



Summary: Autonomous Weapons

Autonomous weapons are here. Weapons that can select and engage targets without further intervention are already here, from [robot dogs with sniper rifles](#), to [automated machine guns](#), to [kamikaze drones](#), likely already having [claimed their first kill](#).

Development is ramping up. The Pentagon has [already put someone in charge](#) of “algorithmic warfare” and has [requested over \\$3 billion](#) for AI-related activities in 2024. The story is the same globally with spending on military robotics estimated at around \$7.5 billion in 2015 and expected to grow to [over \\$16 billion](#).

Autonomous weapons (AWs) bring many different potential harms, including:

- Increased military conflict and death as both starting a war and pulling the trigger become easier and potentially more frequent.
- Putting military grade destruction in the hands of civilians.
- Loss of an ability to resist authoritarianism, as AWs stand to [shift the balance of power](#) away from our ability to bear arms guaranteeing a check on power, giving a small authoritarian group unprecedented power to subjugate the rest.
- Unintended destruction or escalation as a result of losing control of AWs.

Keeping a human in the loop isn’t a panacea. Restricting all autonomous weapons development to systems which require heavy human involvement is not a sustainable solution because adversaries are [unlikely to restrict themselves similarly](#). The U.S. will likely reduce human involvement in response, so as not to lose its tactical advantage.

Guardrails on further development is needed, and there are many options:

- We can start with policies that have broad appeal, like a commitment to thoroughly testing AWs before fielding them in battle, or a commitment to keep [actions surrounding nuclear weapons in human hands](#).
- Binding treaties will be necessary, and the current venue for discussing them (the [UN CCW](#)) is unlikely to achieve one soon. One promising direction is encouraging NATO to take a leading role here, potentially serving as a staging ground for further conversations and developing guidance for further development through its [“principles for responsible use”](#).
- Domestically we’re currently [quite vulnerable](#) to AWs like drones. To better protect our public spaces we’ll need clarity on the jurisdictional mess that currently responds to such incidents, and further research and development towards cheap and scalable ways to detect and respond to potential threats.

Full Report: Autonomous Weapons

“[Senior targeting officer Joey Temple] estimates that, with [an autonomous targeting system’s] assistance, he can now sign off on as many as 80 targets in an hour of work, versus 30 without it. He describes the process of concurring with the algorithm’s conclusions in a rapid staccato:

‘Accept. Accept. Accept.’” — [Bloomberg](#)

Introduction

Alfred Nobel, namesake for the Nobel Peace Prize, invented dynamite hoping unleashing such a potent force upon the battlefield [might end war](#). Richard Gatling invented the Gatling gun in hopes of [making war more humane](#) and reducing the number of people needed to wage war.

In reality, war was [immediately made bloodier](#) and the piles of dead on the battlefield only grew larger following the introduction of new, even more deadly technology. Use of AI to create autonomous weapons comes with similar promises of being a force used for good, and holds similar potential to bring further suffering and death to the battlefield. As autonomous weapon development takes off, we should consider what effect this technology is likely to have, and whether further development is in our best interests.

This report will introduce what autonomous weapons are and their current use cases, followed by their potential risks and the policies that could mitigate them.

Current Autonomous Weapons

For the purposes of this report, **autonomous weapons** (AWs) are **weapons which are goal-oriented and self-directed** rather than merely executing a rigid, pre-programmed, step-by-step plan. Or, as the [DoD defines them](#), “weapons that, once activated, can select and engage targets without further human intervention”. What do these look like?

Consider Turkish weapon manufacturer STM’s [Kargu attack drone](#), pictured below. The drone can autonomously identify targets, track them down at speeds of up to 90 mph, and then execute them, all with no **human in the loop**, *i.e.* **without a human having to intervene anywhere in the process**. The Kargu has likely **claimed the first kill** for AWs, a [UN Security Council report](#) on the Second Libyan War writing: “retreating [enemies] were subsequently hunted down [by government forces] and remotely engaged by... lethal autonomous weapons systems such as the STM Kargu-2... programmed to attack targets without requiring [connection]”.

