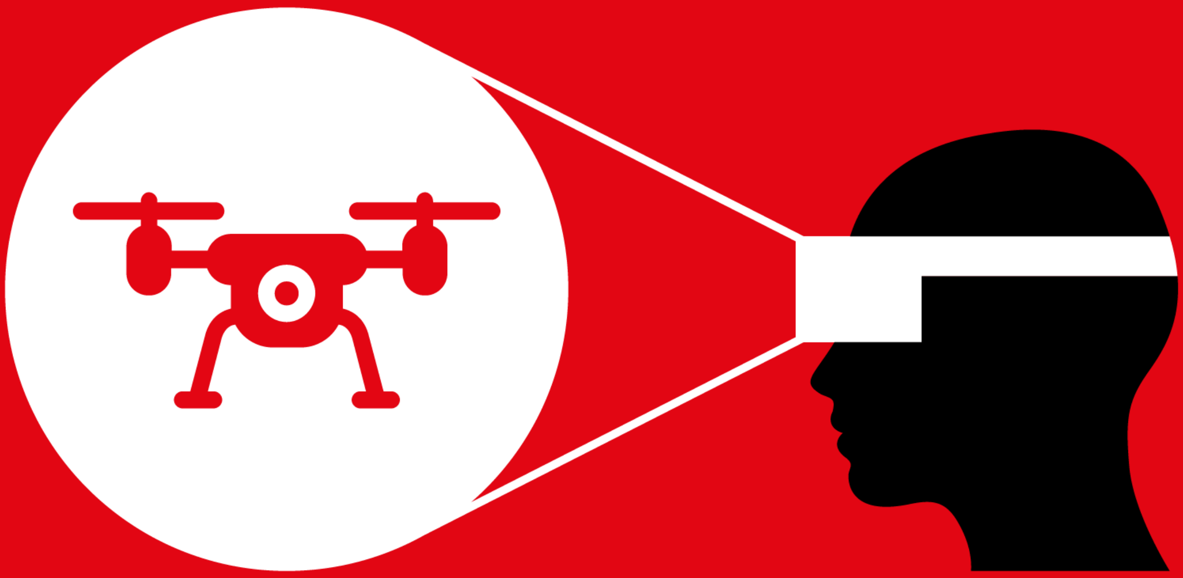
September 2022
APPG AI Evidence Meeting

Artificial Intelligence, National Security & Defence: Autonomous Weapons

PARLIAMENTARY BRIEF



***Artificial Intelligence, National Security & Defence: Autonomous Weapons** is a* Parliamentary Brief based upon the All-Party Parliamentary Group on Artificial Intelligence (APPG AI) Evidence Meeting held in House of Lords: Committee Room 1 on the 7th of September 2022.

This APPG AI is co-Chaired by **Stephen Metcalfe MP** and **Lord Clement-Jones CBE**.

We would like to express our appreciation to the following people for their oral evidence:

- **The Rt. Rev Steven Croft, Lord Bishop of Oxford**
- **Dr. Daniel Clarke,** Head of Applied Research, **Rebellion Defence**
- **Verity Coyle,** Senior Advisor**, Amnesty International**
- **Dr. Sidharth Kaushal,** Research Fellow**, Royal United Services Institute (RUSI)**
- **Taniel Yusef,** Tech Developers Coordinator**, UK Campaign to Stop Killer Robots**
- **Dr. Mariarosaria Taddeo,** Associate Professor and Senior Research Fellow**, Oxford Internet Institute**

Big Innovation Centre is the appointed Secretariat for APPG AI

- CEO, **Professor Birgitte Andersen**
- Rapporteur, **George Farrer**

www.biginnovationcentre.com | Email: info@biginnovationcentre.com | @BigInnovCentre

https://bicpavilion.com/about/appg-artificial-intelligence |
Email: appg@biginnovationcentre.com | @APPG_AI

PARLIAMENTARY BRIEF

Artificial Intelligence, National Security & Defence: Autonomous Weapons

All Party Parliamentary Group on
Artificial Intelligence

APPG AI Sponsors

The Group supporters – British Standards Institution, Brunel University London, CMS Cameron McKenna Nabarro Olswang, Deloitte, Ernst & Young, Innovate UK – UKRI, Osborne Clarke, PwC, and Rialto – enable us to raise the ambition of what we can achieve.

Contents

1. Introduction

In this meeting, the APPG AI discussed issues surrounding Artificial Intelligence (AI), National Security and Defence, with a heavy focus on Autonomous Weapons. The development of military technologies has increased massively in recent decades and AI is being implemented on the battlefields. However, there are concerns as to whether AI has a place in warfare, especially when civilians are involved. Critical questions faced at this evidence meeting include how Autonomous Weapons could represent a turning point in warfare, and whether they align with international humanitarian law and ethics.

Advances in military technology have been a hot-button topic in recent times, especially following the Russian invasion of Ukraine in February 2022, where defence and national security has been on many governments' lips, around the world. Therefore, the APPG on AI considered how, and if, Autonomous Weapons should be used on the battlefields with leading academics and industry experts in this field.

Main questions:

- *Advances in Machine Learning and Artificial Intelligence (AI) represent a turning point in the use of automation in warfare. How is (or can) AI be used in military and Autonomous Weapons?*
- *Are there opportunities for safer military solutions or what are the risks?*
- *What are the emerging standards and conventions?*

List of panellists:

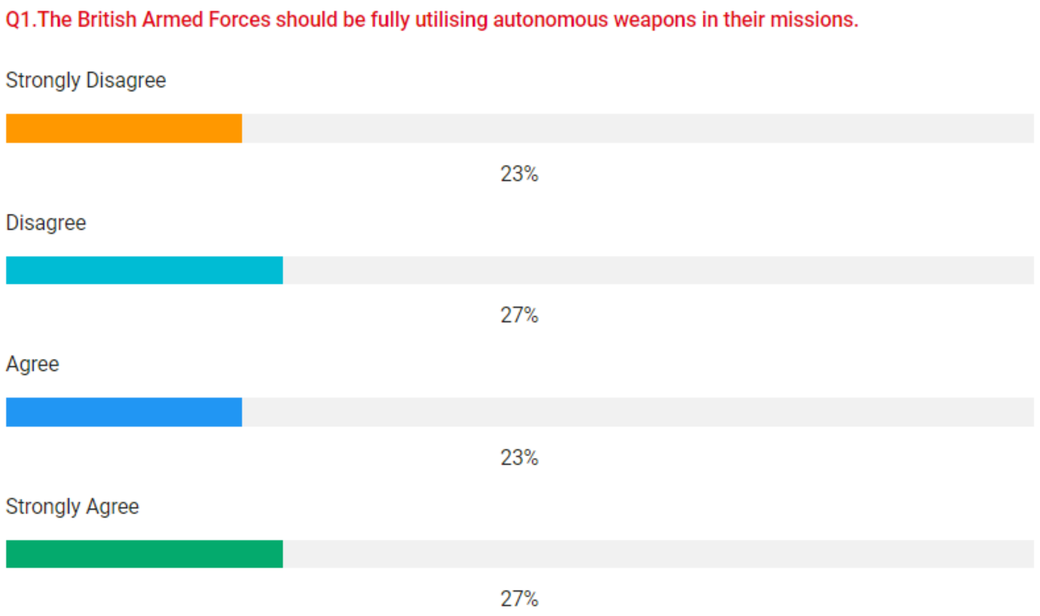
- **The Rt. Rev Steven Croft, Lord Bishop of Oxford**
- **Dr. Daniel Clarke,** Head of Applied Research, **Rebellion Defence**
- **Verity Coyle,** Senior Advisor**, Amnesty International**
- **Dr. Sidharth Kaushal,** Research Fellow**, Royal United Services Institute (RUSI)**
- **Taniel Yusef,** Tech Developers Coordinator**, UK Campaign to Stop Killer Robots**
- **Dr. Mariarosaria Taddeo,** Associate Professor and Senior Research Fellow**, Oxford Internet Institute**

(From L-R: Dr. Sidharth Kaushal, Verity Coyle, Rt. Revd. Steven Croft, Prof. Birgitte Andersen, Taniel Yusef, Lord Holmes of Richmond, Dr. Daniel Clarke, Dr. Mariarosaria Taddeo, Stephen Metcalfe MP)

This meeting was chaired by co-Chairs **Lord Clement-Jones CBE** and **Stephen Metcalfe MP**.

Parliament has appointed Big Innovation Centre as the **Secretariat of the APPG AI**, led by **Professor Birgitte Andersen (CEO)**. The Project Manager and Rapporteur for this meeting is **George Farrer**.

