



Artificial Intelligence in Weapons Systems Committee

Corrected oral evidence: Artificial Intelligence in Weapons Systems

Thursday 8 June 2023

11.15 am

[Watch the meeting](#)

Members present: Lord Lisvane (The Chair); Lord Browne of Ladyton; Lord Clement-Jones; The Lord Bishop of Coventry; Baroness Doocey; Lord Fairfax of Cameron; Lord Grocott; Lord Hamilton of Epsom; Baroness Hodgson of Abinger; Lord Houghton of Richmond; Lord Mitchell; Lord Sarfraz; Lord Triesman.

Evidence Session No. 8

Heard in Public

Questions 109 - 119

Witnesses

I: Richard Moyes, Managing Director, Article 36; Professor Noel Sharkey, Emeritus Professor of AI and Robotics and Professor of Public Engagement, University of Sheffield; Dr Paddy Walker, Senior Research Fellow in Modern War Studies, University of Buckingham.

USE OF THE TRANSCRIPT

1. This is a corrected transcript of evidence taken in public and webcast on www.parliamentlive.tv.
2. Any public use of, or reference to, the contents should make clear that neither Members nor witnesses have had the opportunity to correct the record. If in doubt as to the propriety of using the transcript, please contact the Clerk of the Committee.

Examination of witnesses

Richard Moyes, Professor Noel Sharkey and Dr Paddy Walker.

Q109 **The Chair:** Good morning. Thank you very much for joining us. We have Professor Sharkey remotely and Richard Moyes and Dr Paddy Walker with us in the committee room. Perhaps you could begin by very briefly introducing yourselves.

Richard Moyes: Good morning and thank you for inviting me to give evidence to the committee. I work for an organisation called Article 36, which is an NGO, a civil society organisation, based in the UK. It works on weapons policy and law-related issues. I have a background working first of all in landmine clearance and explosive ordnance disposal operations and subsequently on developing new international policy and legal instruments on a range of weapons systems. I currently co-ordinate Stop Killer Robots, which is an international civil society partnership that is concerned with the issue of autonomy in weapons systems, which, of course, intersects very closely with issues of AI and weapons systems.

Dr Paddy Walker: Thank you for having me. I have a PhD in a very particular part of the field under discussion concerned with the challenges to the deployment of autonomous weapons systems. My postdoctoral work focuses on the behavioural, soft—if you can call it that—leadership implications of removing human supervision in lethal engagements. By way of context, I am an ex-military cavalry officer of no note a long time ago. I am a senior fellow at the University of Buckingham's Humanities Research Institute, and just down the road from here I am an associate fellow at RUSI. Until a couple of weeks ago, I was co-chair of the NGO Human Rights Watch's London committee.

The Chair: I am sure that the Skins provided a very good foundation for your defence knowledge.

Professor Noel Sharkey: Good morning. Thank you for having this committee as well as having me. I am very pleased to see this development at the House of Lords. I am an emeritus professor at the University of Sheffield, a professor of artificial intelligence and robotics. I am also professor of public engagement. I started writing publicly first in the *Guardian*, warning about these autonomous weapons in 2007, and I am now chairman of the International Committee for Robot Arms Control, a group of scholars that started in 2009. We are a founding member of the Campaign to Stop Killer Robots. That has been going for 10 years at the United Nations and it is very depressing, I must say.

The Chair: Thank you very much. As you know, the format of the meeting is that it is being broadcast and after the meeting you will get a transcript so that you can correct any errors that may have crept in. They very rarely do.

Q110 **Lord Grocott:** This is a question about definition. Pretty early on in our inquiry we found that there were a number of different definitions of

autonomous weapons systems. One of our witnesses even said that it did not matter that there was no agreed definition of what was meant by it. However, it certainly seems to me, and probably to the committee, that if you are going to have an international agreement on anything connected with them, you need to know what you are talking about—in other words, you need to have a definition.

Allied to that, because so much of this discussion repeatedly has been to do with hypotheticals, we would be very grateful if you could give us any examples of autonomous weapons systems that are currently in use that you are aware of and some estimate as to what their impact has been.

Dr Paddy Walker: I will start on the definitional piece and then hand over to Noel and Richard on the second piece.