If this weren't worrying enough, the Kargu, and other [“kamikaze” drones](#) STM has developed, are increasingly capable of acting together as a unit, creating what's known as a **drone swarm**. Not unlike current military units, fully capable drone swarms will **involve groups of drones able to coordinate and take various formations and carry out an attack that's dynamic and responsive to changing conditions** on the battlefield. According to the New Scientist, Israel has [already used a rudimentary drone swarm](#) in the war in Gaza, and they're not alone. Further development is underway in militaries across the world, including the U.S. which, according to Forbes, might have [swarms ready to deploy this year](#).



[The Turkish Kargu-2 Kamikaze drone.](#)

AWs are not limited to drones though. As Forbes has reported, the DHS has contracted Ghost Robotics to create autonomous robots which can patrol the border, a model with [a sniper rifle attached](#) recently showcased as seen below. A recent showing at Edwards Air Force Base saw a [fully AI powered F-16](#) take flight for the first time, going on to participate in a mock dogfight¹.

¹ [A dogfight](#) is an “aerial battle between fighter aircraft”



[Ghost Robotics's quadruped armed robot unveiled in 2021.](#)

Defensive AWs are in development too. According to the New York Times, in Ukraine the company Roboneers [has developed an automated machine gun](#) which does everything short of actually pulling the trigger, seen below. These systems aren't far from full autonomy though, as even these weapons which require a human in the loop are often only a software update away. When an executive from Roboneers was asked how long it would take to make it fully autonomous, [he responded "by tomorrow"](#).



[Roboneers automated machine gun.](#)

Efforts are already underway to incorporate these weapons into military functions. [Seen below](#), the US is [conducting training exercises](#) with, [and against](#), AWs, with over [800 live AI projects](#) according to Bloomberg. Israel and Ukraine are even further ahead, having [already fielded AWs](#).



[A U.S. infantryman with an autonomous Ghost robotics dog.](#)

Future Development and Deployment

Development has certainly begun, and already borne fruit, but we're only at the beginning. The Pentagon has [already put someone in charge](#) of “algorithmic warfare” and according to Defense Scoop has [requested over \\$3 billion](#) for AI-related activities in 2024. As US Deputy Secretary of Defense Kathleen Hicks [put it in 2023](#), the goal is “to field...autonomous systems at scale of multiple thousands, in multiple domains, within the next 18-to-24 months...[and] the Secretary has asked me to personally oversee it”. As she put it more succinctly later on in the same speech, “we are not taking our foot off the gas, and in fact we're accelerating”.

The same trend is visible globally, as spending on military robotics more broadly was estimated at around \$7.5 billion in 2015, and expected to grow to [over \\$16 billion](#) by 2025 by Statista researchers, 100% growth in under 10 years. As a Ukrainian cabinet

member [remarked recently](#) in the New York Times: “We need to win first...To do that, we will do everything we can to introduce automation to its maximum”.

The Harms of Autonomous Weapons

Increased Military Conflict and Death

The bar to entry for war is likely to only lower as AWs are increasingly deployed. As countries debate the costs and benefits of waging wars, the prospect of waging war without expending human lives is likely to lure countries into thinking they can achieve their strategic aims without the backlash that normally comes with expending human lives on the battlefield. But these wars are likely to end with human bloodshed all the same. The currency of war is human death, and as Scharre has remarked “violence...is always inflicted on people. It will always be so, because it is the pain and suffering that causes the enemy to surrender”. Further war is also risked by fielding technologies that might serve to escalate things further than they would have gone had humans maintained control².

While conflicts are underway, we’re also likely to have more ruthless killing on the frontline. Those effectively pulling the trigger are no longer frontline soldiers that have to face their victim, so hesitation will become a thing of the past. Studies in psychology tell us that the further away someone is from their target the easier it is to pull the trigger, and the coders and deployers of these systems will be far away both in terms of space and time. When such deployment results in atrocities, we run into the so-called “accountability gap.” It’s not clear who will be legally to blame for deaths caused accidentally by these systems³, leaving the immoral and reckless to profit while others suffer, with no one held accountable for the resulting tragedies.

