander who is in possession of a limited number of platforms, may for example, decide that there would be more efficiently deployed elsewhere, or, with only a limited number of munitions, decide that it would be more effective not to use one in the present circumstances. See; Michael N. Schmitt & Eric W. Widmar, “On Target”: Precision and Balance in the Contemporary Law of Targeting 7 Nat’l Sec. L. & Pol’y 379, 400-404, (2014). [↑](#footnote-ref-40)
41. Schmitt & widmar, *Id.*, at 401. [↑](#footnote-ref-41)
42. However*,* also *see* Schmitt & widmar, *supra* note 39, at 401, who note that sometimes a tactical situation can change after an targeting decision has been made, but beyond the point at which an attack can be called off. [↑](#footnote-ref-42)
43. Corn and Schoettler, *supra*, note 19, at 802, and Schmitt & widmar, *supra* note 39, generally. [↑](#footnote-ref-43)
44. See the four Geneva Conventions and the three additional protocols; Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, Aug. 12, 1949, 75 U.N.T.S. 31 [hereinafter Geneva Convention I]; Geneva Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, Aug. 12, 1949, 75 U.N.T.S. 85 [hereinafter Geneva Convention II]; Geneva Convention (III) Relative to the Treatment of Prisoners of War, Aug. 12, 1949, 75 U.N.T.S. 135 [hereinafter Geneva Convention III]; Geneva Convention (IV) Relative to the Protection of Civilian Persons in Times of War, Aug. 12, 1949, 75 U.N.T.S. 287 [hereinafter Geneva Convention IV]; Protocol (I) Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter Additional Protocol I]; Protocol (II) Additional to the Geneva Conventions of 12 August 1949 and Relating to the Protection of Victims of Non-International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 609 [hereinafter Additional Protocol II]. Note also, Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Adoption of an Additional Distinctive Emblem (Protocol III), Dec. 8, 2005, T.I.A.S. No. 07–908 [hereinafter Additional Protocol III], though it should be noted this latter provision has had only a relatively minor effect upon existing IHL. [↑](#footnote-ref-44)
45. Dinstein, *supra*, note 27, ¶ 442. Schmitt & widmar, *supra* note 39, at 400, who also direct the reader’s attention to; art 3(4)CCCW, but in additionally; Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125, U.N.T.S. 3, UK Reservation; ICRC Customary Rule 15, *supra* note 30; ICRC Customary Rule 17, *supra* note 30, and, for example; Article 3(10) of Amended Protocol on Prohibitions or Restrictions on the use of Mines, Booby Traps, and other Devices, May 3, 1996, 2048 U.N.T.S. 93 [Amended CCW Protocol II]. [↑](#footnote-ref-45)
46. Protocol II on Prohibitions or Restrictions on the use of Mines, Booby Traps, and other Devices, Annexed to; the Geneva Convention on Prohibitions or restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (hereinafter CCCW), (1980). [↑](#footnote-ref-46)
47. Schmitt & widmar, *supra* note 39, at 400 [↑](#footnote-ref-47)
48. Schmitt & widmar, *supra* note 39, at 402 provide: attackers need only select a less harmful means or methods that do not involve sacrificing military advantage and that are feasible. [↑](#footnote-ref-48)
49. ICRC, Commentary on the Additional Protocols to the Geneva Conventions of 12 August 1949, 2198; Schmitt & widmar, *supra* note 39, at 400. [↑](#footnote-ref-49)
50. *Id*. 400-401. [↑](#footnote-ref-50)
51. See the Statement of the U.S. at the CDDH, Official Records, Vol. VI, CDDH/SR.42, 27 May 1977, p. 241, and; Jean-Marie Henckaerts and Louise Doswald-Beck, International Committee of the Red Cross: Customary International Humanitarian Law, Volume II: Practice Part I, Volume II: Practice Part II (2005) (hereinafter referred to as; ICRC Practice Relating to Customary Rules); Practice relating to ICRC Customary Rule 15, ¶175. The U.S. refer to: “that which is practicable or practically possible, taking into account all circumstances at the time, including those relevant to the success of military operations”, rather than mirroring to art 3(4) where the final sentence reads: “including humanitarian and military considerations”. This was very similar to the German stance pre-ratification of API. *See* CDDH, Official Records, Vol. VI, CDDH/SR.42, 27 May 1977, p226, which states that for the sake of Art. 57 API feasible should mean: what is practicable or practically possible, taking into account all circumstances at the time, including those relevant to the success of military operations. [↑](#footnote-ref-51)
52. *Id.* [↑](#footnote-ref-52)
53. [U.K.] § b, Reservations and declarations made upon ratification of the of the 1977 Additional Protocol (1998) and § 532, at note 191; [Germany] § 2, Declarations made upon ratification of the of the 1977 Additional Protocol (1991); [Australia] The Manual of the Law on Armed Conflict’s glossary states that feasible precautions are those which: are practicable or practically possible taking into account all circumstances ruling at the time including humanitarian and military considerations; [Canada]§ 5 Reservations and declarations made upon ratification of the of the 1977 Additional Protocol (1990) and, [Canada] §112.6 of the Use of Force Manual which states: “Feasible” is understood as that which is practicable or practicably possible, taking into account all circumstances ruling at the time, including humanitarian and military considerations. Also *see*  ICRC Customary IHL Database, *practice relating to Rule15*, <https://ihl-databases.icrc.org/customary-ihl/eng/docs/home> (hereinafter ICRC database Rules). [↑](#footnote-ref-53)
54. § 3, Reservations and declarations made upon ratification of the of the 1977 Additional Protocol (2003), also see; ICRC Database Rule 15, *Id.* [↑](#footnote-ref-54)
55. A question that may be raised here is as to whether a commander may discharge his/ her duty by deploying a weapon which is independently capable of doing everything ‘feasible’ in terms of making real-time CDEM assessments. In other words, could deploying an EAI lessen the burden upon the human commander? A full investigation is beyond the scope of the current discussion, however, due to the fluidity of armed conflict, the “fog” and/ or unpredictability of war, this is perhaps unlikely. [↑](#footnote-ref-55)
