

**BAE SYSTEMS**

**WIRED**  
**Consulting.**



## **Robotic weapons are coming**

**– what should we  
do about it?**



**Artificial intelligence is set to revolutionise warfare. Mitigating the risks without sacrificing national security will require careful and urgent discussion...**

**In the final phase of the Libyan civil war**, in 2020, Prime Minister Fayez al-Serraj launched a new offensive against the forces of renegade commander Khalifa Haftar. According to a United Nations report published the following year, al-Serraj's troops deployed weapons of a type that may never have been used against humans before: lethal autonomous weapon systems (LAWS). These devices – which allegedly included kamikaze drones – “hunted down and remotely engaged” Haftar's logistics convoys and soldiers. “The latter's units were neither trained nor motivated to defend against the effective use of this new technology,” read the report, “and usually retreated in disarray.” It's unclear whether these devices killed anyone or were operating in autonomous mode during the attacks, but the incident drew attention to a looming shift in how we fight wars.

LAWS are an emerging category of munition – and they're unlike any other. These weapons aren't just about bigger bangs or greater accuracy, they are themselves decision-making agents: they use artificial intelligence (AI) to identify, select and engage a target without human input. LAWS could take many forms, from swarms of micro-drones to autonomous tanks or submarines, but the quality of their strategic thinking, their capacity for intricate coordination, the speed with which they're able to attack – all of these things have led many experts to argue that those who possess this technology could enjoy a dramatic, asymmetric advantage over those who don't. Some believe they will bring about a revolution in warfare as significant as that of the atomic bomb.

By their very nature, these weapons are contentious, because they throw up profound ethical and safety questions. Since 2013, the campaign group Stop Killer Robots has argued for LAWS to be outlawed. In 2017, 116 experts, including Elon Musk and Mustafa Suleyman, co-founder of Google's DeepMind AI subsidiary, signed an open letter to the UN calling for a ban:

**“We do not have long to act. Once this box is opened, it will be hard to close.”**

Open letter to the UN, led by Elon Musk and Mustafa Suleyman



But Pandora's box is already open. Not only may LAWS currently be in use, but many conventional systems are just a firmware update away from autonomous operation. Right now, states are at loggerheads over what to do about it. In 2021, a majority of the 125 nations that subscribe to the UN's Convention on Certain Conventional Weapons demanded new controls on this technology – but the talks failed to reach unilateral agreement. The likes of Russia, the UK and the US have historically opposed a pre-emptive ban, and are said to be investing heavily in the development of such systems. The war in Ukraine has thrust the LAWS conversation once again into the spotlight this year, with speculation that Russia could deploy drones that operate autonomously.

It seems a new era of warfare could well be about to dawn, and we need to think deeply about the implications. WIRED convened a round-table of experts, in partnership with defence, aerospace and security firm BAE Systems, to consider the most compelling arguments on either side of the debate, and to ask: what should the path ahead look like?

## Why create robotic weapons?

Autonomous weapons could confer a significant military advantage. Sure, they could soon be more cost-effective than human soldiers – they don't require training, clothing, feeding or pensions, and they don't need to sleep or take time off. But the game-changer is the AI element: not only in the sophistication of the decision-making – like a computer playing chess, it seems probable that LAWS would eventually outclass the human mind on the battlefield – but also in the ability of LAWS to swarm. A drone swarm involves potentially thousands of drones moving like a murmuration of starlings, perfectly coordinated but operating autonomously. This is hard to defend against. The drones can be widely dispersed, making them hard to track and shoot down, but can congregate at an opportune moment to unleash their firepower in a concentrated fashion before dispersing again. When you're attacked by a warship, you know what to put in your crosshairs – but a drone swarm has no 'centre of gravity' at which to direct a counter-attack.

There is no doubt that LAWS are in development around the world and in some cases, field-tested. This raises the question of how a nation should prepare for the prospect of conflict with another state which has LAWS in its arsenal. "Some nations are likely to be thinking that if you keep to a very limited view of LAWS, but you then lose the war – is that ethically better than keeping your obligation to the nation to win and protect your way of life?" asks Dave Short, technology director at BAE Systems. "That's what I think a lot of nations are now contemplating: do they need to create their own LAWS so that they can be confident of defending themselves?"

