UKRI Trustworthy Autonomous Systems Hub | The University of Nottingham | University of Southampton | The Policy Institute | King's College London

Trusted autonomous systems in defence

A policy landscape review

Juljan Krause

November 2021

About the TAS Hub

The UKRI TAS Hub assembles a team from the Universities of Southampton, Nottingham and King's College London. The Hub sits at the centre of the £33M Trustworthy Autonomous Systems Programme, funded by the UKRI Strategic Priorities Fund.

The role of the TAS Hub is to coordinate and work with research nodes to establish a collaborative platform for the UK to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice by individuals, society and government. Read more about the TAS Hub here.

Acknowledgement

The author would like to thank Prof Mark Kleinman, Rachel Hesketh and Dr Justyna Lisinska from King's College London, Prof Stephen Meek from the University of Nottingham, Gareth Giles from Public Policy | Southampton and Linda Slapakova and James Black from RAND Europe for their thoughtful and insightful comments on earlier versions of this review

Any errors or omissions are of course the author's responsibility. This work was supported by the Engineering and Physical Sciences Research Council (EP/V00784X/1), and also by the Policy Associates programme at the University of Southampton.

Contact

Juljan Krause is a PPS Policy Associate and PhD researcher at the University of Southampton.

The Series Editor is Professor Mark Kleinman, Professor of Public Policy at the Policy Institute.

Citation: Krause J. 'Trusted autonomous systems in defence – a policy landscape review. Available from: https://doi.org/10.18742/pub01-063

Abbreviations

AI	Artificial intelligence
DARPA	US Defense Advanced Research Projects Agency
DASA	Dstl's Defence and Security Accelerator
DIU	US Department of Defense's Defense Innovation Unit
DOD	US Department of Defense
DSIS	Defence and Security Industrial Strategy
DSTL	Defence Science and Technology Laboratory
ML	Machine learning
MOD	UK Ministry of Defence
NATO	North Atlantic Treaty Organisation
NSIB	National Security and Investment Bill
RAS	Robotic and autonomous systems
R&D	Research and development
SACs	Systems with Autonomous Capabilities
TAS	Trustworthy Autonomous Systems
TAS-Hub	Trustworthy Autonomous Systems Hub

Executive summary

This review discusses some of the challenges to building trustworthy autonomous systems for defence that arise for the UK and its allies. It is part of a series of TAS policy landscape reviews that survey the issues that emerge around designing trustworthy autonomous systems. This review focuses on trust in novel autonomous systems in the defence context, and the policy problems that are associated with this.

What do we mean by autonomy?

- When it comes to autonomy in defence systems, it is a **question of degree** rather than being a case of a simple yes or no. It is therefore more appropriate to talk about Systems with Autonomous Capabilities (SACs). Autonomy raises numerous practical questions about the **trustworthiness** of SACs and the **degree and the quality of human involvement** in human-machine teaming.

- Autonomy in SACs is a complex issue that is driven by (at least) three variables: 1) the **character and quality of the command-and-control relationships** between human operators and machines; 2) the **types of decisions** that are to be fully or partially delegated to SACs; and 3) the **sophistication, maturity, and reliability** of SACs. Autonomy is, therefore, best construed as an umbrella term that encompasses all of these factors.

Legal and ethical challenges

- Presently, the UK and North Atlantic Treaty Organisation (NATO) do not believe that SACs require an amendment to international law or a new treaty. While ethical and legal norms may evolve over time, at present the UK government considers the development and deployment of SACs to be **fully covered under existing legislation** if they are designed in compliance with international humanitarian law.

- While most ethical reflections on SACs draw attention to potentially troublesome implications of delegating decision-making to machines, some commentators suggest SACs will be **more precise** and governed by pre-defined rules and are therefore **ethically desirable**.

- While the Ministry of Defence (MOD) points to **efficiency gains** of SACs and how they are likely to **put service personnel out of harm's way**, they do present several new challenges for policymakers. These include questions of (algorithmic) **accountability** and fears of an **erosion of responsibility**. The deployment of SACs requires strong codes of conduct and robust, transparent processes of decision making.

Public opinion

- Opinion research points to **considerable public opposition in both the UK and the US to the use of lethal SACs**. However, there is a significant fall in public opposition to SACs if their deployment is **contextualised** and not framed in terms of life or death, or human out-of-the-loop decisions. Given the proliferation of autonomous systems in other domains, such as health care and autonomous vehicles, some commentators find an overall increase in public acceptance of autonomy which may spill over to the domain of SACs. However, the empirical picture is **mixed and likely to remain volatile**.

Human-machine teaming

- Trust in SACs among the armed forces is currently underdeveloped. **Closing trust gaps in the military requires consistent long-term engagement**. The empirical picture in this area is **significantly under-researched**, which limits the evidence base.

- Human-machine teaming places new demands on the **education, recruitment, and ongoing training** of military personnel. Human operators of future SACs will require extensive training. Policymakers should think strategically about recruitment and how the armed forces can attract and retain talent against strong private sector competition.

Innovation

- To a significant extent, novel SAC technologies will be developed in collaboration with new industry partners that may have no history of a prior engagement with the MOD. Collaboration may continue to **blur the boundaries between public and privately funded research**. The 2021 Defence and Security Industry Strategy (DSIS) seeks to better align government research and development (R&D) and innovation spending with private sector activity.

- In many cases, SACs incorporate **dual purpose technologies**. This raises questions about the **proliferation of inventions**, **intellectual property rights** and the **involvement of foreign industries**, some of which are addressed in the 2020 National Security and Investment Bill (NSIB).

Introduction

At the Strategic Command Conference in May 2021, Commander General Sir Patrick Sanders called for the UK to develop a "bold and clear" vision if it hopes to retain its competitive edge in defence (UK Strategic Command, 2021, p. 2). Above all, defence must "sense and understand threats and opportunities with greater fidelity and respond more rapidly in the information age" (ibid.). Artificial intelligence (AI) and autonomous systems will form integral elements of this vision to build a more agile and responsive armed forces. In its recent Integrated Review, the government declares advancing AI capabilities as the cornerstone of its defence and security policy in the years to come. It has committed £6.6bn over the next four years to R&D in order to accelerate the "pull-through" of new technologies in this space (HM Government, 2021a, p. 73).

> "With the issue of integration across the armed services being particularly widely discussed at the moment, autonomous systems are likely to add to the complexity of the already mammoth task of multi-domain integration"

In the near term, autonomous systems can be expected to find early applications in domains with limited complexity so that machine learning (ML) techniques can be effectively applied. In situations that are "bounded and fast-moving", the MOD outlines in its recent Joint Doctrine on Multi-Domain Integration, "the emphasis is likely to be on high tempo through automation and autonomy" (Ministry of Defence, 2020, p. 29). It is hoped that autonomous systems will realise efficiency savings and put service personnel out of harm's way. Military strategists in the US seek to "enhance" capabilities such that "fewer soldiers are required for robot control as RAS [robotic and autonomous systems] perform dull, dirty and dangerous tasks on their own" (US Army Training and Doctrine Command, 2017, p. 3). With the issue of integration across the armed services being particularly widely discussed at the moment (Ministry of Defence, 2020), autonomous systems are likely to add to the complexity of the already mammoth task of multi-domain integration.

While there are numerous challenges that AI and automation present in the context of defence and national security policy more broadly, this review focuses on the particular issues around trust and trustworthiness that autonomous systems in defence are likely to pose. It is by no means a complete account of all the complexities that autonomy poses for defence and security policymakers. Rather, it presents a high-level overview of the most pressing concerns facing multidisciplinary teams that are tasked to design trustworthy systems for UK defence, or are concerned with the larger policy implications of these systems.

In collating some of the major policy issues that emerge from building autonomous systems for defence applications, this review draws on desk research and a review of publicly available sources. While this work is exploratory and cursory in nature, and the literature review is mostly limited to the grey literature, it aims to survey the most relevant areas of concern. The introductory section covers relevant definitions that circulate in the literature. Following NATO terminology (NATO Allied Command Transformation, 2016), it develops the term Systems with Autonomous Capabilities (SACs) and argues that autonomy in the context of defence is a question of degree rather than a binary variable. Section two covers cross-cutting policy themes in defence that extend to other TAS policy lanscape reviews, such as the legal challenges, ethical concerns, and public opinion issues around autonomous systems. Section three reviews domain-specific questions and explores how trust in autonomous systems among the armed forces can be improved. It also discusses some of the novel challenges that human-machine teaming presents and how private sector innovation in defence complicates building trust. Section four concludes and cautions against short-term policy solutions, presenting the need for a nuanced approach that takes a long-term view.