Deadly Capabilities Spread Beyond the Military

Further AW development stands to put destructive capacity in the hands of more people, whether it be other countries, terrorists, or everyday citizens with a vengeance. One way this could happen is through military developments being leaked. Cybersecurity for major U.S. defense projects has been insufficient [time and again](#). For example, data from the Joint Strike Fighters project, the US fighter jet program which

² As Scharre recounts in *Army of None*, during the Cuban Missile Crisis, the president’s senior advisors resolved that if a U-2 plane they had sent to fly over Cuba was shot down, the United States would attack Soviet Russia without any further debate. That U-2 was shot down, but despite their earlier decision they reconvened and decided not to attack, likely because the possibility of starting a war became more real in that moment and they reconsidered whether that was really worth it. Had they sent an lethal autonomous drone instead and given it the instructions to fire back if fired upon, war likely would have ensued.

³ As argued by Human Rights Watch, placing blame on the soldier or commander for unintended deaths caused by AWs is not legally viable, but [neither is holding the developer liable](#) as defense contractors are largely shielded from liability during war.

eventually led to the development of the F-35, [was leaked to the Chinese government](#), which is remarkable given this was the most expensive weapon the US has ever developed, costing hundreds of billions of dollars⁴. That leak went to the Chinese government, but other leaks have put sensitive information in the hands of civilians too, like the [2023 Pentagon document leaks](#).

Government leaks are not the only way AWs may fall into dangerous hands. To make an AW, you really just need three simple things: a body (robotics, some way for it to take actions in the physical world), a brain (AI, a way for it to discern between different courses of action and make choices), and a weapon. We already have capable weapons and progress in both robotics and AI [has improved drastically over time](#), such that today, even those with very little experience can make rudimentary AWs.

Two [casual enthusiasts on X](#) were able to build a drone capable of identifying specific targets, tracking, and flying towards them in a matter of hours. Do the same with a larger drone with an explosive attached, and you effectively have an AW. As the New York Times has reported, this is effectively [what the Ukrainians did early on](#), taking code available online, hobbyist computers, and commercial drones available for as little as hundreds of dollars, and strapping on a bomb to finish things off (see below).

⁴ The full lifetime cost of the program, including the development, production, and maintenance of all the jets throughout their lifetimes, is now estimated to cost the US government [over \\$2 trillion](#), nearly 10% of the current (2024) U.S. GDP.



[A Ukrainian Soldier equipping a commercial drone with a grenade.](#)

Worryingly, these weapons stand to empower bad actors to cause even more harm. A man who owns a pistol and wants to fire upon a crowd can kill many, but he's limited physically and also has [a very low chance](#) of making it out alive or without being arrested after the fact. Compare that to someone who has deployed a dozen lethal drones and launched them from miles away.

Part of the risk is our inability to defend against it. An individual citizen is helpless, and the government is not much better prepared. Ranking members of the Senate Armed Services Committee have recently commented that [our radars are broadly incapable](#) of detecting or tracking small aircraft, with nearly all of our domestic military bases at risk. If our military bases lack the capacity to defend against drones, how secure do you think our civilian public spaces are?

Loss of an Ability to Resist Authoritarianism

We cherish the right to bear arms as a check on tyranny that poses a threat to any authoritarian attempt to take away our freedom. However, this power is dependent on a world in which the gun is an effective way to respond, i.e., in which enough people bearing arms actually can resist more powerful forces.

AWs stand to change that calculus. As argued in Quartz, an army of AWs will [shift the balance of power](#), giving a small authoritarian group unprecedented power to subjugate the rest. With AWs, even just one dictator can stand at the helm of an entire army, one that will heed his orders with neither questions nor moral reservations.

Loss of Control

As systems improve, and become more capable of movement, target identification, and further destruction, the stakes are raised in terms of how bad losing control of one of these systems might be. Loss of control could lead to unintended violence on and off the battlefield, as we lose the ability to abort a mission or deescalate a situation.

For example, in 2023, a US Air Force colonel claimed that in a recent test simulation an autonomous drone had [killed its operator](#) after they had tried to stop it from destroying its target. When they trained the system to stop killing the operator, it stopped, but then started destroying communication towers so that it could no longer receive orders to not execute the strike⁵.

These risks further compound as autonomy is incorporated into complex decision systems like our nuclear command and control, raising the stakes even further for what loss of control might represent.