Some contend that there's a wider ethical case for using these weapons. Humans can seek revenge, act sadistically, get tired and make mistakes; AI, they argue, is a precision tool and could potentially be coded to obey the laws of war. Using them instead of humans could also save lives. General Sir Richard Barrons, formerly of Joint Forces Command and now co-chairman of Universal Defence & Security Solutions, puts it starkly: "In counter-terrorism, if you need to breach a compound wall, someone has to go through the breach, find and probably kill or detain the terrorist," he says. "You've got two choices to send through that breach: a dog or a Special Forces operator. It's a really high risk business. In the past I've had conversations with people and I've said, 'Imagine I built a machine that will go through the compound wall quickly, and based on the facial recognition data that we've supplied it with, it will kill the right people – are you happy with that?' They say, 'We're not going to do that, that's a robot killing a human.' And then I say, 'Well, OK, now I'm going to send your son or daughter to that compound. How do you feel about that?' They say, 'You know what, the robot is a really good idea.'"









## Why are people worried?

The multitude of objections to LAWS include technical problems such as algorithmic biases, unreliability and unpredictability. Many people, however, simply recoil at the ethics. As UN Secretary General António Guterres put it in his address to the General Assembly in 2018: "Let's call it as it is. The prospect of machines with the discretion and power to take human life is morally repugnant." And even if we did decide that it was acceptable to allow machines to take lives within the parameters of existing humanitarian conventions, there's a further issue. "Accountability and plausible deniability is a major concern," says Gopal Ramchurn, professor of artificial intelligence at the University of Southampton and director of the UKRI Trustworthy Autonomous Systems Hub. "If the machine makes a mistake – they wrongly classify something and target it – the human can say, 'Well, I didn't have anything to do with it, that was the machine.' That is the key issue for me."

Even if those wars involve fewer humans, that doesn't obviate the ethical considerations. Lucia Retter, research leader in defence and security at RAND Europe, asks: "If you have a machine-on-machine war, where it's basically just a loss of kit rather than a loss of life, what does that mean for how we perceive war?" One possibility is that we may see declaring war as a much less difficult decision.

To Stuart Russell, professor of computer science at University of California, Berkeley – and an advisor to the Future of Life Institute, which seeks to mitigate existential threats to humanity – the primary concern is not ethical but practical. "To me the biggest argument, and it's often omitted, is that autonomous weapons decouple the number of weapons you can deploy from the number of people you need to deploy them," he says. "So you can press one button and kill 10 million people. That seems like a really bad idea." You might argue the same is true with a hydrogen bomb, but only very few states can produce one – that's not the case with autonomous weapons. "You can easily imagine these will be manufactured and sold in the millions or tens of millions."

What's more, the AI devices themselves are imperfect. Developments in artificial intelligence over the past two decades have made it spectacularly more capable than it once was. The old paradigm was that engineers had to manually provide the AI with rules, but of course trying to turn unpredictable real-world scenarios into sets of rules is a time-consuming and potentially endless task. The new approach, machine learning (in which the computer teaches itself from trial and error) using 'deep' neural networks (algorithms inspired by the structure of the human brain), has led to breakthroughs in everything from image recognition to vaccine development. But although in one sense AI is incredibly smart – you try beating it at chess – it's also extraordinarily stupid. After all, how often has your voice assistant misunderstood a perfectly simple instruction? When AI is interacting with AI, that dumbness can be amplified: signals can be misperceived and interactions can get caught in feedback loops.

"The thing that keeps me up at night is escalation," says Kenneth Payne, professor of strategy at King's College London and the author of *I, Warbot: The Dawn Of Artificially Intelligent Conflict* (2021). He's worried about that happening for a subtly different reason: the norms of this new form of warfare are yet to be established. "In war games, it's often quite hard to produce escalation spirals, you don't want nuclear war even in a virtual war game. But a war game last year postulated a human-machine team squaring off against a human-machine team; the scenario was China versus the US and its allies. And it went into a fairly rapid escalation, because each side knew that the other side had outsourced some of its escalation decisions to AI. But they weren't sure where those thresholds were, so you produce massive uncertainty and the real need to get your retaliation in first. It's based on one war game, but it's that dynamic that worries me."