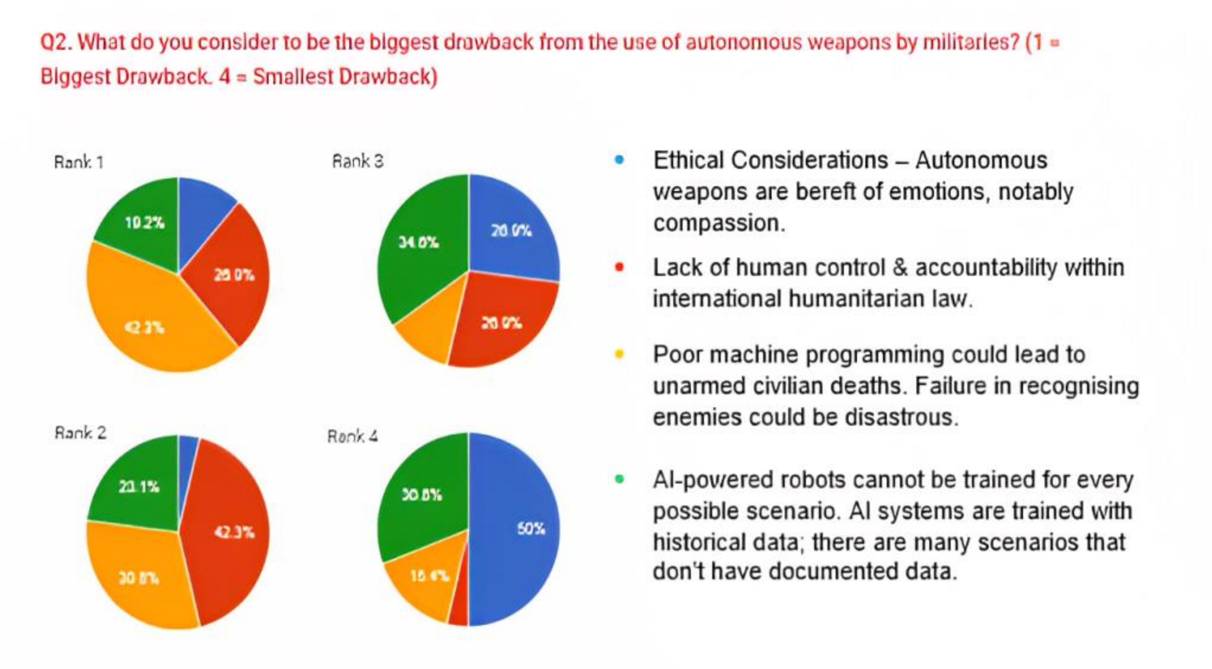
2. APPG AI Pavilion Survey



Q1.The British Armed Forces should be fully utilising autonomous weapons in their missions.

Prior to the APPG AI meeting, a survey was issued on the **APPG AI's Pavilion Platform**. Question 1 asked APPG members whether they believed *'The British Armed Forces should be fully utilising Autonomous Weapons in their missions'*.

The results here were extremely even, with almost a complete split between the four options. The only conclusion that can be drawn from here is that there is no clear consensus on whether Autonomous Weapons should be used. Therefore, it is evident that this is a polarising subject which certainly requires further debate and discussion.



Q2. What do you consider to be the biggest drawback from the use of autonomous weapons by militaries? (1 = Biggest Drawback. 4 = Smallest Drawback)

- Ethical Considerations – Autonomous weapons are bereft of emotions, notably compassion.
- Lack of human control & accountability within international humanitarian law.
- Poor machine programming could lead to unarmed civilian deaths. Failure in recognising enemies could be disastrous.
- AI-powered robots cannot be trained for every possible scenario. AI systems are trained with historical data; there are many scenarios that don't have documented data.

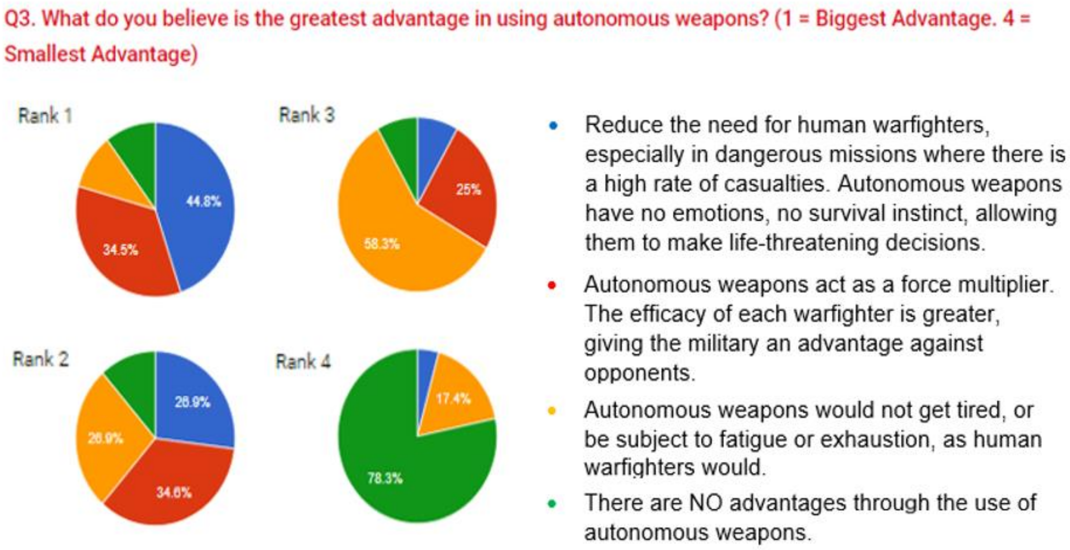
Question 2 questioned what the APPG AI community believed *to be the biggest drawback from the use of Autonomous Weapons by militaries*.

42.3% consider that **'poor machine programming leading to unarmed civilian deaths'** is the biggest negative aspect when using Autonomous Weapons. Furthermore, 30.8% of respondents ranked it the second, so it is something that has to be considered by the lawmakers in power.

Furthermore, a **'lack of human control & accountability within international humanitarian law'** was the second most selected (26.9%) option when respondents ranked their biggest drawback from autonomous weapons. In fact, of all alternatives, (42.3%) ranked it the second largest drawback from using autonomous weapons.

Therefore, these are the two issues that should be seen as most important areas by policymakers and technologists, for making the use of Autonomous Weapons as safe and as ethical as possible.

That **'AI robots cannot be trained for every possible scenario'** was ranked as the third (34.6%) biggest drawback, with **'ethical considerations'** decided as being the least important downfall in the use of Autonomous Weapons by the APPG AI community.



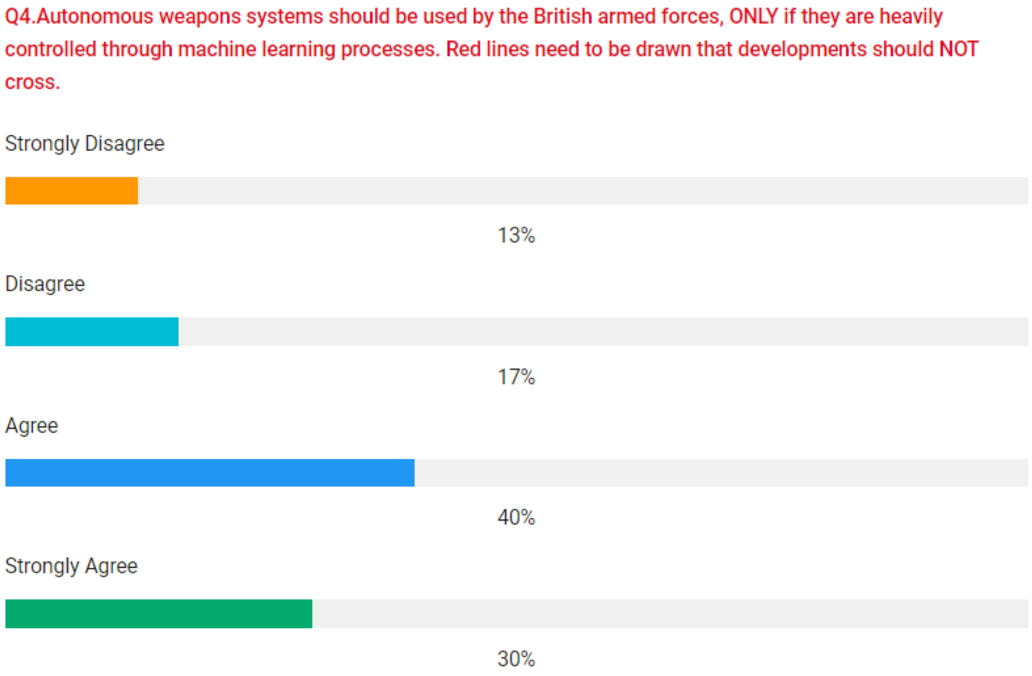
Q3. What do you believe is the greatest advantage in using autonomous weapons? (1 = Biggest Advantage. 4 = Smallest Advantage)

- Reduce the need for human warfighters, especially in dangerous missions where there is a high rate of casualties. Autonomous weapons have no emotions, no survival instinct, allowing them to make life-threatening decisions.
- Autonomous weapons act as a force multiplier. The efficacy of each warfighter is greater, giving the military an advantage against opponents.
- Autonomous weapons would not get tired, or be subject to fatigue or exhaustion, as human warfighters would.
- There are NO advantages through the use of autonomous weapons.

Question 3 asked respondents to rank four choices in order of what they *believe the greatest advantage in using Autonomous Weapons* to be. The idea that Autonomous Weapons would **'reduce the need for human warfighters'** was the leading advantage with 44.8% ranking this first. This was closely followed by **'Autonomous Weapons acting as a force multiplier'** which 34.5% ranked as their biggest advantage.

It was widely considered that **'Autonomous Weapons not getting tired or subject to fatigue'** was the third most important positive of Autonomous Weapons. 58.3% of those surveyed ranked this third in their list.

Finally, the statement declaring that **'there are no advantages through the use of Autonomous Weapons'** was ranked last by over three quarters (78.3%) of respondents. This shows that there are indeed some benefits when using Autonomous Weapons, and the upper hand they have over human warfighters in terms of being emotionless, more efficient, and more durable on the battlefield.



Q4.Autonomous weapons systems should be used by the British armed forces, ONLY if they are heavily controlled through machine learning processes. Red lines need to be drawn that developments should NOT cross.

Question 4 put the statement *'Autonomous Weapons Systems should be used by the British armed forces, only if they are heavily controlled through Machine Learning processes. Red lines need to be drawn that developments should not cross*' to the APPG AI community.

It is clear to see from the results that most people believe this statement to be true. 40% '**agree'** whilst 30% **'strongly agree'** that Autonomous Weapons need to be controlled through machine learning if they are to be used, e.g., by adding precision.

