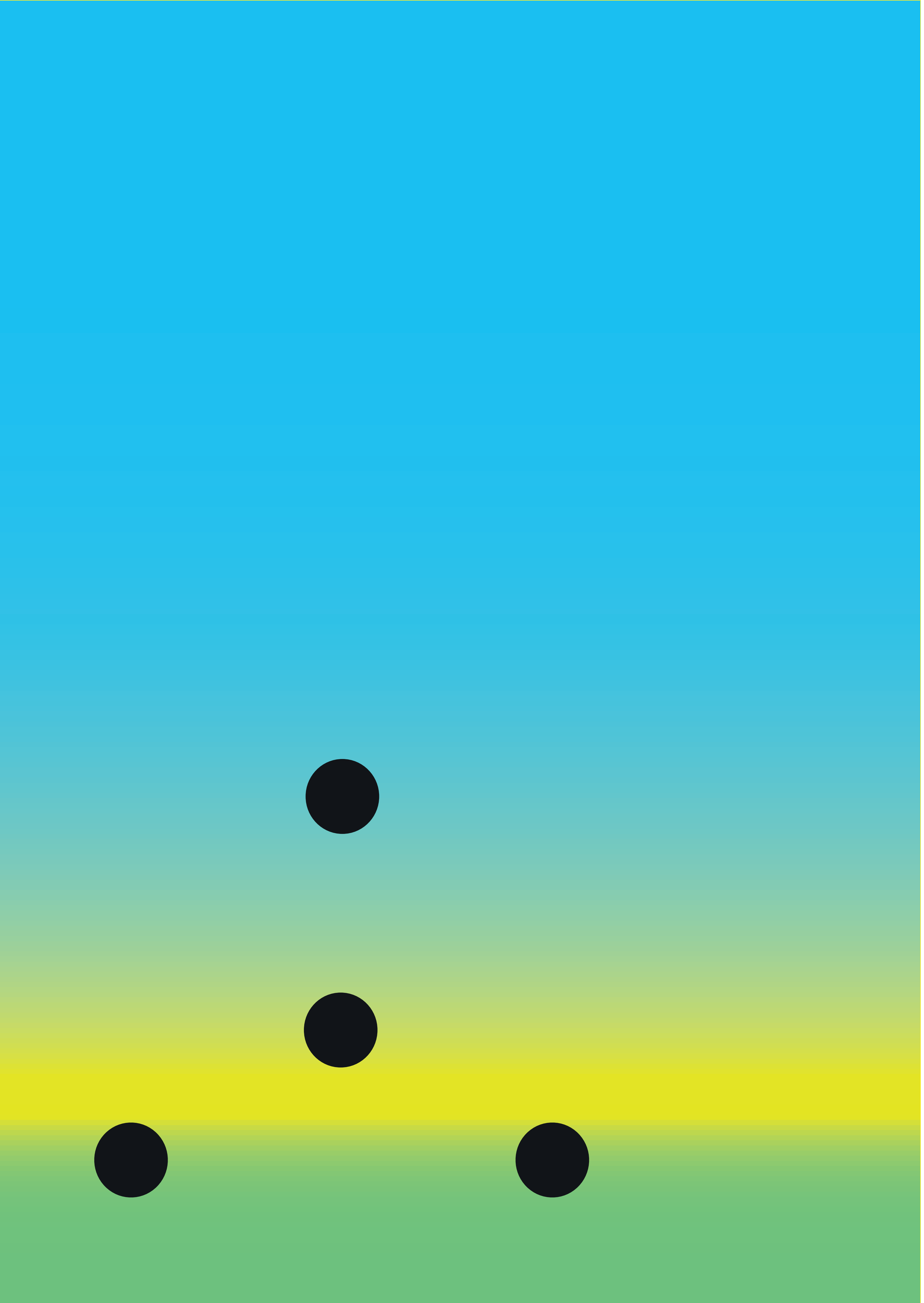
THE LIFECYCLE OF LETHAL / AUTONOMOUS WEAPONS SYSTEMS: OUTSTANDING TECHNOLOGICAL CONCERNS

UK CAMPAIGN TO STOP KILLER ROBOTS
TECHNOLOGY DEVELOPERS WORKING GROUP

MEANINGFUL HUMAN DIGNITY THROUGH MEANINGFUL HUMAN CONTROL

STOP KILLER ROBOTS

MEMBER



WELCOME TO THE CAMPAIGN

We are the **UK Campaign to Stop Killer Robots (UK CSKR)**, a network of UK-based organisations working toward a pre-emptive ban on lethal autonomous weapons systems.
We support the aims of (and are endorsed by) the global campaign of the same name. Our ultimate goal is to establish a legally-binding international treaty which ensures that meaningful human control is retained over the use of force and prohibits the development, production, transfer and use of fully autonomous weapons systems, also known as lethal autonomous weapons systems, LAWS, or killer robots. Our campaign objective is that the UK Government participates in the negotiation of, and joins, such a treaty.

The UK CSKR is a coalition of several organisations, each with its own mission and expertise, demonstrating the diverse set of groups with a strong interest in securing the goals of the campaign. The UK Steering Committee consists of Amnesty International UK (AIUK), Article 36, Drone Wars, Saferworld, The United Nations Association UK (UNA-UK) and The Women's International League for Peace and Freedom UK (WILPF-UK). There are also 18 NGO members of the campaign, as well as student fellows and associate lawyers who facilitate the campaign's work.

The UK campaign is organised into four dedicated work-streams (Technology, Political, Military, and Universities), overseen by a part-time consultant coordinator hosted by UNA-UK, a charitable company limited by guarantee (no.1146016). This paper is written to express the particular concerns of the Technology Developers Working Group of the UK Coalition of the global Campaign to Stop Killer Robots,[1] and does not necessarily represent the views of all coalition members.

[1] https://www.stopkillerrobots.org/

EXECUTIVE SUMMARY

The Tech Developers Group has been alarmed by the lack of understanding of the issues surrounding autonomy; the building, training and implementation of machine learning (ML) computer programme systems in particular. Such systems may be entrusted with all manner of data gathering, sensing, language and behavioural functions, and the range of issues that warrant consideration span from the basic to highly complex. We are concerned that the conversation amongst interested parties and delegates on the present and future capabilities of these technologies is either unrealistically utopian or dystopian, and largely fuelled by vested military or business interests, rather than being based on sound technological understanding as applicable to field use and law. Key players appear to lack a fundamental understanding of the basic structure of programming, training and the sequential processes which computers undergo when making "decisions". Such understanding is crucial if ML is to be developed and implemented in a way that accords with humane values.

The following paper seeks to provide a resource, which first brings together the most pertinent legal precedent and applicable laws, while delivering some business and commercial examples of technology application and regulation, before going into greater technological detail. The paper has been researched and authored by the Tech Developers Working Group Coordinator, and International Representative of WILPF-UK, Taniel Yusef. The paper's tech-cycles chart integrates outstanding issues of concern raised by the group with considerable input by Paddy Walker here. The campaign is grateful to those experts, Laura Nolan, Liz O'Sullivan and Noel Sharky, who leant their time for peer review. We hope that this paper will clarify and highlight critical issues, helping to ensure that discussions about ML in automated weapons systems remains grounded in scientific reality. It should be a compendium for those who may have expertise in one particular area but not necessarily all. It attempts to show where law, industry and complex tech converge in the theatre of war.

Putting the potential for ML based automated weaponry to one side, present-day weapons architecture are already error prone and have lessons to teach us regarding quite primitive levels of spoofability. The much hyped harpie drones are one notable example, having already crashed controversially due to possible hacking. **"The formal cover up story in the news was that its wing broke off. The real story was that the Control Centre lost control over the UAV and 'someone' else flew the UAV and mistakenly crashed it."**[2] Completely removing humans from the loop, and potentially having them merely 'on' the loop, would only magnify such vulnerabilities. More complex weapons architecture encounters additional issues due to the multiple components, and subjective, human-made parameters, all amalgamated into one system. A multiplicity of parts increases error potential, and how components interact to produce the weapon's final act is not satisfactorily predictable. This paper will focus on complex artificial intelligence components, the use of which may already be banned, regulated or are of serious concern[3] in dual-use technologies or other applications outside of LAWS/AWS, or which need robust commercial guidelines. Whether or not these components are currently being developed with the intention of being incorporated into a sophisticated weapon, we believe it important to tackle the issues associated with various individual components, to understand the risk of dangerous and disturbing results if used in LAWS/AWS.

As this paper will discuss, the outcomes of algorithm based machines are really **selections** based on input data filtered and matched by a form of training which is **biased** by definition. The problem at its most basic, is that ML can be extremely efficient at making highly processed mistakes: these mistakes may appear accurate to the computer, and register correct within the parameters of its programme. The mistakes can be further exacerbated when the computer programme updates or repeats according to these false positives, depending on the nature of the learning. Furthermore, out-of-test environments are prey to all manner of issues which can fool, change, blind, intercept, or confuse the sensing capacity and generally corrupt or delete the data. Not only would this render the readings and actions potentially wildly inaccurate, but highly dangerous. The latter applies to an updating model or a static one; where supervised, unsupervised or reinforcement learning models may apply, (these shall be discussed later).

While we delve into these particular issues, it would be a mistake to assert that these are the only barriers, or that there is a technological perfection that, if reached, would make a weapon system so reliable that we would find it ethically or legally tolerable so as to offset decision-making or accountability to them.

Rather, this paper serves to:

- **highlight the misplaced confidence in ML particularly and full autonomy in general,**
- **emphasise the need to legally code meaningful human control over critical functions**
- **ensure this control at crucial moments of target selection and use of force functions,**
- **consider the measures which would need to be implemented at the conception, development, procurement and deployment stages.**

[2] https://www.richardsilverstein.com/2013/10/13/iranian-sabotage-saudi-arabia-stations-israeli-drones-for-iran-at-tack-on-its-territory/

[3] Open letter calling for a global ban on biometric recognition technologies that enable mass and discriminatory surveillance, 07 June 2021 https://www.accessnow.org/cms/assets/uploads/2021/06/BanBS-Statement-English.pdf

Nearly 70 nations have now joined the call for a combination of both prohibitions and regulations in the form of a legally binding instrument.[4] Meanwhile, the International Committee of the Red Cross, ICRC, has formally declared their position and concerns, also supporting a ban.[5] The UK campaign is collaborating to find examples of relevant regulatory standards used in other jurisdictions which should provide inspiration or a template for a ban on LAWS/AWS. Ethical frameworks, industry standards and trade regulations[6] already exist from which the UK government could evolve its own AI strategy and enact protective, responsive legislation. For example, Canada has implemented a directive for autonomy in decision making[7], while in the UK it is mandated that STEM students acquire an Academic Technology Approval Scheme[8] certification in order to study subjects which would provide knowledge "to develop Advanced Conventional Military Technology (ACMT), weapons of mass destruction (WMDs) or their means of delivery". Some areas of controversial or mal-applicable science already have detailed ethical standard and licensing regulations, such as the important but potentially problematic embryology[9] field.

These regulations have safeguarded the use of scientific knowledge rather than limited scientific advancement in that area.

We believe the same is both possible and desirable with respect to the components discussed in this paper.

The Stockholm International Peace Research Institute, SIPRI, has studied the ways in which responsible research can be implemented using a mixture of compliance mechanisms, funding, governance and regulation. They emphasise the interdisciplinary nature of technological development and the participation of multiple stakeholders - universities, civil-society, private companies, and the state all have unique roles and interests - and duly encourage a range of measures to target each sector appropriately. Lessons in both ethics and specific regulation can be drawn here. SIPRI's key findings open by stating;

"The development, diffusion and adoption of military and dual-use applications of AI is not inevitable; rather it is a choice, one that must be made with due mitigation of risks."[10]

It is significant then, that without yet the directive of international law, finance is already shifting. The second largest German Bank, DZ Bank, has published its 2020 Sustainability Report 2020[11] with p45-46 is specifically related to ARMS. It is the third bank after GLS Bank and KD Bank in Germany which responded to the call to install regulations/exclusions in relation to autonomous weapons. The Belgian Banking sector is moving in a similar direction, creating a sustainability label, as many are, with a weapons policy similar to DZ Bank. While it mainly confines to illegal weapons, it looks to "controversial or discriminate weapons"[12], which indicates its direction.

[4] https://www.amnesty.org/en/latest/news/2021/11/global-a-critical-opportunity-to-ban-killer-robots-while-we-still-can/

[5] https://www.icrc.org/en/document/peter-maurer-role-autonomous-weapons-armed-conflict

[6] Regulating Dual-Use Technology Trade, The Economist, Dec 30, 2010

[7] https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

[8] https://www.gov.uk/guidance/academic-technology-approval-scheme

[9] https://www.hfea.gov.uk/about-us/applying-for-a-research-licence/

[10] p.31, Boulanin.V, Brockmann. K and Richards. L, Responsible Artificial Intelligence Research And Innovation For International Peace And Security, 2020

[11] https://www.dzbank.de/content/dzbank_de/de/home/unser_profil/investorrelations/berichte/2020.html

[12] para 3.3 https://www.towardssustainability.be/sites/default/files/files/RevisedQS_Technical_20210531.pdf

More robustly, the Norwegian Sovereign Wealth Fund, worth $1trillion, was recommended by a government appointed committee that companies manufacturing lethal autonomous weapons technology be removed from its portfolio.[13] Considering there had been little by the way of legal shifting, but much by the way of ethical concern among the very population of engineers at the helm of the technology, it bares consideration by the very investors and rating agencies themselves. This has effects on industry and research priorities. This is even without the move of some states toward a discussions of a domestic ban, for example Belgium's recent hearing on whether to move a parliamentary vote[14] into law. This is not to mention the stirring of debate within our own parliamentary houses, both the call for a ban by MP Alan Smyth[15], and the recent House of Lords discussion in response to a question put forth by Lord Clement-Jones.[16]

The UK government's 2020 paper (UK Commentary on the Operationalisation of the LAWS Guiding Principles[17]) reiterates its earlier 2018 submission, in which the ''Lifecycle of a Weapon System[18]" was set out. In this it was recognised that "a compendium would require input from multiple stakeholders across disciplines, including governments, industry and civil society"[19]. Accordingly, we use this paper to scrutinise the entire lifecycle of a weapon, and to present, at each of the six stages, our technological understanding and how this relates to law and military command. The issues covered include procurement, (and therefore issues of verification), practical implementation, (command and control), feasibility (field hazards and reliability), predictability (proportionality and accountability), bias and error, (discrimination and accuracy) and the commercial sector. "Phase 0", defined in the 2020 document as 'national policies, political control' is crucial. These are issues of sovereignty, economy, and security, which straddle international and domestic law, and which are both affected by and affecting of trans-national commerce.[20] The 2018 paper also engages with NATO's Allied Joint Targeting Cycle, a recognised proposition for best-practise. We use this as a useful parallel structure through which to analyse the specific, ''deliberate and dynamic targeting[21]'' foci of concern.

[13] https://ethicsintech.com/2020/06/18/norways-wealth-fund-urged-to-extend-weapons-ban-to-include-killer-robots/

[14] https://paxforpeace.nl/news/overview/belgium-votes-to-ban-killer-robots

[15] https://hansard.parliament.uk/Commons/2020-12-16/debates/531F815F-67EA-4594-A01D-D5CB73D938BF/details

[16] https://parliamentlive.tv/event/index/0f6746fa-3805-478e-bd16-89d7f1d9a225?in=14:54:43

[17] P.1 UK Government, UK Commentary On The Operationalisation Of The LAWS Guiding Principles,2020

[18] P.3 UK Government, "Group of Governmental Experts of the High Contracting Parties to the 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", Second Session Geneva, 27 - 31 August 2018 Item 6 of the provisional agenda Other matters, 2018.