**"You can easily imagine these  
will be manufactured and sold in the  
millions or tens of millions."**

**Stuart Russell**

Professor of Computer Science at University of California, Berkeley and a  
representative of the Future Of Life Institute







## Is banning an option?

Much of the debate around how to treat this issue focuses on the prospect of an international ban. An open letter from 2015 published by the Future of Life Institute, and signed by 4,667 AI and robotics researchers including Stephen Hawking, makes the case:

*"The key question for humanity today is whether to start a global AI arms race or to prevent it from starting. If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow... [This] should be prevented by a ban on offensive autonomous weapons beyond meaningful human control."*

Critics point out that if you can't get all nations to agree to a ban – which seems likely to be the case in the current geo-political climate – then it only takes one unfriendly nation to develop LAWS to put you at an unacceptable disadvantage. "One way out is the partial ban," argues UC Berkeley's Stuart Russell. "This would be a ban on small anti-personnel weapons that could be turned into large swarms and would be easily proliferated. I would agree that if you didn't develop autonomous fighter aircraft, you would lose air superiority, and probably with submarines, something similar."

Russell notes that past agreements banning chemical and biological weapons haven't been perfect but they have been relatively effective. "What we're really trying to do [with chemical and biological weapons] is to prevent large-scale manufacturing," he says. This would lead to cheaper prices, wider availability and "the threat of annihilation on a nuclear scale. So you go after the facilities and companies that can produce chemicals in large quantities and you regulate those and you inspect them." However, RAND's Lucia Retter believes that taking a similar approach with LAWS would be impractical. "With conventional weapons that have been banned, you can more or less physically monitor whether actors have them or not, but you can't do that with LAWS." A drone may not be autonomous right now, but a firmware update could provide it with autonomy tomorrow.

## So, what's the path ahead?

Regardless of whether a ban is practical, there was consensus among the experts WIRED spoke to that there should be careful thought given to how these weapons are developed. First, there needs to be a serious, informed conversation within government and with the public about goals, risks and trade-offs. "Most political decision makers and their supporting officials do not know enough about this topic, and will inevitably have some prejudices and default settings," says General Sir Richard Barrons. "So I think managing this process starts with an investment in creating a common lexicon and understanding about what the issues are."

The conclusions of those discussions then need to be turned into meaningful policy. "It is really important to actually have a recognised national position that's going to last more than one strategy paper and a parliament," says BAE Systems' Dave Short. "Something that creates a framework that allows a deep understanding of what it is that needs to be dealt with." Ramchurn believes that doing this effectively will require international coordination: "So working with your partners, working with your allies, not each country coming up with its own set of principles around ethical AI."

Finally, there needs to be mechanisms to create accountability. KCL's Kenneth Payne advocates for "an AI Commissioner, operating across domains, not just national security, along the lines of the Information Commissioner. Somebody that can give a degree of oversight and public confidence at a time of change and uncertainty."





## Writing the rulebook

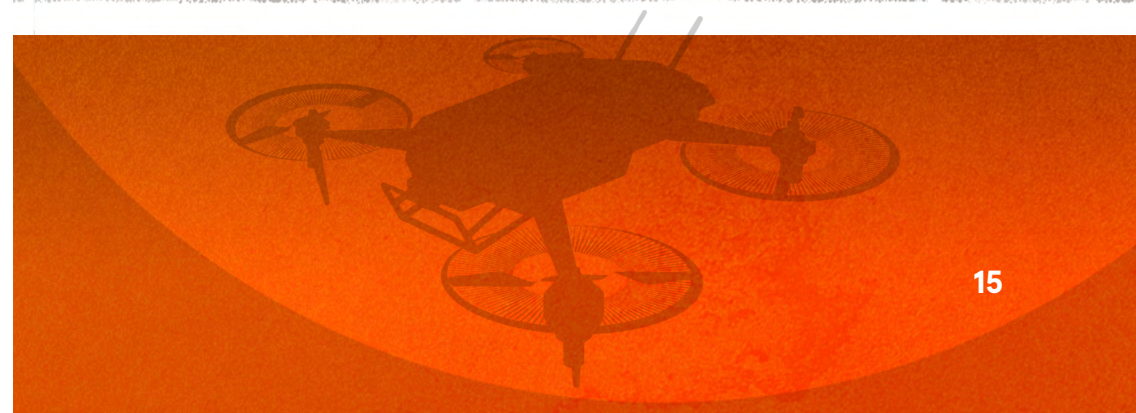
As for the regulations themselves, what should they look like? Clearly, for the same reason that a ban may be challenging, we may not be able to impose these rules on other countries – but some believe there may at least be ways to enshrine our priorities and values without putting ourselves at a disadvantage.