In total 30% either **'strongly disagree'** or **'disagree'** with this statement.

Overall, the results show that regulation and increased Machine Learning processes are required when Autonomous Weapons are put to use by the British armed forces.

3. Recommendations for policymakers

1. For there to be constructive debate around Autonomous Weapons Systems (AWS), **policymakers must set a clear definition.** This is because there is currently a gap in understanding between AWS capability, and public awareness. Parliament has a key role in breaching this gap in understanding and must confer with international partners to finalise an internationally agreed definition of AWS. **This cannot be a closed-door conversation** and must lead to transparent debate over the use of such systems.
2. The UK must connect its own debates regarding AWS with international concerns, and consequently **a national and international agreement over the use of AWS** should come from this informed debate. **The UK must be at the centre of discussions** over the development of international norms and standards, done in consensus with bodies like the UN. The creation of global standards and norms stems from a worldwide agreed definition of AWS.
3. Despite the increased development of AWS, **there are problems with the technology itself**, which could compromise its usefulness in military operations. Artificial Intelligence (AI) is a 'learning technology', based on complex datasets and algorithmic programming. Therefore, the act of surprise in warfare is something which would decrease the effectiveness of AWS – there are only so many simulations that an AI can be trained to expect, and the environment may become too complex. **AWS may never be able to fully replicate human characteristics** and is subject to questions of unexplaininability and uncontrollability – these cause issues with the just war theory[1].
4. The lack of meaningful human control that some AWS do have mean that sometimes they do **contradict the just war framework** – granting AI choice over target selection shifts this ethical framework. Loss of human control and judgement in actions on the battlefield and may mean that the AI does not understand how decisions between life and death are made, additionally would find **distinctions between civilians and combatants hard to judge**. AWS need the ability to detect contradictions between the two otherwise there is no doubt that this contravenes international humanitarian law.
5. If there is a place for AI on the battlefield it must be deployed ethically and safely, and those in control must be **fundamentally certain of a positive impact**. Responsible AI is essential to defence, therefore needs to be implemented sensibly. **The UK needs to take a leading role in responsible AI, ethics, and standards** – these will lead to the technology getting better and more effective itself.

[1] "Just war theory deals with the justification of how and why wars are fought". (Internet Encyclopaedia of Philosophy, https://iep.utm.edu/justwar/)

6. Science and technology can have massive impacts on democracy and humanitarian values. **If developed responsibility, strong AWS can bring more people out of harm's way**. It is paramount to decrease the number of people in danger, and this has been done with previous advancements in military technology. The development of responsible AWS can grant the UK **competitive advantage over its rivals**, in aspects of defence and security.

Our expert speakers at the meeting concurred that there was a need for an internationally-agreed definition of Autonomous Weapons Systems (AWS) – currently there is not one. **An internationally agreed definition will provide certainty about what an AWS system is, and their potential uses on the battlefield**, and help us answer questions over their ethics and applications with humanitarian law. Additionally, a definition of AWS will help further the debate around these systems and narrow the gap in understanding that many of the public have over these technologies. The panel believes that for debate and discussion to go further a definition is paramount.

There was also consensus regarding the fact that AI's usage in military operations **must be responsible in order to comply with internationally regarded human rights frameworks**, such as the just war theory. If target selection is left down to the AI system itself, this shifts the aforementioned ethical frameworks, and a loss of human control, particularly in the taking of life **contravenes the *jus in bello* [2] aspect of just war theory**. Whilst lethal autonomous weapons systems (LAWS) may not comply with such frameworks, new standards need to emerge to allow for AWS to get closer to being consistent with humanitarian values. There was overwhelming harmony between the panel that the **UK should be at the forefront of these international discussions**.

Despite agreement on several of the issues discussed at this evidence session, there was some disagreement about how quickly the potential of this technology would be realised. Arguments were raised that AI and Machine Learning's impact on warfare was huge in terms of **protecting democracy and humanitarian values**, enabling militaries to do more, with greater precision and less potential for unintended damages. On the other hand, it was also suggested that we should be cautious in believing that technology can be revolutionary in warfare. The dynamic nature of warfare means that we should not believe that everything will change because of new developments in technology. **It was said that the technology may be used for more tedious tasks and supporting elements**, before ousting the capabilities it supports.

The Rt. Revd. Steven Croft, Lord Bishop of Oxford starts by detailing that Parliament has a sentinel role when it comes to AWS. Revd. Croft asserts that **Parliament has a duty to bridge**

[2] *Jus in bello* is the part of just war theory that looks at the right and wrong ways for states to behave in times of war, governing the way that war is conducted.

the gap between public awareness of AWS and what is happening. Currently, deployment of these systems is greater than the awareness of them and how they are used, therefore transparent debate is required and further details around the systems themselves should be open to inspection. Revd. Croft states that Parliament needs to help narrow this gap in understanding, and work towards a new public consensus.

Moreover, Revd. Croft refers to ethical frameworks that have been developed over time to limit the damage that war can do. Revd. Croft mentions the just war theory, which discusses the steps which should be taken before going to war and the right and wrong way to behave in times of war, amongst other things. It is in the national interest to deploy any technology effectively and safely, which will help encourage debate, and Revd. Croft states that security reasons are not sufficient for that debate to be refused. Furthermore, **delegating the decision-making process for target selection to an AI does shift the ethical framework** somewhat, according to Revd. Croft. This then raises questions of common humanity and human dignity, a boundary which for many should not be crossed. If effective debate on the implementation of AWS can be had, Revd. Croft believes that this will lead to national and international agreement, and thus a coherent treaty to ban some uses of the technology when it comes to warfare.

Dr. Daniel Clarke, Head of Applied Research at Rebellion Defence, explains that technology has the potential to have a huge positive impact on people, and aid the men and women that put themselves in harm's way in defence of our humanitarian values and democracy. When technology is improved, this is having a constructive impact on those affected by such technology. Clarke claims that the technology must be designed correctly, as **whoever develops the most elegant, deliberate, and effective solution, will gain a competitive advantage over their competitors** – in this case potential international adversaries.

Clarke details that developments in technology such as autonomous terminal guidance and precision guided munitions have made such a difference in making warfare safer, in turn helping people who put themselves in jeopardy. Clarke contrasts this to bombing campaigns in WWII, where cities were bombed indiscriminately with extremely limited targeting, and thousands of civilians were killed. This demonstrates that technology is making warfare safer and more targeted, as is the aim of technologists.

Agreeing with Revd. Croft, Clarke believes that the **UK has an active role to play in the development of ethical frameworks, principles, and standards around technology in warfare and AWS**. Clarke states that this will influence allies and strengthen the ability to safeguard democratic principles worldwide. Moreover, Clarke asserts that **standards will make technology better, as we will have to be more thoughtful about how it is developed**, and more deliberate in what the technology is able to do. Therefore, if those standards are adhered to, the technology will be more effective, elegant, and purposeful than potential adversaries.

Verity Coyle, Senior Adviser at Amnesty International, starts by stating that **the development**

of international norms is vital to counteract the threat to human rights that AWS pose. Coyle argues that implications such as the loss of human control and judgement, lack of understanding of life and death decisions and the unpredictability of outcomes, are all serious risks of the technology that does not align with international human rights law. Thus, Coyle calls for **legally binding instruments** that ensure the state has control in exercising meaningful human control in the use of force. Additionally, Amnesty International is concerned about systems selecting and applying force to targets when activated; they strike in response to information from sensors that are matched against a target profile.

Despite technologies being highly sophisticated, Coyle goes on to explain that they will **never be able to replicate a complete range of inherent human characteristics**, in order to comply with international human rights. For example, AWS may not be able to analyse the intention behind people's actions, respond to dynamic situations or distinguish between civilians and combatants.

Coyle finishes with a strong recommendation for the UK and its international allies. She asserts that **it is essential for countries to step up now**, to prevent AWS from causing unlawful killings. Referring to the **Defence Artificial Intelligence Strategy**[3] (June 2022), Coyle states that the strategy does not support efforts to create an international treaty to address issues caused by AWS. For Coyle, **a new international treaty is paramount,** and the UK and other leading nations should be at the forefront.

Dr. Sidharth Kaushal, Research Fellow at the Royal United Services Institute (RUSI), kicks off by stating that war is inherently an interactive human activity, therefore we should be **careful in believing that this technology is revolutionary in terms of warfare**. Kaushal details the 'revolution in military affairs'[4] that was talked about in the 1990s and 2000s, where people thought the battlefield would be rendered transparent by new strategies such as computers and networking. This did not take place and adversaries quickly worked out ways around the new technology. It is rare that technology, including AI, will change the logic of modern warfare – perhaps in the future, but not necessarily at this moment in time.

However, Kaushal does explain that the introduction of AI and other technologies into the battlefield may create a more **offence-dominated environment**. If both sides in a conflict have a sense of where each other's forces are, this may create incentive towards a first strike. Additionally, increased information gathering, and new forms of deception may come about because of AI. This may include deliberately mis-training an opponent's algorithm, as small contextual changes can lead an algorithm to make some drastic differences in how it classifies a given image – particularly in peace time by consciously feeding them poor data.