56. For the sake of this article, leading nations are considered to be those that have a (global firepower) power index that ranks them no. 25 or above. See; *2020 Military Strength Ranking, Global Fire Power*, <https://www.globalfirepower.com/countries-listing.asp>. See also; Ellen Ioanes, *This is how the US and Iran rank among the world's 25 most powerful militaries*, Business Insider, (Jan. 7, 2020), <https://www.businessinsider.com/most-powerful-militaries-in-the-world-ranked-2019-9?r=US&IR=T>. [↑](#footnote-ref-56)
57. For a useful discussion regarding the state practice and opinion juris, the elements that need to be satisfied in this sense, *see* Dinstein, *supra*, ¶ 43-45. [↑](#footnote-ref-57)
58. *Id.* [↑](#footnote-ref-58)
59. For example, § 556 of Australia’s Defence Force Manual (1994) states: In the conduct of military operations, constant care must be taken to spare the civilian population and civilian objects to the maximum extent possible, also *see*: ICRC Practice relating to Customary Rule 15, *supra* note 50, at 337; § 15 Canada’s LOAC Manual (2001), states civilians are entitled to protection from the dangers arising from military operations. In conducting operations care should always be taken to spare civilians and civilian objects, also *see;* p. 4-2, ICRC Practice relating to Customary Rule 15, *supra* note 50, at 337; France’s LOAC Manual (2000), p. 98 states: In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects, also *see*; ICRC Practice relating to Customary Rule 15, *supra* note 50, at 338; § 404 Germany’s Military Manual (1992) states: The civilian population as such as well as individual civilians…shall be spared as far as possible. also *see;* p. 4-2, ICRC Practice relating to Customary Rule 15, *supra* note 50, and the German Soldiers’ Manual (2006): When attacking a military objective, all necessary precautions shall be taken to spare as far as possible the civilian population located in the area or in the immediate vicinity of the object, also *see* ICRC Database Rule 15, supra note 52; § 5.32 of the UK LOAC Manual (2004) provides that: In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects, also *see* ICRC database Rule 15, *supra* note 52. [↑](#footnote-ref-59)
60. § 8.1 of the US Naval Handbook (2007) states: The law of targeting, therefore, requires that all reasonable precautions must be taken to ensure that only military objectives are targeted so that noncombatants, civilians, and civilian objects are spared as much as possible from the ravages of war. See also; ICRC Database Rule15, *supra* note 52. [↑](#footnote-ref-60)
61. ICRC Practice relating to ICRC Customary Rule 16, *supra* note 50, at 55, states, [e]ach party to the conflict must do everything feasible to verify that targets are military objectives…[and that]… state practice establishes this rule as a norm of customary international law applicable in both international and non-international armed conflicts [↑](#footnote-ref-61)
62. § 457 of Germany’s Military Manual (1992) states: “Before engaging an objective, every responsible military leader shall verify the military nature of the objective to be attacked. [↑](#footnote-ref-62)
63. 8.1 of the US Naval Handbook (2007), also see ICRC database Rule 16, *supra* note 52. [↑](#footnote-ref-63)
64. §13.32 of the UK LOAC Manual (2004) states that: “(a) those who plan, decide upon or execute an attack must take all feasible measures to gather information which will assist in determining whether or not objects which are not military objectives are present in an area of attack, (b) in the light of the information available to them, those who plan, decide upon or execute an attack shall do everything feasible to ensure that attacks are limited to military objectives. Also see; ICRC Database Rule 16, *supra* note 52. [↑](#footnote-ref-64)
65. § 5.53 of Australia’s LOAC Manual (2006). Also see ICRC Database Rule 16, *supra* note 52. [↑](#footnote-ref-65)
66. § 112.2 of Canada’s Use of Force Manual (2008), for example, stating that: All feasible precautions must be taken to verify that the target is a military objective, and not a civilian or a civilian object, and that it is not subject to any of the specialized regimes of protection which prohibit, or severely restrict, attacks on certain persons and objects. Also see; ICRC Database, Rule 16, *supra* note 52. [↑](#footnote-ref-66)
67. French LOAC Manual (2001), p.98. Also see; ICRC Practice relating to Customary Rule 16, *supra* note 50, at ¶ 222. [↑](#footnote-ref-67)
68. *Id*., § 8.3.1 states: must take all reasonable precautions, taking into account military and humanitarian considerations, to keep civilian casualties and damage to the minimum consistent with mission accomplishment… [↑](#footnote-ref-68)
69. § 5.53 of Australia’s LOAC Manual (2006) states that: all reasonable precautions must be taken to avoid injury, loss or damage to civilians and civilian objects and locations. It is therefore important to obtain accurate intelligence before mounting an attack. [↑](#footnote-ref-69)
70. § 711.3 of Canada’s LOAC Manual (2001), and; ICRC Practice relating to Customary Rule 15, *supra* note 50. [↑](#footnote-ref-70)
71. § 4, Rule 2, of the Code of Conduct for CF Personnel, Office of the Judge Advocate General, 2001; ICRC database Rule 15, *supra* note 52. [↑](#footnote-ref-71)
72. France, LOAC Teaching Note (2000), p. 2, and; ICRC Practice relating to Customary Rule 15, *supra* note 50. [↑](#footnote-ref-72)
73. § 510 Germany’s Military Manual (1992), and; ICRC Practice relating to Customary Rule 15. [↑](#footnote-ref-73)
74. § 12.26(i) of the UK LOAC Manual (2004), and; ICRC Database Rule 15, *supra* note 52. [↑](#footnote-ref-74)
75. § 4(b) The Law of Armed Conflict (Pamphlet) D/DAT/13/35/66, Army Code 71130 (revised 1981), and ICRC Database Rule 15, *supra* note 52. [↑](#footnote-ref-75)
76. § 556(e) of the Australia, Defence Force Manual (1994), and, ICRC Database Rule 17, *supra* note 52. [↑](#footnote-ref-76)
77. § 834 of the Australia, Defence Force Manual (1994), and ICRC Practice relating to Customary Rule 17, *supra* note 50, at ¶ 272. [↑](#footnote-ref-77)
78. § 8.3.1 of the US Naval Handbook (2007), also see; ICRC database Rule 17, *supra* note 52. [↑](#footnote-ref-78)