[19] P.3 UK Government, UK Commentary On The Operationalisation Of The LAWS Guiding Principles, 2020

[20] P.2 UK Government, UK Commentary On The Operationalisation Of The LAWS Guiding Principles, 2020

[21] P.3 UK Government, "Group of Governmental Experts of the High Contracting Parties to the 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", Second Session Geneva, 27 - 31 August 2018 Item 6 of the provisional agenda Other matters, 2018.

Lifecycle of a Weapon System

0 **National Policies**
Political Control

1 **Research and Development**
Early Research And Deployment / Pro grame and Project Management / Requirements Definitions / Detailed System Design

2 **Testing & Evaluation, Regulation, Certification**
Test and Evaluation and Acceptance / Regulation and certification

3 **Deployment, training, command and control, Operation And Planning**
Training / Rules of Engagement / Operations Planning / Deployment to operational theatre

4 **Use & Abort**
Targeting decisions and activities / Battlespace management

5 **Post Use assessment**
Battle damage assessment / Lessons learned / In-service feedback

NATO Allied "Joint Targeting Cycle"

1 **Commanders Intent, objectives and Guidance**
2 **Target Development**
3 **Capabilities Analysis**
4 **Commanders Decision, Force Planning and Assignment**
5 **Mission Planning and Force Execution**
6 **Assessment**

INTRODUCTION

At the September, 2021 CCW[22] GGE[23] on LAWS, some states asserted that specifying 'Characteristics' in any written text would be restrictive and problematic, as autonomy in weapon systems would not be limited to the technologies of algorithms. It is true that technologies such as heat sensors, and other long-used weapons technology may also be applicable to AWS. It is the application, legality and **critical function characteristics** that can be similarly precarious regarding full autonomy. It is relevant, for example, that sensor-based systems can be used to sense for proxy indicators; heat-signal shape, movement, biometrics, weight, 'object recognition,' movement or biometrics encoding patterns in order to create target profiles representing humans. These equally necessitate the kind of regulatory language that creates normative frameworks safeguarding behaviour, practise and prohibitions. **Therefore, it would be appropriate that specific regulations could be put in place around certain technologies, which might be subject to positive obligations regarding meaningful human control, while others could be subject to specific prohibitions.** This requires a nuanced approach and a normative operational framework with consensus around targeting and application of force decisions. Moreover, some of the problematic issues which historically endure for more primitive weapons technology should not be ignored but seen as grave warnings in AWS. Those issues of feint for example, are vastly more complex when it does come to the intractable complexity of AWS which might then use machine learning or algorithms in general. For this reason, this paper focuses on this more problematically complex technology both for its ethical concerns and for the command challenges in general, but it does not preclude its relevance to these other technologies.

Those critical functions[24] provide a useful framework for considering the legal regulation of autonomous technology / uses. This is suggested in the 2018 paper where reference is made to the phases of NATO's allied Joint Targeting Cycle;

"Phases 1,2,3,4 and 6 will always include humans in the decision-making process. Certain steps within phase 5 (find, fix, track, target, engage, exploit and assess) may be automated and in several cases, have been for decades (e.g. torpedoes tracking targets) – but only in certain circumstances and in line with the factors outlined at paragraph 8, above..."[25]

Para 8 recommends the GGE specify which functions be deemed critical and need human control[26]. However, para 19's description of the responsibility of the chain of command concerns us regarding the fundamental inability to control certain aspects of AWS, if ML is left to be independent, and to programme with a degree of reliability for ''desired end states''[27].

[22] The Convention on Certain Conventional Weapons

[23] Group of Government Experts

[24] http://www.article36.org/wp-content/uploads/2020/10/Regulating-autonomy-leaflet.pdf

[25] p3, 2018_GGE+LAWS_August_Working+Paper_UK.pdf

[26] p2 2018_GGE+LAWS_August_Working+Paper_UK.pdf

[27] p5, para18. 2018_GGE+LAWS_August_Working+Paper_UK.pdf

WHAT IS A LETHAL / AUTONOMOUS WEAPON SYSTEM (LAWS) / (AWS)

A lethal autonomous weapon system, would be able to survey, target, select, engage and attack, without human oversight, intervention or legal consultation. They are often referred to as having autonomy at stages of critical function. Weapons systems using such autonomy and associated military technologies are used increasingly for reconnaissance and targeting.[28] Existing machinery use varying degrees of some-such autonomy.

The Phalanx Gun, for example, uses radar to scan an area of sky then fires rapidly at ongoing air strikes. While it has an element of autonomy over a limited time and space, it is activated and deactivated by a human operator. It is noteworthy that this weapon has already mistakenly fatally struck an Iranian passenger jet, demonstrating that problems arise even where there is some human control. It is also an example of the tenuous nature of hybridity, where human intervention may have been possible.
To the proposition made by some schollars, that a human could have intervened, does not make the outcome any less fatal where that human failed to intervene.

The definition of LAWS/AWS has long been controversial, and accordingly the August, 2021, session of the CCW GGE, saw some convergence and divergence on definitions and characteristics of Autonomous Weapons. It was encouraging that ''lethality'' had been contested by many as a redundant defining characteristic of AWS. Hereafter, for our analysis, noting that a weapon system can potentially cause or contribute to devastating harm and/or illegal engagements (biological and architectural) not necessarily or limited to lethality, and further encouraged to by the same suggestions in the September GGE, we will refer to Autonomous Weapon Systems (AWS), not limiting to Lethal Autonomous Weapons systems, (LAWS), unless specified, allowing that that both terms can apply

However, we believe that the distinction drawn between "fully" and "partially" autonomous weapons is misleading and potentially dangerous. Defining Fully Autonomous Weapons as ' 'designed'' to operate outside of legal frameworks, while relevant, is a potentially misleading inclusion when differentiated from Partially Autonomous Weapons Systems. When combined with later references to Risk Mitigation and Weapons Reviews this runs the risk of superficially placating those states which have called for a ban for Fully Autonomous Weapons, while in substance regressing to opaque self-regulation and lack of multi-lateral obligations which can tangibly manifest compliance demands and accountability in a meaningful way. It must be understood that any weapons system that carries out critical functions at critical stages without human interaction or oversight, in practice, is Fully Autonomous. It is the nature of this autonomy, the task it carries out, and whether or not it can operate under its own directives which is of paramount importance, particularly in the case of targeting humans. Rather than drawing artificial distinctions between full and partial autonomy, the emphasis should be on specifying critical functions deliberately and consistently. Encouraged by the UK's continued insistence at this session, that human control is relevant in different ways across the Lifecycle, we explore in this paper the realities of some fundamental factors.

[28] https://dronewars.net/wp-content/uploads/2018/11/dw-leash-web.pdf

Finally, the rather tenuous and hopeful argument that AWS might improve the humanitarian cause by increased precision is rather stretched. It assumes that the programs/ML/information upon which the system acts is suitably flawless. As we will see, if the program is left unchecked and a flaw arises, AWS acting on the basis of such inaccuracies may be precise according to its own information, but still result in erroneous outcomes - **errors and precision are thus not mutually exclusive.** Maximising the humanitarian compliance of AWS requires that a human is kept in the loop to supervise and ultimately approve the final actions of the system's basic reconnaissance. This leads us to the **importance of meaningful human control.**

TO ERR IS HUMAN? TO CONTROL IS DIVINE

In both the 2018 paper and the 2020 paper, human control, in some form, are mentioned as both paramount and necessarily relevant throughout the Lifecyle;

"Taking a human-centric through life approach, enables human life to be considered at multiple stages and from various perspectives. This includes across all Defence Lines of Development, the acquisition of weapons systems and their deployment and operation."[29]

It is worth noting that there was a period wherein a number of states were contesting the very language of human control, pointing to the mandate of the Group of Government Experts. It is useful, therefore, that the UK's 2020 paper specifically explains;

"III. Human-machine interaction
Human control is an enabler of military effectiveness and can help avoid undesirable unintended consequences. It is not a simple concept – it can be distributed in nature, affected by context and must be considered across the lifecycle of the whole system. We believe discussions on this are central to the continued success of the group; they should be carried out in tandem with work on a compendium on good practice.
We believe this to be one of the most important areas of focus for the group, and one that may allow the group to make the most meaningful headway in the discussions on LAWS."[30]

[29] P.3 UK Government, "Group of Governmental Experts of the High Contracting Parties to the 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", Second Session Geneva, 27 - 31 August 2018 Item 6 of the provisional agenda Other matters, 2018.

[30] P.3 UK Government, UK Commentary On The Operationalisation Of The LAWS Guiding Principles, 2020

At the September, 2021, GGE on LAWS, the UK delegation reinstated their view that human control is relevant in different ways across the weapon's lifecycle. The 2018 Paper urges the development of a **"A compendium of good practice mapped against a weapon lifecycle [which] would provide a clear framework for the operationalisation of the guiding principles by states."**[31] By mapping those issues which make operationalisation and reliable development fundamentally problematic in technical and practicable terms, we suggest in this paper that

there is simply no ''good practice'' method of reliably implementing LAWS/AWS without a new, legally binding framework on autonomy in weapons systems regarding their ultimate intended use starting from development stage.

This must ensure certain compliance using positive obligations and other specific legal coding which set international normative standards for the behavioural and procedural methods of engaging and operating critical functions within the system. To assume a version of ''best practise'' and hope for the best among states, with a technology so changeable and inconsistent would be deeply foolish and impractical. Here we analyse the technological issues relating to the 6 stages of the weapon Lifecycle, to highlight that the very idea of accountability, principles and mechanisms implied by this earlier development stage, have inherent, often insurmountable problems which appear later in the cycle.

We note that in a recent letter[32] from the Foreign, Commonwealth and Development Office to our Campaign, as well as comments in the recent September, 2021, session of the GGE, the importance of monitoring human-machine-interaction (and most recently human control) was described as critical across the lifecycle. However, we are disturbed by the looseness of favoured language like ''sufficient'', and the subjectivity prone "risk-assessment", over the more concrete obligatory or prohibitive language of legal regulation or normative operational frameworks. For such reasons, we find any notion that the commercial sector can be left without guidance, while problematic technology would be faithfully 'caught' at the procurement phase, protected by Article 36, to be deeply naive, and possibly dangerous. It must be remembered that any high risk applications of current technology are likely to fall under existing export control frameworks; including end-use provisions on the most egregious cases. However, the strength of licensing applied will have necessitated international and/or national agreements to guide these decisions. Any emerging technologies will require similar frameworks, particularly relating to meaningful human control.

More reassuringly, at the most recent September session, the need for human-machine teaming with responsibility throughout the life-cycle, including at the development stage, was re-iterated by the UK. Here, we provide an academic assessment of features which require constant oversight and others which are fundamentally rogue in their capacity. Initially, we provide a basic explanation of rudimemtaary concepts of AI and ML training and function. Then, we apply some legal context and precedent providing the landscape for the following section. We then apply some industry references, both commercial examples of similar component technology and financial possibilities around regulation or investment appetite. Next, the body of the paper sets out our main technological concerns, which we find under each phase of the Lifecycle. Lastly, the more detailed technological issues are delineated in a tabular form to show their interactions, both with the corresponding phases of Lifecycle, and with the phases as set out in NATO's Joint Targeting Cycle.[33]

[31] P.2 UK Government, "Group of Governmental Experts of the High Contracting Parties to the 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", Second Session
Geneva, 27 - 31 August 2018 Item 6 of the provisional agenda Other matters, 2018.

[32] Letter from the FCO to the UK Campaign to stop Killer Robots. Jan 2021

[33] P.30 Fig. 2.2 NATO STANDARD AJP-3.9, Allied Joint Doctrine For Joint Targeting, Edition A Version, 1 April 2016

What the table shows is that almost every process interacts with command and control, targeting, accuracy, and the ability to comply with law. The modular nature of this technology creates additional concerns with the 'dual-use' problems involved in both development, application, and rogue use. This reiterates the fact that oversight and the need for military control (both political, legal and strategic) is required due to the nature of computers. Meanwhile, there is a parallel command which is required, the technological one, which is required to keep the programme functioning and running in the correct or desired manner.

In addressing phase '0' of the lifecycle; the national, development and legal phases, we must broadly understand the legal context before briefly framing the business context. These lay a foundation to which the outstanding tech-specific concerns must refer. It is worth noting that NATO's Allied Joint Publication paper states that under Phase 1; Commanders Intent, objectives and Guidance; ''the Targeting process is conducted within political strategic direction and guidance."[34]

[34] p.30 NATO STANDARD AJP-3.9, Allied Joint Doctrine For Joint Targeting, Edition A Version, 1 April 2016

0/ NATIONAL POLICIES
POLITICAL CONTROL

CODIFYING CODE

Code and Context: Legal Precedent[35] for the extra codification of human control and dual use provisions, relating to national policies and political control.

The Technological community are and have been concerned that AI, and ML especially, could be used to track, select, target and then engage a lethal strike. They signed an open letter in July 2018 to the UN which increased to 2000 signatories and has since reached well over 3 000. This pledge[36] has signatories including famous scientists and academics like the late Stephen Hawking, Noam Chomsky and tech giant Elon Musk, as well as leading AI companies like Google DeepMind and academic institutions such as University College London, UCL. Legal push-back from some states asserts that banning the technology would stifle technology or that a pre-emptive ban of non-existent technology is unnecessary, and in any case current IHL is suffice.
However, the sector is audibly concerned that without the safeguarding of international law, the technology will develop beyond the ability to wrestle it from bad actors. They also assert a concern around a lack of understanding of its limitations.

There is an argument that developing technology or potentially dangerous applications of technology cannot be banned pre-emptively[37]. Blinding Lasers, however, were pre-emptively banned in 1995, (First Review Conference, CCW, Protocol IV). This was very much a drive of scientists themselves, similarly alarmed at the potentially horrific secondary use of this positive laser technology, to instantly blind soldiers or civilians in war, and the tributary potential of policing or rogue uses that could cascade from there. This is also a positive example of pure science being protected by law not stifled by it, (we have laser surgery today). Meanwhile, far from prohibiting scientific proliferation, examples like the Biological Weapons Convention allows states parties to divert their stockpiles to "peaceful purposes[38]," while the Treaty to Prohibit Nuclear Weapons permits "research, production and use of nuclear energy for peaceful purposes without discrimination[39]." Both ban treaties encourage the collaboration of states and sharing of scientific knowledge for peaceful technological advancement.

It is relevant to note that the Chemical Weapons Convention, Mine Ban Treaty and Convention on Cluster Munitions ban prohibit "under any circumstances"[40] the use, development, production, acquisition, stockpiling, retention, and transfer as well as assistance with those prohibited activities. The significance here is that verification around stockpiling or development, and the fact that assistance has elsewhere been interpreted as including financial, means there is great scope to impose soft and hard law on states who fail to comply, while restrictions can be applied at the funding level. Moreover these treaties apply in peace time. Therefore are applicable to law enforcement, border control and at development levels.

[35] https://www.hrw.org/report/2020/10/20/new-weapons-proven-precedent/elements-and-models-treaty-killer-robots#_ftn16

[36] https://futureoflife.org/lethal-autonomous-weapons-pledge/

[37] Summary Record (Partial) of the 13th Meeting, Review Conference of the States Parties to the Convention on the Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects, CCW/CONF.I/SR.13, May 3, 1996, para. 69

[38] Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (Biological Weapons Convention), signed April 10, 1972, entered into force March 26, 1975, art. II

[39] TPNW, pmbl., para. 22

[40] New Weapons, Proven Precedent, Elements of and Models for a Treaty on Killer Robots, Human Rights Watch, Oct, 2020

The Ottawa Declaration on Cluster Munitions noted the humanitarian impact as; "unacceptable harm to civilians." These were existing technologies, already in the field, 40 million of which have been cleared to date. While the technology has not disappeared into the ether, the stigma around the weapons is such, that any further use would draw international repercussions, investigation into use, chain of command, supply chain, finance, possible involvement of ICC/ICJ, sanctions and so on. It is significant that in recent GGE sessions on LAWS, International Criminal Law has been increasingly referenced, despite the limited past engagement.