A Human In The Loop Isn't a Panacea

A common point of agreement across the many people Scharre interviewed was that keeping humans in the loop would make a difference. Unfortunately, this isn't a reliable long-term solution. Militaries are incentivized against having a human in the loop, both to reduce bodies on the battlefield and to gain a strategic advantage through quicker responses. As this process unfolds, demands for making decisions more and more quickly are likely to make oversight nothing more than a formality.⁶

This increasing speed of battle will also lead to a tradeoff between oversight and military superiority, and it's doubtful our enemies will choose restraint. During a 2023 UN debate on AI arms control [a Russian diplomat remarked](#) "we understand that for many delegations the priority is human control. For the Russian Federation, the priorities are somewhat different."

⁵ The Air Force quickly [denied the incident](#), and the officer retracted his comments, calling them a "thought experiment" instead. But the officer still recognized the risk, [remarking](#) "we've never run that experiment, nor would we need to in order to realize that this is a plausible outcome"

⁶ Scharre mentions how many operators came to blindly trust the targeting suggestions of autonomous targeting systems in the Iraq war, 25% of the engagements of one unit resulting in friendly fire.

As adversaries throw off the shackles of keeping a human in the loop, their actions will force reactionary development to maintain military superiority. As Henry Kissinger and Graham Allison put in [a recent Foreign Affairs piece](#): “never in history has one great power fearing that a competitor might apply a new technology to threaten its survival and security forgone developing that technology for itself”. Keeping humans in the loop is helpful for now, but it leaves many harms unanswered and isn’t a sustainable solution for those it does address. Additional policy solutions will be needed to fully address the relevant risks.

Policies Directed at the Issue

In 2019, Israel’s delegation to UN negotiations on lethal autonomous weapons compared regulating AWs now to [trying to regulate the internet in the 1970s](#). Five years later, AWs are being developed in armies across the globe and we’ve already had reports of AWs deployed on the battlefield – we’re well past the realm of the theoretical. The weapons, and their potential risks, are here and growing quickly, and we can’t afford to wait any longer to begin defining effective policy solutions.

Commonsense Policy

We should start by building consensus around basic policies most stakeholders are likely to agree to. One such policy is requiring AWs to be thoroughly tested before being deployed on the battlefield. The [DoD directive on autonomous weapons](#), for example, requires “realistic system developmental and operational test and evaluation” which ensures weapons function as anticipated even against “adaptive adversaries taking realistic and practicable countermeasures”. The DoD directive also recommends strengthening cybersecurity around AW development, which would be an effective way to counter the risk of government leaks mentioned earlier.

Another such policy is taking actions to prevent AW development spreading to civilians or terrorists. Countries could take actions to better track the sale of AWs by companies, potentially instituting some sort of know your customer requirement for the sale of AWs above a certain capability. They could also commit to strengthening cybersecurity around AW projects, perhaps setting forth basic cybersecurity standards for both commercial and governmental developers of AWs to adhere to.

Finally, the U.S. could attempt to make diplomatic inroads in support of its policy of keeping decisions concerning nuclear weapon deployment under human control. This policy was initially part of the U.S. [“Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy”](#), but was later removed due the reticence of new, non-nuclear weapon holding signatories who didn’t want to legitimize nuclear weapons,

[according to Lawfare](#). The US's position remains unchanged⁷, and now likely needs to look elsewhere to build international consensus. The U.S. could [engage the other members of the security council](#), the P5, on this measure, potentially proposing it at a future P5 conference. If this doesn't work, they can seek further signatories for the [2020 joint statement with France and the UK](#), which commits to human control of nuclear weapons and would allow them to reach out to other relevant nuclear powers outside the P5. Israel, though it isn't the biggest nuclear superpower, might be a relevant country to engage here given [its ambitions](#) to become an autonomous weapon superpower.

Further Multilateral Options

Easily agreed on measures won't be enough to keep the world safe from the threat of AWs. To do that, real, binding treaties that govern the responsible use of the technology and set reasonable standards in place will be necessary.