79. § 112.4 of the Use of Force for CF Operations, Canadian Forces Joint Publication, B-GJ—005-501/FP-001 (2008), for example states: [a]ll feasible precautions must be taken in the choice of means and methods of attack to avoid, and in any event to minimize, incidental civilian loss and damage (i.e., collateral damage). Also see; ICRC Database Rule 17, *supra* note 52. [↑](#footnote-ref-79)
80. § 5.2 of the French LOAC Manual (2001), also see; ICRC Practice relating to Customary Rule 17, *supra* note 50, at ¶ 279. [↑](#footnote-ref-80)
81. § 457 of the German Military Manual (1992), also see; ICRC Practice relating to Customary Rule 17, *supra* note 50, at ¶ 280. [↑](#footnote-ref-81)
82. § 5.32.4 of the UK LOAC Manual (2004) states that there is an: obligation to select the means (that is, weapons) or methods of attack (that is, tactics) which will cause the least incidental damage commensurate with military success. [↑](#footnote-ref-82)
83. *Id.* 5.32.4. [↑](#footnote-ref-83)
84. *See* Schmitt & widmar, *supra* note 39. [↑](#footnote-ref-84)
85. Indeed, the art. 57 obligations extend further than is considered for the current analysis. This includes, for example, an duty to provide an advance warning of an attack, where possible [i.e. where feasible] (art. 57 (2)(a)(iii)(c)), and; to choose the target that leads to lesser civilian harms when presented with two or more targets that offer a similar military advantage art. 57(3). [↑](#footnote-ref-85)
86. For the sake of the present article, the terms “ROE” and “military doctrine” are used interchangeably. The DoD Dictionary of military and associated terms, for example states: ROE [are] directives issued by competent military authority that delineate the circumstances and limitation under which US forces will initiate and/or continue combat engagement with other forces encountered. (Doctrine is undefined). [↑](#footnote-ref-86)
87. It having already been examined how DTP fits within the wider body of IHL. [↑](#footnote-ref-87)
88. *See* Paul Scharre, Army of None, 31-32 (2018), where, identifying the automatic/ automated/ autonomous spectrum, the author notes, “[a]s machines become more sophisticated, they become more capable and able to accomplish more complex tasks in more open-ended environments”. [↑](#footnote-ref-88)
89. Gurgen Petrosian, *Elements of Superior Responsibility for sexual violence by subordinates,* 42 Man L.J.123 127 (2019). [↑](#footnote-ref-89)
90. Article 44 Lieber code states: [a] soldier, officer or private, in the act of committing such violence, and disobeying a superior ordering him to abstain from it, may be lawfully killed on the spot by such superior. [↑](#footnote-ref-90)
91. *See e.g.,* Article 12 U.K. Armed Forces Act (2006), which states: (1) A person subject to service law commits an offence if— (a)he disobeys a lawful command; and (b) he intends to disobey, or is reckless as to whether he disobeys, the command. (2)A person guilty of an offence under this section is liable to any punishment mentioned in the Table in section 164, but any sentence of imprisonment imposed in respect of the offence must not exceed ten years. [↑](#footnote-ref-91)
92. Article 8 of the Nuremberg Charter, otherwise known as Charter of the International Military Tribunal, annexed to Agreement for the Prosecution and Punishment of Major War Criminals of the European Axis (London Agreement) August 8, 1945, 59. Stat. 1544, 82 U.N.T.S. 279, states: “The fact that the defendant acted pursuant to order of his Government or of a superior shall not free him from responsibility, but may be considered in mitigation of punishment if the Tribunal determines that justice so requires”. [↑](#footnote-ref-92)
93. *See e.g.,* The Manual for Courts-Martial, United States (2016 Edition), (hereinafter US Courts-Martial Manual), Rule 916(d) which states: It is a defense to any offense that the accused was acting pursuant to orders’, also see; Keith A. Petty, *Duty and Disobedience: The Conflict of Conscience and Compliance in the Trump Era*, 45 Pepp. L. Rev. 55 (2018), 103, in turn citing; Osiel, *supra* note 2, at note 21, stating: “Virtually all military codes include some provision on due obedience…” [↑](#footnote-ref-93)
94. Osiel, *supra* note 2, at 950, and Petty, *Id*., at 103-104. [↑](#footnote-ref-94)
95. US Courts-Martial Manual, *supra* note 91*,* at Rule 916(d). [↑](#footnote-ref-95)
96. *See* ICRC Customary Rule 154, *supra* note 30. [↑](#footnote-ref-96)
97. *Id.* [↑](#footnote-ref-97)
98. *See* ICRC Customary Rule 155, *supra* note 30, which states: “Obeying a superior order does not relieve a subordinate of criminal responsibility if the subordinate knew that the act ordered was unlawful or should have known because of the manifestly unlawful nature of the act ordered.” [↑](#footnote-ref-98)
99. *See* Petty, *supra* note 92, at 97, who in turn notes that the U.S. Army Field Manual states: [a]lthough this manual recognizes the criminal responsibility of individuals for [crimes against peace, crimes against humanity, and war crimes], members of the armed forces will normally be concerned only with those offenses constituting 'war crimes.’ [↑](#footnote-ref-99)
100. Osiel, *supra* note 2, at 961. Note also that the author supports the notion that the duty upon a U.S. soldier is only to refuse to follow a manifestly unlawful order. Thus the U.S. soldier does not need to exercise “situational judgement”. Osiel, *supra* note 2, at 971. [↑](#footnote-ref-100)
101. Note for example that military doctrine accepts that the ‘obligation’ to follow a superiors orders is what separates members of the armed forces from their civilian counterparts. *See e.g.,* New Zealand, Interim Law of Armed Conflict Manual, DM112, New Zealand Defence Force, Headquarters, Directorate of Legal Services, Wellington, November 1992, Annex C, § C12(1), which states: an “obligation, and the one which clearly sets a member of a military force apart from his civilian counterparts, is the obligation to obey lawful commands of a superior officer.” [↑](#footnote-ref-101)
102. See generally. ISAAC ASIMOV, I, ROBOT 40 (1950). [↑](#footnote-ref-102)
103. Grimal/ Pollard, *supra* note 25, at note 35. [↑](#footnote-ref-103)