Particularly in the case of the TPNW, ''unacceptable suffering''[41] was asserted, as was the disproportionate harm to women and girls and indigenous peoples, highlighting discrimination, (due to geographical testing and ionising radiation's effects on women and girls). This also emphasises the wide-scale and indiscriminate nature of the harm. The inability to control or intervene is further exemplified here. The very concept of human control, or lack there-of, and the condemnation of autonomy in weapons systems, have been pertinent in weapons bans previously, as has the excessive and untenable human suffering[42].

Anti-personnel mines are sensor based and autonomous, in that they are triggered by the victim. They were banned partly for this reason; their indiscriminate harm without intervention, including after battle has ended. Similarly, the United Nations General Assembly resolutions noted that, chemical weapons´; **"effects are often uncontrollable and unpredictable and may be injurious without distinction to combatants and non-combatants, "**[43] and thus deemed to be in violation of international law. These were than banned in 1993. Biological Weapons are similarly banned and an excellent example of the indiscriminate nature of a weapon. The feeling was strong enough that IHL-extraneous treaties were drafted, to some degree because some weapons discussed were deemed weapons of mass destruction, due to the indiscriminate mass harm they can cause. There are overlaps which, with en-masse application and the indiscriminate harm that some AWS could inflict, bares some consideration.

Looking forward to the technology and legal frameworks, the Martens Clause guides our tone with the basic tenet that;
"Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity and the requirements of the public conscience."[44]

The Declaration of Human Rights and the general obligations of the Geneva Convention are similarly present to guide and inform the work of the GGE. Significantly, the latter's Protocol 1 states that parties to conflict must have obligations to either do, or prevent certain actions or activities, such as indiscriminate attacks.

[41] TPNW, pmbl., para. 6

[42] See TPNW, pmbl., para. 9 as an example.

[43] 1969 UN Year Book, https://cdn.un.org/unyearbook/yun/pdf/1969/1969_40.pdf

[44] https://www.icrc.org/en/doc/resources/documents/article/other/57jnhy.htm

The Geneva Convention's articles on Proportionality raise some specific issues which are pertinent for this paper. Proportionality, (Article 51)[45], requires awareness of possible harm, which until now has required an assessment of possible civilian casualties, potential harm to the other party, to infrastructure and accountability as to whether the gain was proportionate to the threat posed. These require reliable reconnaissance for calculations, reliable prior knowledge. In the case of technology use, it requires that the technology itself, gathering that intel, or applying that force, not be a potential known cause of disproportionate or incidental harm itself, which would immediately be a crime.

Or rather, the inability to ensure that the technology can operate within an acceptable degree of predictability and therefore harm threshold, or that it is providing accurate data (we assert that it almost definitively cannot), is beyond certain rules of war and law.

This is highlighted by **Article 51. 5 (b), which prohibits indiscriminate attacks,** partly describing indiscriminate attacks as;

- " (b) 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."[46]
- **Article 57. 2 (a) (iii) meanwhile, proposes precautions of care to be taken in attack**, requiring states to,
- " (iii) refrain from deciding to launch any 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;"[47]

Nor are we satisfied that AWS could satisfy;

- **Article 3, adverse distinction, between combatant, injured combatant and civilian, which is requisite to protecting those peoples states have a duty to protect under IHL,**
- **Article 48, which obligates that "Parties to the conflict shall at all times distinguish between the civilian population and combatants[48]," or**
- **Article 85, which makes it a "grave offence[49]" to attack indiscriminately in the knowledge of anticipated excess of harm.**

Similarly, the false testing and problematic nature of identifying useful and dynamic data modelling to create appropriate thresholds, with the opaqueness of the ML's matching protocol, false positives are guaranteed, which makes lab testing before the noise of war, matched with a potentially wide scope for error in the field of the unknown, a horrifying notion. Accurate readings may be wildly erroneous, which could be a legal accountability and Article 36 nightmare.
The knowledge that these are suboptimal and potentially wildly in violation at the point of conception fundamentally undermines Article 57, the obligation to take precautions to spare civilian population or civilian objects, which requires ability to distinguish them. As long as these cannot be assured, this is fundamentally a failed instrument. After a time, which companies would want to pour their investments into failing or publicly unpopular products?

RISKY BUSINESS

It is worth remembering, that where there is buzz in all directions around AI, investments are speculative and follow the money. Asset managers are acutely aware of their fiduciary responsibility, which comes with legal as well as financial ties. There are historical lessons to be learned about unpopular products from a PR, consumer and client-accountability point of view. More so, toxic investments, **if predictably undesirable and therefore predictably unprofitable** due to various political, legal or other weather, could make areas of a portfolio uninvestable and their share-price plummet, (the previous giants of tobacco and fossil fuels have taken a tumble recently for example). The obligation to invest responsibly from a financial point of view falls to asset managers, which can have hefty repercussions from unhappy investors if they are seen to invest avoidably poorly, especially to invest in products knowingly involved in harms. This equally goes for public money like pension funds and the increasing call for transparency around how these funds are managed. Equally, there is a growing trend for ethical investment and rating agencies are catering for this by increasingly looking into funding and company behaviour, catering to all manner of social, environmental and arms related activities awareness. An ICAN/PAX report showed $63billion losses in investment in nuclear weapons in 2021 alone, with 127 financial institutions ceasing their investments, many citing the TPNW as an influencing factor. Share prices are falling and investors are concerned about associated PR as well as stock viability.

It is wise, then, for financiers to see such coming trends and mark them. This is already beginning with ethical AI investors, venture capitalists and hedge funds, the fact that people are asking questions of their banks' investing behaviour, private investors having control over ethical opt-outs and the fact that the banking sector is starting to shift. It is commendable then, that the 2020 UK commentary paper acknowledges the uniquely leading role of investors and industry.

"Dialogue between governments and industry is particularly important given the intersection with industry standards and the fact that investment in research and development by private technology companies tends to dwarf that of governments." [50]

This crucial dynamic of technology development also affects commercial security, workforce and finance flow. As mentioned, leaders in the field have raised their concerns publicly. Tesla's Elon Musk and Alphabet's Mustafa Suleyman led the group of more than leading robotics experts in their 2017 call to the international community to protect humanity from these weapons before it is too late:

"Lethal autonomous weapons threaten to become the third revolution in warfare. Once developed, they will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend. These can be weapons of terror, weapons that despots and terrorists use against innocent populations, and weapons hacked to behave in undesirable ways. We do not have long to act. Once this Pandora's box is opened, it will be hard to close."[51]

[45] Protocols To The Geneva Conventions of 12 August Additional to the Geneva Conventions of 12 August 1949

[46] p.38 Protocols To The Geneva Conventions of 12 August Additional to the Geneva Conventions of 12 August 1949

[47] p.41 Protocols To The Geneva Conventions of 12 August Additional to the Geneva Conventions of 12 August 1949

[48] p.46 Protocols To The Geneva Conventions of 12 August Additional to the Geneva Conventions of 12 August 1949

[49] p.46 Protocols To The Geneva Conventions of 12 August Additional to the Geneva Conventions of 12 August 1949

[50] P.3 UK Government, UK Commentary On The Operationalisation Of The LAWS Guiding Principles, 2020

[51] Elon Musk and tech leaders call for UN ban of 'killer robots' and AI weapons

In his comprehensive paper on command issues, Paddy Walker terms the varied actors involved in the development and application of a weapons system, the "Delivery Cohort[52]," which he notes as inevitably convoluted and cross-sector. These include; politicians, generals, local commanders, soldiers on the ground, maintenance and service personnel, those in manufacturing and procurement (including procurement lawyers and those involved in Article 36 reviews), those in design and programming, heading commercial entities, regulators and lawyers, the Press, the Third Sector and some might argue, the public. The trans-boundaried nature of commerce, and sometimes collaborative nature of code itself, (eg. multi-engineer, many-labeller), means a company might sell its technology anywhere and that technology may be repurposed. **The technology may then be delivered by engineers unfamiliar with the previous modular processes. The final product can be something quite unfamiliar to the end user, and more difficult to access, fix or reconfigure when there is a problem.**

Regulation, International Property and other standards, can be more robust in the business sector than the opaque ''best practise" of the international review arena, while commercial patenting vs. product protections in-house, are jealously guarded. Standards can become normative where they provide sector access, entry standards to regional trade or raise business best practise for the avoidance of litigation, poor public profile or the loss of labour, clients and investor ratings. Such normative standards have led to changes in law, which is significant for the opacity of this sector and its problematic tech. Where there have been issues with AI technologies there are already calls for commercial improvements. Self-regulation by industry, however, in any dual use tech which could end up in weapons, should be deeply worrying to any government.

We know, for example, there have been many recorded instances of racism and gender bias[53] in facial recognition, voice recognition, and so on, due to the imbalance of data making the training sets from which machines learn and run. Facial recognition is already used in border control and failing. These have implications for civilian safety and direct implications for AWS[54]. Examples in the UK alone give us some telling and worrying examples. Big Brother Watch's report on the accuracy level of the NEC's NeoFace Watch used by the UK police to identify faces among crowds showed a **95% failure rate, that is to say only 5% accuracy**[55]. The same technology was even less effective when used at the Notting Hill carnival, which is an Afro- Caribbean festival in London,

with an accuracy rating of only 2%.

It is worth noting the comparative, and rather embarrassing example, of the Amazon and Microsoft software which alerted US lawmakers to the issue of inaccurate technology. 28 congress members were matched with criminal mugshots in an exercise by the American Civil Liberties Union (ACLU).[56] These are not comforting examples about the capacity of the technology, nor the quality of data or their weighting.

WIRED with Element AI found that only 12% of the leading researchers working on AI Were women, while Mines action Canada asserted that 0.0004% of the population have the expertise to build autonomous weapons systems[57]**.**

Which means that potentially, only 0.000048% of those working on AWS may be women.

[52] Leadership Challenges from the Deployment of Lethal Autonomous Weapon Systems How Erosion of Human Supervision Over Lethal Engagement Will Impact How Commanders Exercise Leadership, Paddy Walker RUSI Journal, 6 May 2021 Technology

[53] https://reachingcriticalwill.org/resources/publications-and-research/publications/13601-a-wilpf-guide-to-killer-robots

[54] Sharky, Noel. 'The Impact of Gender and Race Bias in AI' August 2018

[55] https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf

[56] https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

[57] p.20 https://www.stopkillerrobots.org/wp-content/uploads/2021/09/Gender-and-Bias.pdf

This rather exacerbates the issue of limited subjective input building programmes, labels, parameters and data sets. Research, civil society and the commercial sector are differently recording the problems of computer bias, the missing face problem and other contextual bias assumptions made by computers and the assembly of available data collated through the social prism. Disability[59], skin colour, gender, vocal recognition and facial patterns as well as sexual orientation and socio-economic classifications[60] are made with data labelling and decisions around selectively classed, and 'weighted' data currently favouring the white North American male and ''visibilising'', by replication, various social stereotypes. It leaves women of colour, with non-American accents and without visibly fully abled bodies the most vulnerable to algorithmic bias. **As most wars are not fought in North America, it leaves little comfort that the civilians there would be accurately detected and protected. When algorithms reduce to people clusters of data by stereotype and similiarity it can disappear them all together.**

In a recent visit to the United States, The Special Rapporteur for Poverty noted the;

"Orwellian side to CES, [the coordinated entry scheme], ... A ranking algorithm gives the homeless respondent a vulnerability score between 1 and 17 and a second, matching algorithm, matches the most vulnerable homeless to appropriate housing opportunities."[61]

Noting its inefficiency as a whole, the report also noted the invasiveness of the data collection and the then transferability to other sectors to be similarly used, or rather, misused. More so, other uses of AI were noted, those for pre-trial risk assessment tools, in order to set often prohibitive and discriminatory bail conditions. These compare individuals to a "population of individuals who share certain characteristics".[62]

It has been observed by NGO and academic sources, that fines for misuses of such data, as well as for problematic algorithms[63] in business and governance, from insurance to finance, housing allocation to bail and parole processing within the justice system, have been incurred where racism or other forms of bias in the design or output of ML based programmes exist.

These fines not only exhibit the legal as well as financial costs of such ill-used or ill-equipped technology, but they suggest a trend of increasing ramifications for companies where lack of oversight can lead to various violations of domestic laws and trading standards.

It is concerning then that such a small demographic is responsible for creating the data sets and labelling parameters which can create targeting profiles, so wide reaching are the failures.

[59] AI Now, "Disability, Bias and AI." Nov 2019

[60] p7-8 Acherson. R, A WILPF Guide to Killer Robots. Jan 2020

[61] 4. I Statement on Visit to the USA, by Professor Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights, Washington, December 15, 2017, https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?News-ID=22533&LangID=E

[62] 4. II Statement on Visit to the USA, by Professor Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights, Washington, December 15, 2017, https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?News-ID=22533&LangID=E

[63] How does automated decision-making effect our daily lives? Algorithm Watch, 2021 https://algorithmwatch.org/en/stories/

There are many examples of the problematic nature of machine learning within the commercial sector alone. Research has tracked fines and lawsuits based on bias within systems[64] and abuses. The New York state's attorney has already brought cases against companies which violated domestic obligations, for example, against major insurers, bestowing vast fines as a result of harmful or discriminatory algorithms creating computer bias[65]. Facebook, for example, now uses AI-powered facial recognition software as part of its core social networking platform to identify people. Meanwhile, law-enforcement agencies around the world have experimented with facial recognition surveillance cameras to reduce crime and improve public safety[66]. A Chinese study showed that a search using mug-shots to search for ''criminality'' was actually using the expression of a smile as the common defining factor. Not very promising. Incidentally, Facebook was the largest accruer of fines for IT based violations alone in 2019[67] with additional suits for discrimination.[68]

While the usual weapons developers are working on AI, we know that both dual use, and dual industry (military and commercial) players exist. Commerical actors have developed aircraft with AI integrated systems. However, in the commercial sector automated components have already inflicted devastating results. The airbus A320 and modified 737 had new engines with new sensors (a new software system). Shortly after the planes launched, multiple system failures led to fatal crashes.[69]

Boeing whistleblower and retired US Navy Captain, Mr Pierson testified before Congress about the repetitive faults that the sensors and software displayed indicating erratic SPD (speed) and ALT (altitude) flags. The new size engine lifted the nose; the new sensor, fitted outside of the plane, fed back that plane was rising at an acute angle and heading for a stall. This created a feedback loop for the sensor. The pilots weren't able to override the plane's actions and both RyanAir and Boeing had 2 crashes and people died. In fact the Boeing testimony to Congress stated that, **"one in 25 Max airplanes had already experienced a safety incident within the first year of being in service, two of which happened to be fatal crashes. This track record is unprecedented in modern-day aircraft. The FAA's recertification fixes fail to adequately address these issues."** The software system went wrong, producing mass failure and even with pilots, the results were deadly. A written paper describes a litany of factory issues unchecked.[70] Some may find a certain amount of error in new systems tolerable, but the error margin in domestic drones appears far narrower/stricter than in large aircraft, and rather, it is the nature of the failure itself which is of import here. More significant than the system failure, which some concerningly argue is the ''tolerable'' nascent error of emerging technology, is the fact that the system was not appropriately designed for humans on the loop. This emphasises the particularly precarious nature of such design and the need for robust, rigorous regulation which has IHL compliant mechanisms for human control, trigger and abort systems (critical function specificity) even at conception, in the form of specific obligations and prohibitions.