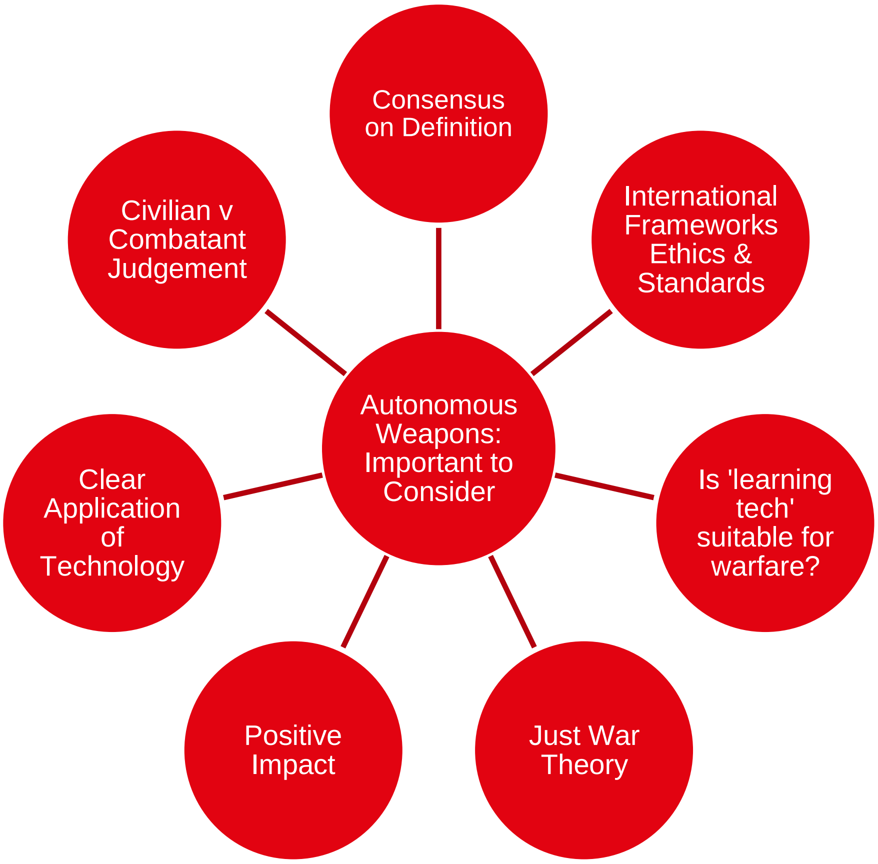
Kaushal finishes by explaining that he believes that we are unlikely to see an immediate

[3] UK Ministry of Defence – **'Defence Artificial Intelligence Strategy'** (2022**).** https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy

[4] A **revolution in military affairs (**RMA) is a theory about the future of warfare, often connected to technological and organisational recommendations for military reform.

revolution in warfare. **We may initially see AI and other technologies utilised for more mundane tasks and as a supporting element**, before then displacing the technology it supports. Kaushal gives the example of aircraft originally being used as spotters for battleships. With there potentially not being any immediate changes to how warfare is conducted, Kaushal asserts that it is likely to take 10 or more years for this technology to have a substantial impact.

Summary Illustration:
Important things to consider when discussing Autonomous Weapons


Consensus on Definition
International Frameworks Ethics & Standards
Civilian v Combatant Judgement
Autonomous Weapons: Important to Consider
Clear Application of Technology
Is 'learning tech' suitable for warfare?
Positive Impact
Just War Theory

Taniel Yusef, Tech Developers Coordinator at the UK Campaign to Stop Killer Robots, presents that AWS and incorporating AI into the use of force, raises fundamental questions in terms of **uncontrollability, unpredictability and unexplainability**. Additionally, that an advanced AI system could not explain their decision-making process to people. Yusef further questions whether an AWS can be said to be under meaningful human control. Humans can deal with faint and unexpected changes in direction, whereas datasets and algorithms may not be able to and subsequently get confused. These limitations in the use of algorithms on the battlefield would only be seen as the environment becomes too complex, with little time to fix the issues. Furthermore, Yusef states that it is essential in a military situation to detect contradictions and evaluate significance. Humans can certainly do these things, but could the same be said for AI systems?

Moreover, Yusef explains that to detect civilians, we must be able to distinguish them from combatants. However, technology which can show us the presence of humans is changeable and is based on context – for example, the existence of a mobile phone. There is a **high chance of false positives** when it comes to AWS being able to differentiate between those involved in the conflict and those that are not.

Yusef finishes by arguing that the unknown future use capacity of AWS, means that **present human rights and humanitarian law is insufficient for controlling its usage**. Yusef suggests that an international instrument will help develop international standards, and safeguard industry, academia, and research. Thus, harmonising with the rest of the expert panel that something needs to be done internationally in order to control the development and application of AWS.

Dr. Mariarosaria Taddeo, Associate Professor and Senior Research Fellow at the Oxford Internet Institute, details that AI's usage in just one part of a growing trend which is the **embracement of AI in defence institutions** for a plethora of purposes. In addition to AI in the context of weapons, AI is also being used in sustainment, cyber and supporting purposes. Taddeo describes how all of these areas come with risks, including the lack of transparency, a lack of accountability, a lack of explainibility and bias. These issues all affect AWS, along with contradictions with just war theory.

Taddeo explains that it is essential to understand what AWS actually are. However**, there is no internationally agreed definition of such systems**. Taddeo cites her comparative analysis[5], which found 12 definitions of AWS provided by international and state actors, some of which are quite diverse. Some definitions that were found by Taddeo's team were unrealistic, and do not apply the key aspects of AWS, which is the learning abilities of the systems. An internationally agreed definition of AWS will help those in positions of responsibility identify the object that we want to regulate as precisely as possible.

The issue of a lack of predictability is due to the very nature of AI, which is that AI systems learn and are based on algorithms and datasets. Taddeo agrees with some other expert speakers who mention that AWS will be susceptible to changes in the environment, however elusive. Furthermore, **this lack of predictability makes it extremely hard to grant more responsibility to humans for the action of these systems, as it separates intentions from actions**. Taddeo explains that ascribing more responsibility to humans is a key element for maintaining the variety of war.

[5] Taddeo, M – **'A Comparative Analysis of the Definitions of Autonomous Weapons'** (2021). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3941214

4. Evidence statements

The Rt. Revd. Steven Croft, Lord Bishop of Oxford

I would like to set the scene for this debate. I have been a member of the House of Lords since 2013. I was a member of the **Select Committee on Artificial Intelligence**[6], which Lord Clement-Jones chaired a few years ago, and I am a founding board member of the UK Government **Centre for Data Ethics and Innovation (CDEI)**[7]. I have been watching this space as a non-specialist and as somebody viewing it from the representation of civil society for some years.

I want to begin with what to me is an enormously powerful quotation from the book of Isaiah in the Hebrew Bible. God says, according to Isaiah, *"upon your walls there O Jerusalem, I have posted sentinels all day and all night, they shall never be silent"* (Isaiah 62.6).

The image I want to offer you is that Parliament has a role as a sentinel in relation to lethal autonomous weapons systems (LAWS). It is a really significant role, and we need to fulfil it more rigorously and vigorously.

[6] **House of Lords Select Committee on Artificial Intelligence**. https://www.parliament.uk/external/committees/lords-select/ai-committee/news/2018/ai-report-published/

[7] **Centre for Data Ethics & Innovation.** https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation

Gap between LAWS Capability and Public Awareness

There is an incredibly significant gap between LAWS capability on the one hand, and public awareness and debate about what is happening on the other. This is a very common experience in my observation of the development of AI: the deployment of narrow AI in a number of fields, typically outstrips awareness of its use which creates challenges. For example, with facial recognition technology or the listening capabilities of smart speakers. The gap is even more serious with LAWS. Understandably there is security and secrecies around the development of new weapons systems, but that can further disguise the significant risks and inhibit the identification of ethical issues and the development of governance, boundaries, and safeguards.

Therefore, the debate needs to be more transparent. Definitions need to be clearer, lines of intent and guidance for deployment, if not details about the systems themselves, must be more open to scrutiny. The ethical reasons for that are essentially political rather than military, they're rooted in what it means to exercise legitimate authority and force on behalf of a democracy, and the scrutiny which that entails.

There is quite a sophisticated public consensus on what it means to exercise lethal military force: to declare war; the ownership and deployment of a nuclear deterrent; the dangers of chemical weapons; the ethical conventions of war. This public consensus has been formed over generations and is continually debated and revised. However, this consensus does not yet apply to these significant new technologies, in particular, the delegation of the decision to use lethal force in particular context to an AI. Parliament has a key role in addressing and narrowing this gap in understanding and working towards a new public consensus as technology evolves.

Responsible AI is Essential to Defence

My second point is that responsible AI is essential to defence. War is a terrible thing, but it can sometimes be, most would acknowledge, the least bad alternative. In order to protect society and the world, ethical frameworks, particularly the just war theory have evolved across many generations to seek to limit the damage done by war.

The framework addresses…

- The steps which must be taken before going to war - *jus ante bellum*.
- The right and wrong ways to behave in time of war – *jus in bellow*.
- Ways for a nation to conduct herself after a conflict – *jus post bellum*.
- The just use of force in measures short of war, such as no fly zones or peacekeeping – *jus ad vim*.

AI is changing and will change all kinds of defence capabilities, intelligence gathering, logistics, communications, and every other area of modern warfare by land, air, and sea. Each of these capabilities needs transparent governance and monitoring, as with the use of AI in the police or the NHS. Some deployment may help prevent war, other deployment may enable the more effective use of legal force and change the balance of power.

However, it must be in the national and international interest to deploy these new technologies effectively and safely, and therefore to encourage debate. Security reasons are an inadequate reason to refuse that debate. It is possible to debate the ethical deployment of AI in self-driving cars ("the trolley problem"[8]) without disclosing commercially sensitive new technology. In the same way, it should be possible to debate the ethical deployment of AI without disclosing new technologies in ways which raise security concerns.

AI & Weaponry

The third area is on AI and weaponry, which raises the following ethical questions which are of a different order to the deployment of AI in policing or healthcare or non-lethal military support.

The first is, delegating the decision-making process for selecting the target for the use of lethal force to an algorithm or AI fundamentally shifts the ethical frame profoundly. This raises big questions about how common humanity and human dignity, that for many people represents a boundary, should not be crossed.