104. Corn and Schoettler, *supra*, note 19, at 795. Indeed, as noted in a previous piece, it is not unusual for Asimov’s rules to appear in the debate surrounding AWS. *See* Grimal/ Pollard, *supra* note 25, at note 35 which, for example, identifies; Rebecca Crootof, War Torts: Accountability for Autonomous Weapons, 164 U. PA. L. REV. 1347, 1372 n.135 (2016); Andrew Figueroa, License to Kill: An Analysis of the Legality of Fully Autonomous Drones in the Context of International Use of Force Law, 31 PACE INT’L L. REV. 145, 156 n.71 (2018). [↑](#footnote-ref-104)
105. Article 77 of a Draft version of API did state that “no person shall be punished for refusing to obey an order of his government or of a superior which, if carried out, would constitute a grave breach of the provisions of the Conventions or of the present Protocol”, see; CDDH, Official Records, Vol. I, Part Three, Draft Additional Protocols, June 1973, p. 25. However, as noted by the ICRC, the provision failed to gain the required two thirds majority that it need to pass as law. *See* ICRC Practice relating to Customary Rules, *supra* note 50, at 3799; Also *see* Article 33(1) of the Rome Statute of the International Criminal Court (1945) (hereinafter Rome Statute) relating to Superior Orders and Prescription of Law, which states: The fact that a crime within the jurisdiction of the court has been committed by a person pursuant to an order of a Government or of a superior, whether military or civilian, shall not relieve that person of a criminal responsibility unless: (a) The person was under a legal obligation to obey orders of the Government or the superior in question; (b) The person did not know that the order was unlawful; and (c) The order was not manifestly unlawful. [↑](#footnote-ref-105)
106. Having identified that following an instruction would lead to the commission of a “manifestly” criminal act. [↑](#footnote-ref-106)
107. See e.g.; ICRC Practice relating to Customary Rule 154, *supra* note 50, at 3804, citing §15-4(d) United States, Air Force Pamphlet, 110-31, International Law – The Conduct of Armed Conflict and Air Operations, U.S. Department of the Airforce, (1976), which states: [m]embers of the armed forces are bound to obey only lawful orders, and; United States, Your Conduct in Combat Under the Law of War, Publication No. FM 27-2, Headquarters Department of the Army, Washington, November (1984), 26, stating: Although you are responsible for promptly obeying all legal orders issued by your leader, you are obligated to disobey an order to commit a crime. [↑](#footnote-ref-107)
108. As is discussed further below, and, as can also be seen in art. 33(1) Rome Statute, *supra* note 104, the existing standard is one of “manifest” unlawfulness. [↑](#footnote-ref-108)
109. See; Section 10, § 1, United Kingdom, The Law of Armed Conflict, D/DAT/ 13/35/66, Army Code 71130 (Revised 1981), also; ICRC Practice relating to Customary Rule 154 *supra* note 50. [↑](#footnote-ref-109)
110. Section 2.3, Armed Forces of the Philippines code of Ethics, (1981), and; ICRC Practice relating to Customary Rule 154, *supra* note 50. [↑](#footnote-ref-110)
111. France, Disciplinary Regulations as amended (1975), Article 8(3), and ICRC Practice relating to Customary Rule 154, *supra* note 50, at 3802. [↑](#footnote-ref-111)
112. § 1306, Australia, Defence Force Manual (1994), also see; § 13.7, The Manual of the Law of Armed Conflict, Australian Defence Doctrine, Publication 06.4, (2006), and; ICRC Practice relating to Customary Rule 154, *supra* note 50, at 3800. [↑](#footnote-ref-112)
113. Article 21, Congo, Disciplinary Regulations (1986), and; ICRC Practice relating to Customary Rule 154, *supra* note 50, at 3802. [↑](#footnote-ref-113)
114. Canadian Code of Conduct (2001), Rule 11, § 4, states: …in accordance with military custom you should…obey and execute the order – unless – the order is manifestly unlawful; DRC constitutional law that: No one is required to execute a manifestly illegal order, see; Article 25, [DRC] Constitution of the Transition (2003), and; ICRC practice relating to customary Rule 154, *supra* note 50. And, §§ 143 and 145 of the German Military Manual (1992) that: it is expressly prohibited to obey orders whose execution would be a crime. See also, ICRC Practice Relating to Customary Rule 154, *supra* note 50, at 3802. [↑](#footnote-ref-114)
115. As confirmed by the ICJ in the Nuclear Weapons Advisory Opinion, *supra* note 27, at ¶ 25, while International Human Rights Law is applicable in armed conflict, IHL must take precedence where there is a conflict due to the concept of *lex specialis.* For further discussion *see* Dinstein, *supra* note 27, at ¶ 89-92, and; Marko Milanovic, *The Lost Origins of Lex Specialis: Rethinking the Relationship Between Human Rights and International Humanitarian Law, in* Theoretical Boundaries of Armed Conflict and Human Rights, (Jens. D. Ohlin ed., 2014). [↑](#footnote-ref-115)
116. Canadian Code of Conduct (2001), Rule 11, § 4; ICRC Practice relating to Customary Rule 154, *supra* note 50, at 3800. [↑](#footnote-ref-116)
117. Canadian Code of Conduct (2001), Rule 11, § 5, which continues with the examples that: mistreating someone who has surrendered or beating a detainee is manifestly unlawful, though this is not, of course, intended to be exhaustive; ICRC Practice relating to Customary Rule 154, *supra* note 50, at 3800-3801. [↑](#footnote-ref-117)
118. For a useful discussion, see; Erin Blakemore, ‘*Why German Soldiers Don’t Have to Obey orders’*, History, (17 November, 2017), <https://www.history.com/news/why-german-soldiers-dont-have-to-obey-orders>. Blakemore notes, for example, that from August 1934, German soldiers swore an oath not to the state, but the Fuehrer himself, *see* Jewish Virtual Library, *Adolf Hitler: The Fuehrer Oath*, <https://www.jewishvirtuallibrary.org/the-fuehrer-oath>, which provides the translation: I swear by almighty God this sacred oath: I will render unconditional obedience to the Fuehrer of the German Reich and people, Adolf Hitler, Supreme Commander of the Wehrmacht, and, as a brave soldier, I will be ready at any time to stake my life for this oath. [↑](#footnote-ref-118)
119. § 142, Germany, Military Manual (1992); ICRC Practice relating to Customary Rule 154, *supra* note 50, at 3802. [↑](#footnote-ref-119)
120. Blakemore, *supra* note 117. [↑](#footnote-ref-120)