[64] Good Jobs First Violation Tracker; Class action lawsuits alleging discrimination against customers

[65] Violation Tracker Individual Record, United Health Group
https://violationtracker.goodjobsfirst.org/violation-tracker/ny-unitedhealth-group-inc

[66] Lindsey. N, Facial Recognition Surveillance Now at a Privacy Tipping Point, Feb 2019

[67] https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf
https://violationtracker.goodjobsfirst.org/violation-tracker/-facebook-inc-0

[68] https://www.aclu.org/other/summary-settlements-between-civil-rights-advocates-and-facebook

[69] https://www.msn.com/en-gb/lifestyle/travel/737-max-e2-80-98unexplainable-electrical-anomalies-e2-80-99-and-other-faults-claims-former-boeing-manager/ar-BB1d4iqD

[70] https://img1.wsimg.com/blobby/go/ec12e28d-4844-4df3-a140-ca706a04c0f7/downloads/737%20MAX%20-%20Still%20Not%20Fixed.pdf?ver=1611532831723%20

There are other high profile examples of automation failing, like Tesla's infamous self-driving car driving straight into an 'invisible' truck, which had rolled onto its side, killing the driver. The truck, being on its side was an unknown entity to the car, no longer a truck, unrecognisable. The passive driver was too late to intervene[71]. Notable to industry; Tesla face a wrongful death litigation.[72] It would be a mistake to confuse the idea that some accidents are tolerable, and this merely replaces one cause of accident (human error) with another.

Not withstanding the ethics of removing human accountability, to make such an argument fundamentally misses the problem that such accidents highlight; the very basic errors of complex system processing and recognition that can be encountered, and which are often only evident in practise upon deployment. These, machines can be fazed by quite simple interruptions of their predicted parameters; hence any proposition that deploying static, non-updating models as safer and more predictable being a highly erroneous one as an alternative. (The Tesla will have been shown trucks before, it hadn't factored for the irregularity of a truck on its side). Further, these are commercial examples of the evident failure of hybridity at critical junctures.Moreover, the question of accountability here is further complicated. If a human is on the loop, but struggling to intervene effectively, how is legal accountability applied appropriately?

Notably, any academic attempt to create 'Hybrid Autonomy' where humans intervene at critical moments, which are inevitably moments of high stress, time pressure and/or limited information is not promising. Research thus far suggests the attempt to delicately move between human and machine in this fashion has produced erratic results.

Global supply chain and market forces have different priorities. The presence of the UN Global Compact and the Business Human Rights Forum, which takes inspiration from the UK Modern Slavery Act, recognises the need to operate within ethical standards and finally link business standards to human rights standards beyond the previously tokenistic ESG, (environmental, social and governance), and CSR, (corporate social responsibility). Within the work of the COP, (United Nations Climate Change Conference) and SDGs, (sustainable development goals), data archiving tools are becoming ways of supplying vital information about crucial scientific records on the ground as well as rights abuse mitigation and remedy, from the environmental to human. The old, 'out of sight' model will not be tolerated and the public are less forgiving of companies which supply every day items while simultaneously profiting from harmful activities elsewhere, and they are prepared to act on this conviction.

Public opinion on this lesser known technology is already shifting. A global poll has shown that **61% of people opposed the development of AWS in 2018**, a jump from 56% the year before.[73] Divestment/boycotting campaigns have been successful in other areas, and led to assistance and finance clauses within treaties gaining teeth, while investment rating and Bond regulators have started to measure all manner of violations according to new trends; highlighting the fact that investors like to future proof. The example of stock price decline in fossil fuels, nuclear weapons, or more so, the very speed of divestment from fissionable material (Deutsch Bank), or from all nuclear weapon related material, (Norwegian Government Pension Fund) from their portfolios so shortly after the adoption of the TPNW shows financial institutions' movement with the warning signs of the profit climate, as well as with ethics.

[71] https://www.forbes.com/sites/bradtempleton/2020/06/02/tesla-in-taiwan-crashes-directly-into-overturned-truck-ignores-pedestrian-with-autopilot-on/

[72] https://electrek.co/2019/08/01/tesla-faces-wrongful-death-lawsuit-fatal-crash-autopilot-semi-truck/

[73] https://www.stopkillerrobots.org/2019/01/global-poll-61-oppose-killer-robots/ [Accessed 11 Aug.2019].

It is a significant indicator then, that, the very same and largest Sovereign Pension fund in the world, announced as far back as 2016 that regarding dual use technologies,

"If you think about developing technology for recognising cancer, that is fine. But if you are adapting it to track down a certain type of individual in a certain environment, and cooperating with others to make an autonomous weapon out of it, don't be surprised if we take a look at you."[74]

As mentioned in the introduction, we have since learned that this ''look'' has become a decided policy leer. These are all hopeful for the scope of regulation, the UK technology sector and international legal incentives. Elsewhere, the EU and UK have led the world in Financial Regulation by creating and implementing legislation, such as MIFID II,[75] Markets in Financial Instruments - Directive, and MAR[76] Market Abuse Regulation. This has led to the recent boom in the coveted, uniquely European regulatory technology (Reg Tech), leaving the typically US-based tech finance companies far behind in this advanced area. This indicates a hugely positive potential for regulation of sensitive tech-based industry applicable to various associated fields, including autonomous weapons.

The commercial opportunities of leading in reg tech, the associated compliance-based infrastructure and the technological advancement, advantage and labour, could lead a particular market which straddles sectors, while influencing the ethical and legal directives of national and global standards.

Non-subscription to standards nationally means falling behind as normative behaviour accedes to the needs of trade demands. Cross-border standardisation together with the associated litigation, stigmatization or profit loss, soon incentivises those regions to comply. This coerces normative practise and often law. International law can guide regions and states to do so, while making certain applications, transfer of technology, behaviours and uses internationally illegal. This guides national and regional legislation as well as sector standards. **Being the regulatory leader in an industry, sets the ethical and legal standard while placing the region/ nation at an industry advantage in the sector and in the field of regulation technology itself.**

More specifically, by incentivising tech companies to work from the development stage toward ethical and legal frameworks, with the appropriate consultation with government and civil experts, the products created by such companies can be designed with certain specifications of restricted use at the outset. **Legal and ethical codes can similarly make the practise of intended use transparency, and at product / sale stage, more attractive for those working on the products.** Smaller start-ups, especially, have shared with us a desire to ethically code, or draft documentation for their products, but lacked the resources or legal guidelines to do so. **By collaborating with private companies, (and the associated development/academic institutes) at the pre development stage, the UK government, GGE and civil society experts can provide UK technology companies and collaborators, with protections against unintended uses and dangerous reverse engineering. It can also make the UK industry a flagship of cutting edge science and a leading standard bearer for trust, efficiency and reliability.**

[74] Reuters."Exclusive-Norway wealth fund's ethics watchdog warns firms not to make killer robots", Joachim Dagenborg, Gwladys Fouche, 2016

[75] https://ec.europa.eu/info/law/markets-financial-instruments-mifid-ii-directive-2014-65-eu_en

[76] Market Abuse Regulation; https://www.fca.org.uk/markets/market-abuse/regulation

It is crucial that in doing so, the nature of such tech - legal collaboration is directed toward upholding and advancing the ethical and legal standards which comply with the UK's own obligations; regarding discrimination, privacy, data protection, but more significantly the knowing development of weapons systems and critical function components which could violate International Human Rights, International Humanitarian Law, Laws of Armed Conflict and the Arms Trade Treaty, for example.

In deed, by working with these companies on pre-development, industry-wide standards can be built into the initial designs and concepts, with a fundamental understanding of the potential hazards of lack of oversight, unpredictability, the iterative nature of the programme when independent, bias and data corruptibility by noise, and by spoofing. In all these cases, meaningful human control at various stages in a programme (especially weapon specified systems should be factored into sub-agent and holistic design for crucial moments of their operation and throughout their development cycle. Baring in mind, that at this stage we can discuss technology safeguards before they become misused domestically or transferred, hacked, downloaded and developed by bad actors. This is an important stage at which to be thoughtful.

It is pertinent to note that the variety of state and non-state researchers and investors cross-sector developing technologies (from cloud storage, to aerial mapping) are not protected by commercial or national security interests. This is because, according to the Tshwane Principles[77], which balances the public right to information against the need to protect people from national security threats, the interests of human rights and the transparency of data / information takes precedence. The Principles therefore mean that even if one were to agree that current IHL, sharing of best practice and Article 36 Weapons Reviews are enough to control the development and use of AWS, some data sharing for certain parameters might be required. Some states would undoubtedly find this a disturbing concession when it comes to best practise, but it would be a necessary standard when it comes to industry especially and a particular public right in terms of data-scraping technologies. It would be hard to imagine either party happily publicising their data.

Finally, at some point the rewards of implementing any new technology must out number those of the one preceding it, even beyond the obvious ethical considerations. The cost and upkeep (which would require interaction itself) is rarely considered. One assumes that any new technology is used because the advantages outweigh the risks. However, it is not discussed how and how often, nor the costs or risks involved, of fixing and refuelling different machines and their parts, especially those that may be struggling and incommunicado. AWS may require a number of different expertise from mechanical to software engineers and not all at the same time to maintain the many errors or breakdowns and services that may be required. For a live battle-field situation, this could be a complicated and dangerous affair; several people interacting with one system.
Far more important, is the cost of lowering the threshold for war. While we fundamentally contend the moral and technical viability, the very argument presented that fewer soldiers and civilians would be harmed, would not only make the decision to go to war uncomfortably easier, but escalate an arms race and the asymmetry of power that would follow. This could lead technologically disadvantaged states to proliferate out dated and illegal weapons, weapons of mass destruction, or resort to more pernicious methods of national infrastructure violation like cyber and financial warfare.

[77] Global Principles On National Security And The Right To Information "The Tshwane Principles"finalized in Tshwane, South Africa issued on 12 June 2013
https://emea01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.justiceinitiative.org%2Fuploads

1/ RESEARCH AND DEVELOPMENT
EARLY RESEARCH AND DEPLOYMENT / PROGRAMME AND PROJECT MANAGEMENT / REQUIREMENTS DEFINITIONS / DETAILED SYSTEM DESIGN

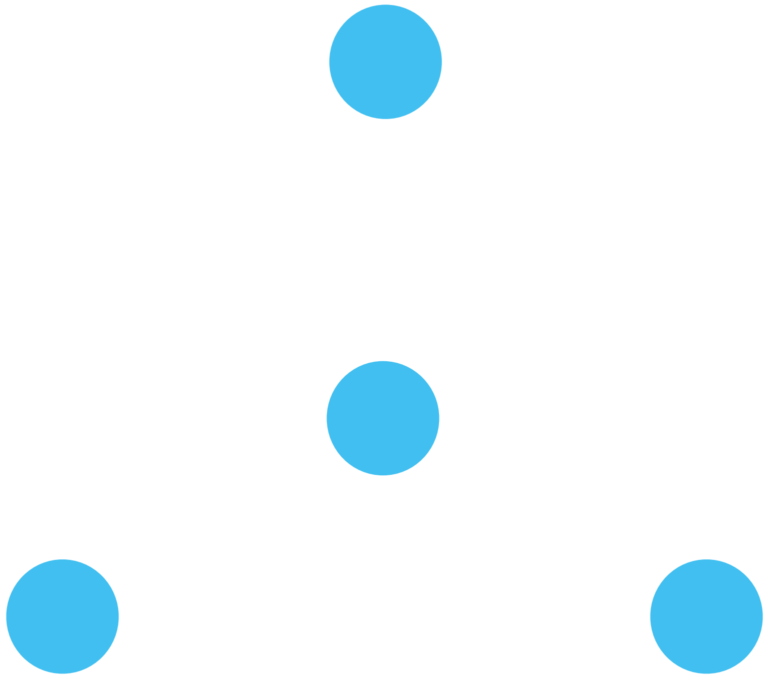
In order to establish an understanding of the issues we must consider the way programmes are built, developed, updated, labelled and self-determine by responding to the real-time present scenarios they are faced with.

ML, AI AND THE ''INTELLIGENCE'' QUOTIENT

Quite different from other weapons' architecture, AWS' operation is based upon extraordinarily complex and unpredictable machine learning routines. AWS are defined as robotic weapons that have the ability to sense and act unilaterally depending on how they are programmed. Such human-out-of-the-loop platforms will be capable of selecting targets and delivering lethality without any human interaction. This weapon technology may still be in its infancy, but both semi-autonomous and other pre-cursor systems are already in service. There are several drivers to a move from merely automatic weapons to fully autonomous weapons which are able to engage a target based solely upon algorithm-based decision-making. This requires material step-change in both hardware and software and, once deployed, posits a significant change in how humans wage war. But complex technical difficulties must first be overcome if this new independent and self-learning weapon category can legally be deployed on the battlefield.

The issue for AWS is that machines are enduringly incapable of matching the human brain upon which it is modelled.

Rather, the autonomous weapon must make statistical approximations that are based upon learned patterns (the training set) that are forever being fed by its own data-polling. Sensed data (once labelled and categorised) is then statistically compared to representations of available training data in order to produce accurate predictions of outcomes within a previously configured set of thresholds.



It is therefore useful to consider how statistical tools might apply in a battlefield context. Using the example of a neurual network here, within the weapon's network, it is envisaged that each neural unit be connected with innumerable others with such statistical links either having an enforcing or an inhibitory effect on the activation state of the weapon's neural units. In this manner, each individual neural unit will have a broad summation function, a threshold function or a limiting function on each connection and on the unit itself. In this way, a battlefield signal must exceed a limit that, in theory, has been defined by the Delivery Cohort before being able to propagate on to other neurons. **It is thus envisaged that the autonomous weapon system will become 'trained' rather than being explicitly programmed.**

A constraint, however, is that the autonomous weapon must minimally have its architecture fixed before training starts. In other words, training cannot subsequently improve the weapon's architecture.

The training set, after all, is a known case. After a sufficient number of practise iterations, it is expected that the computer will be able to reconcile present case sensed data to the training set, the known case. In this manner, the scenario becomes filtered, matched and turned into information (language) and/or action. The architectural model in this case is for such iteration to continue repeatedly until the weapons network's weights find the global minimum of the error function averaged over all of that training data.

Booch, chief scientist on IBM's Watson programme, concludes that such "reasoning and learning are the litmus test to defining an AI[78]." Autonomous weapons' reasoning capabilities must therefore encompass not less than an understanding of a known case whose relationships can then be carried over to the present case. In considering this issue, practitioners argue that two-plus-two bananas should be analogous to two-plus-two apples. The nub of this paper and supporting graphics is to evidence why this will be particularly complicated for AWS routines.

Currently, system routines in machines are 'handcrafted' whereby human programmers are responsible for defining tasking and the way that solutions are to executed. Human beings decide the labels and set perameters.

It may be that many human beings decide labels differently across components, even across companies, which could eventually create the one weapon system. The labels, then, are neither consistent, nor reliable, failing to offer what industry calls ''ground truth''; the technology is subject to the fallibilities of the human building and training it, much to contradict the cleanliness and precision it is assumed it will create. AI, as a science, is a definitively conservative technology- its seeks to preserve the conditions of past into future, and the imperfections therein.

[78] Grady Booch, Chief Scientist, IBM Watson/M, Department of Embodied Cognition, RUSI/Institute for Life Conference Collaboration, in conversation with the author, 8 November 2017

The issue for LAWS deployment is that autonomy's inherent limitations are only revealed as an environment become too complex to be captured in such models' programming. It is for this reason that the those supporting such weapons' deployment must then rely upon ML to underpin that deployment. Humans, after all, are evolutionarily capable of reasoning when available information is imperfect, formulating deductions that are based on knowledge that is 'generally true'. This is not the case in machine coding. In the case of AWS, working with incomplete information will at best require cascading through multiple routines in parallel (each with their own filters, error bias screens, weightings and confidence predictions). Weapon outcomes must forever be inappropriately transient as weapon sensors contribute new data to refine previously available information (without necessarily contradicting it). AWS' operations, moreover, must facilitate counterintuitive capabilities such as detection of contradictions, evaluation of significance and, complicatedly, the efficient rejection of those alternatives that leave the weapon with foreseen unsatisfactory outcomes.