There's clearly a sliding scale in this process of delegation, not an absolute line, that in itself creates an issue. It depends on different uses: from remote weaponry to loitering weaponry; to defensive systems of a human agent in control; to a remote autonomous system.

For those reasons, thirdly, questions of definition of what constitutes an AI system are key in public debate, and a key entry point to international dialogue. The development of new weaponry always carries a risk of proliferation, particularly if weapons are developed and sold are physically small, easily accessible, highly transportable, and quickly deployable. The evidence thus far indicates that AI, like all new weaponry, is likely to be more effective as a weapon of war or subjugation against non-combatants.

Conclusion

The rules for use of LAWS need to be clear and subject to international agreement. My encouragement is to fellow parliamentarians to raise the level of debate around these issues and connect the UK debate with international concerns, leading to a coherent treaty to ban some uses of AI in warfare. The process of establishing national and international agreement depends, unavoidably, on an informed public debate.

[8] The trolley problem is a thought experiment in ethics about a fictional scenario in which an onlooker has the choice to save 5 people in danger of being hit by a trolley, by diverting the trolley to kill just 1 person.

Dr. Daniel Clarke, Head of Applied Research, Rebellion Defence

I am the Head of Applied Research, for a company called **Rebellion Defence**[9]. I also teach at the **Defence Academy of the United Kingdom**[10], teaching a lot of emerging technologies on the Battlespace Technology course, to our young and developing officers. I have got quite a breadth of experience of developing the techniques and methodologies, which make up AI and Machine Learning across a number of domains including, industry, academia, and working in government.

Operation Herrick

I'd like to start with a small anecdote about my own experience working as a Ministry of Defence (MoD) scientist. So back in the days of Operation Herrick[11] I was deployed as one of the MoD scientific advisors to theatre, and I saw first-hand the impact that science and technology had on the men and women that put themselves in harm's way in defence of our humanitarian values and our democracy. I've always approached science and technology with the idea that when you make things better, it is having an impact on people, and when I develop that technology, it's about ensuring that its impact is positive; positive to the people, such as the men and women who put themselves in harm's way.

AI, Machine Learning & Defence

When we look at AI and Machine Learning, which are overly broad terms, I always like to think that AI and Machine Learning are series of techniques and methodologies, which have the

[9] **Rebellion Defence.** https://rebelliondefence.com/

[10] **Defence Academy of the United Kingdom.** https://www.da.mod.uk/

[11] **Operation Herrick** was the codename under which all British military operations were conducted in the war in Afghanistan from 2002 to 2014.

potential to accelerate almost any technology to which they are applied. Whether that's logistics, autonomous vehicles, the internet and social media, or defence and security. The defence and security part is really important, because the other thing I like to say is that whoever develops the most elegant, the most effective and the most deliberate solution will gain a competitive advantage over their competitors. When we think about this from a nation state perspective, from a defence and security perspective, our competitors are our international potential adversaries.

It’s also important to realise that AI and Machine Learning have been around for quite a while. Let us say that AI and Machine Learning have been used in defence in the past, actually we already used these in a wide number of applications, the most prominent being precision guided munitions. We have been using things like the tomahawk land attack cruise missile[12] since the 1990s, and when we talk about the terminal phase, we talk about autonomous terminal guidance. The idea of AWS is not something that has emerged in the last few years, it's something that has been considered in the technology domain for decades, and probably even further back.

Benefits of AI & Machine Learning in Warfare

To look at the benefits that this technology has brought, we can go back to the Second World War. When we think about Britain's participation, we had thousands of bombers per night dropping hundreds of thousands of tonnes of high explosive on civilian cities, almost indiscriminately – there was some targeting, but it was very broad. If we fast forward to recent military operations and the use of precision guided munitions, autonomous terminal guidance has helped us to do more, much more precisely, and with much less potential for causing indiscriminate damage. This is from a technologist's perspective, what we do. We want to make things better, we want to make it safer, and we want to help those people who put themselves in harm's way.

In terms of the UK’s role in this, the UK has a rich and diverse history of developing AI. We have led its academic development – you look at the fantastic universities and world leading research institutes around the UK. You look at the firms which use this technology and apply this technology, we have a number of world leading firms that are ensuring that this technology is effectively deployed.

What is changing though, is this perception of how the power of AI can add value across a number of these different domains, a number of these different technologies that we want to accelerate. This is where it becomes particularly important because these technologies are gaining potency, they're becoming more effective, and they're allowing us to do a lot more than we were able to do in previous decades.

[12] **The Tomahawk Land Attack Missile (TLAM)** is a long-range, all-weather, subsonic cruise missile. It is 5.5m in length, with a range of 1,000 miles and a speed of 550 mph. (http://www.armedforces.co.uk/Europeandefence/edequipment/edmis/edmis1a2.htm)

The UK Must Play a Leading Role in Developing Standards

I must agree with the Rt. Revd. Steven Croft's statements in that the UK has, as one of the world's leading democracies, an active role to play in the development of ethical principles, ethical frameworks, and the standards by which companies like Rebellion must achieve as we are developing this technology. I genuinely believe that by doing this, the UK will help to steer not only internal development and adoption of these frameworks and standards, but actually we'll influence our allies, and we'll influence our partners in developing those. They will see the UK taking a leading role and they will collaborate with us to do this.

The UK, and I say the UK very colloquially, must take a leading role in the development of those standards, principles, and frameworks. We have to take a highly active role in enforcing them. For example, when we purchase technology, we must ensure that that technology has been developed within those frameworks, standards, and principles.

One of the questions I often hear is that, is there the potential that if we have to adhere to a set of rigorous standards, does this make our technology less potent? I would argue that it does not. I would argue that it makes our technology better because we have to be more thoughtful about how we develop it, we must be more deliberate in what we are trying to get the technology to achieve, and we must be more considerate in the way that we develop that technology during the development process. Bringing this full circle, I genuinely believe that if we develop and we adhere to those standards, our technology will be more effective, more elegant, and more deliberate than our potential adversaries. This will underpin our democratic values and the human rights that we uphold as a democracy.

Verity Coyle, Senior Adviser, Amnesty International

I'm going to talk about AWS and why **Amnesty International**[13] believes that the development of additional international norms is vital to meet the potentially catastrophic threat to human rights that they pose.

New Technologies

New and emerging digital technologies developed for the use of force by military and other security forces, including the police are increasingly being used in a wide range of contexts including but not limited to:

- Armed conflicts.
- Law enforcement operations.
- Border security.
- Management of immigration enforcement.
- Private security.
- Counterterrorism measures.

These technologies often rely on data sets, algorithm-based programming, and Machine Learning processes that have serious implications for compliance with international human rights law and international humanitarian law, including:

- Loss of human control and judgement in the use of force.
- The difficulties of understanding how life and death decisions are being made.
- The resulting unpredictability of outcomes.

[13] **Amnesty International**. https://www.amnesty.org.uk/

- Uncertainty over accountability for human rights violations.
- Discriminatory effects of algorithmic biases.

There are also serious risks that these technologies could proliferate to non-state actors, including organised crime, private security companies and individuals.

For all these reasons, Amnesty International, along with thousands of AI scientists and roboticists, and a growing number of states – not the UK yet – advocate for international legally binding instruments that requires states to ensure the exercise of meaningful human control over the use of force, prohibits AWS that select and target human beings without meaningful human control, and strictly regulates all other AWS.

Autonomous Weapons Systems (AWS)

So, when we're talking about AWS, what are we talking about? A pink eye terminator in the future? Drones? Landmines? The point I want to make is that the call for binding international regulation and prohibitions should focus primarily on a process. We are concerned with systems that when activated select and apply force to targets without human intervention.

These are systems that trigger a strike in response to information from sensors being matched against a target profile. There are systems where the human operator is not setting the specific time and place where that strike will occur. However, the human operator may be setting the parameters for that process, for example, limiting the area or duration over which a system can operate, or the types of target the system is permitted to strike or not.

The technical characteristics described here relate to the process by which the specific time and place of an application of force will be determined, and this raises two specific points:

- Such a process might occur across a number of physical units that are connected to comprise the system – i.e., the process does not need to be embedded in just one weapon or object.
- Some pieces of machine hardware can fall within this scope in one configuration, but not in another.

A human pilot traditionally remotely flies a current predator armed drone. Strikes by the drone involve a human operator, viewing a potential target through the drone’s camera, sometimes cross referencing that against other sources of data, selecting the target and choosing to undertake a strike against it. In this configuration, we do not have the process of machine target selection on autonomous application of force. However, the same predator drone could theoretically receive its electronic instructions, not from a person, but from another computer, one that is analysing the camera footage and perhaps information from other sensors on the battlefield. If that computer were now identifying a potential target and activating a strike against it without meaningful human control, we would have the process of autonomous target selection – an application of force that we are very concerned with.

Technologies will not Replicate Human Characteristics

While future technologies may evolve to be highly sophisticated, we believe they will never be able to replicate the full range of inherently human characteristics necessary to comply with and apply international human rights or humanitarian law and standards. This includes the ability to analyse the intention behind people's actions, to assess and respond to an often dynamic and unpredictable situation or make complex decisions including about the proportionality or necessity of the use of force in a conflict situation to distinguish between civilians and combatants.

Conclusions

What do we think UK Parliamentarians can do? We welcome the support that has been taken already and the action on AWS from some Parliamentarians and hope that these efforts continue. The UK government are active in the UN discussions - they're putting forward papers, they're engaging with the discussions at the **Convention on Certain Conventional Weapons (CCW)**[14] – I attend those with my colleagues. However, the UK does not support the call for additional international norms to govern AWS and are ultimately content to continue talks within a consensus-based forum that has no hope of moving forward.