So far, policy on LAWS has centred around the idea of retaining meaningful human control. The US requires “human judgement over the use of force” in deploying LAWS; the UK demands “context-appropriate human involvement”. This leaves some ambiguity: are we talking about human involvement and judgement at the design stage, for instance, or on the battlefield? It may be more useful to come up with specific guidelines for how we use AI in war. Short suggests defining “limiters” for how these platforms operate. “If it’s a learning system, it will be influenced by the environment it’s fighting in, and it just knows that it needs to find a way to win. If nations consider building these systems, they need to determine what are those absolute red lines that the system cannot go outside of. Ultimately, BAE Systems believes, like its customers, that there needs to be meaningful human input in the weapon command and control chain.”

Payne has ideas for what those limiters might be. In his book *I, Warbot*, he proposes taking a cue from the ‘laws of robotics’ devised by Isaac Asimov in the 1940s, and devising ‘rules for warbots’. He argues that other states “might themselves arrive at something similar – driven by the same uncertainties and anxieties.” His suggested rules are as follows:

1. A warbot should only kill those I want it to, and it should do so as humanely as possible.
2. A warbot should understand my intentions and work creatively to achieve them.
3. A warbot should protect the humans on my side, sacrificing itself to do so – but not at the expense of the mission.

These principles, he writes, “are designed to hold humans responsible for the machines they build”.





## Regulations require innovation

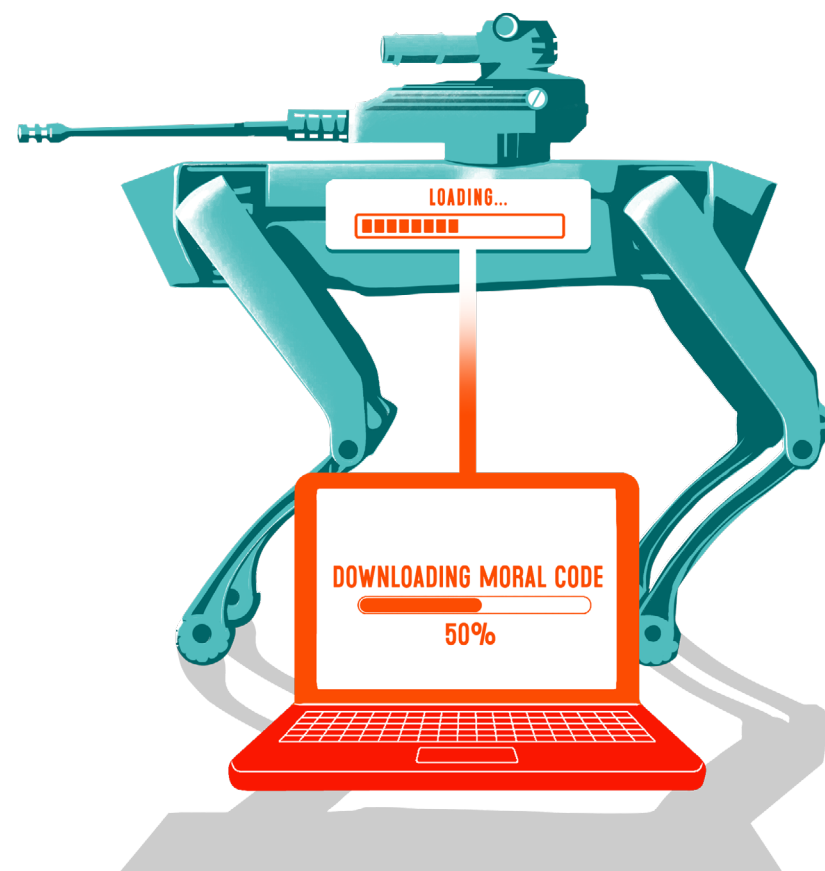
Creating 'laws for warbots' is not as straightforward as it might sound. It entails some major technical challenges. For one, you might give the system rules – but who's to say that someone else won't reprogram it? Russell suggests that a possible option might be to require that LAWS are only sold with application-specific integrated circuits, or 'ASICs', that cannot be reprogrammed. "With ASICs, you basically take your software and you burn it into the hardware, so the hardware just does one thing." Could people not simply create a new ASIC and replace the chip? "Manufacturing ASICs is actually a fairly big ask for a non-state actor."