Such learning instability is therefore a characteristic of autonomous machines and, in a battlefield setting, is further compromised by data noise. It is well understood that minute distortion in AWS' classification of its sensed data will likely lead to different data classes being inseparable in the space where such variables are processed.

In other words, if a weapon's dataset is noisy, the class boundary that separates different class examples is almost impossible for the weapon to define and separate for ongoing statistical analysis.

A trained model which is deployed complete, without updates (and therefore expected to work first time, every time), would be problematic due to its lack of correction. Meanwhile on-board ML which could update according to data input, would be in constant flux, independently self-modifying, similarly without review and differently problematic. This is where we need to understand reinforced and unsupervised learning.

MYTH BUSTING

The field of Deep Learning (DL) provides ML with some undoubtedly impressive seeming advances due to extra layers and complexity, making it appear to some extent improved and highly functioning. As such, DL has created buzz, for example around its capacity to enable performance of tasks unsupervised, creating somewhat of an unbalanced religiosity around the capacities of this processing function's ability. However, its capacity is invariably limited by the parameters of its data sets, as well as its inability to code infinitely. That is to say, it cannot be coded for every eventuality, (particularly in real-time), nor can its pre-deployment sets remain immune from the molestation of its value setting and weightings by real world scenarios, this is termed ''generalisation''. In fact, in its purest sense, some researchers believe "deep learning may well be approaching a wall."[79]

Let us broadly distinguish between the terms **reinforcement, supervised and unsupervised** learning, all of which have their inherent flaws in the battle field, for the sake of clarity and to remove a potentially false hope.

Reinforcement learning; whereby the algorithm is given parameters to experiment with, finding and retaining responses according to feedback (accurate or inaccurate). This method is quite dangerous because it can easily retain errors and the opacity of the language means we cannot tell what aggregation of data it is associating with the object. Interestingly, the term comes from the area of behavioural psychology relating to environmental interaction;

Supervised Learning; human beings annotate data creating labels, using the methods of Classification, (training from limited data) and Regression (training data from data which has a set of variable outputs). This essentially develops a function by learning to ''predict'' outputs based on previous labelled examples; i.e the algorithm learns to mimic a pattern the human teaches it.
This is where the majority of deep learning would apply;

Unsupervised learning, which continues learning with minimal direct human involvement in the field is deeply problematic. The task is more akin to pattern finding in data. It does this by; reducing data from higher to lower dimensions, 'Dimensional Reduction', looking for clusters of similar data, 'Clustering', and 'Density Estimation' which uses estimated parameters for the distribution/density of data. Here, the machine looks for labels in previously unknown data, deciding what is to be deemed useful or not.

These terms can be easily confused. The definitions are not straightforward as there can be shades of overlap in each. Reinforcement learning, for example, which essentially 'crafts' experiments on its own, can be trained and deployed as a static model or learn in the wild. In some cases, it can be useful to simply consider ''continuous learning'' as a feature applicable to any form of AI, which is not fully contained under these three categories.

However, it is pertinent to emphasise clarification of another potential misconception.
A closed, or static, system would be pre-trained, deployed, and then not updated. It would still respond to its surroundings according to its training data, with potential problems. It would, be deeply naive and misleading, to think that such a machine would be sufficiently more predictable. (Loosely, one could say **that its behaviour reacts to its environment only in a way which validates its hypothesis)**. We will argue that due to all sorts of limitations of the system, like data profusion and others, this could not be predictable. For example, it would be a complete illusion that, were the technology limited to some replica of a militarily sanitised, unpopulated setting, potentially finding a real life lab-equivalent-ideal in which a machine could mirror its data set, that the data processing and associated behaviour would remain without noise corruption, for reasons we will explore. Moreover, it would be very dangerous and hugely problematic to try and imply that if there were limitations of use to unpopulated areas, then technology could operate better or legally, potentially shifting the responsibility to civilians to keep themselves safe from state actors, and not the other way around as the law demands.

[79] P3. Gary Marcus, Deep Learning: A Critical Appraisal Gary Marcus, New York University, https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf

Secondly, while we emphasise that the autonomous selection and targeting of humans must be prohibited, the technology which signifies the presence of human beings is changeable, can be abstractly context-based (presence of a mobile phone signifies presence of a human being, for example), does not account for civilian versus combatant judgements, and while a machine does not qualitatively comprehend the difference between civilian infrastructure or military, the technology used to isolate these is inherently the same. It is only meaningful human control, which can oversee, intervene, decide, engage and abort machine surveillance and use of force, which can make such calculations. Further, the very training of any pre-deployment set can fail and the final machine can be wholly suboptimal because of the nature of its settings (weighting, values and so on). The real world scenarios can prove to be far more confusing to anything remotely beyond its boundaries as we will see. Here are a few problematic fundaments of the unsupervised model misrepresented in the public sphere;

Deep Learning facilitated Machine Learning and other AI architecture do have some impressive PR. However, whether within the computer habitat of various gaming exercises, or in stunt-like exhibition exercises like the AI vs aircraft DARPA-TopCatchallenge, some of the seemingly impressive results obscure some fundamental issues. The celebration of these belie a fundamental misunderstanding of the technology on display, the mechanisms, the maths, the problems therein, or a willing attempt to distract from these issues.

Highly limited rules and controlled exercises do not allow for the very surprise-scenarios warned against, nor can computers, evidently, process them. The TopCat example, while high speed and accurate in some specific areas, had none of the noise of war or surprises which could throw an ML system into an algorithmic or labelling spin. There was a notable lack of healthy criticism of the exercise, particularly where such cost and force may be applied. No scepticism was was given to the lack of real-world dissembling and other issues that would, in reality, plague this pristine display. In a real world dog-fight, with weather, behaviour, the unpredictability of the programmes or even manned aircraft, there has always been an issue of surprise, a defining element of tactical warfare. This evidenced no capacity to deal with 'feint', a sudden unexpected change in direction intended to confuse, for example. 'Spoofing' (confusing) limited data sets, as we will discuss, could be a remarkably simple affair. **The simple act of setting off a flare behind heat seeking missiles has historically had around a 90% spoofing success rate.** There are only so many situations a machine can be programmed to expect.

It is frequently asserted, and often conflated, that increased processing, or a high capacity for learning (reliably repeating a task with a degree of accuracy, even finding patterns giving the impression that meaning has been created) is the same as cognitively understanding the task. A gaming exercise, for example DeepMind's Breakout system, presented the false notion that the system had 'figured out' a way of beating the system by tunnelling through a wall, but this too had been learned, not cognitively understood. This was little more than an exercise in object recognition and pattern finding, attaching mathematical modelling to joystick moves and an elaborate mathematical from of trial and error.

The programme had not inferred through an act of creative thinking, or a qualitative understanding of walls and tunnels. The system had a limited set of clearly definable rules, entirely different from human behaviour. In this case, there were finite moves before a possible win. When simple individual changes were thrown at the more superior Antari system, for instance altering a Y coordinate relating to height, or the addition of a wall (usefully representative of a physically manifest change that might appear in real-world scenarios), it failed consistently.[80]

Apply such academic failures from such minimal changes to the TopCat exhibition, with the force and speed behind them, without human intervention, and the results could be utterly devastating. Nor does the aircraft have the opportunity for live trial and error. It is, rather, trial by error. Similarly, abstract contextual cognition, associative heuristics and appropriate deduction, eg. the presence of a pram or a toy possibly indicating the presence of children, is beyond a programme's calculation. Certain situational awareness is experience-based, infinite and may be impossible to programme.

Upon data gathering, **weighting is required to allocate importance to certain sensed information and this weighting must be learned.** To reorder this weighting in live situations with competing information and therefore responsive manoeuvring of priorities, would be unpredictable and questionable at least, as it would be impossible to code for all the possible scenarios encountered. (Moreover, to process this adequate "infinity" in realtime, even if conceivably possible, would surely take the kind of processing time and power that would negate the utility of the machine). An example would be coding for action in robotics, ie. the physical response to sensed data.

The difficulty here is that the weapon's confidence weightings are themselves time dependent and this imprecise measure must influence which rule the AWS fires first and therefore the output response.

A further challenge then arises from the several measures of quantity with which the weapons' key data inputs are measured (the weapon's voltages, temperatures, flow rates and so on) requiring real-time translation and additional 'learning latitude'. If a requisite threshold is programmed to correspond with specific action response and several were to occur simultaneously, how would the AWS prioritise, and how changeably?

For rudimentary example; a perceived 'threat' might trigger the action response of retreat from, let us say, a flash of light; an opening door could trigger advance/forwards incrementally, 'curiosity,' while silence could be perceived as no threat, 'proceed'; (which a human on the ground could equally read as a threat and choose to retreat, in certain contexts). If all three were to happen at the same time, how would a programme respond? This could freeze/stun the AWS. Hierarchicalising labelling as well as sequencing, thus, have different repercussions, as do the actions associated with the ''information'' derived from them. Moreover, the very presence of those three instances in situ, change and inform the meaning of each other, they both give and derive context and meaning. Silence has one value in the middle of night, and another following a relentless bombardment during the day. ML based AI, in its complexity, is neither intuitive nor ''intelligent,'' it is comparatively dumb, but possibly quick.

[80] Kansky, K., Silver, T., Mély, D. A., Eldawy, M., Lázaro-Gredilla, M., Lou, X. et al. (2017). Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics. arXlv, cs.AI

2/ TESTING & EVALUATION, REGULATION, CERTIFICATION
TEST AND EVALUATION AND ACCEPTANCE / REGULATION AND CERTIFICATION

THE TECHNICAL-DEBT COLLECTOR

Machine Learning is "the high-interest credit card of technical debt".[81] The incrementality of the systems is problematic, while the predictable performance of simpler component parts in traditional engineering is more reliable it is still prone to issues. Complexity as a modular component of complexity leads to greater issues, as well as serious problems relating to replicability.[82]

Consider some of the features providing opacity, predictability and consistency problems. In classical computer programming, code language reads instructively. Conversely, the labelling of deep learning AI, does not appear visibly readable in a decipherable manner. At best, it is a case of visibly showing nodes' geographical location within a potential neural network. The most common language of DL, recurrent neural networks (RNNs), superficially, is somewhat linear as opposed to the hierarchical structure of natural language. Largely, the systems are beyond visibility which, in itself, presents huge issues for recording or understanding issues of bias[83], as well as correcting or accounting for issues regarding to potential violations of law. How can one correct errors that cannot truly be seen? Meanwhile the lack of ability of AI to link causality makes it very difficult to apply the kind of differentiation context which much of international humanitarian law is centred around, particularly when, but not limited to, deciphering between combatants, civilians, injured persons and so on, (Articles, 3, 48, 57, 85, Geneva Convention Protocol 1). This makes transparency and ``black box´´ issues some-what of a conundrum without a meaningful human chain of command. Remember our different kinds of learning, we cannot know if accurate means correct in some cases, or if retained information was correctly so.

Basic functional problems which present as promisingly solvable in basic lab settings, like local minima (trapped in a suboptimal equilibrium[84]), prove much more problematic and complex when it comes to the wider scale problems of field interaction of noise. As we know, we do not have infinite data, nor infinite representations of objects, nor the infinite possible encounters with those objects or their movements, contexts and interactions with the environment.
Therefore, simulations in isolation or with limited modelling will always be insufficient.

This leads to "filling in" of missing or partial data, or disappearing that which has incomplete data all together as anomalous, irrelevant (below threshold or beyond boundary), which in potential conflict contexts is alarming. For a commander, this would mean missing potential targets. More fundamentally, however, this means that civilians, wounded combatants (protected under Article 3) and any civilian infrastructure could be missed all together, because of sub-par technical modelling, or rather the nature of modelling itself.

Both the limiting nature of Supervised learning and the reductive pattern searching of Unsupervised learning, present obvious flaws in their narrow reach.

[81] Sculley, D., Phillips, T., Ebner, D., Chaudhary, V., & Young, M. (2014). Machine learning: The high-interest credit card of technical debt. Proceedings from SE4ML: Software Engineering for Machine Learning (NIPS 2014 Workshop). Google 2014

[82] Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2017). Deep Reinforcement Learning that Matters. arXiv, cs.LG.

[83] O'Neil, C. (2016). Weapons of math destruction : how big data increases inequality and threatens democracy

[84] p.5 Gary Marcus, Deep Learning: A Critical Appraisal Gary Marcus, New York University, https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf

Meanwhile, an human being can fill in data gaps (decipher a whole face even if mostly obscured or angled), and make other heuristic inferences due to our sophisticated evolutionary advantage even as babies[85]. If a machine ``sees´´ half a tank, or miss-reads / miss-records the data for person, or child because they are obscured[86] by something unusual or standing near something unexpected (a relational obscurity), or insufficient data was available to train the machine in recognising it, then the lack of data references needed for an accurate reading could render the object or individual invisible all together, or read as something entirely other. [If the machine generalises and makes a best-guess approximation of the image sensed, it either does this by "interpolation between known examples, [or] extrapolation, which requires going beyond a space of known training examples[87]," neither of which are accurate and equally concerning in our context.]

Further, the order of labelling is as significant as the mechanism for it. Abstract associations and inferences beyond specified criteria are difficult. While the human brain is much more advanced, with researchers proving abstraction to be possible even in new borns[88], the criteria decided for the labelling (by the coder) is performed in a certain order. This cascade has its own repercussions both being contingent as well as causal. Sabour and colleagues[89], posit the likelihood of the most common DL neural network architecture causing innate problems due to labelling. When presented with live sensed data, the DL might exhibit;

"Exponential inefficiencies that may lead to their demise. A good candidate is the difficulty that convolutional nets have in generalizing to novel viewpoints... we have to chose between replicating feature detectors on a grid that grows exponentially ... or increasing the size of the labelled training set in a similarly exponential way."

In fact, in order to aggregate information based on explicit definitions, the systems would need to have been "fed" an almost infinite number training sets (or have ongoing access to constant changing data somehow modulating the training sets?). **In reality, computers are better at size and speed, with applicable rules, humans are better at complexity** (Lake, Salakhutdinov, & Tenenbaum, 2015; Lake, Ullman, Tenenbaum, & Gershman, 2016).

If in the context of AWS which continue learning, ML's intended role is to facilitate onboard deduction and independent prediction whereby all immediate experiences should inform the weapons future expectations and update its capacity to predict (compute and respond appropriately to them), **data parameters which provide the first framework are quite definitively, necessarily, predictive and limiting.** They work on the presumption of certain rules which can extrapolate out and evolve based on a tug between this training and sensed data expanding the node positions or other parameters exponentially out, whether accurately or otherwise. In either case, they will always inform the direction of this latter action; learning (if unsupervised), or deduction, if trained then deployed 'as is'.

AWS' mutability will arise from the choosing and modification of description parameters that are used to train the weapon's behaviours ('parameter profusion').

[85] Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. Science, 283(5398) (5398), 77-80

[86] Plamen Angelov and Alessandro Sperduti, 'Challenges in Deep Learning', in ESANN 2016 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (Bruges: ESANN, 2016), pp. 489–91. See also Soroush Nasiriany et al., 'A Comprehensive Guide to Machine Learning', University of California at Berkeley, 18 November 2019, pp. 82–88, ←http://snasiriany.me/files/ml-book.pdf→, accessed 19 June 2020.

[87] Marcus, G. F. (1998a). Rethinking eliminative connectionism. Cogn Psychol, 37(3)(3), 243-282

[88] Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: the newborn brain detects identity relations and sequential position in speech. J Cogn Neurosci, 24(3)(3), 564-574

[89] Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic Routing Between Capsules. arXiv, cs.CV

It is also further problematic that the ML spine of the AWS does not only assess data points and parameters nodes from the external world against its training set. There are further, opaque, 'undeclared consumers'. The outcomes of certain routines can become the kernels (input) to another constituent of that sequence. This not only displaces original kernels, but creates unintended feedback loops between the weapon's algorithm and external world. It makes accurate assessment further difficult and the possibility of the commander to alter firmware almost impossible.
This makes the very spine an erratic operator acting independently within the environment.