Let's not wait until AWS end up causing unlawful killings, it is time for countries to step up now. As part of the **Defence Artificial Intelligence Strategy**, the UK Government has unveiled a new policy on Autonomous Weapons. They recognize that systems which identify, select and attack targets without context-appropriate human involvement would be unacceptable - this is significant.

It’s concerning though that the policy does not support ongoing efforts to create a new international treaty to address the novel issues raised by the development of AWS. Also, the policy bases its arguments on the overly vague term context-appropriate human involvement, which does not adequately address the core problems related to the maintaining of meaningful human control over the use of force.

In the international discussions on Autonomous Weapons, there is widespread recognition that certain factors are necessary for meaningful human control, or sufficient predictability over autonomous systems, including:

- Human control over the duration and geographical scope of an autonomous system's operation – these are vital to making judgments about the possible effects of a system's use.
- Human understanding of the target profile the system uses – this is vital to understanding what will trigger the system to apply force, including the possibility of false positives, targeting things that are not the intended targets. This is also an area

[14] **United Nations Convention on Certain Conventional Weapons.** https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/

where AI and Machine Learning raise particular concerns.

These factors are not mentioned in the Government's paper, please ask them why. We also urge you to ask if the UK considers it acceptable to allow machines to identify and automatically fire on human beings, because they are silent on that issue.

Dr. Sidharth Kaushal, Research Fellow, Royal United Services Institute

I am going to approach a slightly different facet of this issue, which is what I believe the impact of AI and lethal autonomy could be on the battlefield of the future from a military standpoint.

Revolution in Military Affairs

War is always an interactive and human activity, and there is a risk in believing that technology can be revolutionary. One might think back to the so-called ‘Revolution in Military Affairs’ people talked about in the 1990s and 2000s – the idea of a battlefield rendered transparent by computers, by networking, by proliferating senses. Then the ways in which it was unpicked by responding adversaries were who at the low end of the spectrum taking shelter in complex terrain like cities, at the higher end of the spectrum, the level of state adversaries unpicked, the very networks upon which the ‘Revolution in Military Affairs’ depended, with things like electronic warfare capabilities or the ability to hold expensive aircraft at risk with the ground-based air missiles.

So, the first point I would like to point out is that it's relatively rare that technology in and of itself is militarily revolutionary. Secondly, in many ways AI and legal autonomy, what we viewed within the context of a broader package of technological shifts that are associated with the ‘fourth industrial revolution’, is that while they may change the grammar of how warfare is conducted, they may not change the logic as much.

Technology Having Impact

Beginning at the strategic level there's the possibility that a combination of increasing computing power, the ability of things like neural networks to track patterns in data, combined with the ability to draw more and more information from the environment with proliferating sensors can create an offence dominated environment. If for example both sides in an interstate conflict can probabilistically have a sense of where each other's forces are, does that create an incentive towards a first strike or the risk of suffering a catastrophic first strike if one fails to do so? Here a combination of the computing power of neural networks, but also the fact that you can increasingly deploy distributed sensors on autonomous capabilities in areas where it would be either too politically risky or too likely to cause a casualty, if a human operator we use probably lends itself to this.

For example, recent events where a Chinese unmanned underwater vehicle was captured by the Indonesian state[15] or the recent Iranian attempt to seize an American unmanned vessel in the Persian Gulf[16], both of which passed with far less incident than would've been the case had manned assets been used.

The other side of this is, does these same capabilities solve the so-called 'Wohlstetter Problem'[17]? This problem was named after the intelligence analyst who described it, who when examining cases of surprise attack like Pearl Harbour noted that the big problem was never a deficit of information, it was always a surplus – the fact that the signal was drowned out by the noise. Whether new technologies create a strategically offence-dominated environment, or not very much depends on the human perception of what their impact is. I think that the human element should never be removed from this.

Incentive for Persistent Information Gathering

The second point is that these technologies create both an incentive for persistent information gathering in peace time and perhaps new ways of doing so, like deploying autonomous capabilities, which even if they are shot down or seized can occur at relatively low risk in political cost. It also creates incentives for new forms of deception, many forms of Machine Learning and neural networks are quite easy to spoof or mislead in certain ways. For example, small contextual changes can lead an algorithm to make some substantial differences in how it classifies a given image in, in many notable experiments. Therefore, there may be incentives to consciously mis-train an opponent's algorithm and an opponent's AI in peace time by consciously feeding it poor data. Again, the dynamic we are describing, that of information gathering on the one hand and deception on the other, is as old as warfare itself. The tools may change but the basic structure remains the same.

Tactical Opportunities of Autonomous Weapons

Moving down to a more tactical level discussion, I would suggest that there are three trends

[15] **'China's underwater drones seized in Indonesia expose tech, routes, and potential submarine plans'** (2021). https://www.scmp.com/news/china/military/article/3117076/chinas-underwater-drones-seized-indonesia-expose-tech-routes

[16] **'US navy intervenes after Iran seizes American sea drone'** (2022). https://www.theguardian.com/us-news/2022/aug/30/us-navy-drone-iran-persian-gulf

[17] **Roberta Wohlstetter** was an American military intelligence analyst most known for her work: 'Pearl Harbour: Warning & Decision'. She was awarded the Presidential Medal of Freedom by President Ronald Reagan in 1985.

that are worth noting when it comes to autonomous capabilities.

The first, is that it is not just that increases in the processing power held on board small assets that can make it possible for increasingly small and cheap assets to classify their own targets. It's that this interacts with other trends such as the way in which additive manufacturing is driving down the costs of creating things like airframes, as well as the emergence of increasingly powerful explosives like aluminium nano-fibre which burns at several times the rate of TNT. Accumulatively, it creates a dynamic where it is not just possible to feel these capabilities, but to feel them in mass. In many ways, what we may be seeing is the return of relatively large masses of quite cheap capabilities to the battlefield. We have already seen a taste of what this might look like in the way in which quite poor non-state actors have used autonomous capabilities. You can see the way in which the Houthis[18] have used expendable unmanned aerial vehicles (UAVs) against the Saudi Arabian air defences. Yes, the air defences may shoot down these things with great frequency, but they cost much more than what they are shooting down, and even a single hit makes the model worth it. One proposition I would make is the future looks less like a high tech set of very sophisticated capabilities and more like masses of relatively cheap autonomous capabilities being fielded in large numbers.

The second point I'd make is that because as I said, a number of Western state adversaries have worked very hard to unpick the Western way of using things like long range, electronic warfare capabilities. Russia is a leader to unpick the networks upon which Western systems of systems depend or long-range surface air missiles to halt key enablers at reach, you do find a dynamic where effective long-range control may be exceedingly difficult to achieve in a contested battle space. This creates a strong military incentive to leverage autonomy, though the ethical questions regarding this are of course, highly relevant, so that there is a bit of a fork in the crossroads between the military realities of confronting a near peer, which absolutely do incentivize autonomous rather than remote capabilities and some of the very pertinent ethical considerations raised.

No Immediate Revolution?

I believe that we are not likely to see an immediate revolution in the conduct of warfare. Initially the adoption of these capabilities may be for more mundane tasks, whether that's logistical support or as decoys de-stimulator an opponent system, but it often is the case that a revolutionary technology is first adopted as a supporting element before eventually displacing the capabilities it supports. Aircraft carriers were originally meant to be spotters for battleships, which is quite a good example. In the 10-year horizon, we might expect these capabilities to be more of a supporting arm for traditional capabilities, that's what we are seeing in Ukraine – they're acting as spotters for very traditional artillery. After 10 years we might see more substantial changes. We should always bear in mind that for every move in war there is a countermove, we perhaps do not want to repeat the 'Revolution in Military Affairs' fallacy twice.

[18] The **Houthi** Movement is an Islamist political and armed movement that emerged from Saada in North Yemen in the 1990s.

Taniel Yusef, Tech Developers Coordinator, UK Campaign to Stop Killer Robots

I'm the Tech Developers Coordinator for the **UK Campaign to Stop Killer Robots**[19] and I'm lucky to work with a lot of leading researchers, academicians, and roboticists.

AI's Decision-Making Capability

Weapon systems that apply force based on processing sensor data, incorporating AI into the use of force, raise fundamental issues of unpredictability, uncontrollability, and unexplainabilitly. An advanced AI system wouldn't be able to explain their decision-making process to people, raising fundamental questions about whether it could really be said to be under meaningful human control. It is worth noticing that research points to erratic performance when humans are required to intervene in moments of high stress or in situations of limited information or too much information.

Ample evidence shows humans tend to over trust and misplace delegation to machines. Recent exhibitions like **DARPA's** (Defense Advanced Research Projects Agency)[20] **Top Gun AlphaDogFight**[21], or games like **AlphaGo**[22] and Atari – a human being was famously beaten by a machine – have gathered much excitement. However, both ignore some basic truths. These are highly structured restrictively rule based scenarios with nothing like the surprise complexity and data confusion or cognitive requirements for war. In a real-world dog fight with

[19] **UK Campaign to Stop Killer Robots.** https://www.stopkillerrobots.org/
[20] **Defense Advanced Research Projects Agency.** https://www.darpa.mil/
[21] **'AI vs. Human Fighter Pilot: Here's Who Won the Epic Dogfight' (**August 2020). https://www.popularmechanics.com/military/aviation/a33765952/ai-vs-human-fighter-pilot-simulated-dogfight-results/
[22] **AlphaGo.** https://www.deepmind.com/research/highlighted-research/alphago

weather, behaviour unpredictability, the programs of other aircraft, there is always by tactical desire and design an issue of surprise. This a dividing element of tactical warfare, AI is evidencing no capacity to deal with faint, a sudden unexpected change in direction intended to confuse or spoofing tricky limited data sets, which could be a remarkably simple affair. A pink painted tank is camouflaged to AI. There are only so many situations a machine could be programmed to expect. The Atari game was famously confused by a simply shifted Y-axis, and the addition of a wall.