Definitions and concepts

What is autonomy in defence?

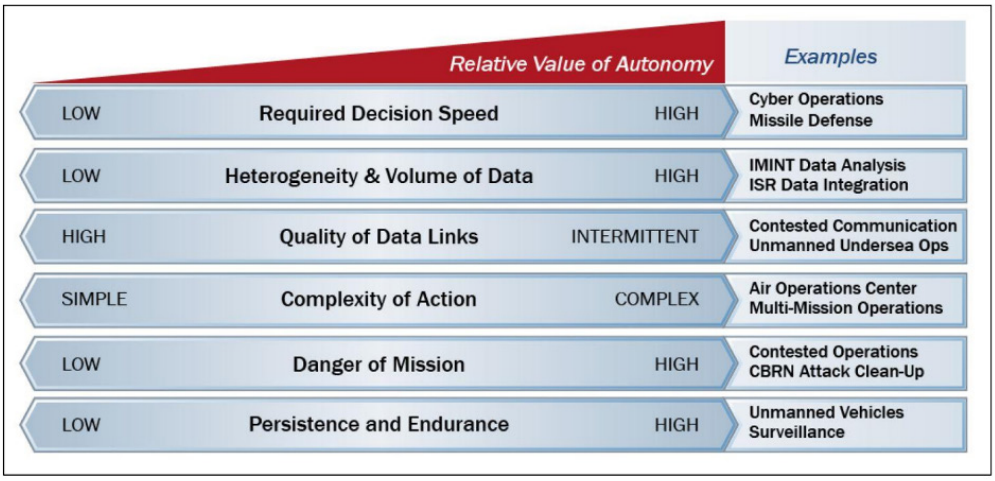
Autonomy is a laden term, which matters in a discussion of how trustworthy autonomous systems can be designed and trust levels improved. The Trustworthy Autonomous Systems Hub (TAS-hub) defines an autonomous system as “a system involving software applications, machines, and people, that is able to take actions with little or no human supervision” (TAS-Hub, 2020). Traditionally, as far as human behaviour is concerned, the term autonomy suggests a capacity for independent decision-making and rational reflection. It captures the competency of agents to ground their actions in moral and ethical principles (Schneewind, 1998). Autonomy in humans, therefore, suggests a capacity for rational behaviour and a reflective attitude. Machines are not wired this way, so this is obviously not what “autonomy” seeks to convey in the context of defence.

Given that, in ordinary language, the term “autonomy” applies to minded creatures, when we talk about “autonomous systems” in public discourse, it often brings to mind sci-fi horrors of grotesque anthropomorphic machines that inevitably turn on their creators. This misperception among parts of the public has prompted the US Navy to publicly rule out a “Skynet and a Terminator scenario” for future warfighting (Larter, 2019). More generally, public discussion of autonomous systems is often conflated with discussion of artificial intelligence.

In the context of defence, “autonomy” has a much more precise meaning. The MOD defines autonomy primarily in terms of the capacity of autonomous systems to pursue a course of action with limited human intervention: an “autonomous system is capable of understanding high-level intent and direction” (Ministry of Defence, 2017a, p. 13). For the MOD, this enables an autonomous system to “understand” and “perceive” its environment such that “it is able to take appropriate action to bring about a desired state” (ibid.). Ultimately, autonomy involves the capacity to decide “a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present” (ibid.). While the exact levels of human involvement will vary depending on the character of autonomous systems and their task and mission environment, the development of radical human-out-of-the-loop architectures that remove human oversight entirely do not seem high on the MOD’s agenda.

Realistically, autonomous systems are hoped to save resources, protect the armed forces, and extend capabilities. For the MOD, “cheap, smart” autonomous systems “can provide resilience, greater persistence, mass and political choice at reduced cost” (Ministry of Defence, 2017b, p. 13). Similarly, the US Army aims for “RAS to perform higher risk missions for longer duration, expand operational depth and stand-off distance, and allows soldiers to focus on those missions humans do best” (US Army Training and Doctrine Command, 2017, p. 3). Examples of applications include intelligence collection and analysis, the clean-up of contaminated environments, or clearing the routes of explosive devices. In general terms, the US Department of Defense (DOD) suggests that the value of autonomous systems rises sharply with the speed, complexity, and danger of missions (see Figure 1).

FIGURE 1: VALUE OF AUTONOMY TO THE DEPARTMENT OF DEFENSE (DOD) DEFENCE SCIENCE BOARD



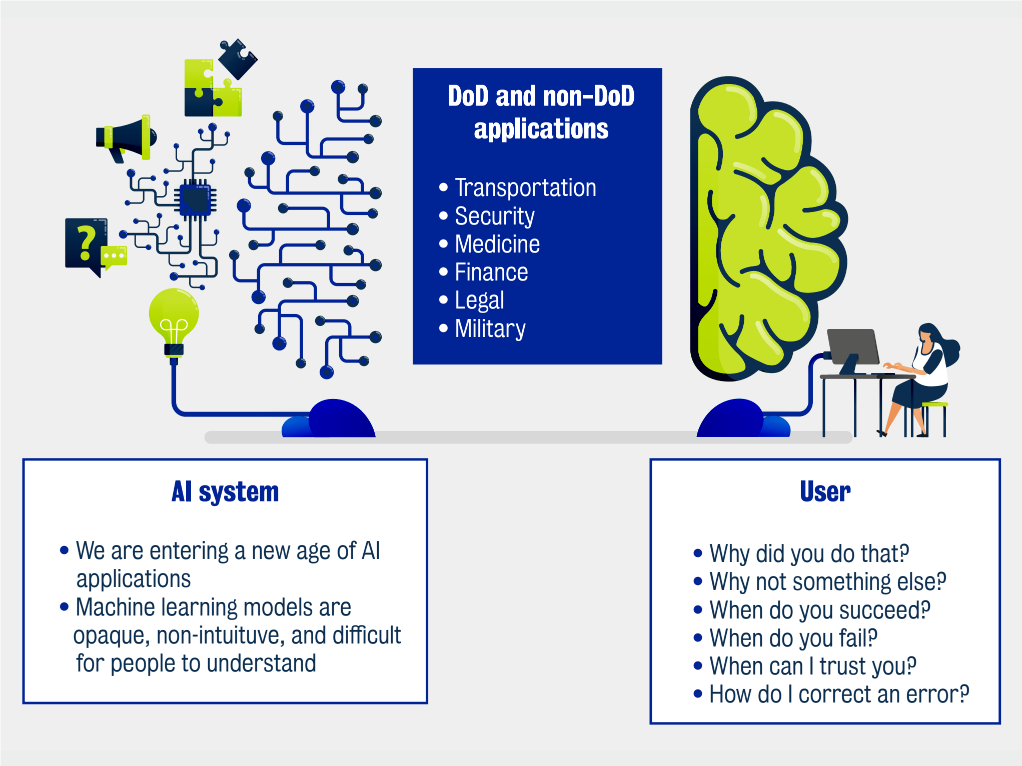
Source: (Sayler, 2020, p. 28)

> "The multi-layered and non-linear structure of the machine learning that powers many autonomous systems can significantly erode the traceability and explicability of decision-making processes"

Building trustworthy autonomous systems that can achieve such missions is not an easy task. The issue is amplified by the lack of transparency of many ML models. The idea of the "black box" in machine learning is where the decision-making process of the AI is completely opaque, to the point where even its designers can't explain why an AI arrived at a specific decision (Boulanin and Verbruggem, 2017). By contrast, Explainable AI is artificial intelligence in which the results of the solution can be understood by humans. Recent experimental successes in Explainable AI – such as DARPA's XAI programme – notwithstanding, the multi-layered and non-linear structure of the machine learning that powers many autonomous systems can significantly erode the traceability and explicability of decision-making processes.