Researchers assert that AWS would be 95% deployment stage 'glue code' and 5% executive. This means that there is little room for further fundamental iteration. The more wide the application, the more redundant the glue code could be, or the more incompatible with the small changes. Live changes, which must surely take place instantaneously to be useful or on an ongoing fashion, need to be compatible across the entire system and integrate fully, which will be a hungry affair. How will this updated version be verified 'in field'?

At what stage is a constantly updating weapon, no longer the weapon that was verified for deployment?

We have introduced the idea that AWS architecture, for the sake of discussion, is potentially based on artificial neural networks which learn by wrote. This requires an harmonious and perfectly timed integration of the phases of sensing, labelling, classification and action allocation.
The vulnerability of this superficial structure can fracture the integrity of the whole system with a domino effect, with an error in one component it effects the function of the others. We already see issues in the testing and verification phase which overlap with deployment and operational command. These must be explored further.

3/ DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING. TRAINING / RULES OF ENGAGEMENT / OPERATIONS PLANNING / DEPLOYMENT TO OPERATIONAL THEATRE

CODES OF CONDUCT

Consider Phase 1 of the Joint Targeting Cycle; Commanders Intent, Objectives and Guidance, linked to the political, strategic and technological directives, deployment and operational management as equally linked to Phase 0/ National Policies, Political Control and international law. **A field commander needs to translate their intent and the vision of their battle strategy to their troops, with a degree of predictability.** This requires behavioural compliance, relatability and reliability, hence the heavily indoctrinated codes of conduct. The manner of command is for specific enterprises to create an overall goal rather than a whole field operation. Feedback and mitigation, or anticipation, of issues on the ground, (civilians, things in the way) are imperative. This requires constant interaction with people and systems on the ground, as well as feedback from exercises which did or did not work. The same is true of AWS which by their nature are independently reacting to their surroundings. Trusting the reliability of performance is an issue which starts at development stage, in the misleadingly clean environment of high-performance. **Orders that will be predictably followed and have feedback in a clear chain of command are essential for strategy as well as accountability in the field.**

More realistically, a problem for the commander upon deployment is that a system cannot be an unknown entity, whether for data collection only or, especially, engaging in attack, it must work first time and every time. The prospect of ''learning on the job'' or improving over time where lethal and legal mistakes could occur exponentially could be catastrophic and untenable. Remember our TopCat experiment. This does not work as a technical spine.

There are several components which link issues of ML training and its difficulties with predictable deployment. These speak to the very proposition of "Modification to systems to comply with Theatre Entry Standard (TES)"[90] proposed as key requisites by the government as not merely necessary for deployment to theatre, but possibly requiring adjustment between the assessment and deployment stages. In the following section we shall lay out some of the ways in which intrinsic features of ML training and computer kernel iterations upon deployment challenge and obstruct the faithful or even, in some cases, plausible, manifestation of this mechanistic, behavioural and procedural intent.

Weapon actions must comprise the appropriate reaction to every relevant sensed stimulus. AWS cannot offer intermittent or erratic performance where only specific sensed inputs lead to weapon outputs.

Processes are required to manage challenges around AWS 'attention', prioritising one data string over other sensed information. The 'cocktail party effect' isolates particular data while ignoring seemingly irrelevant information that may be key to that weapon's compliant and useful operation, which a human may not.

[90] P.9 UK Government, "Group of Governmental Experts of the High Contracting Parties to the 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", Second Session
Geneva, 27 - 31 August 2018 Item 6 of the provisional agenda Other matters, 2018.

Conversely, if there is buzz (data extraneous to the parameters) this could be 'tuned out'. That very buzz, however, could be indicative of something significant. Therefore sensitivity is relevant here. This partly relates to data-polling, partly sensitivity and partly to profusion.

After all, if ML reads data constantly, and 9/10 of war is quiet, the 1/10th of relevant (signalling) chaotic noise of theatre might be smoothed out as an anomaly, into formlessness; irrelevance.

This relates directly to all the Joint Targeting Cycle phases and the following specific subsets of the Lifecycle of a Weapon System; Phase 1- Requirements Definitions, Phase 3- Targeting decisions and activities / Battlespace management and Phase 5- Lessons learned / In-service feedback. Moreover, it has the obvious aforementioned legal ramifications if potentially significant data is disappeared.

Also related, is the previously mentioned 'undeclared consumer'. This is the unmonitorable issue where the computer generated output of one process will consume into the required inputs for the next process, and if this is erroneous then the next one is so biased. **This creates an opaque and problematic feedback loop, whereby the computer kernels as well as the sensed data become influencing data input. This necessitates constant human supervision, to tweak and guide alterations in order to keep the updated version aligned with the commander's intent and specific targets, or it could degrade accuracy.** This relates to Intent, Targeting Decisions and Activities, Battlespace Management and verification. All these require intervention and transparency.

Let us see where this fits in with the most significant phase of the Lifecyle; Phase 4- Use & Abort, Targeting decisions and activities / Battlespace management, and of the Nato Targeting Cycle; Phase 2- Target Development and Phase 5- Mission Planning and Force Execution, the latter of which the UK paper reiterates can have elements of autonomy.

4/ USE & ABORT
TARGETING DECISIONS AND ACTIVITIES / BATTLESPACE MANAGEMENT

WHO'S LINE IS IT ANYWAY?

This paper asserts that coding is systemically ill-equipped to deal with ambiguity, context or situational awareness. It fails to compile language in the way that a human brain can, nor can it "see" and ascribe meaning (value) in any conceptual or qualitative sense. The AWS' requirement to manage a visual sensor reduces visual data into stable descriptions. Visual reconstruction is a complex local problem. The first challenge is data's dynamic classification into continuous regions and discontinuous boundaries. AWS hardware must also detect objects regardless of the weapon's environment, target appearance, target position and motion pattern. Hardware must determine the weapon's position in relation to that environment, a complex requirement given volatile direction-of-gaze and the AWS' fluctuating line of sight. All the while, the AWS' hardware must deal consistently with data disorder and noise.

Automatic Target Recognition (ATR), for example, requires the successful integration of complex components such as, but not limited to; structured-light 3D scanning, thermography optics, LIDAR scanning, MRI scanning, hyper-spectral imaging, radar and synthetic aperture sonar. The performance of these at a supra-systemic level will be programmed at the integration / design level and certain compromising factors must be weighed at specification. **Data-polling rates, parameters, threshold levels for accuracy as well as force application, sequencing decisions and integration of data into a whole must be programmed by definitive methods of 'information unification'. Adding to this the fact that the data may still be compromised by the varying quality of uptake and, therefore, condition of target objects, makes components of this integrated system turgid in their processing ability.**

A more-is-more, approach to hardware is not necessarily a superior approach to ATR. We have described the different methods of human cognition versus computers, and the corrupting nature of data processing in its necessary aggregation of data to 'smoothing' out of a statistical line for pattern finding. A soldier is trained to make targeting decisions (carrying out a commander's intent) using abstractions created by context and experience (both legal and historically based judgements) as well as following specific orders. Intrinsically here, the considerable 'technical debt' inherent within machine vision processes, from 'infobesity' to 'data smog' corrodes the ability to create a clear visual edge by its data points, mapping transposal and action translation creating faithful follow-through of intent.

Rather it is likely to be intoxicated by incoming information and overwhelmed by its own anomalies (or inability to process) and exponential likelihood of misinterpreting sensed data. More tech will not mean more accuracy.

LINES OF SITE

It is important to remember that the function of the system is to aggregate new data in order to categorise it as closely to learned sets and, with unsupervised learning, update these sets (assimilate the new information) in situ looking for its own labels. This act of statistical analytics and ordered labelling is an assumption in prescribed sequences. Interruptions of the sequence or deviations from the presumed pattern make mapping the real world onto the learned cartography of the world hugely problematic. In his excellent essay evaluating the superficiality of the supposedly super abilities of deep learning networks, Marcus explains his experiment using simple binary numbers, odds and evens, and training sets with wide parameters. They performed well within those parameters; "but they could not extrapolate beyond that training space[91]." The manner, and often cluttered, urban nature of conflict, exists very much beyond the training space.

More specifically in the world of ATR, the very process of matching could be frustrated by the simple passing of time, movement of the sensor or the object of interest. **This 'correspondence problem' between edge-detected images and memory based models is a recurring one of noise confusion and matching confusion.** The establishment of the sharp edges and pixel-definition associated with light capture, for example, can be easily confused by equally stark changes in light, shadows across the sensor or movements and the complexity of matching itself. This issue of **'Class boundary'** within ATR is significant. Once the data is partially obfuscated i.e whatever specific correlation the machine has come to recognise as the particular object – length, shape etc. with some mathematical room for variation- is obfuscated either by another object, a tree, a building, people, weather, dust; the data noise, (data smog) one would expect, or if the sensor is blocked; the extraneous pixels may be ignored or the partialised object detected unfamiliar rendering the image indistinct to the extent that the computer no longer understands the object at all. This could be similarly true for reflectants and an anomalous recognition of light. **ML is utterly reliant on class boundaries in its training sets in order to match clusters of data to these sites. A key weakness for the ML model is captured in the acronym CACE ('change anything, change everything'); bearing in mind, it only takes a change of half one percent of a photo's pixels in order to turn a general into a cucumber.** This makes Phase 3, Capabilities Analysis almost impossible to verify stably, as its reliability in the face of noise will be untested and its 'accurate' readings unknowable giving false positives.

PINK IS THE NEW GREEN

In times of conflict those on the ground will be sensitive to irregularities rather than scanning for comfortable patterns. As discussed, instances of extraneous/ anomalous data can be dangerously smoothed out or disappeared. Remembering that war is one tenth chaos, a system trained for patterns, would not do well with tailored surprise or 'feint'. Using an example from the commercial sector, Google automated fine predictives based on flu data trends, but could not account for the (rather more relevant) spike in flu season[92]. This failure was delivered by some of the most advanced data processors, (currently working with the Department of Defence on AI and surveillance processing), rather dampening the claims that automation could improve accuracy and enhance the human life protections entrenched in humanitarian law. On the contrary, they are a window into the fundamental problem with AI's intrinsic inability to cope with the precise anomalies that define warfare. It is exactly the unusual behaviours, the light reflecting off of a piece of metal over the horizon, (our reflectant example), a partly obscured or usually green tank painted pink (camouflage in an AI world of warfare), which should be alarming.

[91] Marcus. G, Deep Learning: A Critical Appraisal New York University https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf#page=5&zoom=auto,-199,146

[92] (Lazer, Kennedy, King, & Vespignani, 2014)

While easy spoofing could be an advantage to the spoofer[93], it has far more pernicious and complex ramifications for how widely AI can be fooled, mis-target or miss targets. If 3D printed turtles can be misread as rifles[94] and stripes as school buses[95]; in the field of war, civilians and infrastructure can be struck while AI can easily be misled and courted to fratricide. It gives the AWS and AI camouflage[96] much more dangerous repercussions, making harms and humans, invisible in plane sight.

MURDER BY NUMBERS

The partial data and iterative learning leading both to potential confirmation bias97 and the problematic confirmation of erroneous classifications due to additionally erroneous weighting or poor aproximations (either from restrictive base rates or too widely encompassing parameters), could lead to false identifications in the field. The nature of ML design also infers that the known state of the AWS and the desired state are consistent. This makes transfer and ongoing learning a complicating interruption in its development between unlearned and learned states, from old to new environments, as does the expected immutability of certain target profiles / objects, in a changeable world.

In this sense, there is an inherent contradiction between the AWS need to be front-facing and determine its system state ahead of time, and its ability to be respond to changing environments.

The incremental nature of the AWS' collection of data points makes the application of classifications and response fundamentally one associated with timing and sequencing. This is directly significant to the link between identification and action (including potential use of force). In addition, in its acclimation to its environment and prioritising, the labelling, sequencing and attribution of action will apply to scenic, state and event assessment. This dynamic data management will require live response to located prioritised areas, as well as responsiveness to emergent activities, each requiring follow-up, check-in and the technological adjustment throughout the weapon's system in the component parts and as a whole. **Part of the issue will be the difficulty in assigning threat level or likely intent behind any object or scenario.** Meanwhile, the nearby associative data which can confuse or delete the semantic labelling of the object, 'the object recognition conundrum' can be a key problem in target recognition or battlefield assessment.

[93] Reim. G, US Air Force grapples with vexing problem of AI spoofing, September 2020 https://www.flightglobal.com/defence/us-air-force-grapples-with-vexing-problem-of-ai-spoofing/139973.article

[94] (Athalye, Engstrom, Ilyas, & Kwok, 2017)

[95] (Nguyen, Yosinski, & Clune, 2014)

[96] Plamen Angelov and Alessandro Sperduti, 'Challenges in Deep Learning', in ESANN 2016 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (Bruges: ESANN, 2016), pp. 489–91. See also Soroush Nasiriany et al., 'A Comprehensive Guide to Machine Learning', University of California at Berkeley, 18 November 2019, pp. 82–88, ←http://snasiriany.me/files/ml-book.pdf→, accessed 19 June 2020.

[97] Amos Tversky and Daniel Kahneman, 'Judgement Under Uncertainty: Heuristics and Biases', Science (Vol. 185, No. 4157, September 1974).

Sequencing is partly affected by the very modular nature of the system itself and the independent agency nature of the contributing components. Different ``sub-agents''[98] which contribute to the collection, selection, aggregation and translation of data into information and then associate into action, (for example allocation of threat, classification, target signature, cartography and readability of field clutter) reduce data which falls beyond boundaries to a mathematical mean. This smoothing disappears the most vital elements which lie beyond a computer's qualitative cognisance, heuristic experience and the modification of supervision. It is important to note that the sequencing of a routine can deliver utterly different outcomes. In parallel, the extent to which new data is programmed to influence the changes in the data set or to what extent the programme is to wed itself to the learned programme in its incremental versions (anchoring) will effect the extent to which it evolves in the face of its external factors and/or the potential for variability in both accuracy and sensitivity to the relevance of sensed data.

When all data is collected and ordered into final action, and thresholds assigned to decide on that action, how must proportionality be assessed mathematically in order to assign weight to input data? The extent to which, as explained, the simple sequencing order can apply widely variant outcomes in both intent and accuracy leaves worryingly arbitrary output possibilities beyond the ethical concerns of such a task.

When computer kernels, eliminating possible crucial data, sense for the risks and gains of the use of force, who and what decides the weight of risk and gain?

What mathematical weight would be applied to the cost of lives and infrastructure for civilians versus combatants?

Would we be in the potentially alarming world where programmers assign literal values to human life and military gain, mathematically calculating collateral damage, blaming error codes and the infinite possibilities of independent sequencing trial and error for possible outcomes in the field

(where abstracts and unknowns definitively cannot have been accounted for or pre-learned) and any excess harm? In this sense, goals and values could be both subjective, cultural, nefarious, opaque, shifting, anachronistic, computer and human bias generated, down to the labelling and sequencing set by the coders and then disrupted by the independent cascade of the machine's reactivity and decay.

Furthermore, the error-correction of the systems, (the observed difference between current and desired state[99] would be largely dependent on speed of computation and feedback loops. Firstly, feedback loops are distinctly less effective at high-level firing sequences. Secondly, choices must be made about the rate and frequency of the readings taken by the system (data-polling) to have an up to date reading of its surrounding, which has its own limitations. This updating of information may happen by the second, hourly, daily and change the information traceability and factorabilty; plus ability to process, and equally degrade data. Moreover, a notion that multidirectional intel may be applied by some state actors for corroboration is of little comfort if other states do not engage similarly, or if the tools for this nexus of confirmed intelligence is provided by equally flawed technology. These questions lead us to stages 4 and 5 of the Targeting Cycle.