I have some sympathy with your previous participant, going perhaps at 40,000 feet, that we should not get too hung up on precise definitions. I will try to occupy that 40,000-feet position by saying that my definition of a robotic, autonomous, weapon would be that it has the ability to sense and to act unilaterally depending upon how it is programmed.

For me, the issue is agency and where that agency lies. Agency, as far as I am concerned, is the action or intervention, the control, the responsibility of the in-field commander, wherever that is and whoever has allowed these machines to be used, producing a specific effect. I am sure we will return to the issue later on, but it is that agency profusion—a weapons system that perhaps has agency with its automatic target recognition or somewhere within the whole piece of the weapon; you may have meaningful human control up top, but down below you have that profusion—and that agency dilution where we start to have problems because the user cannot comprehend necessarily what effects are to be delivered by that weapons system.

The Chair: Professor Sharkey, would you agree with that?

Professor Noel Sharkey: Yes, I would. I would take a slightly simpler definition, because as a scientist I go along with Einstein's dictum that we should make our explanations as simple as possible but no simpler. The United States started this off with a kind of one-liner, which is a weapons system that, once activated, can select targets and kill them without further human supervision. Most of the states now have some version of this with a little bit of elaboration, including the ICRC. The important thing here for us all is the critical functions; it is not just autonomy of the weapons systems but the critical functions of target identification and delegation of the decision to kill. For me, it is this delegation of the decision to kill that is a key factor.

The Chair: Mr Moyes, are you on the same page with this?

Richard Moyes: Yes, broadly. Noel mentioned the International Committee of the Red Cross. It has published a definition on this, and our orientation to this subject matter would follow the same basic conceptual terms. When we talk about autonomous weapons systems, specifically

that form of words, we are talking about weapons systems that use sensors to determine specifically where and specifically when force should be applied. The human commander having put that weapon into operation in an area, specifically when and where force is applied is then determined by the weapons system, because that system has identified something using its sensors that matches the on-board target profile. That basic mechanical structure of sensors determining where and when force is applied is, for us, the foundational question here.

There are all sorts of other potential areas of what might be called autonomy in weapons systems, but for autonomous weapons we are talking specifically about the target identification process or the location and time of force occurring being determined by the sensor array.

Coming on to the other part of the question on systems that work in this way, we believe that there should be additional prohibitions and regulations in these areas, but we are not suggesting that all systems that work in this way should be prohibited. There are already a large number of systems that function broadly within this structural framework, particularly defensive systems such as missile defence systems, where they are monitoring an area of sky and, when a radar signature matches the programmed profiles of the weapons system, they respond to it. That kind of missile defence system falls within this structure of operation, but for us it is being done in an effectively controlled manner.

An area of more concern is systems like the HAROP loitering munition perhaps, which can operate over a significant period of time flying over an area and has the potential within that area to seek out targets that match a certain profile. Our concern is that as those operations expand over greater areas of space and longer periods of time, and where the target profiles themselves are increasingly complex, a user's ability to predict effectively what that system will do becomes diffuse. That is where we think we need constraints to ensure that human control can be applied.

Professor Noel Sharkey: I had not finished that part of the question about which weapons are available today. I just left that for the definitional bit.

Since I started on this, a number of platforms have been touted as being autonomous, although they are still being developed today. There were systems, tanks for instance, in China and Russia. I will not go into the details too much because you want to know what is happening now, but there are a number of platforms on the sea and in the air. The United States developed the X-47B, for instance, a fighter jet with a lot more reach than its usual F-16s and F-35s on aircraft carriers for use in the Pacific. So many of these have been developed, and particularly in China now the developments are towards swarms of small autonomous robots. It is very difficult to find out how far they are developed. We get intelligence, but we do not know for sure.

One system that was developed that is worth noting was a sniper detector called the REDOWL, which would detect a sniper using sound differential between two speakers. It could swing around immediately and shoot that sniper, but it has not been deployed because the problem has been ricochet—it will fire at anything when you get ricochet. This is one of the dangers of warfare. It is all very well in theory, but once you get ricochet it will start shooting. If you have one of your own people fire at the sniper, it will immediately possibly shoot them as well, so that has not been deployed.

Today, the Kargu-2 has been developed in Turkey and is apparently fully autonomous. I have not tested these things myself. Has it been tested in Syria? We believe so. We believe that it has been tested in Turkey itself, but we have no real evidence for this as yet.