Another problem is the capability of AI to interpret the rules. Say, for example, that you want your warbot to comply with international humanitarian law. "At the moment, we can't build AI systems that can autonomously respect the laws of war," says Russell. "We can probably do discrimination but we can't do proportionality and necessity."

We also have to address a perennial difficulty in any framework that seeks to regulate the behaviour of AI, which is a challenge often referred to as the 'black box problem'. We see the inputs and the outputs of AI systems, but we often don't know how or why the system has made the decision that it has. Short says nations need to consider whether to, and if so, how much blind faith to place in the AI system. He references a famous game of Go between grandmaster Lee Sedol and Google DeepMind's AI player, AlphaGo. In game two, AlphaGo's 37th move was completely counterintuitive. "It looked like a suicide move, but ultimately won the game. However it's one thing moving a piece in Go, but it's different to moving your aircraft carrier. If the system that you know is nearly always right tells you that you've got to move that major asset from one point to another, and you have no idea why you're doing it, what's the trust level that's needed?"

The University of Southampton's Gopal Ramchurn thinks it's imperative we innovate to solve this problem. "These systems can plan millions of steps ahead," he says. "You need to be able to design interfaces, things that could help people understand why the machine is telling them to do that."

For a technology that seems to be emerging so rapidly, but which raises such profound issues, it's perhaps troubling how many questions around the development, nature and use of LAWS are yet to be answered. Resolutions will need to be found, and fast. But in doing so, we need to ensure we plan for the technologies of tomorrow rather than those of today – because this space is likely to evolve rapidly. "In 10 years' time," says Ramchurn, "I think we will have a variety of these systems – and we'll have them in forms that we've never thought about before."





## BAE SYSTEMS

At BAE Systems, our people develop technologies to help Governments, their armed forces and security services defend countries, in the air, at sea, on land, in space and in the cyber domain. In the UK we build submarines, warships and some of the world's most advanced military aircraft, and integrate defence and security systems across all domains. We also maintain and upgrade the Royal Navy's warships at Portsmouth Naval Base and service the Royal Air Force's Typhoon fast jet for 24/7 security operations.

We make a significant economic and social contribution – delivering £10 billion to UK GDP and in the region of £4 billion in exports each year. In the UK we employ 35,300 people, and we support 143,000 full time jobs, investing £3.8 billion with 5,000 supplier companies. We are committed to levelling up regional communities by transferring knowledge, skills and technology.\*

We invest £1.1 billion in emerging and sustainable technologies in the UK every year, both directly and on behalf of the Government and we are targeting net zero greenhouse gas emissions across all our operations by 2030. With 72 percent of our UK workforce in highly skilled engineering roles, we spend more than £90 million annually on training our people and with schools, colleges and universities.

Discover more at [www.baesystems.com/en/home](http://www.baesystems.com/en/home)

We're always looking for talented people to join our team and help protect the UK and our allies. Scan the QR code to go straight to our recruitment site.



Discover more at [www.baesystems.com/en/careers](http://www.baesystems.com/en/careers)

\*Data from 'BAE Systems' Contribution to the UK Economy' published 2022

## WIRED Consulting.

WIRED is known for its fresh thinking and deep expertise on the technological, scientific and societal trends that are shaping our world.

Consulting is a division of WIRED dedicated to taking the unique WIRED network, knowledge and brand to commercial organisations – helping them to build internal knowledge, develop strategy and create thought leading content that positions their brand at the cutting edge of trends.

Discover more at [consulting.wired.co.uk](http://consulting.wired.co.uk).