[98] p8 Leadership Challenges from the Deployment of Lethal Autonomous Weapon Systems: How Erosion of Human Supervision Over Lethal Engagement Will Impact How Commanders Exercise Leadership | RUSI

[99] Douglas Allchin, 'Error Types', Perspectives on Science (Vol. 9, No. 1, March 2001), pp. 38–59

NATO TARGETING CYCLE:

4 COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT
5 MISSION PLANNING AND FORCE EXECUTION

It is crucial to remember that it is within Phase 5 of the Joint Targeting Cycle, Mission Planning and Force Execution that the NATO paper suggested pure autonomy could exist. But Phase 5 is somewhat directed by Phase 4- Commanders Decision, Force Planning and Assignment and therefore are considered together in terms of technological practicalities and implications. A commander needs to know how up to date the data they receive is. Too little and its stale, out of date data- high frequency data polling means too much to assess relevance. The command has to trust the weight (relatability) and the equivalency level of the present case to the training set. They must be close enough to make an adequate decision. Both of these command, target and infield management (including abort) issues straddle Phases 4 and 5 of the Lifecycle, while having both development, verification and post assessment ramifications.

In their independence, AWS have to be attributed a process to manage and govern Goals and Value settings.

Goals prompt an independent weapon to develop plans of action while values enable it to assess the comparative merits of such plans. If this process is stunted or inappropriately undertaken, the weapon will either be illegal or useless.

Goals concern what must be undertaken at once, what should be undertaken next, the resumption of a task that was previously discontinued and, more complex for the weapon, what actions should subsequently take place in order to capitalize on battlefield opportunities. None of these factor the ethical weighting.

We know from the nature of ML architecture that in the evolution of its programme it retains and off-loads information. This problem of 'habituation'- i.e what a programme forgets and what it retains, presents a widening gyre error issue. **If it retains a mistake but forgets valuable information that is problematic.** As the system is opaque, we can not see its decision making process, nor can we 'see' how it comes to learn, by what association of data points does it come to associate mathematical patterns to create its labels and class boundaries as previously suggested. Its mistakes can be beyond reach until manifest as exterior actions. It may start to correctly identify 'tree' but by a completely unknown statistically associated definition. As such, if this starts to mis-associate these boundaries or disassociate them with 'tree', encompassing other objects, we cannot correct the appropriate issue. Meanwhile, in its unsupervised state, if it retains the exact edges which are creating the errors but relinquishes the most 'tree' correlating aspects of its class matching, this leads to exponential errors or **"Catastrophic forgetting".**[100]

[100] Catastrophic Forgetting Still a Problem for DNNs https://deepai.org/publication/catastrophic-forgetting-still-a-problem-for-dnns

Factors arising from International Humanitarian Law (IHL) involve evaluative and contextual judgement for which the local commander must remain responsible and accountable. Sufficient control must therefore be retained to enable situation-specific judgement to comply with those rules[101]. If the commander (as well as the programmer) cannot reliably know what the machine is responding to, then how can this be assured? Moreover, how can all those involved in the decision making be held accountable to the IHL / LOAC, Law of Armed Conflict, rules on proportionality, precautions-in-attack and distinction? Only human assessment and critical control can assure oversight and associated accountability.

The ability to ensure protections, proportionality and known outcomes requires precision, predicability, feedback, assessment and reliable data. The conundrum of data-polling frequency presents a problem here.

Too little and the data is either sparse or stale, too much and it is too complex to decipher while demonstrably leading to diminishing returns in efficacy; increasing sensor feedback affects incrementally smaller weighting variation and a less dynamically responsive instrument[102].
This further complicates the proposition that multiple instruments will decrease an error potential.

The 2018 UK paper optimistically states that, "ROE [rules of engagement] will be tailored to the specific mission and operational environment and will take into account national and international law[103]." The survey of operational environment, OE, is something which has been explained in previous sections to be corruptible in the data set translation, both externally and within the kernels themselves. **In practicable terms, the iterations of an evolving programme in response to external and learned parameters can both retain and forget valuable or problematic data. Highly processed errors can occur without oversight or flagging.** Besides the issues of false positives and misleading lab testing, (undermining Article 36 and procurement testing confidence) without the noise and feint issues associated with war, the OE itself manifests differently in terms of ML. The computer of the AWS, with the various issues of unpredictability and more complex issues set out in this paper, becomes an OE in and of itself, increasingly so when differently trained systems meet and respond to each other.

More specifically in this case, as listed in detail below, the nature of ML interaction with external factors and partial data (eg CACE) renders the learned programme and subsequent "independent" programme unreliable and therefore ineffective, thus anathema to the "tailored"[104] behavioural and target-intent requisites for ground command. It is worth noting that the recent Norwegian rules of engagement is more than 1,000 pages. How is such nuance to be programmed and understood? How are the subjective views of states and differing national agendas to be resolved without the specificity coded within an operational normative framework?

The information contained within a command must be coupled to previously given information as well as to information that may follow. AWS must factor for nested structures and conditionals which regularly characterize complex instructions. While it may be human practice to understand what has been directed without having to figure out exactly the meaning of the words, this does not translate across in machine coding. Command and analysis both use several categories of facts within their syntax. **Instances here include indexical facts, normative facts, strong convictions, observations and hints, clarifications, reinforcements as well as basic ontological factual statements. All of these sub-types inform human decision but must now be precisely recognised only in AWS code.**
The challenge is also that such categorizations are volatile and change unpredictably according to new intelligence, new feedback and local input from the weapon's sensors.

[101] Mark A Staal, 'Stress, Cognition and Human Performance: A Literature Review and Conceptual Framework', NASA/TM—2004–212824, NASA STI Programme, August 2004. See also R D Campbell and M Bagshaw, Human Performance and Limitations in Aviation, Revised Edition (Oxford: Blackwell, 2008); Gyanesh Kumar Tiwari, 'Stress and Human Performance', Indo-Indian Journal of Social Science Research (Vol. 7, No. 1, 2011), pp. 40–49.

[102] Xavier Girot and Yoshua Bengio, 'Understanding the Difficulty of Training Deep Feedforward Neural Networks', in Yee Whye Teh and Mike Titterington (eds), 'Volume 9: Proceedings of the 13th International Conference on Artificial Intelligence

[103] P.9 UK Government, "Group of Governmental Experts of the High Contracting Parties to the 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", Second Session
Geneva, 27 - 31 August 2018 Item 6 of the provisional agenda Other matters, 2018.

[104] Ibid P.9 UK Government, "Group of Governmental Experts of the High Contracting Parties to the 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", Second Session, Geneva, 27 - 31 August 2018 Item 6 of the provisional agenda Other matters, 2018.

LIFECYCLE: 5/ POST USE ASSESSMENT
BATTLE DAMAGE ASSESSMENT / LESSONS LEARNED / IN-SERVICE FEEDBACK

NATO TARGETING CYCLE: 6 ASSESSMENT

We have noted repeated flaws within the machinery creating behavioural and legal compliance, reliable outcome and predicability issues as ubiquitous to every stage of the lifecycle of the weapon. None are so acute and so stark as at the nexus of the planning and execution of force (targeting).

The issues which affect class boundaries and profusion can create both under or hyper sensitivity, creating both false detection and false positives.[105]

However, the same issues which this creates for targeting become apparent for assessment, accuracy monitoring, verification and validity of legality during and post use.

It is worth pondering whether a reliable Abort mechanism in a hackable, spoofable world of autonomous software with entirely onboard self-contained software, is a dichotomous complication in and of itself. While entirely necessary to comply with IHL, and the proposed frameworks, they create a fundamentally problematic weakness for a commander. Similarly, somewhere between the Joint Targeting phases of force planning, assignment and force execution; additional routines will be necessary to manage the AWS' failure modes (outright veto, 'fail safe', 'fail too safe', which is a commercial standard, 'fail dangerously' and 'fail deadly') as well as the integration of otherwise independent assets into the commander's wider portfolio of battle plan assets.

The very degradation that would occur between lab-testing and field outcomes potentially resulting in false positives which could report as accurate but in reality be wildly off target and possibly constitute war crimes, are the very intractable nature of the AWS which would make in operation delegation almost impossible. In such instances, the problem, which makes post assessment near futile, is that high accuracy scores in testing could pass threshold tests to satisfy both procurement lawyers as well as opaque Article 36 reviews, and more worryingly, the public's perception of the activities and responsibilities of our military and weapons systems by technically reporting successes which belie the truth on the ground. If this is the case in post-assessment, the same is true of live feedback which influences not only the data received by the AWS for the commander's benefit, but the reporting of the AWS on its own performance, making the Delivery Cohort's supervision of an autonomous system even more difficult. While a remote system may independently monitor its efficacy in terms of targeting 'hits' and sensor based input, the huge demand of the pre-known, ready-mapped model with in field processing of possible paths, prioritised paths and optimal paths, for example, in a fluctuating operational environment and fluctuating operating system, reassessing its goals and values, is a strange one to entrust to the kinds of flaws we have highlighted.

The design of autonomous componentry would have to recognise and manage possible coding errors at inception stage and the challenge of re-factoring code in an AWS that is likely out of communication.

Errors that occur while incommunicado (which inherently they would) could make live feedback, correction and recording, as well as post assessment difficult.

It appears, that for an automated weapon system to stay faithful to commander intent, reliable to initial parameters and without violations of law, it would require much supervision of its autonomy, just to function as a useful piece of operational kit. This is somewhat of a contradiction in terms, and in AWS.

The following table is visually indicative of our concern with AWS' inadequacy mapped across the Lifecycle of a Weapon System and NATO's Joint Targeting Cycle. The ubiquity of this concern is evident in the densely concentrated interactions between technological issues and targeting phases as well as the telling overlap all the way from pre-development (political and international) to post use assessment in its macro-narrative.

[105] Leadership Challenges from the Deployment of Lethal Autonomous Weapon Systems: How Erosion of Human Supervision Over Lethal Engagement Will Impact How Commanders Exercise Leadership | RUSI

OUTSTANDING TECHNICAL ISSUES INTERACTION WITH FRAMEWORK PHASES

	LIFECYCLE OF A WEAPON SYSTEM						NATO ALLIED JOINT TARGETING CYCLE					
AWS' MACHINE LEARNING (ML) SPINE	NATIONAL POLICIES 0	RESEARCH AND DEVELOPMENT 1	TESTING & EVALUATION, REGULATION, CERTIFICATION 2	DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING 3	USE & ABORT 4	POST USE ASSESSMENT 5	COMMANDERS INTENT, OBJECTIVES AND GUIDANCE 1	TARGET DEVELOPMENT 2	CAPABILITIES ANALYSIS 3	COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT 4	MISSION PLANNING AND FORCE EXECUTION 5	ASSESSMENT 6
Weapon architecture based upon neural networks **involves intractable complexity.** AWS' data input requires three characteristics (a weighting or synaptic value, a summing function and a threshold-based output function) but the model's veracity requires that all such **inputs are received free from corruption, noise and arrive in sufficiently full detail.**	X	X	X	X	X	X	X	X	X	X	X	
The weapon's picture-building requires that **sensed signals are labelled, matched with training sets and then weighted before intermediate output can be calculated** in order to establish **patterns**. Model performance **dramatically deteriorates** in lockstep with the **polling frequency** of the weapon's primary sensors.	X	X	X	X	X	X	X	X	X	X	X	X
The AWS' picture-building **requires compensation routines to manage for variable intensity** in this primary sensed data.		X	X	X			X	X	X	X	X	X
Picture-building is compromised by AI's inherent 'Exclusive-Or Problem' (when **no combination of weighting values triggers thresholds** that have been set during AWS configuration).		X	X	X	X	X	X	X	X	X	X	
AWS' neural networks comprise signal paths traversing from front to back but such ML architecture empirically learns at different, unbalanced and unknowable rates. An ancillary constraint here is that **the unsupervised weapon must minimally have its architecture fixed before training starts thereby constraining the degree of improvement possible** through subsequent training.	X	X	X	X	X	X	X	X	X	X	X	
Routines will be required to manage the AI characteristic of 'data discounting' whereby **battlefield features with only a small number of data examples may be smoothed to the extent of formlessness** despite those data-points' possibly critical importance (the notion of war being 9/10th inactivity and 1/10th chaotic activity).		X	X	X	X	X	X	X	X	X	X	X
Data discounting may also be caused by an overwhelming number of **learning examples in one set** of the weapon's data planes **undoing the training effect on the learning examples in a different dataset.**		X	X	X	X	X		X	X		X	
Data discounting may be caused by erroneous model sensitivity (the AWS' detection rate) **or an incorrectly set model specificity** (here, the AWS' false alarm rate). Neither of these characteristics can be recognised or tested upon deployment	X	X	X	X	X	X	X	X	X	X	X	X
The weapon's network connections will require management to prevent interaction in non-linear and complex fashions, the forming of new connections and even new neural units while disabling others. **Current ML models restrict learning to the network's top layer** while lower layers remain random transformations that do not exhibit much input capture.	X	X	X	X	X		X	X	X	X	X	X

OUTSTANDING TECHNICAL ISSUES INTERACTION WITH FRAMEWORK PHASES

	LIFECYCLE OF A WEAPON SYSTEM						NATO ALLIED JOINT TARGETING CYCLE					
	NATIONAL POLICIES 0	RESEARCH AND DEVELOPMENT 1	TESTING & EVALUATION, REGULATION, CERTIFICATION 2	DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING 3	USE & ABORT 4	POST USE ASSESSMENT 5	COMMANDERS INTENT, OBJECTIVES AND GUIDANCE 1	TARGET DEVELOPMENT 2	CAPABILITIES ANALYSIS 3	COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT 4	MISSION PLANNING AND FORCE EXECUTION 5	ASSESSMENT 6
Descent gradients (the weapon's first-order optimization algorithm for finding a local minimum of a differentiable function in the data returned by its sensors) are prone to inherent dilution in these lower layers **providing unpredictable and weak guidance to the overall learning process of the weapon.**		X	X	X	X		X	X	X	X	X	X
Descent gradients also **demonstrate plateauing of performance** as the gradient reduces.	X	X	X	X	X	X	X	X	X	X	X	X
Management of ML connections remains an intractable challenge. **Configuration routines must balance freezing connections once routines are learnt** (resulting in the AWS being a 'one trick pony') or having them remain open in a state of perpetual learning (resulting then in a unstable ever-learning weapon that the local commander cannot understand)	X	X		X		X	X	X	X	X	X	X
ML decision-making's 'satisfying' approach is **systemically inappropriate** whereby, for instance, a 90% threshold (here, a data match of 90%) is the defined **trigger point for AWS engagement.**	X	X	X	X	X	X	X	X		X	X	X
DATA POLLING AND MANAGEMENT												
Data receipt is critical to this AWS model but, as **noise increases in datasets**, class boundaries that separate different class examples become **impossible for the weapon to define** and then separate for ongoing statistical analysis.			X	X	X	X	X	X		X	X	X
Repeatedly sensed data may also push the weapon's neurons into saturation which then **desensitises neurons to all inputs.**			X	X	X	X	X	X		X	X	X
It is also systemically challenging for the **AWS' sensors, the likely sole source of inputs** for its decision processes, to garner consistent information. **Smoke, reflectance, image echo** as well as issues around data intensity intractably complicate processing. Here, class boundaries separating **different data examples resist definition where that data is partial, noisy or indistinct, requiring human intervention** if the weapon is to designate data strings for further statistical analysis.	X	X	X	X	X	X				X	X	X
Data mismatch against its training set, anything that is statistically out of the ordinary (whether the result of feint, by enemy surprise or by inadequate data separation) **will compromise on-board data analysis.**	X	X	X	X	X	X	X	X	X	X	X	X
Error in the weapon's sensing of its current state **must carry forward** in the machine's future learning and future battlefield actions.	X	X	X	X	X	X	X	X	X	X	X	X