One of the most worrying things is that most people just develop the platforms. Ever since the campaign started and the press picked up on it, everybody says, “Of course it will have human control”. However, Russia is different. Kalashnikov has been openly testing a targeting system for a few years now, perhaps four or five, and the *New York Times* was allowed film that system. It works on machine learning, so it learns targets and fires directly on them. This is a generalised system that can appear in anything. I know that Kalashnikov has developed a couple of drone systems that are being used in Ukraine at the moment. It touts them as having these systems on board. Its words are that they were good for target acquisition and hitting targets. Those are its words, whatever the Russian for that is. Let me just look at my notes to tell you what they are called, because I am not too good at remembering these. They are the KUB-BLA drone and the Lancet drone. Ukraine also has the—excuse my pronunciation—Bayraktar drone, which is also supposed to have fully autonomous modes.

The Chair: Hansard may need to make further inquiries of you after our session.

Professor Noel Sharkey: Yes, certainly.

Lord Grocott: I still am not clear. Richard Moyes said that there are weapons systems in existence at present that have degrees of autonomy and which you would not want to ban or get some agreement on banning. Can you give us any examples of existing weapons systems with degrees of autonomy that you would want to see banned? I will put the same question to Professor Sharkey. I want to get some clarity. I want to see something, a piece of kit at the moment, that is unacceptable or at least that you hope would become unacceptable. Otherwise, it is very difficult to get a grip on precisely what we are talking about.

Richard Moyes: From our perspective, we would want to see prohibitions on systems that cannot be used with meaningful human control. And on antipersonnel systems functioning in this way.

In terms of systems that cannot be used with meaningful human control, I am not sure that there are any straightforwardly now, but if we had systems that used machine learning where they were continuing to learn their target profiles after they have been put into operation, that for us would straightforwardly be incompatible with a system being effectively controlled by the operator. Similarly, if a system could not be constrained in its duration and its area of functioning, that would be prohibited. Systems like the HAROP and other loitering munitions systems are on the boundaries of what we would consider to be acceptable. We need to draw prohibitions now on the controllability to prevent a movement into that area in the future.

On the antipersonnel side, there are already sentry gun systems on the Korean border that have the capability of targeting people autonomously. They are not used in that mode at present, but as far as we are concerned we should put in place a line against allowing systems that would be used in that mode in the future. These days, there has not been a general adoption of systems that automatically, autonomously, target people, and if we as a society can put in place a barrier to that now, we should do so. It does not bode well for the protection of humanity in conflict, for troops as well as for civilians, if we allow machines that can automatically process people and kill them. If those sentry guns were used in an autonomous antipersonnel mode, we would call for them to be prohibited. If the target profiles of certain existing systems were developed in particularly complex ways, we would also consider that those should be prohibited.

Dr Paddy Walker: Can I introduce one other line here? To my thinking, there is a material step change in both hardware and software that is required before these machines may be deployed, from a laws of armed combat perspective but also from a user perspective. There are gaping chasms in the technology, be it automatic target recognition—I am sure we will come on to data later—ways of backfilling for missing data, duplicated data, contradictory data, and data that is obsolete. I listened briefly to your previous speaker in the last session talking about stochastic models and prediction and statistics backing these things up. They are incredibly fragile.

I still think there is generally a revolution in expectation. We have seen this thing on “The Terminator” or whatever, or in a presentation or a publication, so it exists. They do not necessarily exist. The other danger that comes from that is that that is norm changing as well. Suddenly it becomes acceptable to start thinking that we are going to have lethal engagements with no meaningful human control. As Noel said, that point of delegating the decision to kill to an algorithm suddenly becomes possible because “I’ve seen it in a film”, or whatever, “So these things exist”. I am not sure that is so.

Professor Noel Sharkey: If we look at the existing weapons, we can perhaps throw some light on the question that is being asked here. There are a number of what you can call SARMO weapons—weapons that sense

and react to military objects. The Phalanx, which I am sure you all know, is one of those, as is the Iron Dome system. Systems that are used to shoot down military objects or prevent swarming have gone wrong at times.