121. Ilja Baudisch, *Freedom of Conscience and Right to Conscientious Objection - Refusal to Obey to Military Orders - Legal Ban on the Use of Force (Article 2(4) UN Charter) - Neutrality of States in Armed Conflicts*, 100 Am. J. Int'l L. 911 (2006), 912. For a similar, but perhaps a more convivial discussion, also see; John Ford, ‘When Can a Soldier Disobey an Order?’ (War on the Rocks, 24 July 2017), <https://warontherocks.com/2017/07/when-can-a-soldier-disobey-an-order/>. [↑](#footnote-ref-121)
122. See; United States v Huet-Vaughn, 43 M.J. 105 (C.A.A.F.1995) at 114; Watada v. Head, No. C07-5549BHS, 2008 U.S. Dist. Lexis 88489, at 8-10 (W.D. Wash. Oct. 21 2008) (Hereinafter referred to as the Watada Case); ICRC Practice relating to Customary Rule 154 *supra* note 50. Petty, *supra* note 92, at 100-101. [↑](#footnote-ref-122)
123. In fact, Watada was not permitted to enter his preferred defense, because the decision in Huet-Vaughn meant that: [t]he order to deploy soldiers is a non-justiciable political question … an accused may not excuse his disobedience of an order to proceed to foreign duty on the ground that our presence there does not conform to his notions of legality *See*, Murdoch, *supra* note 2, Ford, *supra* note 120, and; Petty, *supra* note 92, at 100-101. [↑](#footnote-ref-123)
124. Petty, *supra* note 92, at 101, citing; U.S. Dep’t of the Army, Regulation 600-8-24, Officer Transfers and Discharges at 6, (13 September 2011). [↑](#footnote-ref-124)
125. *See* Petty, *supra* note 92, at 129-130, citing; Army Doctrine Reference Publication (ADRP) 6-22. [↑](#footnote-ref-125)
126. *See* Petty *supra* note 92*,* and; Baudisch, *supra* note 121. [↑](#footnote-ref-126)
127. Baudisch, *supra* note 121, at 911. [↑](#footnote-ref-127)
128. Note, at ¶ 4.1.2.1 the German court states: (“human dignity”) also includes the protection of the freedom of conscience according to Article 4, ¶ 1 of the Basic Law. In any case, it does not reduce that protection. Translation from; ICRC Practice relating to Customary Rule 154, *supra* note 50. Note also; Baudisch, *supra* note 121 at 911-912, at notes 2 and 5 citing; Article 4 of The German Basic Law (*Grundgesetz*) provides: “freedom of faith and conscience, and freedom to profess a religious or philosophical creed, shall be inviolable ....(3) No person shall be compelled against his conscience to render military service involving the use of arms. Details shall be regulated by a federal law." [↑](#footnote-ref-128)
129. Baudisch, *supra* note 121 914, and; ICRC Practice relating to Customary Rule 154 *supra* note 50. [↑](#footnote-ref-129)
130. ¶ 4.1.2 The Limits of Obedience to Superior Orders Case; Baudisch, *supra* note 121, at 911; ICRC Practice relating to Customary Rule 154 *supra* note 50. [↑](#footnote-ref-130)
131. ¶ 4.1.3.1.1 The Limits of Obedience to Superior Orders Case states: that a soldier…has to act with all the diligence and responsibility possible to him and has to act accordingly. An “unconditioned” or “unconditional” obedience is not compatible with this normative imperative. Requested is rather a “thinking” obedience, an obedience “reflecting” the consequences of carrying out the order – especially also with regard to the limits of the applicable law and the ethical “yardsticks” of the personal conscience; ICRC Practice relating to Customary Rule 154, *supra* note 50. [↑](#footnote-ref-131)
132. *See e.g.,* Paola Gaeta, *The Defence of Superior Orders: The Statute of the International Criminal Court versus Customary International Law,* EJIL 10 (1999), 172–191, and; Yoram Dinstein, The Defence of 'Obedience to Superior Orders' in International Law(OUP, 2012). The current authors also note that these two cases are focused more upon the *jus ad bellum*, than they are in the *jus in bello*, which is the intended subject matter of the present article. Nevertheless, they do provide an obvious and clear example of the narrow and wide obligation. [↑](#footnote-ref-132)
133. For example, an order to commit genocide, or torture. [↑](#footnote-ref-133)
134. *E.g.,* torture, sexual violence and mistreatment of POWs. [↑](#footnote-ref-134)
135. Higher ranking members of the armed forces may be particularly aware of a need to consider the current political environment. Consider, for example, the Captain of the USS Theodore Roosevelt, Brett Crozier, who was recently removed from his position, and looks set to face charges for failing to respect the chain of command. See; BRADLEY PENISTON, ‘*Navy to Punish Fired Captain of the USS Theodore Roosevelt*’ Defense One, (19 June 2020), <https://www.defenseone.com/threats/2020/06/navy-punish-fired-uss-roosevelt-captain/166300/>. [↑](#footnote-ref-135)
136. *Id*., there may also be financial implications if a combatant was to have a criminal conviction imposed upon them, as this may prevent them from gaining further employment. See; [↑](#footnote-ref-136)
137. Petty, *supra* note 92, at 128, citing; J. Carl Ficarrotta, *Selective Conscientious Objection: Some Guidelines for implementation,* *in* When Soldiers say no: Selective Conscientious objection in the Modern Military (Andrea Ellner et al. eds. 2016), 199, and; H.L.A. Hart, The Concept of Law, (Peter Crane et al. eds., 2nd ed. 1994), 55-57. [↑](#footnote-ref-137)
138. Petty, *supra* note 92, at 109. [↑](#footnote-ref-138)
139. Whether the order is consistent with the concept of human dignity, human rights, public liberties, morals, the conscience of reasonable right think people, and even whether the order is of service to the state (which might be referred to as being consistent with policy), and finally, whether the soldier has a reasonable chance of executing the order successfully (reflected by the wide obligation). [↑](#footnote-ref-139)
140. At least strategically speaking, a nuclear strike with a single warhead would in many cases be unwise, *see e.g.,* Gray, supra note 3. The point is, that if an ambiguous, or erroneous order was sent to a large number of launch sites, it would clearly be more difficult to prevent launch at several sites (if that was the desired course of action), than as opposed to an order that was sent to a single location. [↑](#footnote-ref-140)
141. Dinstein, *supra*, note 27, ¶ 25 [↑](#footnote-ref-141)
142. *Id.* ¶ 26. [↑](#footnote-ref-142)