OUTSTANDING TECHNICAL ISSUES INTERACTION WITH FRAMEWORK PHASES

	LIFECYCLE OF A WEAPON SYSTEM						NATO ALLIED JOINT TARGETING CYCLE					
	NATIONAL POLICIES 0	RESEARCH AND DEVELOPMENT 1	TESTING & EVALUATION, REGULATION, CERTIFICATION 2	DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING 3	USE & ABORT 4	POST USE ASSESSMENT 5	COMMANDERS INTENT, OBJECTIVES AND GUIDANCE 1	TARGET DEVELOPMENT 2	CAPABILITIES ANALYSIS 3	COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT 4	MISSION PLANNING AND FORCE EXECUTION 5	ASSESSMENT 6
Similarly, much of what the weapon has **recently learnt may be invalid** if its environment or its **combat task changes,** a trade-off between a weapon that is 'constantly learning' as opposed to one that is using what is already known to work at the cost of missing out on further improvement.		X		X		X	X		X	X	X	
Inappropriate variability arises from ML's systemic characteristic of 'dropout' whereby learning routines must **omit randomly selected neurones** in order to reduce over-fitting and false correlation	X	X	X	X	X	X	X	X	X	X	X	
It is **unknowable from the outset** if a weapon's training data is both **sufficiently relevant** to its y-function (the task that the commander has for each weapon) or of **sufficient size** appropriately to train that weapon network.	X	X	X	X	X	X	X	X	X			X
The efficacy of ML is also inescapably dependent upon the fit between image classification and the weapon's training: a marginally different set-up or a **marginally different training dataset** to the AWS' sensed data empirically leads to **substantial discrepancies in output.**	X	X	X	X	X	X	X	X	X	X	X	X
Relevant learning data sets with which to 'teach' an unsupervised weapon are **rare**; existing military datasets are either **restricted**, particular to a setting (and therefore **irrelevant**) or **very narrow** task.	X	X	X	X			X	X		X		X
A key weakness for the ML model is captured in the acronym **CACE ('change anything, change everything')**; increasing the breadth of training parameters empirically leads to inappropriately random outputs (eg. **manipulating a fraction of an image's pixels** in order to defeat current recognition routines).	X	X	X	X	X	X	X	X	X	X	X	X
AWS efficacy will depend upon **management of rapid but varied obsolescence** within its sensed data. Data obsolescence **leads to systemic instability.** For example, AWS' movement and all relevant navigable space must be i**dentified, processed and made 'map-ready**' for each of the weapon's representations. The resulting dataset must dynamically be searched **in real-time** to evaluate every available path, chose the optimal path and, finally, goals, values and action selection **must all be revised** to account for that newly selected path.	X	X	X	X	X	X	X	X	X	X	X	X
Polling frequency (the rate of recurrence that the AWS polls new data from its sensors) will determine the **weapon's ability to handle data**, its memory management and processing efficacy. Polling frequency also governs rates of data decay and is complicated by itself being a dynamic and changing function.	X	X	X	X	X	X	X	X	X	X	X	X
Processes are required to **manage the issue of data saturation** in order to prevent the model's **desensitising**. Arbitration contributes to an appropriate model for determining how one sensor input is preferred over others; the configuration issue of settling input intensities. Two variables that may be useless by themselves can be useful together. Similarly, a single variable that is useless by itself can be instrumental with others.	X	X	X	X	X	X	X	X		X	X	X

OUTSTANDING TECHNICAL ISSUES INTERACTION WITH FRAMEWORK PHASES

	LIFECYCLE OF A WEAPON SYSTEM						NATO ALLIED JOINT TARGETING CYCLE					
	NATIONAL POLICIES 0	RESEARCH AND DEVELOPMENT 1	TESTING & EVALUATION, REGULATION, CERTIFICATION 2	DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING 3	USE & ABORT 4	POST USE ASSESSMENT 5	COMMANDERS INTENT, OBJECTIVES AND GUIDANCE 1	TARGET DEVELOPMENT 2	CAPABILITIES ANALYSIS 3	COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT 4	MISSION PLANNING AND FORCE EXECUTION 5	ASSESSMENT 6
Processes will be required to manage AWS' **data processing order** (and the management of different outcomes according to which data string is processed in which order), managing both the **conundrum of 'signal intensity' as well as 'data habituation'** (the decreased response of ML to repeated stimuli).	X	X	X	X	X	X	X		X		X	X
Processes will be required to manage **overfitting of sensed data** to that agent's training data or its initial representation upon deployment (the machine's Day One state following configuration).	X	X	X	X			X			X	X	X
Processes are required to deal with AWS' **stale data.**		X		X		X			X	X	X	X
Routines will be required to identify and then manage loose data dependencies (the creation of **inappropriate associations based on mistaken correlations** during propagation forwards and backwards in the AWS' neural network).		X	X	X	X	X	X	X	X	X	X	X
Integration of otherwise **separate proprietary coding routines** (that together will comprise AWS' operation) will require routines to manage resulting glue code, pipeline jungles and dead code.	X	X		X		X	X	X	X	X	X	X
It will be necessary to manage and **arbitrate 'un-learning routines'** in AWS data processes.		X		X	X		X	X	X	X	X	X
Processes will also be required to **manage 'partially observed states'** in the AWS' sensed data and how the weapon backfills appropriately for missing, broken or unexpected data in its matching and decision processes.	X	X	X	X	X	X	X	X	X	X		X
ADDITIONAL CODING CHALLENGES												
The **'smoothing routines'** required to prepare and then manage data handling necessitates that the coding basis for AWS' operation is that of an inappropriate statistical 'approximator'.	X	X	X	X	X	X						
Autonomous componentry must recognise and manage coding errors, the challenge of **re-factoring code** in an AWS that is likely **out of communication.**		X	X	X	X			X		X	X	

OUTSTANDING TECHNICAL ISSUES INTERACTION WITH FRAMEWORK PHASES

	LIFECYCLE OF A WEAPON SYSTEM						NATO ALLIED JOINT TARGETING CYCLE					
	NATIONAL POLICIES 0	RESEARCH AND DEVELOPMENT 1	TESTING & EVALUATION, REGULATION, CERTIFICATION 2	DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING 3	USE & ABORT 4	POST USE ASSESSMENT 5	COMMANDERS INTENT, OBJECTIVES AND GUIDANCE 1	TARGET DEVELOPMENT 2	CAPABILITIES ANALYSIS 3	COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT 4	MISSION PLANNING AND FORCE EXECUTION 5	ASSESSMENT 6
AWS' coding requires the **artificial** imposition of start and end-states to avoid AWS being in an inappropriate state of 'perpetual intermediacy'. Such coding choices must invariably be constrained by earlier choices. Unless AWS **tasking is very restricted**, critical pathways will remain undiscovered, ignored or misunderstood.		X	X	X	X	X	X	X	X	X	X	X
Coding is systemically **poor to deal with ambiguity, context or situational awareness,** like the 1,000 page Norwegian rules of engagement. The information contained within a command **must be coupled to previously given information as well as to information that is to follow.** AWS must factor for **nested structures and conditionals which regularly characterize complex instructions**. Computers cannot build the same hierarchical complexity of human language. The challenge is also that such **categorizations are volatile and change unpredictably** according to new intelligence, new feedback and local input from the weapon's sensors.	X	X	X	X	X	X	X	X		X	X	
Processes are required to **manage AWS' goal-setting and value-setting**. Goals are associated with AWS independent plan of action, while values assess the viability of these plans.	X	X	X	X	X	X	X	X	X	X	X	X
Errors in goal and value setting may have quite unforeseen battlefield consequences that include 'infrastructure profusion' where an **independent weapon** might unexpectedly allocate disproportionately large parts of its **reachable resources into the service of some inappropriately set goal.**	X	X	X	X	X	X		X	X		X	X
Routines and bias filters will be required to manage the programming issue of 'value-loading' in the AWS, the means of directing actions in an AI agent. The current absence of established mechanism to manage this foundational problem (explicit representation, evolution by selection, associative value accretion, use of motivational scaffolds or reinforcement learning) demonstrates the degree of invention still required if ML is to provide an appropriate deployment spine for AWS. Human supervision.	X	X	X	X	X	X	X	X	X	X	X	
AWS data-points (for example, in targeting sequences) are invariably revealed incrementally; the systemic issue is therefore when (in that sequence) the AWS can make **an engagement decision** that is efficient and compliant.	X	X	X	X	X	X	X	X	X	X	X	X
Dampening protocols will be required to avoid the AWS oscillating around a desired state and becoming inappropriately **paralysed in its decision-making.**	X	X	X	X	X		X	X	X		X	X
Human intervention is also required to mediate between locking-in AWS' deployment **assumptions versus modifying** those assumptions in light of **new** instructions, **new** priorities or **newly** sensed data.	X	X	X	X	X	X	X	X	X	X	X	X

OUTSTANDING TECHNICAL ISSUES INTERACTION WITH FRAMEWORK PHASES

	LIFECYCLE OF A WEAPON SYSTEM						NATO ALLIED JOINT TARGETING CYCLE					
	NATIONAL POLICIES 0	RESEARCH AND DEVELOPMENT 1	TESTING & EVALUATION, REGULATION, CERTIFICATION 2	DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING 3	USE & ABORT 4	POST USE ASSESSMENT 5	COMMANDERS INTENT, OBJECTIVES AND GUIDANCE 1	TARGET DEVELOPMENT 2	CAPABILITIES ANALYSIS 3	COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT 4	MISSION PLANNING AND FORCE EXECUTION 5	ASSESSMENT 6
Human intervention is currently required to manage and validate download and enactment of software patches in what is otherwise an independent (and thus **incommunicado**) weapon system.	X	X	X	X	X	X	X	X	X	X	X	
Supervision is currently required to establish **attribution** in weapon performance and behaviour.	X	X	X	X	X	X	X	X	X	X	X	X
Processes will be required to **manage 'anchoring'**, (the degree by which a weapon's initial representation is amended in light of newly sensed information from the platform's sensors.)		X	X	X		X	X	X	X	X	X	X
Computers continue to **struggle to interpret context**: vision software may identify a soldier walking but is **unable to determine why** the soldier is walking. This also renders autonomous systems particularly **vulnerable to trickery**.	X	X	X	X	X	X	X	X	X	X	X	X
ACTION SELECTION IN AWS DEPLOYMENT Processes are required to manage challenges around autonomous weapon **'attention'**, prioritising one data string over other sensed information (the **'cocktail party effect' of seemingly irrelevant information that may be key** to that weapon's compliant and useful operation).	X	X	X	X	X	X	X	X	X	X	X	X
Weapon actions must comprise the **appropriate reaction to every relevant sensed stimulus**. AWS cannot offer **intermittent or erratic performance where only specific sensed inputs lead to weapon outputs**.	X	X	X	X	X	X	X	X	X	X	X	X
Routines are required to **manage the firing sequence** for all weapon instructions. Each routine may result in quite **different outputs being triggered depending upon the order** in which command instructions are processed by the unsupervised weapon.	X	X	X	X	X	X	X	X	X	X	X	X
Processes are required to manage the characteristic of the **'undeclared consumer'** whereby interim data decisions are propagated forward to calculate next values in a decision routine. The systemic issue arises from **that next value becoming the undeclared consumer of a prior data decision** which, **not itself a primary datapoint directly taken from sensed data,** may or **may not have been correct.**	X	X	X	X	X	X		X	X	X	X	X
Protocols are required to deal with the **AWS' temporal framing,** the time element of an engagement routine (for instance, a sequential engagement decision).	X	X	X	X	X	X	X	X	X	X	X	
Protocols are required to manage AWS' **fatigue processes** allowing, for instance, **change of decision policy** within that weapon.	X	X	X	X	X	X	X	X	X	X	X	

OUTSTANDING TECHNICAL ISSUES INTERACTION WITH FRAMEWORK PHASES

	LIFECYCLE OF A WEAPON SYSTEM						NATO ALLIED JOINT TARGETING CYCLE					
	NATIONAL POLICIES 0	RESEARCH AND DEVELOPMENT 1	TESTING & EVALUATION, REGULATION, CERTIFICATION 2	DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING 3	USE & ABORT 4	POST USE ASSESSMENT 5	COMMANDERS INTENT, OBJECTIVES AND GUIDANCE 1	TARGET DEVELOPMENT 2	CAPABILITIES ANALYSIS 3	COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT 4	MISSION PLANNING AND FORCE EXECUTION 5	ASSESSMENT 6
Processes are required to **manage the complexity around a veto** including the management in the AWS of a **partial or delayed response, hand-off and withdrawal routines** and the onward communication of states to the local command	X	X	X	X	X	X	X	X	X	X	X	X
Protocols are required to ensure **post-engagement damage assessment.**	X	X		X		X						X
Routines are required to manage weapon **verification, validation and testing** in a **communications-denied** environment;		X	X	X	X		X	X	X	X	X	
Extensive Red-Teaming **(attack simulation)** is required for AWS processes.	X	X	X	X								

LEADERSHIP CHALLENGES TO AWS DEPLOYMENT

	LIFECYCLE OF A WEAPON SYSTEM						NATO ALLIED JOINT TARGETING CYCLE					
	NATIONAL POLICIES 0	RESEARCH AND DEVELOPMENT 1	TESTING & EVALUATION, REGULATION, CERTIFICATION 2	DEPLOYMENT, TRAINING, COMMAND AND CONTROL, OPERATION AND PLANNING 3	USE & ABORT 4	POST USE ASSESSMENT 5	COMMANDERS INTENT, OBJECTIVES AND GUIDANCE 1	TARGET DEVELOPMENT 2	CAPABILITIES ANALYSIS 3	COMMANDERS DECISION, FORCE PLANNING AND ASSIGNMENT 4	MISSION PLANNING AND FORCE EXECUTION 5	ASSESSMENT 6
The difference between a weapon's current state and its desired states is the weapon's **observed** error, the object of the AWS' action selection being to minimize that error. Two technical issues arise. The **pace of this error correction** is not obvious. It depends, for instance, on **how often the error is computed and how much correction is then made on each feedback loop**. Feedback loops, are significantly less effective at modifying a weapon's higher-level action selection such as prioritisation, coordination and collaboration.		X	X	X	X		X	X	X	X	X	X
The AWS must also be **appropriately front-facing** and determine its system state ahead of time.	X		X	X	X		X	X	X	X	X	X
An unsupervised weapon must not generally forget acquired skills (the notion of **'catastrophic forgetting'**), a recognised limitation of neural network models.	X	X	X	X	X	X	X	X	X	X	X	X
'Hybrid autonomy' suggests the toggling of command between human and machine. Research points to **erratic performance when humans are required to intervene in moments of high stress** or in situations of limited information.	X	X	X	X	X	X	X	X	X	X	X	
As weapon autonomy is introduced, slack time in battle processes is reduced, scope for rule-bending and initiative is removed and, in the case of weapons based upon ML, the **leader's ability to predict and influence outcomes is lessened.**	X			X	X		X	X	X	X	X	
Factors arising from International Humanitarian Law (IHL) involve evaluative and contextual judgement for which the **local commander must remain responsible and accountable.** Sufficient control must therefore be retained to enable **situation-specific judgement to comply** with those rules.	X	X	X	X	X	X	X	X		X		
Routines will be required to **manage the AWS' failure** modes (outright veto, 'fail too safe', 'fail dangerously' and 'fail deadly') as well as the integration of otherwise independent assets into the commander's wider portfolio of battle plan assets.	X	X	X	X	X	X	X	X	X	X	X	X

Author: Taniel Yusef
Contributor: Paddy Walker
Designer: Aline Nassar

Thanks:
Peter Burt
Ben Donaldson
Maiara Folly
Matt Mahmoudi
Leyla Manthorpe
Elizabeth Minor
Laura Nolan
Liz O'Sullivan
Noel Sharky

WILPF
UNITED KINGDOM

STOP KILLER ROBOTS
MEMBER