A constraint of an AWS that ignores technical spikes is that the weapon must minimally have its architecture fixed before training starts. In other words, training cannot subsequently improve the weapons architecture, the training set in this case of Machine Learning is a known case. After a sufficient number of practice iterations, it is expected that the computer will be able to reconcile, present case sensed real world data to the training set – the known case.

The issue for AWS deployment is that autonomy's inherent limitations are only revealed as an environment becomes too complex to be captured in such models programming. It is well understood that minor distortion in AWS classification of sense data will likely lead to different data classes being inseparable in the live space where those variables are processed, so maybe false positives. Meanwhile, known data and social bias becomes embedded in decision making and the associated action selection in robotics and other such machine systems.

Civilian Impact

In order to protect civilians, we have to be able to recognize them, distinguish them from combatants, and protect injured combatants too. Technology which signifies the presence of human beings is changeable and can be abstractly context based, say the presence of mobile phones signifies the presence of human beings. That does not account for civilian versus combatant judgments. While humans are evolutionarily capable of reasoning when available information is imperfect and partial, formulating deductions that are based on knowledge that is generally true based on heuristic experience or patients, this is not the case in machine coding.

Limitations of AWS

Weapon outcomes must forever be inappropriately transient, as sensors contribute new data to refine previously available information without contradicting it. AWS operations must in fact facilitate counterintuitive capabilities, such as detection of contradictions, evaluation of significance, and complicatedly the efficient rejection of those alternatives that may leave the weapon with unsatisfactory outcomes. Goals and values must first have been set allocating importance to certain sensitive information and thus waiting must be learned, thresholds for action agreed upon pre-programming. Learning instability is therefore a characteristic of autonomous machines and in a battlefield setting further compromised by data noise, hacking, jamming, interference by opponents. In other words, if a weapon's data set is noisy, the class boundary that separates different class examples is almost impossible for the weapon to define and separates for ongoing statistical analysis.

AWS cannot offer intermittent or erratic performance where only specific sense inputs lead to weapon outputs, ignoring or invisibilising others, which could be legally and tactically significant. Nor can it improve accuracy using the urban war environment as a testing ground, it cannot learn on the job and satisfy law. Rather considerable technical debt inherent within machine processes from info to data smog, which is to say the likely intoxication by incoming information and overwhelmed by its own anomalies and the exponential likelihood of misinterpreting sensed data corrodes the ability to create faithful follow through of commander's intent and thus accountability to chain of command and law. More tech will not mean more accuracy.

International Frameworks & Regulation

Unknown future use capacity means current humanitarian and human rights law will be inadequate for controlling such use. Remembering that technological developments are led by a trans-boundary commercial sector, as is historically true the military, prohibitions, and positive obligations under the framework of an international instrument can both shore up international standards and safeguard research, academia, and industry. After all, the UK mandates the granting of academic technology approval scheme certification for STEM students who may acquire knowledge contributing to weapons of mass destruction or their means of delivery.

We have created an entire industry of regulation using Machine Learning. Along with the European Union, we are leaders in reg-tech; years behind which sits the US and with which it must comply. The Federation of German industries (BDI) have called for the EU to protect industry by creating regulation on AWS. They say companies are specialists for their products, they should not be asked to conduct autonomous risk assessments and carry the substantial risk associated, but in precise regulation would do damage to the export control environment as a whole.

It appears the UK Government understands the need for legal structures to be formalised for the protection of industry. In its response to the **House of Lords International Relations and Defence Select Committee Report** [23] on the law of the sea – despite their international posturing on autonomous maritime vehicles paralleling that on AWS domestically – the Government[24] states that the new legal framework should be developed, when parliamentary time allows. So regulatory hurdles will not prevent innovation.

However, the **Foreign Affairs Committee report 'Encoding Values, Putting Tech at the**

[23] House of Lords International Relations and Defence Committee – **'UNCLOS: the law of the sea in the 21st century'** (March 2022). https://publications.parliament.uk/pa/ld5802/ldselect/ldintrel/159/15902.htm

[24] **Government Response to UNCLOS: the Law of the Sea in the 21st Century** (May 2022). https://committees.parliament.uk/publications/22581/documents/168699/default/

Heart of UK Foreign Policy'[25], to which we also gave evidence, stated to the following effect, and I think this puts it best:

"There is a real risk that UK companies may find themselves at a disadvantage relative to China's growing market power when it comes to defining standards for critical technologies, such as AI and autonomous systems. Should authoritarian governments achieve and sustain disproportionate influence in global standard setting bodies, there is a significant risk that the design specifications and standards underpinning the technologies that we rely on in our everyday lives will not be aligned with the fundamental principles of democracy, privacy, and human rights. These regulations have safeguarded the use of scientific knowledge rather than limited scientific advancement in that area."

I will finish on that, as we ponder, that is the technology that's creating our militaries.

[25] House of Commons Foreign Affairs Committee – **'Encoding values: Putting tech at the heart of UK foreign policy'** (July,2022). https://committees.parliament.uk/publications/22998/documents/168554/default/

Dr. Mariarosaria Taddeo, Associate Professor & Senior Research Fellow, Oxford Internet Institute

I have to say that what I am going to say is a reflection of my own views as informed by my research and not mirroring any of the views of the institutions I work for. I am also a philosopher and geneticist working on the ethical governance of digital technologists, particularly in defence and security – I've been doing this for about 15 years now.

AI's Usage in Defence

I want to start with giving some context. The use of AI in the context of weapons is just one element of a bigger trend, which has to do with growing embracement of AI in defence institutions for any sort of purpose, whether it is sustainment and support, adversarial or non-kinetic, cyber, or whether it is the kinetic. All these uses come with some important technical risks. Some of these risks also for AWS are germane to what we are seeing in other domains: healthcare, security, finance, lack of transparency, lack of accountability, lack of explainibility, and bias.

When it comes to AWS, we have a specific set of issues, on which we must focus. These are related to just war theory, which is the ethical theory that underpins international humanitarian laws. It has to do with human dignity, and it also has to do with military virtue.

Autonomous Weapons Systems (AWS)

Now to understand where these issues come from, we have to focus on what AWS are. I am going little bit philosophical here, but this is an important issue because at moment, there is no internationally agreed definition of AWS. My research team ran a comparative analysis, and we found twelve definitions provided by state actors and international actors – they are as diverse as you might imagine. Some of those are unrealistic: attributing and identifying AWS as systems who have intent, nothing foreseeable, nothing realistic and nothing concrete. Then others don't mention the key aspects that we are discussing here today, the learning abilities of these systems.

Lack of Predictability

The ethical questions that come with AWS emerge because we apply force to machines which are autonomous and learning. This learning element is crucial because it prompts another issue, which is the lack of predictability of the outcomes of these systems. We cannot predict with absolute certainty what these systems will do for all the reasons. For example, this is not a new issue. Nobert Wiener, the father of cybernetics, mentioned this in the 1960s, but we kind of forget about it.

The lack of predictability is something that emerges because of the very nature of AI – AI systems learn. Even when we can assume there were no mistakes, no errors, nothing went wrong, the system might develop new outcomes, which are logically sound yet unforeseen, unintended, and possibly unwanted. A lack of predictability is also a consequence of the vulnerability of this technology, which is technologies that are very susceptible to minor changes in the environment. It takes one pixel to change or alter the behaviour of an image-recognition system and to make it sure that it will mistake rivals for targets.

This lack of predictability is very problematic when we think about the ethics of AWS, because it hinders the control, not of the machines, but of the efforts of those machines, which is what we all care about. For example, the lack of predictability makes it impossible to transcribe more responsibilities to humans for the actions of those systems, because it separates intentions from actions. Ascribing more responsibility to humans is a key element for maintaining the variety of war. It was one of the key points stressed in the papers of the Nuremberg Trials.

The same lack of predictability makes it impossible for these machines to respect the principles of just war theory. Distinctions, for example, one of the key principles not putting non-combatants in harm's way. If we can operate on what the machine will target, what the machine will select, we can make sure that the outcomes will not be indiscriminate. This is why the Red Cross determined and defined the efforts of these systems indiscriminately called for ruling out of these machines.

Lethal Autonomous Weapons Systems (LAWS) vs. Autonomous Weapons Systems

My last point is that these issues are very pressing in both cases when we look at or think about LAWS and non-lethal AWS, but the outcomes of these two problems or for these two categories are different. When we think about LAWS, these problems show that these machines are morally impermissible - there is no way. However, when it comes to non-lethal uses, which are uses we see increasingly more commonly, well we can find compromises and trade-offs. Non-lethal weapon systems do pose ethical problems, they pose risks to proportionality, to distinction, they pose risk and more responsibilities. We need to find ways to find solutions to address these risks and these problems, and so far, we have not done that.

Recommendations.

My first recommendation is to do as much as it is possible to inform the policy debate. Let's try to separate the effort to define AWS from the political decisions you make about them. We first need to understand what the object is we want to regulate to identify it as objectively, and as precisely as possible. Then we can make whatever decisions we desire. Without that step the decisions are going to be flawed.

We need to have in place a process, and this is my second point, to understand what is morally permissible, but also socially acceptable when it comes to AWS. This is what allows us to define the threshold. This process has to be transparent and must be open, we must involve experts, we have to involve societies. We are democratic societies; the values are upheld by everybody in our societies. This cannot be a closed-door conversation. There should be a way of establishing an ethics review committee, which acts independently to monitor all possible uses of AI in defence. In particular when it comes to lethal LAWS and AWS. Looking at the nitty details, the procurement, the level of competitive skills and expertise of people who deploy them. We have a set of principles provided by the United Nations Convention on Certain Conventional Weapons about the use of these systems, which should provide guidelines that the UK should consider implementing on a daily basis to avoid the most atrocious or immoral users of these technologies.