143. *See* Dinstein, *supra*, note 27, ¶ 22 where the author identifies “[if]f military necessity were the sole beacon to guide the path of armed forces in wartime, no meaningful constraints would have been imposed on the freedom of action of Belligerent Parties…[but]…[i]f benevolent humanitarianism were the only factor to be weighed in hostilities, war would have entailed no bloodshed, no human suffering and no destruction of property; in short, war would not be war. [↑](#footnote-ref-143)
144. In addition, due to the fact that when programmed to do carry out a specific task, EAIs will not ‘forget’. However, this is not always the case when an EAI or AI is “repurposed”. As is discussed in the following section, for example, one of the difficulties that AI experts do currently face, is that when a self-learning system is given a new task, i.e.; learn to play the ancient Chinese strategy game “Go”, instead of the one you have been playing i.e.; Chess, the system, placing all of its ‘attention’ upon the former, does forget to how play the latter. The concept is referred to as catastrophic forgetting, see e.g., Anthony Robins, *Catastrophic Forgetting, Rehearsal and Pseudorehearsal,* Connection Science, 7:2, 123-146, (1995) [↑](#footnote-ref-144)
145. Corn and Schoettler, *supra*, note 19, at 821-822. Indeed, ROE may contain vital information on joint-force tactics and operation procedures, which could vitally important in future conflicts in which armed forces will be required to consider how ant will affect all services, across all domains. See e.g. John Amble (in conversation with Dave Stephenson, the director of the Joint Staff’s Office of Irregular Warfare and Competition), Modern War Institute PODCAST: COMPETITION, CONFLICT, AND THE FUTURE OF IRREGULAR WARFARE, (22 July 2020) <https://mwi.usma.edu/mwi-podcast-competition-conflict-and-the-future-of-irregular-warfare/>. [↑](#footnote-ref-145)
146. This might be, for example, where an EAI identified that an order was legal, but was otherwise inconsistent with ROE regarding certain targets. *See* Corn and Schoettler, *supra*, note 19, at \*\*\*who note, for example, that ROEs commonly reflect policy that imposes: “restrictions on combat power above and beyond LOAC…[and that] …are often adopted in hopes of avoiding alienation of the civilian population. [↑](#footnote-ref-146)
147. Though one might argue that if a machine was able to display the “nuanced reasoning” that is discussed, for example by Krupiy, *supra* note 13, it could potentially apply track 3. Also see note regarding the possibility of introducing an ethical governor, *infra* note 151. [↑](#footnote-ref-147)
148. ICTY Prosecutor v. Furundzija, ¶ 183. [↑](#footnote-ref-148)
149. Jeremy Rabkin, *What We Can Learn about Human Dignity from International Law*, 27 Harv. J.L. & Pub. Pol'y 145, 146 (2003). [↑](#footnote-ref-149)
150. *See e.g.,* Sassoli, *supra* note 18. [↑](#footnote-ref-150)
151. Adam Saxton, *(Un)Dignified Killer Robots?: The Problem with the Human Dignity Argument,* Lawfare (March 20, 2016) <https://www.lawfareblog.com/undignified-killer-robots-problem-human-dignity-argument>. [↑](#footnote-ref-151)
152. Although the current authors do not intend to support the current discussion on such a believe, one author suggests it will one day be possible to programme future EAIs with an “ethical governor”, which will allow for them to display moral judgement when making decisions. *See* Ronal Arkin, Governing Lethal Behaviour in Autonomous Robots (2009). [↑](#footnote-ref-152)
153. The point here is that the greater protection that is offered by ROE is typically subject to policy. *See e.g.,* Corn and Schoettler, *supra*, note 19, at 822, who argue “a ROE-based decision to forgo an attack, even if motivated by an effort to mitigate civilian risk, is not legally mandated. This fact is an important aspect of contemporary military operations, and the difference between legal and policy-based courses of action should be constantly emphasized.” [↑](#footnote-ref-153)
154. Supra, § 5 Canadian Code of Conduct (2001), Rule 11. [↑](#footnote-ref-154)
155. Which would, of course include the other obligations contained within Article 57 API, such as the obligation to attack the target that will cause the least amount civilian harms in instances where two or more targets offer a similar military advantage. Though a nation may have to be party to the treaty if the provision is not considered customary in nature (*see* note 156). [↑](#footnote-ref-155)
156. The latter of which is likely to include information regarding the interpretation of IHL. Corn and Schoettler, *supra*, note 19, at 822 identify, for example, that U.S. department of Defence, Law of Manual, 51 (2015), at 241, note 303, states that the U.S. does not consider Article 57(3) API to be customary in nature. In addition, and by the way of offering a caveat, while the authors have proposed that the test should carry certain boundaries, in practice any nation choosing to apply such a test, would do so according to which ever model suited their particular state practice – this could be more akin to either the narrow or the wide model previously identified. [↑](#footnote-ref-156)
157. *See e.g.,* Scharre, *supra* note 87, at 50-51. [↑](#footnote-ref-157)
158. See e.g.; Kevin Bonsor, How Landmines Work (How Stuff Works) <https://science.howstuffworks.com/landmine.htm>. [↑](#footnote-ref-158)
159. According to all relevant legal provisions, including in particular: the Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction, (the Ottawa Treaty); Protocol II to the Convention on Certain Conventional Weapons, and, for example; ICRC customary Rule 81, *supra* note 30, which states that: “[w]hen landmines are used, particular care must be taken to minimize their indiscriminate effects… State practice establishes this rule as a norm of customary international law applicable in both international and non-international armed conflicts. This rule applies to the use of anti-vehicle mines. It also applies in relation to anti-personnel landmines for States which have not yet adopted a total ban on their use”. [↑](#footnote-ref-159)