Let us look briefly here at the MANTIS NB, which is a German weapon used for shooting down mortar shells, which is obviously very useful in warfare. It can shoot down mortars very quickly. If you look at the manufacturers' small print—I am a person who reads small print—it says that the same system could also be used for locating where the mortar came from. So it could immediately autonomously target and fire a rocket or whatever weapon it wanted to at that person.

In one case, you have the SARMO weapon that is shooting down military objects. That is okay. That is defensive. We need that protection. But then we turn to the other side, just firing a weapon autonomously at where the mortar was—I do not know how fast they get out of the way—without looking to see whether there are civilians present or whether where it was fired from is in the middle of a village or a town. This would be an indiscriminate use of a weapon, which is where you draw the line completely between an autonomous weapon and a SARMO weapon.

It is the same with the loitering munitions, which are what I would call a cusp weapon: they use radar to determine where an anti-aircraft installation is and take it out. The assumption here is that that radar is connected to an anti-aircraft installation. It could be on top of a hospital roof or anything else, so it is not entirely discriminatory. The MANTIS NB with that extension would be in the realm of an autonomous weapon. It is being delegated with the decision to kill.

The Chair: Thank you. In a moment, I will ask Lord Clement-Jones to move us into a new area of criteria, but Lord Hamilton and Lord Sarfraz have a very brief question each first.

Q111 **Lord Hamilton of Epsom:** This is for Richard Moyes. He says that he would certainly ban an autonomous system that killed invading troops on the Korean border. Would he still hold that view if the North Koreans developed an identical system that did the same thing?

Richard Moyes: Yes, I think so. The issue for us here is drawing a line against adopting any systems that automatically just process people for being killed. There is something fundamentally dehumanising about reducing people to simply points of data and—I will use the word again—processing them for death without a human being involved in engaging with the decision to apply force to a specific person.

As a society, we have the technological capability to have systems that function in that way now and we have had it for quite a long time. We have managed not to choose to operate in that way. With all these things, if we can establish the boundaries now that we want to protect and preserve, it is better to do it now than to try to pull back from the use of systems once they have been adopted, whether that is

normatively or through technological investment in certain systems. I feel that strengthening that normative line now would be socially preferable for us all.

From a military perspective, of course, you can make a case for why this would be a useful asset, but it is very much the same case as the antipersonnel landmine. We are looking at a similar dynamic of functioning there. In the end, there is a grave social risk in deciding that machines can simply determine who to apply force to or simply to apply force to everybody within a particular area. There is a risk that that just pushes the burden on to the civilian population to exclude themselves from those zones.

Q112 **Lord Sarfraz:** I think this question was answered, but we have had a lot of evidence, particularly written evidence, on the risks and why these things can be bad. You gave the example of a landmine, which is an autonomous weapons system. If you apply AI to it and make it a smart landmine, which is much safer and much better, is that an example of this technology being used to make autonomous systems better and safer for the world?

Dr Paddy Walker: Not my bailiwick, I am afraid.

Professor Noel Sharkey: I think Richard should answer this, since he was the culprit.

The Chair: I can see that the buck is passing very efficiently from hand to hand.

Professor Noel Sharkey: He has good answers.

Richard Moyes: Antipersonnel landmines are prohibited under the mine ban treaty, and in those discussions certain lines of argument about adding additional technological capabilities to landmines in order to make them somehow acceptable were rejected. It was decided with respect to antipersonnel mines that they should be prohibited.

On the other hand, I am not arguing that within the space of autonomous weapons in general there is no capacity for AI to perform functions that are useful. Our concerns about AI functions specifically in this space derive more from issues of the complexity of target profile development and the potential of systems to use target profiles that cannot be effectively understood by the operator of the system, or that change subsequent to the system being used, which is another version of the same thing, because the operator then does not fully understand the implications of that system.

I do not think that I am arguing for a complete removal of AI capabilities from all aspects of weapons systems. It is simply that the users of systems need to be able to sufficiently understand the system that they are using and what will trigger an application of force by that system in order to make reasonable determinations about the likely outcomes of using that system in a specific context.

In relation to the issue of antipersonnel weapons systems specifically, if the argument is that AI will allow us to distinguish between civilians and combatants in some way, that is extremely fraught and dangerous ground. It is dangerous on legal terms, because ultimately the legal targetability of individual people is contextually determined. It is not based on some intrinsic physically identifiable characteristics. Even if you are carrying a gun, it does not make you automatically targetable in combat.