Opacity is not the only issue that stands in the way of building trust for autonomous systems. Training data may reflect selection biases and therefore may reproduce discriminatory practices (Hajian et al., 2016; Kirkpatrick, 2016; Skiena, 2011). Trust in a particular system will also depend on the degree to which its proprietor, inventor and issuing organisation are trusted (Rossi, 2019). At the other end of the spectrum, over-reliance also poses challenges of its own. A senior commander in the US Navy sums up the sentiment: "We have to be very careful… that we don't over-trust" (Larter, 2019).

FIGURE 2: THE ISSUE OF MACHINE LEARNING OPACITY FOR DEPARTMENT OF DEFENSE APPLICATIONS



Source: Recreated from https://www.darpa.mil/program/explainable-artificial-intelligence

> "Potentially catastrophic outcomes when machines are literally left to its own devices in combat would surely erode trust in autonomous systems quite significantly among combatants"

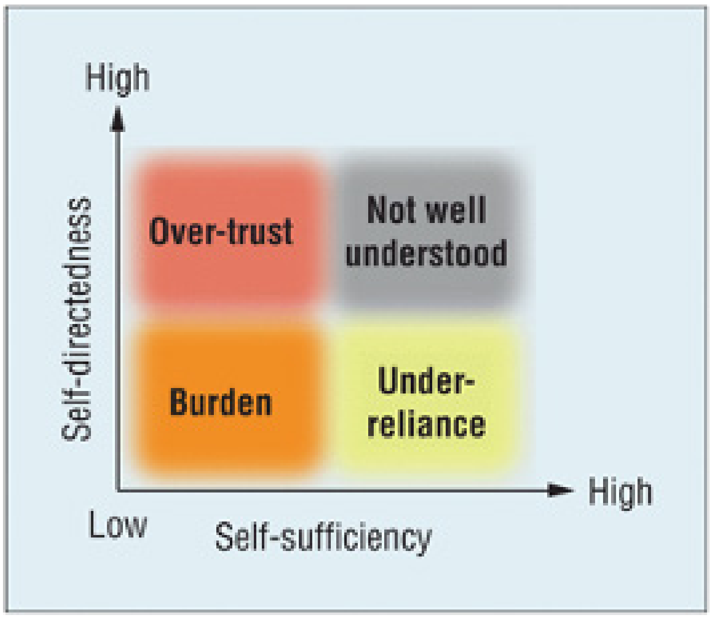
Considering the various policy challenges that autonomous systems present – as discussed throughout this review – engineers and designers face a trade-off (Bradshaw et al., 2013). If an autonomous system displays high degrees of self-sufficiency in that it can execute independently varied and complex decisions in challenging environments, it reduces the demands on human operators. When perceived as effective, machines that display high levels of self-sufficiency should generate trust, even if this relationship will not be linear (ibid.). However, a fully developed system of this kind that is self-directing – in the sense that it can make combat-relevant decisions on its own and therefore does not require much human oversight – could potentially have disastrous consequences in the case of error or malfunctioning. Potentially catastrophic outcomes when machines are literally left to their own devices in combat would surely erode trust in autonomous systems quite significantly among combatants.[1]

On the other hand, an autonomous system that lacks self-sufficiency yet enjoys large degrees of freedom to make (minor) decisions itself may cause issues of over-trust. Conversely, granting a capable and well-developed autonomous system too little leeway in making decisions quickly translates into under-reliance and missed opportunities for deploying the system in the best possible and most effective way. While autonomy is certainly not the only variable that is integral to generating trust in an autonomous

1 This is not to suggest that human combat that does not involve autonomy and AI would be superior, ethically or otherwise. The particular problems of flawed human decision-making during combat – ranging from oversight and negligence due to fatigue to outright abusive, criminal or vengeful behaviour and war crimes – are not the topic of this policy review. Against each of the particular challenges of SACs, as outlined in the sections that follow, counterarguments can be made that purely human decision-making would certainly be problematic as well. While this is no doubt the case, this review focuses on the challenges of SACs and therefore omits a detailed evaluation of arguments about the perils of human decision-making.

system it is a paramount one, and striking the right balance between self-sufficiency and self-directedness to generate trust will be a challenging task (Abbass et al., 2018).[2]

FIGURE 3: CHALLENGES THAT EMERGE IN THE DESIGN OF AUTONOMOUS SYSTEMS



Source: Bradshaw et al., 2013, p. 55

Degrees of autonomy

The nuances in definition and variety of challenges suggest that autonomy in defence is not a binary concept. Autonomy is very much conditional upon the timescale under consideration. In practical terms, it is a question of degree rather than a yes-or-no binary. NATO points out that, at the current stage of development, truly autonomous complex systems do not yet exist (NATO Allied Command Transformation, 2016). For this reason, autonomy seems best described in terms of the overall capabilities of a system. Therefore, it may be more appropriate to speak of Systems with Autonomous Capabilities (SACs) rather than autonomous systems (NATO Allied Command Transformation, 2016). This review follows this definition.

Autonomy in SACs is a continuous function of (at least) three variables (Scharre, 2017, p. 9): i) the character and quality of the command-and-control relationships between human operators and machines; ii) the types of decisions that are to be fully or partially delegated to SACs; and iii) the sophistication, maturity and reliability of SACs.

These entanglements make SACs complex sociotechnical systems and arrangements that involve military and civil organisations, policies, control systems and, of course, human behaviour. The MOD is optimistic about SACs and suggests the following "defence illustrations" in the near- and long-term future (Ministry of Defence, 2019):

- Replacing human operators with machines in **high-risk environments**.
- Delegating **simple and low value tasks** to machines.
- **Exceeding the performance of a human operator** by acting autonomously.

2 The Partnership on AI provides a comprehensive literature review on wider research on trust issues in relation to AI.

- **Generating physical mass** in the battlespace through resilient swarms of low-cost systems.
- **New ways of operating**, including through integrated human-machine teams.
- **Supporting an active military presence** in areas where it would not traditionally be possible.

Given these high expectations, many new policy issues emerge around increasing the trustworthiness of specific applications, as well as driving the acceptance of SACs in society at large.

Cross-cutting policy challenges

Legal and ethical questions

For policymakers and advocacy groups, SACs cause widespread concerns that cut across policy domains. While specific challenges are likely to materialise quite differently subject to the degree and actual character of autonomy involved, and human warfighting has always presented a plethora of ethical and legal questions, novel challenges may be captured broadly in the following diagram (see Figure 4).

FIGURE 4: TAXONOMY OF CHALLENGES FACING SYSTEMS WITH AUTONOMOUS CAPABILITIES

Source: Partially adapted from Article 36, 2020, p.5

> Algorithmic biases may extend to military operations such that targeting/killing could be based on (unknowingly) encoded indicators of gender, race, or other identities"

Digital dehumanisation[3]

Algorithmic biases may extend to military operations such that targeting/killing could be based on (unknowingly) encoded indicators of gender, race, or other identities. The issue of bias in coding has attracted considerable academic attention over the past couple of years. A large body of empirical research points to instances of discrimination based on race and gender that were encoded in automated systems, eg with regards to loan and job applications and facial recognition (Baeza-Yates, 2016; Kirkpatrick, 2016; Knight, 2017; O'Neil, 2017). There is no reason to assume that the design and programming of military applications would be free of any bias of this kind. In fact, senior US intelligence officers have warned against inadvertent and accidental bias in military AI in particular (Pomerleau, 2020).

New dangers to civilians

SACs may further displace violence from militaries onto civilians. They may erode norms and marginalise compassion and human judgement in human-out-of-the-loop

3 Dehumanisation captures the process of depriving an individual or groups of people of their special human qualities, attributes and rights. For critics of SACs, autonomous systems pose the risk of further reducing human beings to mere numbers.

“

SACs may... erode norms and marginalise compassion and human judgement in human-out-of-the-loop operations”

operations. With little or no human oversight, new sources of harm to civilians may go unnoticed and evade oversight and public scrutiny (Human Rights Watch, 2020). Many military strategists, however, disagree that SACs will necessarily undermine and erode international humanitarian law. For the US Army, for instance, SACs may act “more humanely” as ‘they do not need to be programmed with a self-preservation instinct, potentially eliminating the need for a ‘shoot-first, ask questions later’ attitude” (Etzioni and Etzioni, 2017, p. 74). Since SACs respond unclouded by emotions, they may implicitly reinforce international humanitarian law.