160. As the authors are currently examining elsewhere, certain nations may wish to indoctrinate military manuals, ROE, and other guidance to “ensure” that EAIs ensure that certain orders are followed, as opposed to refused. This may be the case, for example, where decision-making responsibilities are not generally delegated to the lower echelons of the military hierarchy. This may perhaps be due to issues with trust, but, the point is, an EAI could potentially ensure orders that are very prescriptive in nature are followed to a T. For a useful discussion see (podcast) Franz-Stefan Gady, *Future warfighting: placing doctrine before technology,* Sounds Strategic, (20 July 2020), at minutes 29.09 to 33.15. <https://www.iiss.org/blogs/podcast/2020/07/future-warfighting>. [↑](#footnote-ref-160)
161. Clearly, the central argument contained in the present thesis is that this will not always be the case. EAI tech is improving all the time. Take for example the U.S. Long Range Anti-Ship Missile (LRSAM) which, by utilizing an “intelligent navigation an direct route” feature, is already able to carry out rudimentary distinction assessments. *See* <https://www.lockheedmartin.com/content/dam/lockheed-martin/mfc/pc/long-range-anti-ship-missile/mfc-lrasm-pc-01.pdf> Future technology may improve upon this in a number of ways. For example, as previously discussed, the weapon itself may contain tech that allows for it to carry out its own complex assessments. Or, an EAI could act as a commander, or even as a kind of military police officer that is used to ‘enforce the law’. In such an instances, nations would clearly have to decide upon, and programme their EAIs accordingly, with the necessary codes of conduct regarding matters such as the issuing of cease and desist commands, and guidance on the use of force when arresting and detaining “friendly” soldiers. [↑](#footnote-ref-161)
162. One additional discussion that could be had here is what of the system was giving an order to fire, as opposed merely proving a direction? In other words, could authorization be delegated to a machine so that it could “indicate” that it had calculated the following attack offers a definite military advantage, that it is lawful, ethical, and very likely to positively influence the course of the battle – Take the shot! Arguably there could be consequences if the human pilot failed to follow such an “order” - because he/ she thought it better in the circumstances – perhaps only due to gut instinct. However, what if, due to not taking the shot, lives were lost, or ultimately, the battle was lost. Could/should it lead to a court martial, or criminal proceedings? Or, should it be of no consequence because the human pilot/ operator used their ‘superior’ (or hierarchical) judgement? In other words, should the ‘plane’ be assigned a higher-rank? The cyclical argument; the problem with threat of criminal sanctions is that the EAI does not change the status quo. EAI should be an improvement of the current balance of being able to refuse an order only in very limited circumstances, juxtaposed with military conditioning and potential criminal charges for insubordination. [↑](#footnote-ref-162)
163. Note that armed UAVs are becoming much more capable, for example than a Mk I Predator Drone. *See e.g.,* Valerie Insinna, *Boeing rolls out Australia’s first ‘Loyal Wingman’ combat drone*, Defense News (May 4, 2020), <https://www.defensenews.com/air/2020/05/04/boeing-rolls-out-australias-first-loyal-wingman-combat-drone/>. [↑](#footnote-ref-163)
164. Note that with the increased use of unmanned technologies in armed conflict, and the possibility of further advances in fully autonomous tech, it is arguable that such systems could and perhaps should hesitate before deploying force, due entirely to the fact that no human life is in direct risk. *See* e.g. Sassoli, *supra* note 18, at 327-328. [↑](#footnote-ref-164)
165. For a useful discussion of the facts of the USS Vincennes and that fateful incident, *See,* Scharre, *supra* note 87, at 169-170. Similarly, the same test may have also prevented the recent shooting down of Ukraine Air Flight 752 by Iranian armed forces on Jan. 8 2020. See Michael Safi, *Iran admits it fired two missiles at Ukrainian passenger jet,* The Guardian (Jan. 21, 2020) <https://www.theguardian.com/world/2020/jan/21/iran-admits-it-fired-two-missiles-at-ukrainian-passenger-jet>. [↑](#footnote-ref-165)
166. Article 1 of the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (hereinafter the Convention Against Torture) defines torture as: any act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person for such purposes as obtaining from him or a third person information or a confession, punishing him for an act he or a third person has committed or is suspected of having committed, or intimidating or coercing him or a third person, or for any reason based on discrimination of any kind, when such pain or suffering is inflicted by or at the instigation of or with the consent or acquiescence of a public official or other person acting in an official capacity. It does not include pain or suffering arising only from, inherent in or incidental to lawful sanctions. Nevertheless, as noted by the ICRC, see text on ICRC Customary Rule 90, *supra* note 30, noting that Article 8(2)(a)(ii) of the ICC Statute removes the need for the suffering to be inflicted: “by or at the instigation of or with the consent or acquiescence of a public official or other person acting in an official capacity”. [↑](#footnote-ref-166)
167. See Richard D. Rosen, *America's Professional Military Ethic and the Treatment of Captured Enemy Combatants in the Global War on Terror,* 5 Geo. J.L. & Pub. Pol'y 113 (2007), at note 31 citing: Memorandum from Jay S. Bybee, Office of Legal Counsel, to Alberto R. Gonzales, Counsel to the President, Subject: Standards for Interrogation under 18 U.S.C. §§ 2340-2340A (Aug. 1, 2002). [↑](#footnote-ref-167)
168. Article 2 of the Convention Against Torture states: 1. Each State Party shall take effective legislative, administrative, judicial or other measures to prevent acts of torture in any territory under its jurisdiction. 2. No exceptional circumstances whatsoever, whether a state of war or a threat of war, internal political instability or any other public emergency, may be invoked as a justification of torture. 3. An order from a superior officer or a public authority may not be invoked as a justification of torture. [↑](#footnote-ref-168)
169. Robert Bejesky, *Pruning Non-Derogative Human Rights Violations into an Ephemeral Shame Sanction*, 58 Loy. L. Rev. 821, 824-825 (2012). [↑](#footnote-ref-169)
170. Subject to the caveats identified *supra* note 160. [↑](#footnote-ref-170)