The current de facto place for multilateral negotiations is the UN Convention on Certain Conventional Weapons (CCW). The group first met to specifically consider autonomous weapons [officially in 2017](#), convening [informally](#) for the three preceding years. Their [overall mandate](#) is to "ban or restrict the use of specific types of weapons that are considered to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately" and their work on AWs seeks to explore the issues they pose and what action should be taken in response.

But even after 10 years of informal and formal meetings the group has failed to even settle on a definition of autonomous weapons, and [according to AI expert Stuart Russell](#) who has testified at some of these meetings, it could easily be another 10 years before they find any agreement. Though defining the bounds of AWs is certainly no easy task, the continued conversation on the lack of an ability to define AWs⁸ seems to instead flow from an effort by countries who don't want regulation to impede further progress. As Nobel Laureate Jody Williams, known for her work securing a ban for landmines, [has said](#), "the so-called 'difficulty of definitions' continues to be willful obfuscation."

The reason this strategy works is that any resolution passed by the group must be [by consensus](#), meaning unless a given policy has buy-in from all the countries represented it has no chance. The stakeholders have fundamentally different incentives and preferences. Those with the ability to develop AWs (e.g. the U.S., Russia, Britain, Israel) want [nearly no no regulation](#), while those who may struggle to develop AWs (e.g.

⁷ It's already committed to human control of nuclear weapons in its [2022 Nuclear Posture Review](#)

⁸ It's worth noting that many of the countries who have claimed autonomous weapons are an issue for the future because they haven't been developed yet are the same countries [actively developing them](#) (e.g. Russia and Israel).

Argentina, Palestine) [want a full ban](#). Compromise solutions could still be possible, but none have surfaced yet, and while it's certainly worth continuing to advocate for responsible regulation in the CCW, we'll likely need to turn to other fora in the meantime.

NATO, for example, could set the precedent here with its policy on AWs. Policy set to limit development by NATO could help demonstrate to other countries that NATO member countries consider the threat serious enough that they're willing to take the first step. NATO has already [outlined its artificial intelligence strategy](#), and some of its "principles of responsible use" will certainly help with AWs. Importantly, they dictate that "clear human responsibility shall apply" to AI systems and that there must be a way to disengage these systems when they demonstrate unintended behavior. NATO can also serve as a staging ground for important discussions on AWs that actually weigh the pros and cons and move towards solutions that might meaningfully reduce the risk, like those mentioned above. NATO's already on this track as it has [already committed](#) to providing a transatlantic forum for discussing "best practice in the field of autonomous systems".

Improving Defensive Capabilities

Finally, slowing down the AW arms race is necessary to avert the risk, but in the meantime the U.S. can take steps to better prepare itself defensively for the risk of AWs to come. As the [ranking members from the Senate Armed Services Committee recently wrote](#) in the Washington Post, the U.S. is defensively vulnerable to AWs like drones as we can "hardly track anything other than commercial aircraft". This comes from both a technological and coordination problem.

While there are [already systems in place](#) at a small number of U.S. bases that can purportedly detect and respond to drones, indicating that this might partially be a procurement problem. But the US Army has paid out [\\$447 million](#) to develop and deploy the limited number of systems currently in place, so these systems are not cheap. Moreover, they seem to be designed for a military context, leaving at risk public spaces without adequate protection. Thus there's a need for further innovation in drone detection and response technology.

The Senators claim [there's a need for more coordination for protecting public spaces](#) because "U.S. agencies lack clear lines of authority about which agency is responsible for stopping these incursions". We need a clear delegation of who's in charge of what spaces, and we need either one agency to take charge in coordinating those efforts or a forum for these agencies to come together to coordinate their responses. Given that the problem is not just setting clear jurisdiction but also figuring out how to adequately

address the threat, an interagency working group coordinating progress on both would likely be the best path forwards.

To Conclude

We face down an incredibly worrying threat. Previous agreements on chemical, biological and nuclear weapons have shown that we can come together, even as adversaries, to limit the destructiveness of a new technology. In a [2018 Ipsos poll](#), 52% of Americans indicated they were opposed to “lethal autonomous weapons”. Even if Americans at large aren’t familiar with the strategic situation that is unfolding around autonomous weapons, we owe it to them to develop these weapons with as much caution as possible, and take the actions we can to steer the world away from a runaway arms race that would benefit no one.