If we start to develop target profiles based on datasets and algorithms, there is the potential for bias to come into that such that we start to identify, perhaps accidentally, perhaps deliberately, people of certain skin colour or age or gender characteristics as being combatants. Already in the US drone programme we have seen people killed in the vicinity of a drone strike if they are men between the ages of 16 and 70, say they may be considered to be military casualties rather than civilian casualties. That does not align with the legal determinations as to whether people are targetable or not.

Once we bring AI into this process of dividing up between civilians and combatants, I think we run a very grave risk of prejudicially bringing to bear social biases in those determinations, and anyway we are undertaking a process that I think is legally extremely fraught.

Q113 Lord Clement-Jones: We have started answering the question about what it is about the use of AI in weapons systems that warrants additional concern or scrutiny above what is required for non-AI-powered weapons. I just want to provoke you with the comments of some of our earlier witnesses. One said that AI is really just the enhanced use of algorithms. He also said that AI is simply applied statistics. Another witness said that use of AI in weapons is not as significant as nuclear. In that context, what are we really concerned about with AI? I think one of them also said that this is only in circumstances where you hand over command and control to AI. What is the essence of the concern?

Professor Noel Sharkey: I do not have any objection to what you have just said that other people have said. I do not see how that has bearing on it. At the moment, yes, AI is governed by statistical pattern recognition, which, trendily, is called machine learning. The word "learning" causes a lot of problems, but it is statistical pattern recognition. In fact, it is parameter estimation for statistical pattern recognition, but unfortunately that just does not catch headlines, which does not help at all with the uncertainty in this.

I think the last major upgrade of the laws of war was in 1977, and that was Geneva Convention Additional Protocol 1. In those days, we were getting things like the little personal computers. There was no idea at that time that computers might be used and that artificial intelligence might be developed to take control of weapons. When you read the Geneva convention you will find that it does not talk about humans doing this or humans doing that or about humans being in control of this or that. It does not talk about it, because it is a default. It is completely

considered to be human, so at least it needs that much upgrade, if nothing else, before it is IHL compatible.

The problem here is whether we can guarantee compliance with the laws of war, with proportionality, with the principle of the distinction, with the principle of proportionality and precaution? I would say absolutely not. We cannot guarantee this. One reason for this, if you look at machine learning, is that it is trained from examples, and you can give it billions of examples. Where the examples for warfare are coming from I do not know. It will not be trained on the battlefield, because it can take millions of iterations to learn something and you do not want that many people to die while it is learning.

You train it off the battlefield and then you test it, but you test it on generalisation. Generalisation essentially means examples that it has not seen before. These examples will fall in with the general remit. For instance, if you have a chess game, you will train it on many games and then you will give it novel games. It will work on those, but they are inside the same domain. The trouble with the battlefield is that it is replete with unanticipated circumstances. When I say replete, possibly an infinite number of things can happen: a number of tricks, a number of spoofs—a number of different things. It is so obvious, like US marines in Afghanistan catching a lot of insurgents in an alleyway, for example. They raised their guns and were about to shoot them, but noticed that they were carrying a coffin, so they lowered their guns, took off their helmets, bowed their heads and let them pass. An autonomous weapon would probably have just killed them all. If you trained an autonomous weapon not to kill people carrying coffins, all insurgents would be carrying coffins. It is quite an odd thing. So we have this idea. I could go on about this for hours and I must not because we will run out of time.

The Chair: We are quite short of time.

Professor Noel Sharkey: Okay. Let me just give you one more thing. When we look at the laws of war and at what big thing is being pushed by the United States, and rightly so, it is Article 36 reviews. Review of the reviews has shown that only about 12 nations are actually capable of carrying out Article 36 reviews. Article 36 reviews—I am sure you all know what they are—are for testing weapons. Say you have a new missile or a new bomb. You will look at the kinetic energy it produces, that kind of thing.

I really promote the use of Article 36, but using it for an autonomous weapon is almost impossible. You can test it as much as you want in the laboratory, but you need to do empirical testing. There is currently no formal way of testing an autonomous weapon and none on the horizon, although a lot of people have tried. This applies to autonomous systems in general. You can test it all you want in simulation, but you will not know how it will behave on the battlefield, particularly against other military systems. I will leave it there, because there is a time shortage.