171. Note that an investigation into the extent of the soldiers duty to refuse to obey a manifestly unlawful order to launch a nuclear weapon has already been carried out by at least one author, *see*  Anthony J. Colangelo, *The Duty to Disobey Illegal Nuclear Strike Orders*, 9 Harv. Nat’l Sec. J. 84 (2018). The purpose of the current discussion is to consider how an EAI might help to ensure that duty is adhered to. [↑](#footnote-ref-171)
172. The point being that the circumstances that are presented in the following scenario are intended to be examined under DTP, and not in relation, for example to Articles 2(4) and 51 of the UN Charter, the latter of which the present authors are in the process of addressing in an accompanying piece. [↑](#footnote-ref-172)
173. It could, for example, be due to the command and control suffering an overwhelming attack, or due to a devastating cyber-attack. [↑](#footnote-ref-173)
174. This is of course true whether considering the scenario with or without the introduction of EAI tech, *See* Colangelo, *supra* note 170, at 90. [↑](#footnote-ref-174)
175. Indeed, in the U.S. there is a “two-man rule”, which, in short, means that the order to launch a nuclear attack is verified by two separate individuals, at every stage in the chain of command, *see* Colangelo, *supra* note 170, at 114-115. It is at least conceivable that an EAI could replace one of those humans at every stage of the process. [↑](#footnote-ref-175)
176. *See supra*, note 139. [↑](#footnote-ref-176)
177. Though this is, of course, subject to the condition that the tech each stage is functioning correctly. Nevertheless, because an EAI cannot authorise force, the impact of malfunctioning equipment is lessened, as compared, for example, to AWS. [↑](#footnote-ref-177)
178. *Supra* note 174. [↑](#footnote-ref-178)
179. Indeed, as identified by Colangelo, *supra* note 170, at 92, due to the unique destructive nature of nuclear weapons, “where conventional weapons can be used in proximity to civilians to achieve the same or similar objectives…an order to use a nuclear weapon would be manifestly unlawful and anyone with sufficient factual knowledge regarding the circumstances of the order should know it”. [↑](#footnote-ref-179)
180. Nuclear Weapons Advisory Opinion, Nuclear Weapons Advisory Opinion, *supra* note 27, at ¶ 97. *Also see* UK LOAC manual “6.17. There is no specific rule of international law, express or implied, which prohibits the use of nuclear weapons. The legality of their use depends upon the application of the general rules of international law, including those regulating the use of force and the conduct of hostilities.”, and The US Naval Handbook (2007) states: There are no rules of customary or conventional international law prohibiting nations from employing nuclear weapons in armed conflict. In the absence of such an express prohibition, the use of nuclear weapons against enemy combatants and other military objectives is not unlawful. Employment of nuclear weapons is, however, subject to the following principles: the right of the parties to the conflict to adopt means of injuring the enemy is not unlimited; it is prohibited to launch attacks against the civilian population as such; and distinction must be made at all times between combatants and civilians to the effect that the latter be spared as much as possible. Also *see* ICRC IHL database, Practice Relating to Nuclear Weapons, <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_nuwea>. [↑](#footnote-ref-180)
181. *See e.g.,* Michael N. Schmitt, *Humanitarian Law and Direct Participation in Hostilities by Private Contractors or Civilian Employees* Chi. J. Int'l L. 5 511 (2004-2005); Ido Kilovaty, *ICRC, NATO and the U.S. - Direct Participation in Hacktivities - Targeting Private Contractors and Civilians in Cyberspace under International Humanitarian Law,* 15 DUKE L. & TECH. REV. 1 (2016); Michael Anderson, *If It Looks like a Duck: Reining in Private-Military Contractor Conduct Through the Amended* *UCMJ*, 50 CASE W. RES. J. INT’L L. 307, (2018); Lindsey Cameron, *Private military companies: their status under international humanitarian law and its impact on their regulation,* 8 Int'l Rev. Red Cross 573 (2006). [↑](#footnote-ref-181)
182. Grimal/ Pollard, *supra*, note 25 at 555-556. [↑](#footnote-ref-182)
183. Note, for example that even in 2004, PMC’s serving in the Iraq war were earning up to $20,000 per month. See Schmitt, *supra* note 180, at 514-515. [↑](#footnote-ref-183)
184. The point is, owing to such financial rewards as those noted *Id.,* a PMC might not want to get themselves a reputation for being insubordinate. [↑](#footnote-ref-184)
185. Neil J Beck, *Espionage and the Law of War*, American Intelligence Journal 29, 1, 126-136, (2011). *Also See* note 12 at 372 in Stefan Kirchner, *Beyond Privacy Rights: Crossborder Cyber- Espionage and International Law*, 31 J. Marshall J. Info. Tech. & Privacy L. 369 (2014) citing; Simon Chesterman, *The Spy Who Came in from the Cold War: Intelligence and International Law*, 27 MICH. J. INT’L L. 1072, 1072 (2006), and; John Kish, Int’l Law And Espionage vii (1995). [↑](#footnote-ref-185)
186. Article 30 Hague Regulations (1899 & 1907) merely state: “[a] spy taken in the act cannot be punished without previous trial”, and Article 31 that: “[a] spy who, after rejoining the army to which he belongs, is subsequently captured by the enemy, is treated as a prisoner of war, and incurs no responsibility for his previous acts of espionage”. *Also see* ICRC Customary Rule 107, *supra* note 30. Article 29 Hague Regulations (1899) identifies “[a]n individual can only be considered a spy if, acting clandestinely, or on false pretences, he obtains, or seeks to obtain information in the zone of operations of a belligerent, with the intention of communicating it to the hostile party. And similarly, Article 29 Hague Regulations (1907) provides that “[a] person can only be considered a spy when, acting clandestinely or on false pretences, he obtains or endeavours to obtain information in the zone of operations of a belligerent, with the intention of communicating it to the hostile party”. [↑](#footnote-ref-186)
187. Metaphorically speaking, “removing it from the power socket, and holding the reset button for 10 seconds” may suffice. [↑](#footnote-ref-187)