4. Speaker Bios

EVIDENCE MEETING:

AI, NATIONAL SECURITY & DEFENCE: AUTONOMOUS WEAPONS

WEDNESDAY 7 SEPTEMBER 2022 5:30 PM UK PARLIAMENT - HOUSE OF LORDS

EVIDENCE GIVERS FROM LEFT TO RIGHT

- **The Rt. Rev Steven Croft, Lord Bishop of Oxford**
- **Taniel Yusef,** Tech Developers Coordinator, **UK Campaign to Stop Killer Robots**
- **Dr. Mariarosaria Taddeo,** Associate Professor and Senior Research Fellow, **Oxford Internet Institute**
- **Verity Coyle,** Senior Adviser, **Amnesty International**
- **Dr. Sidharth Kaushal,** Research Fellow, **Royal United Services Institute** (RUSI)
- **Dr. Daniel Clarke,** Head of Applied Research, **Rebellion Defence**

https://bicpavilion.com/about/appg-artificial-intelligence

The Rt. Rev Steven Croft, Lord Bishop of Oxford

The Rt Revd Dr Steven Croft became Bishop of Oxford in 2016 and was previously the Bishop of Sheffield. He has been a member of the House of Lords since 2013 and a member of the House of Lords Select Committee on AI in 2017-18 which produced the report AI in the UK: ready, willing and able, in 2018 he was appointed as one of the founding Board members of the UK Government's Centre for Data Ethics and Innovation and in 2019 became part of the Ada Lovelace Institute project on Rethinking Data. In 2021 he was appointed to the Lords Environment and Climate Change Committee. He is author of a number of books including *Ministry in Three Dimensions (1999 and 2008),* "Rooted and Grounded", faith formation and the Christian Tradition. (2019). His most recent book is Comfortable Words: a call to restoration: Reflections on Isaiah 40–55 (March 2021).

Publications:

- 'Diocese of Oxford – Bishop Steven's Blog' (2022). https://blogs.oxford.anglican.org/

Dr. Daniel Clarke, Head of Applied Research, Rebellion Defence

Dr. Daniel Clarke is a technology leader who has been involved in developing advanced technologies for over 15 years. Daniel has worked on a diverse range of technologies in a variety of domains and has experience in applying these across government, industry, and academia. Specifically, Daniel has deep technical expertise in the fields of sensor signal processing, artificial intelligence and Machine Learning, and cyber-electromagnetic activities. Throughout his career, Daniel has successfully developed advanced technologies across the defence and security, and automotive space and has seen his work demonstrated at large international trade fairs, and multi-national military exercises.

Daniel will soon join Rebellion Defence as the Head of Applied Research, driving the development of advanced AI techniques, and ensuring their effective application in support of the UK Armed Forces and her allies. Additionally, Daniel is a Lecturer at Cranfield Defence and Security at The Defence Academy of the United Kingdom, helping to educate the UK armed forces in advanced and emerging battlespace technologies. Previously, Daniel has worked as Technical Manager for Autonomous Driving at Siemens, and as a research scientist at the UK MOD's Defence Science Technology Laboratory (Dstl). In 2013, Daniel deployed to Afghanistan as the UK MOD's Scientific Advisor to Task Force Helmand, where he was awarded a commendation for his technical work.

Publications:

- 'Daniel Clarke' (2017) - https://blogs.sw.siemens.com/automotive-transportation/author/danielclarke/

Verity Coyle, Senior Advisor, Amnesty International

Verity works as a Senior Advisor on military, security, and policing at the International Secretariat of Amnesty International in London, where she covers various arms control issues, including the Arms Trade Treaty, Autonomous Weapons Systems and other emerging weapons and artificial intelligence technologies. Verity sits on the global Steering Committee of the Campaign to Stop Killer Robots, of which Amnesty International is a member.

While a non-resident fellow at the DC based think tank, the Stimson Centre, Verity's research focused on Gender-Based Violence as it relates to armed conflict and armed violence, resulting in a practical guide for licensing officials and their information sources, a factsheet developed with UNIDIR and the International Gender Champions, a regional training for officials in Eastern and Central Europe and numerous presentations to groups of State and civil society throughout the Conference of States Parties and Preparatory meetings and UNGA 1st Committee.

Publications:

- ''Killer Robots' – Coming soon to a battlefield near you? – Discussion chaired by Jeremy Khan' (2022). https://www.youtube.com/watch?v=ANjF_bbHGkQ

Dr. Sidharth Kaushal, Research Fellow, Royal United Services Institute (RUSI)

Sidharth Kaushal's research at the Royal United Services Institute (RUSI) covers the impact of technology on maritime doctrine in the 21st century and the role of sea power in a state's grand strategy.

Sidharth holds a doctorate in International Relations from the London School of Economics, where his research examined the ways in which strategic culture shapes the contours of a nation's grand strategy.

Publications:

- 'All Strategies Short of War: Getting the Most out of the Gray Zone' (2022). https://mwi.usma.edu/all-strategies-short-of-war-getting-the-most-out-of-the-gray-zone/
- 'RUSI Missile Defence Conference Report 2022' (2022). https://rusi.org/explore-our-research/publications/conference-reports/rusi-missile-defence-conference-report-2022
- 'Episode 18: The War in Ukraine & Taiwan's Defence Planning' (2022). https://rusi.org/videos/adversarial-studies/episode-18-war-ukraine-and-taiwans-defensive-planning

Taniel Yusef, Tech Developers Coordinator, UK Campaign to Stop Killer Robots

Taniel Yusef received two MAs before her LLM in International Economic Law, Justice, and Development, with associate study in Laws of Armed Conflict, Trade, Peacebuilding, and others. She is UK International Representative for Women's International League for Peace and Freedom. WILPF, the oldest women's peace and security organisation in history, with two Nobel Prizes, is the designated NGO coordinator of several United Nations' meetings: First Committee, Non-Proliferation Treaty and Cyber Security. It has been instrumental in numerous treaties across disarmament, and other, fora. In this capacity, Taniel advocates in the UK, Brussels and UN on Trade, Economics and Disarmament Affairs, including political economy, weapons technologies and business and human rights. As Technology Developers Coordinator for the UK Campaign to Stop Killer Robots, she coordinates and produces research highlighting problematic technological issues across the Lifecycle of Autonomous Weapons Systems (AWS) and associated behavioural and commercial concerns.

Taniel is Visiting Lecturer in Humanitarian Intervention MSc, (University of East London), contributing editor to the European Women's Lobby-Feminist Economics Working Group, specialising in human security and militarism, and has researched women's access to resources on the ground. As member of the International Campaign to Abolish Nuclear weapons (ICAN), she works on nuclear disarmament, including negotiation and adoption of the Treaty to Prohibit Nuclear Weapons (TPNW), for which ICAN received the Nobel Peace

Prize in 2017. Current research includes nuclear winter fallout; autonomy and quantum's risks to nuclear; complex systems, supply-chain-shock, and emergent economies; and emergent unpredictability in AWS and their interaction with law.

Publications:

- The Lifecyle of Lethal / Autonomous Weapons Systems: Outstanding Technological Concerns – UK Campaign to Stop Killer Robots (2021). https://www.wilpf.org.uk/wp-content/uploads/2021/11/UK-CSKR-THE-LIFECYCLE-OF-LETHAL-AUTONOMOUS-WEAPONS-SYSTEMS-OUTSTANDING-TECHNOLOGICAL-CONCERNS.Nov2021.pdf

Dr. Mariarosaria Taddeo, Associate Professor and Senior Research Fellow, Oxford Internet Institute

Mariarosaria Taddeo is Associate Professor and Senior Research Fellow at the Oxford Internet Institute, University of Oxford, and is Defence Science and Technology Fellow at the Alan Turing Institute. Her work focuses mainly on the ethical analysis of artificial intelligence (AI), ethics of AI for national defence, cybersecurity, cyber conflicts, and ethics of digital innovation. Her area of expertise is Digital Ethics. Her research has been published in major journals like Nature, Nature Machine Intelligence, Science, and Science Robotics.

Since 2019, Professor Taddeo leads a Dstl (Defence Science Technology Laboratory, Ministry of Defence UK) funded research project on the Ethics of AI in National Defence. Since 2020, Taddeo serves in the Ethics Advisory Committee to the UK Ministry of Defence.

She has received multiple awards, the 2010 Simon Award for Outstanding Research in Computing and Philosophy; the 2016 World Technology Award for Ethics. In 2018, InspiringFifty named her among the most inspiring 50 Italian women working in technology. ORBIT listed her among the top 100 women working on Ethics of AI in the world. She is one the twelve 2020 "Outstanding Rising Talents" named by the Women's Forum for Economy and Society. Since 2016, Taddeo serves as editor-in-chief of Minds & Machines (SpringerNature) and of Philosophical Studies Series (SpringerNature).

Publications:

- Taddeo, M and Blanchard, A (2022) ‘Ascribing Moral Responsibility for The Actions of Autonomous Weapons Systems: A Moral Gambit’. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4096934
- Taddeo, M and Blanchard, A (2021) ‘A Comparative Analysis of the Definitions of Autonomous Weapons’. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3941214

5. Contact

APPG AI Secretariat

Big Innovation Centre
14-16 Dowgate Hill
London EC4R 2SU
United Kingdom

info@biginnovationcentre.com
www.biginnovationcentre.com

appg@biginnovationcentre.com
https://bicpavilion.com/about/appg-artificial-intelligence

www.biginnovationcentre.com