Richard Moyes: I have made a few comments already that cover aspects of this. I would just note again the distinction between autonomous weapons systems and AI weapons systems, which I noted at the beginning.

The key problem for me is the issue of control. Under the laws of war, as I think Noel pointed to, international humanitarian law applies to people. It is people who have the legal obligations to fulfil the legal requirements and who will be held accountable for the outcomes of their actions. The concern for us is the ways in which, within autonomous weapons systems, certain AI functions in the development of target profiles or the identification of targets can impede a user's ability to adequately predict what that system will do in a particular context of use. It is that human user who has the obligation to determine the likely balance of civilian harms and military utilities that will come from an attack. It is that human user who has the responsibility and who ultimately will be held accountable for what happens from their action.

We need to be able to maintain that chain of accountability through the technology that is used, and there is a risk that certain AI systems in certain functions lose track of that. I mentioned issues of potentially targeting people and the processing there, so I will not reopen that section.

Dr Paddy Walker: I would agree with your previous speakers about the enhanced algorithms and applied statistics, but the key distinction here was brought out by Noel: you are slaving that to the delegation of the decision to kill. I have a handful of technical problems that I will address.

First, everything in an autonomous weapon, because it is autonomous and you cannot speak to it, is doing its own thing. All the commander's vision and battle plan has to be captured in code, and it is that problem of coding to capture ambiguity, intangibles, context and situational awareness, all those things, that are very difficult to abstract and do not have nice, clean edges that make machine learning a difficult technical spine upon which to base autonomous weapons. After all, the whole issue, as we have somewhat discussed, revolves around data. We have talked about obsolete data and the fact that it does not arrive at the right time and it is not nice and clean and is very easy to spoof, as Noel said.

Training these weapons systems will require unbelievably specific data training sets. These do not exist. To the extent that they do exist and a country will have them, they are restricted and confidential. Like the piece that Noel mentioned about how a battlefield is incredibly fluid, dynamic, adversarial and prone to spoof, data as a result is a very inappropriate backbone. If that training dataset thinks that a tank moves obliquely—I am thinking of the 5th Royal Inniskilling Dragoon Guards, Lord Chair—is grey, has a large turret coming out of it, has a particular RF signature, how do I defeat that on my right flank? I will paint my turret pink or whatever. We have all seen examples where less than one half of one half of a per cent of a pixelated picture, once changed, it will not be a banana; it will become a pineapple or whatever.

These are enduring technical challenges where some progress looks to have been made by way of GUI chips and enhancing further the processing power of computers. There look to have been some improvements over the last couple of decades, but we are still essentially using the same backbone to try to crack these problems.

We have talked a lot about sensed data. The machine is going along, the weapon is going along, and it is sensing its data. There are really difficult issues here. How often is it taking data from its sensors? How often is it polling its sensors? This, after all, is incredibly processing-heavy. On the other hand, if you do not do it every second, every minute or whatever, there is the issue of obsolescence. No work is being done on any university bench at the moment on anchoring, the degree to which you change the initial representation of your autonomous weapon with regard to the newly sensed data that has come in. If a machine is going down a road, comes to a pothole and has to move 30 degrees to the right, does it always have to move 30 degrees to the right? It is the mediation of newly sensed data into representations which no one is particularly talking about.

The last thing I would note is some of the operational issues with these autonomous weapons. You set them off on a Tuesday. One goes behind a hill and one goes to your right flank. The following day they will be very different machines, because they are benefiting from experience. They are machine learning. They are updating their representation, as I talked about before. How as the local commander are you to understand the assets that you have at your disposal at that point? One has had to be patched, it has a new software update, but that one was behind the hill. You do not know if it received that. To Noel's point about verification, validation, testing and predictability, do you have a robust system? These are some of the technical issues that I think come from that question.

Q114 Lord Browne of Ladyton: Professor Sharkey has twice made reference to Additional Protocol 1 of the 1949 Geneva Convention. In particular, he did not mince his words about Article 36 and the requirement to conduct legal reviews of all new weapons systems, means and methods of warfare to determine whether the use is prohibited by international law. He does not believe that that is possible to the standard that would be necessary.