Erosion of responsibility

Personal accountability may be avoided due to ever more diluted understandings of where, when and to whom force will be applied and to what degree. In opaque command-and-control relationships, it is not obvious where accountability for life-or-death decisions will ultimately lie. This issue does not arise only from the introduction of autonomous systems: the effective integration of complex non-linear “kill webs” of sensors and shooters in existing distributed kill chains already presents ample challenges (O’Donoughue et al., 2021).

Risk of escalation

The development and deployment of SACs is likely to trigger a response from adversaries to increase their efforts in this domain in a similar fashion. In a scenario of SAC-on-SAC combat, the speed of interaction between SACs may leave little room for meaningful human decision making. This may escalate conflict unnecessarily.

Deterrence

Automation may increase the appetite for armed conflict and invite automation in response. Blending human and machine decision-making complicates traditional assumptions about signalling the potential use of force, which has considerable ramifications for deterrence altogether (Wong et al., 2020).

Legal challenges

The UK government frequently points to international law as the foundation of its foreign policy and military operations (HM Foreign and Commonwealth Office, 2017; HM Government, 2018; Ministry of Defence, 2019). In 2017, then Attorney General Jeremy Wright confirmed that “the UK should and will only use armed force, and will only act in self-defence, where it is consistent with international law to do so” (Wright, 2017). The UK government affirms this position in its 2021 Integrated Review where it details its “commitment to leadership in NATO, supporting its adaptation to threats above and below the threshold of war under international law” (HM Government, 2021a, p. 72). Some of the £24bn increase in defence spending over the next four years announced in the Review is earmarked for the development of “near-peer, high-tech warfighting – and a ‘digital backbone’ to enable multi-domain operations and interoperability with allies and partners” (ibid.). The Review points to the UK’s renewed commitment to the development and deployment of SACs.

The UK’s “clear position” is that international law “is the applicable legal framework for the assessment and use of all weapons systems in armed conflict”, which includes SACs (UK government, 2015). The UK government adds that “there must always be human oversight and control in the decision to deploy weapons" (ibid.), which effectively rules out fully-fledged human-out-of-the-loop SACs that would autonomously select, cue and kill human targets.

Presently, the US government seems to have no desire to develop lethal, fully human-out-of-the-loop SACs: "it is DoD policy that autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force" (US Department of Defense, 2017, p. 2). However, this picture may change. In early 2021, the US National Security Commission on Artificial Intelligence recommended to restrict human oversight to the deployment of nuclear weapons. Given the risk that adversaries may wish to develop lethal, fully human-out-of-the-loop SACs, their future deployment should not be ruled out. As a minimum, the US government is recommended to "develop international standards of practice for the development, testing, and use of AI-enabled and autonomous weapon systems" (National Security Commission on Artificial Intelligence, 2021, p. 10).

The renewed commitment of the UK government to the development and deployment of SACs as outlined in the Integrated Review, alongside the planned rollout of new offensive cyber warfare systems, presents some legal challenges. Referring more broadly to automation, rather than autonomous systems, the Australian government, for instance, warns that the legality of automated warfare "may trump technical issues in terms of impediments to automation" (Pilling, 2015, p. 2). More narrowly, the legal challenges as far as SACs are concerned may be summarised in the following list (see Figure 5).

FIGURE 5: OVERVIEW OF LEGAL CHALLENGES

- **The legality of weaponised SACs per se**
- **The legality of the use of such systems in different kinds of engagement and combat**
- **The legality of these systems in non-military domains (law enforcement, self-defence, etc)**
- **State responsibility for harms caused by SACs**
- **Criminal responsibility for harms caused by SACs**

With no specific international treaty on SACs in sight, commentators (such as Wyatt, 2021) warn against the prospect of diffusion of lethal autonomous weapon systems among terrorist groups and rising middle countries that have so far been ignored by the great-power centric literature. While the UK, the US and NATO maintain that SACs are commensurable with existent legal frameworks and thus require no new international treaty (NATO Allied Command Transformation, 2016), some academics argue that "in this age of extraordinary technological change" automated weapon systems do require a new "international dialogue" (Allenby, 2014), which may or may not translate into a new legal framework.

"Advocates of SACs point to higher degrees of precision and reliability that autonomous systems are likely to afford given that human judgement may be clouded by mental stress, prejudice, or emotional impulse in high-stress scenarios"

Ethical concerns

Advocates of SACs point to higher degrees of precision and reliability that autonomous systems are likely to afford given that human judgement may be clouded by mental stress, prejudice, or emotional impulse in high-stress scenarios. Overall, SACs may therefore reduce collateral damage and the number of civilian casualties (National Security Commission on Artificial Intelligence, 2021). In practice, SACs may be "more humane than humans" as they are designed to "put themselves in harm's way (and so are less inclined to shoot first), and do not have a tendency to rape, to act in revenge or panic" (Pilling, 2015, p. iii). While this line of argument supports SACs and constructs them as ethically desirable, it may pose challenges to the government's communication strategy as it assumes a "tendency" of armed forces towards impermissible behaviour or outright war crimes and/or accidental casualties or conflict escalation.

Opponents of SACs, on the other hand, fear that remote-controlling or distant oversight create a "PlayStation" mentality among human operators (Lee, 2017; Walsh, 2010), which is said to induce a propensity to apply excessive force. This problem could be exacerbated by a lack of clarity on areas of individual responsibility, as pointed out above. Among commentators, there is considerable unease about delegating life-or-death decisions to machines. For advocacy groups, such as Human Rights Watch, the issue constitutes a "moral and legal imperative" to ban fully automated SACs (Human Rights Watch, 2018).

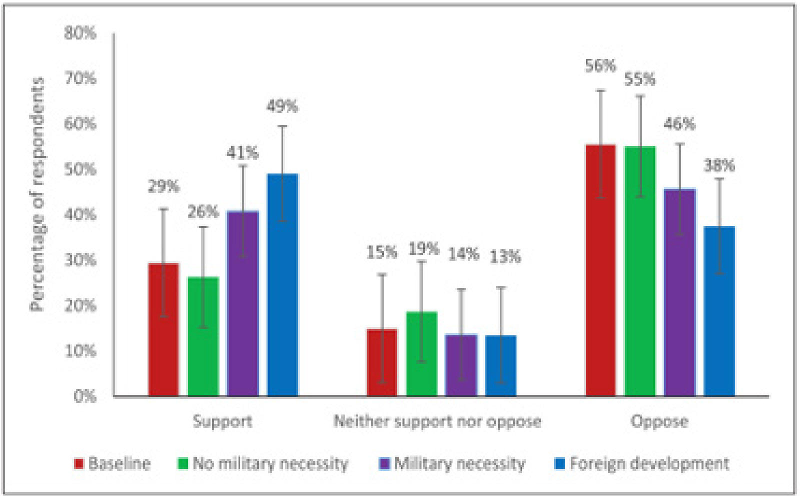
Any specific ethical challenge will depend on the horizon and timescale under consideration. In the near term, a blurring of the domains of individual (personal) responsibility and accountability in partially automated systems where human operators retain critical functions can be expected to present the most pressing challenges. In the long run, the implications of fully automated SACs that identify, trace, cue and fire at targets with no human involvement at all are likely to cause concern. Near term, hybrid SACs can therefore be expected to present more pressing ethical challenges because complete and complex human-out-of-the-loop systems, even if considered desirable, seem largely aspirational at the current stage of development.

Public opinion

Developing trustworthy systems in defence requires a sound understanding of the public opinion landscape on the issue, as public attitudes will shape policy positions. In the US, impressionistic evidence and early research suggested widespread opposition to SACs (Carpenter, 2013). However, the picture appears to have shifted over time. While opposition is still considerable, it is found to be highly context-dependant. In general terms, the US public seems to oppose SACs, in particular if they are framed as lethal weapons. Yet opposition to SACs falls significantly if a blanket ban is considered to be putting US armed forces at risk or if other countries are feared to develop lethal SACs (Horowitz, 2016).

Public opinion research in this area cautions against the view that the US public would uniformly support an outright ban of SACs, as the data from Horowitz (2016) indicate (see Figure 6). Opposition to the development of domestic SACs drops by nearly 20 percentage points if adversaries develop these systems too. Opposition drops by ten percentage points if SACs are deemed a "military necessity" (see Figure 6).

FIGURE 6: PUBLIC OPININ ON SYSTEMS WITH AUTONOMOUS CAPABILITIES IN THE US IF ADVERSARIES DEVELOP LETHAL WEAPONS

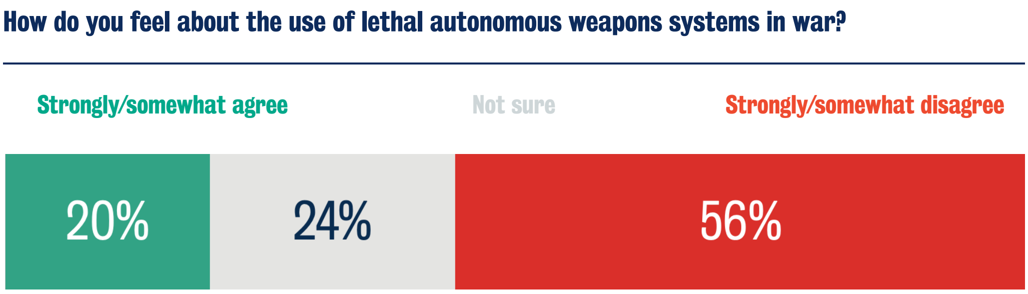


Source: Horowitz, 2016

In the UK, an [Ipsos MORI poll] published in February 2021 suggests that 56.3 per cent of people in Britain either strongly oppose or "somewhat oppose" lethal SACs. This represents a small decline from earlier polls.

Respondents cited the following reasons for their opposition (in order or magnitude): moral and ethical concerns, lack of accountability, concerns about technical failures and the view that such systems are illegal (Ipsos, 2021). The survey therefore suggests a wide range of challenges for policymakers if public opinion is a decision variable for the scale of future deployments of SACs. Among the British public, trust in SACs is by no means assured (see Figure 7).

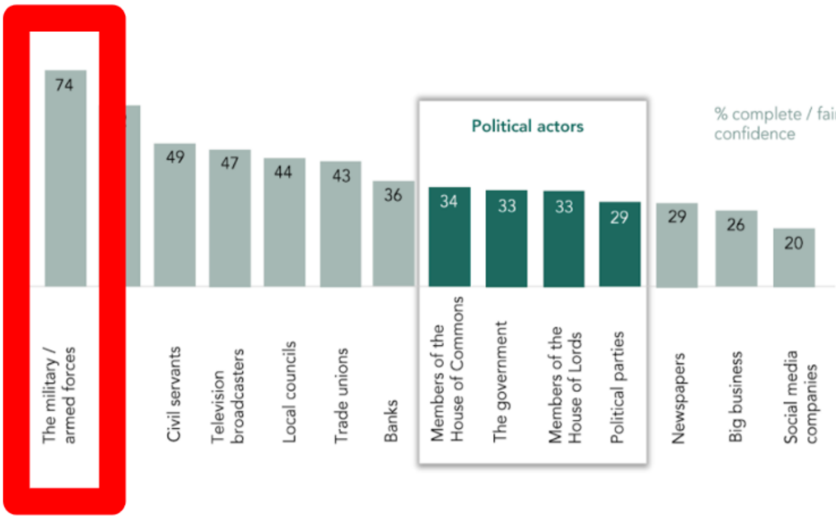
FIGURE 7: PUBLIC OPINION ON SYSTEMS WITH AUTONOMOUS CAPABILITIES IN THE US IF ADVERSARIES DEVELOP LETHAL WEAPONS



Source: Ipsos MORI 2021

The data suggest that the relative lack of trust is SAC-specific and is not due to a general opposition to the military. On the contrary, the UK's armed forces enjoy overwhelming support. According to a 2019 survey published by the Hansard Society, the armed forces enjoy by far the greatest support of the public in comparison with other institutions (Hansard Society, 2019). 74 percent of respondents have "complete" or "fair" confidence that British Armed Forces act in the best interests of the public (see Figure 8).

FIGURE 8: CONFIDENCE IN BRITISH INSTITUTIONS



Source: Hansard Society, 2019

When it comes to generating public trust for SACs, the data, therefore, suggest that opposition is likely to be specific to the use of autonomous capabilities, which points to a considerable trust deficit towards SACs. However, some commentators believe that opposition is likely to erode over the years to come. A recent Chatham House report, for instance, suggests that the diffusion of autonomous systems in other domains will gradually increase levels of support for SACs among the general population. The expected speed at which commercial systems will become available "could normalise the acceptance of autonomous systems for the military and the public" (Cummings, 2017, p. 1). A gradual change in public attitudes towards autonomy "could encourage state militaries to fund the development of such systems at a level that better matches investment in manned systems" (ibid.).

Defence-specific policy themes

Future SACs are hoped to support the military across a wide range of missions as outlined in the recent Defence Technology Framework (Ministry of Defence, 2019): delegating personnel-intensive low-level tasks to machines should make for more efficient resource consumption, while replacing human operators with machines in high-risk environments puts service personnel out of harm's way. Where replacement is not possible, SACs will be tasked to improve the performance of human operators. Novel forms of integrated human-machine teaming should make for a more agile and responsive force.

For this scenario to become reality, building trust among current and future users in the armed forces will be of paramount importance. This is likely to present challenges to internal military training policies. As complex SACs require novel skillsets, the rollout of SACs places new demands on education, training and recruitment policies and requires considerable integration with academia and industry. Such integration is not new, but SACs involve novel dual-use technology platforms and interfaces that require teaming up with new technology partners the MOD may have had no previous relationships with, or that have little appetite to collaborate with UK defence. For instance, the public protest of senior engineers at Alphabet against contracts with the Pentagon "became an identity crisis for Google" (Shane, 2018), which prompts questions about the risks of these new forms of collaborative engagement.

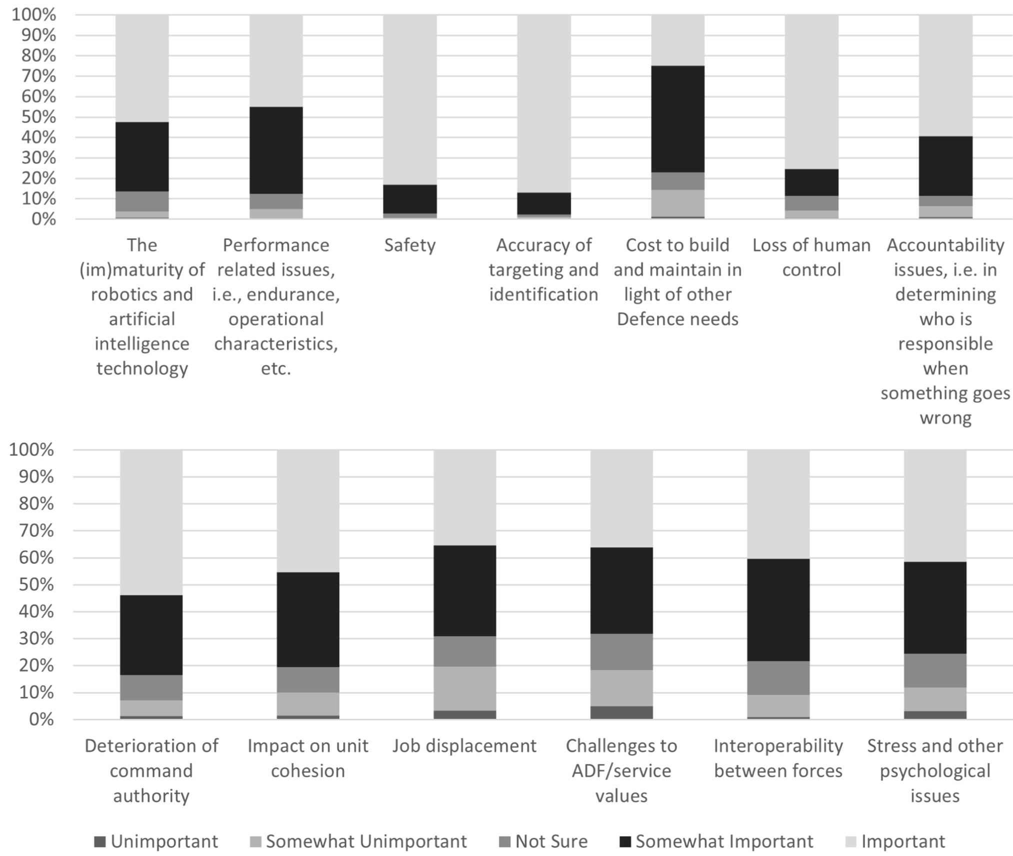
Building trust among the armed forces

As discussed above, most of the empirical research on trust perception regarding SACs focuses on public opinion and attitudes. Empirical data about the views of the users of SACs in the military are not yet widely available – the perception of trust in the armed forces is a considerably under-researched field. A notable exception is a recent survey among officer cadets and midshipmen at the Australian Defence Force Academy (ADFA), which enquires into the willingness of officers to deploy human-machine teams (Galliott and Wyatt, 2020).

The survey identifies 13 variables that inform the perception of trust among future officers who were asked to rank them on a discrete scale of five intervals (see Figure 10).

FIGURE 10: VARIABLES THAT INFORM TRUST PERCEPTION TOWARDS SYSTEMS WITH AUTONOMOUS CAPABILITIES AMONG CADETS

From a scale of unimportant to important, please rank the following dimensions of using systems with autonomous capabilities in military operations.



Source: Galliot and Wyatt, 2020

> “In combat, service personnel are likely to prefer less efficient and slower weapons, vehicles or support systems that work one hundred percent of the time over a novel gadget that may prove unreliable”

Not surprisingly, safety concerns depress overall trust levels, as do concerns about the accuracy of targeting and identification. Reduced operational costs, usually touted by senior officials and defence policymakers as one of the chief benefits of SACs, do not seem to be an important factor.

In light of these findings, the authors recommend:

- incorporating SAC “acclimatisation training” at all levels of the officer training process;
- tactics training should involve robotic units, and
- units at company level should be encouraged to “experiment and innovate” with SACs in war games.

Ultimately, in combat, service personnel are likely to prefer less efficient and slower weapons, vehicles or support systems that work one hundred percent of the time over a novel gadget that may prove unreliable, so generating trust across the above dimensions is likely to be an uphill struggle. Deploying SACs requires an incremental development process of trial and error so it seems reasonable to work towards an iterative “trust but verify” approach (Roff and Danks, 2018) in military training to increase acceptance levels. While trust is built gradually but rapidly taken away if a SAC fails – in particular where it compromises the mission and/or causes injury – any reform to training programmes and processes should be mindful that building trust is a long game. Trust gaps in the military "won't be solved by code but by conversation" (Hartig and Vanhoose, 2019).

Human-machine teaming

For the MOD, military personnel should not consider SACs tools but "partners" that complement human cognitive capabilities. This places high demands on future education and training policies as human operators will need to be able to understand, operate and control highly complex machines yet simultaneously consider them allies and companions. The MOD develops these points in several Joint Concept Notes, most notably in JCN 1/18 (Ministry of Defence, 2018).

Near-term scenarios of human-machine teaming are likely to manifest in greater military mobility. The DOD is presently testing an AI programme that "synthesises drone feeds, robot sensors, small arms fire detection, and ground-based radar" to "out cycle" enemy decision-making (Osborn, 2019). Semi-automated engagement is hoped to move US forces faster across unstructured environments.

In order to increase trust in SACs in human-machine pairings of this kind, the MOD says it will develop trust along the following four dimensions: technology, predictability, familiarity and context (see Figure 11) (Ministry of Defence, 2018, p. 48).

FIGURE 11: COMPONENTS OF TRUST IN HUMAN-MACHINE TEAMING

Technology	The MOD assumes that the better military personnel understand the technological dimensions of a system, the more trusted it will be. Developing a "mechanical understanding" is therefore essential for the armed forces to become comfortable with deploying SACs.
Predictability	Consistent anticipation and rational expectations about the behaviour of SACs will support building trust. Over time, growing confidence in SACs is hoped to allow for relaxing error and fault tolerances.
Familiarity	The MOD recognises that "trust is emotional". Increased familiarity with how SACs work in practice is likely to increase confidence.
Context	As with any system, trust in SACs is context-dependant. Weather conditions, the unstructured character of the battlefield – in particular the "Last Mile," the highly contested and hostile environment where soldiers are actively engaged in combat – and other factors all impact the trustworthiness of SACs.

Certification and assurance programmes

The DOD recognises that insufficient trust in SACs is likely to lead to "high-regret, unintended consequences" which may erode trust to a point where SACs will not be adopted "except in extreme cases such as missions that cannot otherwise be performed" (Defence Science Board, 2016, p. 37). The World Trade Organisation's Dispute Settlement Body (DSB) calls for standardised trust assessments at all stages of the design, development and operations process and therefore recommends the development of assurance programmes.

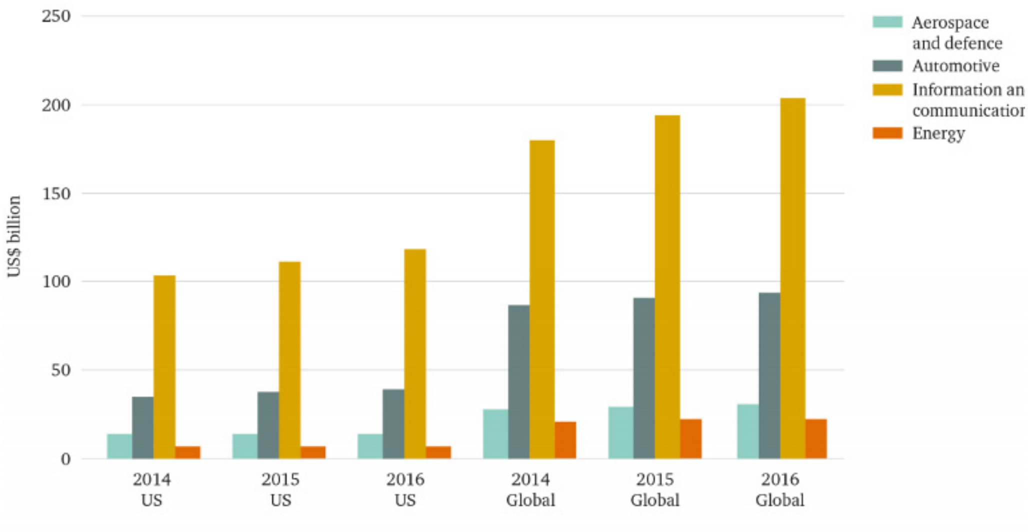
In response, the Defence Advanced Research Projects Agency (DARPA) is currently testing technologies for the "continual assurance of Learning-Enabled, Cyber Physical Systems" (LE-CPSs) (Neema, 2021). The programme develops assurance certificates of safety and functional correctness that are to be issued at the design stage and updated/evaluated during operations in the field. The programme is largely built around popular machine learning approaches, such as supervisory and reinforcement learning. Compromises to safety constraints are detectable through shifts in probability distributions over performance scenarios. Machine certificates are to complement, not replace, "human assurance" (Cummings, 2019). Iterative scenario training to test for the "behavioural repertoire" (Lyons et al., 2017) of a system could be used to continuously improve the trustworthiness of an SAC.

Industry-driven innovation

Previous cycles of innovation in defence were largely internal to the military. Stealth and precision weapons, for instance, were developed under classified programmes (Maddock, 2021). However, the new wave of automation happening now is largely driven by private sector innovation in artificial intelligence, machine learning, optoelectronics, and mechanical engineering. Examples of relevant "spin-ins" include Amazon Web Services Cloud technology, driverless cars and cooperative warehouse and logistics robots (Knox, 2020; Weinberger, 2019).

Over the past couple of years, spending in the aerospace and defence sector, which is home to most SAC research projects, has not kept up with the significant rise of R&D spending in other relevant areas, most notably in the automotive and ICT sectors. For Chatham House, this development suggests that private sector innovation in these thriving sectors will be of growing importance in developing future SACs. As Cummings (2017, p. 11) noticed, this raises the question of "whether defence companies will have the capacity to develop and test safe and controllable autonomous systems, especially those that fire weapons" if some of the technology that drives these systems has been developed in other sectors.

FIGURE 12: RESEARCH AND DEVELOPMENT SPENDING IN FOUR SECTORS RELEVANT TO SYSTEMS WITH AUTONOMOUS CAPABILITIES 2014-16



Source: Cummings, 2017

Military bureaucracies are considered to be facing disadvantages in the fast-paced environment of innovation today due to a perceived lack of agility and responsiveness (Amyx, 2020; Araya, 2020; Knox, 2020). The question of whether military SAC research is responsive, effective and efficient enough – as it can also capitalise on research outside the military – creates urgent policy problems around long-term strategy and R&D funding. Further issues arise from foreign industries' first-mover advantages, knowledge transfer, proliferation, and intellectual property rights, all of which complicate the development of trustworthy SACs.

“Military bureaucracies are considered to be facing disadvantages in the fast-paced environment of innovation today due to a perceived lack of agility and responsiveness

In a January 2020 memorandum, then US Secretary of Defence Mark Esper sketched the development of a new Joint Warfighting Concept for the US “to align personnel, equipment, training, and doctrine to win on any battlefield” (US Department of Defense, 2020). To achieve this, the US would need to advance the “development of crucial emerging technologies, such as hypersonic weapons, directed energy, artificial intelligence, and autonomous platforms” and “champion cloud initiatives” (ibid.). This move requires novel, collaborative forms of engagement with commercial industry partners.

President Obama's Third Offset Strategy (US Department of Defense, 2015), later abandoned by the Trump Administration, had already begun to make headway towards better aligning the Pentagon's R&D and procurement decisions with private sector strategy and operations. The initial phase of this continuing alignment, which includes reformed procurement and alternative contracting mechanisms, is largely judged to have been successful (Knox, 2020). The DOD set up the Defence Innovation Unit (DIU) in 2015 which is tasked with “fielding and scaling commercial technology” (Defence Innovation Unit, 2021) beyond traditional partners in the defence industries to help solve critical national security challenges by involving a much larger group of civil and commercial stakeholders. The DIU invites commercial partners to submit innovation proposals for “mission critical challenges” and subsequently awards contracts.

In the UK, the Defence and Security Accelerator (DASA) pursues a similar strategy. Both established industries and SMEs are invited to respond to open calls and themed competitions. Defence Digital, part of Strategic Command at the MOD, and Jhub, the innovation centre of Joint Forces Command and the British Army Innovation Team also operate in this area. NavyX is the Royal Navy's "Autonomy and Lethality Accelerator" and the RAF's Rapid Capabilities Office manages several large-scale industry partnerships, such as Tempest for "bring[ing] a 'plug and play' approach" to systems design. Many further initiatives exist (Mehdian-Staffell, 2021).

The 2021 Integrated Review considers science and technology "an increasingly important metric of global power" (HM Government, 2021a, p. 30). The Defence and Security Industrial Strategy that followed is tasked to "support innovation and convert it into deployable national security capabilities" (HM Government, 2021a, p. 38). The Industrial Strategy promises to develop "specific cross-sector innovation campaigns" which "exploit the strengths of the UK civil and defence sectors" (HM Government, 2021b, p. 62). The National Security Strategic Investment Fund (NSSIF) is the government's corporate venturing arm to develop dual-use technologies in collaboration with the private sector.

As such, as efforts to link military R&D with the private sector gain momentum, new trust issues are likely to surface where core technologies emerge outside the military domain. SACs that are not developed by domestic industries alone may face resistance from the armed forces or present additional security challenges, eg with regard to data sharing or property rights. The larger the network of stakeholders involved in designing, developing, testing, and operating SACs, the more complex the issue of trust is likely to become.

Conclusion

More than six years ago, then US Deputy Defence Secretary Bob Work made a robust case for pursuing the development of SACs with urgency given the advances that adversaries had been making: "we still believe we have a margin, but the margin is steadily eroding and it's making us uncomfortable" (US Department of Defense, 2015). In an era of growing great-power competition and increasing fragmentation, this urgency has certainly not diminished. To keep up with her rivals, RUSI recommends the UK establish "an appetite for risk" and recognise "the need for experimentation and the inevitability of regular failure" (Louth et al., 2017, p. vii) in the domain of autonomous systems. Payne (2021) suggests that, in the years to come, liberal states will have to find a new balance between maintaining a competitive edge over authoritarian rivals – who have greater regulatory flexibility – and the fundamental values that make them liberal.

Notions of urgency and risk could complicate the issue of trust further. Indeed, as the above discussion demonstrates, trust in autonomous systems in the context of defence is a complex, multidimensional affair. The UK's position, in line with that of the US and other NATO countries, is to say that the development of fully-fledged, lethal human-out-of-the-loop SACs with no human oversight at all is not on the horizon. In any case, SACs are considered to be covered by existing legal frameworks and require no new legislation or new international treaties. Should this position change, however, and lethal, independently operating SACs be developed, changes to international law may become necessary after all. Such a radical shift in attitude would surely invite considerable opposition from civil society actors and create significant new policy challenges.

This review cautions against the view that the public would uniformly oppose SACs while the armed forces would be most eager to deploy them. The empirical picture is more nuanced. It suggests that public opinion about SACs should not be considered in isolation but in the context of broader trends in attitudes towards larger developments in automation and AI more generally. Further progress in automation – ie in the design and development of processes that do not require human oversight – will have an impact on more specific challenges in the development of SACs that need to perform well under significant levels of uncertainty. Given the nascency of so many applications in this space, acceptance levels are likely to rise and fall around real-life events of success and error, which are also likely to impact the ways that SACs are perceived. It should also be noted that acceptance and trust are not the same. Moreover, while many military strategists and technologists are optimistic about SACs, this attitude does not necessarily reflect the views of the armed forces generally, where reservations and caution about autonomy remain considerable.

> Autonomy complicates the relationship between the reliability and predictability of technological artefacts, which are being elevated from the status of a tool or support system to that of a "partner" in combat"

Generating trust in human-machine teaming presents one of the most significant near-term challenges for designing trustworthy SACs. Autonomy complicates the relationship between the reliability and predictability of technological artefacts, which are being elevated from the status of a tool or support system to that of a "partner" in combat. This shift places additional demands on service personnel. More research is required to assess how the armed forces come to trust new technologies since "a failure to understand how humans can or cannot trust [SACs] has direct consequences for military strategy, the conduct of hostilities, and even long term economic impacts from investment and procurement choices" (Roff and Danks, 2018, p. 3). This line of enquiry should be mindful that, just like any other technology, SACs are socio-technical systems that should not be defined in technical terms alone. For this reason, assurance programmes that employ machine learning to certify the trustworthiness of an SAC should be complemented by a rigorous "test and verify" programme that involves robust human input. Above all, generating trust for SACs among the public and the armed forces alike is a long-term game that escapes short-term policy cycles.

References

Abbass, H.A., Scholz, J., Reid, D.J. (Eds.), 2018. Foundations of Trusted Autonomy, 1st ed. Springer.

Allenby, B.R., 2014. Are new technologies undermining the laws of war? Bulletin of the Atomic Scientists 70, 21–31. https://doi.org/10.1177/0096340213516741

Amyx, S., 2020. The Technological Innovation Challenges In Defence. Forbes.

Araya, D., 2020. Transforming The US Military For The 21st Century. Forbes.

Article 36, 2020. Regulating autonomy in weapons systems.

Baeza-Yates, R., 2016. Data and algorithmic bias in the web. ACM Press, pp. 1–1. https://doi.org/10.1145/2908131.2908135

Boulanin, V., Verbruggem, M., 2017. Mapping the development of autonomy in weapon systems. Stockholm International Peace Research Institute, Stockholm.

Bradshaw, J.M., Hoffman, R.R., Woods, D.D., Johnson, M., 2013. The Seven Deadly Myths of "Autonomous Systems." IEEE Intelligent Systems 28, 54–61. https://doi.org/10.1109/MIS.2013.70

Carpenter, C., 2013. How Do Americans Feel About Fully Autonomous Weapons? – Duck of Minerva. Duck of Minerva. URL https://duckofminerva.com/2013/06/how-do-americans-feel-about-fully-autonomous-weapons.html (accessed 3.21.21).

Cummings, M.L., 2019. Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification? [Opinion]. IEEE Technology and Society Magazine 38, 20–26. https://doi.org/10.1109/MTS.2019.2948438

Cummings, M.L., 2017. Artificial Intelligence and the Future of Warfare. Chatham House Research Paper 18.

Defence Innovation Unit, 2021. Accelerating Commercial Technology for National Security [WWW Document]. URL https://www.diu.mil/ (accessed 3.22.21).

Defence Science Board, 2016. Summer Study on Autonomy. Department of Defense, Washington DC.

Etzioni, A., Etzioni, O., 2017. Pros and Cons of Autonomous Weapons Systems. Military Review: The Professional Journal of the US Army.

Galliott, J., Wyatt, A., 2020. Risks and Benefits of Autonomous Weapon Systems: Perceptions among Future Australian Defence Force Officers. Journal of Indo-Pacific Affairs 17–34.

Hajian, S., Bonchi, F., Castillo, C., 2016. Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. ACM Press, pp. 2125–2126. https://doi.org/10.1145/2939672.2945386

Hansard Society, 2019. Audit of Political Engagement 16: the 2019 Report. Hansard Society, London.

Hartig, L., Vanhoose, K., 2019. Solving One of the Hardest Problems of Military AI: Trust. Defence One.

HM Foreign and Commonwealth Office, 2017. Rules-based international system 2016 to 2017.

HM Government, 2021a. Global Britain in a competitive age: the integrated review of security, defence, development and foreign policy. HM Stationery Office, London.

HM Government, 2021b. Defence and Security Industrial Strategy: A strategic approach to the UK's defence and security industrial sectors. Crown Copyright, London.

HM Government, 2018. Rules Based International System Conference [WWW Document]. GOV.UK. URL https://www.gov.uk/government/news/rules-based-international-system-conference (accessed 2.16.20).

Horowitz, M.C., 2016. Public opinion and the politics of the killer robots debate. Research & Politics 3, 2053168015627183. https://doi.org/10.1177/2053168015627183

Human Rights Watch, 2018. Heed the Call [WWW Document]. Human Rights Watch. URL https://www.hrw.org/report/2018/08/21/heed-call/moral-and-legal-imperative-ban-killer-robots (accessed 3.21.21).

Human Rights Watch, 2020. Stopping Killer Robots Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control. Available from: https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and

Ipsos, 2021. Global Survey Highlights Continued Opposition to Fully Autonomous Weapons, Ipsos Market Research.

Kirkpatrick, K., 2016. Battling algorithmic bias: how do we ensure algorithms treat us fairly? Communications of the ACM 59, 16–17. https://doi.org/10.1145/2983270

Knight, W., 2017. Google's AI chief says forget Elon Musk's killer robots, and worry about bias in AI systems instead. MIT Technology Review.

Knox, T.J., 2020. US Military Innovation In The 21st Century: The Era Of The "Spin-On." Wharton Research Scholars 194, 57.

Larter, D., 2019. The US Navy says it's doing its best to avoid a 'Terminator' scenario in quest for autonomous weapons. Defence News.

Lee, P., 2017. The Drone Operator and Identity: Exploring the Construction of Ethical Subjectivity in Drone Discourses. Critical Approaches to Discourse Analysis across Disciplines 9, 17.

Louth, J., Taylor, T., Tyler, A., 2017. Defence Innovation and the UK. RUSI Occasional Paper 40.

Lyons, J.B., Clark, M.A., Wagner, A.R., Schuelke, M.J., 2017. Certifiable Trust in Autonomous Systems: Making the Intractable Tangible. AIMag 38, 37–49. https://doi.org/10.1609/aimag.v38i3.2717

Maddock, I.A., 2021. DARPA's Stealth Revolution. DARPA Resources.

Mehdian-Staffell, S., 2021. The Integrated Review: A Technological Revolution at the Heart of UK Defence and National Security. RUSI Commentary.

Ministry of Defence, 2020. Joint Concept Note 1/20: Multi-Domain Integration (Development, Concepts and Doctrine Centre Publications No. Joint Concept Note 1/20). Ministry of Defence, Bristol.

Ministry of Defence, 2019. Defence Technology Framework. Ministry of Defence, London.

Ministry of Defence, 2018. Joint Concept Note 1/18: Human-machine teaming (Development, Concept and Doctrines Centre). Bicester.

Ministry of Defence, 2017a. Unmanned Aircraft Systems (Joint Doctrine Publication No. 0-30.2). Development, Concepts and Doctrine Centre, Swindon.

Ministry of Defence, 2017b. Joint Concept Note 1/17: Future Force Concept (No. Joint Concept Note 1/17), Development, Concepts and Doctrine Centre. Ministry of Defence.

National Security Commission on Artificial Intelligence, 2021. Final Report. National Security Commission on Artificial Intelligence, Washington DC.

NATO Allied Command Transformation, 2016. Autonomous systems: íssues for defence policymakers. NATO HG SACT, Norfolk, Va.

Neema, S., 2021. Assured Autonomy [WWW Document]. Defence Advanced Research Projects Agency. URL https://www.darpa.mil/program/assured-autonomy (accessed 3.21.21).

O'Donoughue, N.A., McBirney, S., Persons, B., 2021. Distributed Kill Chains: Drawing Insights for Mosaic Warfare from the Immune System and from the Navy. RAND Corporation, Santa Monica CA.

O'Neil, C., 2017. Weapons of math destruction: how big data increases inequality and threatens democracy. Penguin Books, London.

Osborn, K., 2019. Soldiers use AI to fire precision grenades, guide drone attacks. Armed Forces.

Pilling, M., 2015. Issues Regarding the Future Application of Autonomous Systems to Command and Control (C2) (No. DSTO–TR–3112). Joint and Operations Analysis Division, Canberra.

Pomerleau, M., 2020. Top intel official warns of bias in military algorithms [WWW Document]. C4ISRNET. URL https://www.c4isrnet.com/artificial-intelligence/2020/11/18/top-intel-official-warns-of-bias-in-military-algorithms/ (accessed 3.21.21).

Roff, H.M., Danks, D., 2018. "Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems. Journal of Military Ethics 17, 2–20. https://doi.org/10.1080/15027570.2018.1481907

Rossi, F., 2019. Building Trust In Artificial Intelligence. Columbia SIPA Journal of International Affairs.

Sayler, K.M., 2020. Artificial Intelligence and National Security. US Congressional Research Service 43.

Scharre, P.D., 2017. The opportunity and challenge of autonomous systems, in: Autonomous Systems: Issues for Defence Policymakers. NATO Capability Engineering and Innovation Division, Norfolk, Va., pp. 3–26.

Schneewind, J.B., 1998. The Invention of Autonomy: A History of Modern Moral Philosophy. Cambridge University Press, Cambridge.

Shane, S., 2018. How a Pentagon contract became an identity crisis for Google. CNBC.

Skiena, S.S., 2011. The Algorithm Design Manual, 2nd ed. 2008 edition. ed. Springer, London.

UK government, 2015. Statement to the Informal Meeting of Experts on lethal Autonomous Weapons Systems.

UK Strategic Command, 2021. Welcome Note.

US Army Training and Doctrine Command, 2017. The US Army Robotic and Autonomous Systems Strategy. US Army, Fort Eustis VA.

US Department of Defense, 2020. Memorandum for Secretaries of the Miitary Departments.

US Department of Defense, 2017. Directive Number 3000.09.

US Department of Defense, 2015. The Third US Offset Strategy and its Implications for Partners and Allies.

Walsh, D., 2010. Leading UN official criticises CIA's role in drone strikes. The Guardian.

Weinberger, S., 2019. Meet America's newest military giant: Amazon. MIT Technology Review.

Wong, Y.H., Yurchak, J., Button, R.W., Frank, A., Laird, B., Osoba, O.A., Steeb, R., Harris, B.N., Bae, S.J., 2020. Deterrence in the Age of Thinking Machines, Research Reports. RAND Corporation, Santa Monica CA.

Wright, J., 2017. Attorney General's speech at the International Institute for Strategic Studies.

The Policy Institute

The Policy Institute at King's College London works to solve society's challenges with evidence and expertise.

We combine the rigour of academia with the agility of a consultancy and the connectedness of a think tank.

Our research draws on many disciplines and methods, making use of the skills, expertise and resources of not only the institute, but the university and its wider network too.

Connect with us

@policyatkings kcl.ac.uk/policy-institute