Given that we have Richard Moyes here, I want to explore with him some other things about Article 36. We were told by witnesses from the Ministry of Defence that Article 36 and our consistent compliance with it was one of the guardrails that would ensure that our weapons systems would be compliant with international humanitarian law. The history of Article 36 is that 174 countries, I think, signed Articles 35 and 36, and only eight of them appear to carry out these processes to ensure conformity. Thankfully, the United Kingdom, which is the country that we are most concerned about, is one of them. Of the six of them, two of them do not publish anything about the regulation and the review. The US does, and we and France are the two that do not publish anything. There is nothing in the public domain that shows whether we have conformed or not, how we did it or what the regulation was.

I am interested, Richard, in whether you agree with Professor Sharkey that essentially AI-enabled weapons, particularly autonomous ones and even more particularly lethal autonomous ones, present a challenge that we cannot meet in relation to our Article 36 obligations. Can you confirm that hitherto it has not been possible for anybody to independently assess whether the United Kingdom is conforming with this, because it publishes nothing about it?

Richard Moyes: There is a lot that could be said. Broadly, yes, I think that Article 36 reviews are important. In addressing issues of autonomy in weapons systems, we will need to continue to have states reviewing at a national level the systems that they develop.

I do not want to push to one side weapon review processes altogether, but Article 36 weapon reviews are obligations to assess whether your weapons system would be legal or illegal under the existing law. From an orientation where I do not think that the existing law provides the clarity of guidelines that are needed to shape the future of technologies on this issue, those review processes themselves are automatically insufficient, because, I believe, the existing legal framework is insufficient.

They are insufficient in a few key areas, particularly because there is no clarity that we need to have adequate understanding of the systems that we use and that that has some benchmarks to it, or that the users of systems need to be able to sufficiently constrain the duration and the geographical area of the systems that they use. These tests are fairly simple conceptually but they would give states at a national level an augmented ability to evaluate whether a type of AI challenges or enables the understanding of the user in relation to the operation of that system.

Without those additional tests, we are not interrogating those AI systems effectively. We might be at our own national level just because of good practice on our part—we do not know, because it is not public—but certainly no other states know and we have no ability to ensure that other states are doing the same, which is another aspect of this: that because this is purely national and often secret, there is no norm-building mechanism at work here within the wider international community. We are not setting shared standards; we are all just going off and doing our thing. It is an appeal 'to trust' us at a national level, and we may have more or less trust for different states around the world as to how they would orientate to these questions.

Q115 The Lord Bishop of Coventry: My questions were on a very similar line, so I do not want to repeat them. It sounds, Mr Moyes, as if you are saying that international humanitarian law needs further interpretation—I think as the MoD says—in order to apply it to AI weapons systems, guidelines and so on. Am I right in understanding that, rather than saying that it needs another significant development in it, it is sufficient as it is with a layer of interpretation? Do our other witnesses think the same?

Richard Moyes: I would say yes. We think it needs additional legal augmentation, but in relation to the obligations on commanders in

attacks, we think there should be an obligation on the users of systems to sufficiently understand the systems that they use and to understand what will trigger an application of force by that system. There should be an obligation to be able to sufficiently limit the duration and area of a system's functioning so that they can apply their existing obligations under the law.

Those rules would be rules on use of systems that could be interpreted as simply enabling what the law already says. It is just that the law does not articulate these terms within itself. As Noel said, when the law was drafted in 1977 these issues were not a concern. In civilian society we have acknowledged that we need new rules on automated decision-making in the civilian space, because our existing legal structures do not fully rise to that.

The only additional area is the antipersonnel line. That is a more distinct line that is less of a straightforward interpretation of existing rules and more of a precautionary, socially based orientation to say, "We want to prevent a development of harm in these areas". That rule would be on a slightly different footing in relation to the existing law.

Dr Paddy Walker: Can I add one point to Richard's point about rules on use, which I think is a useful construct? It is not just constraints, therefore. It might also be positive obligations on the in-field commander that they know what is happening at the target end, or whatever it is, and some of the targeting profile considerations that Richard has already detailed. So constraints but also positive obligations would be part of that rule set.

The Lord Bishop of Coventry: May I just seek one clarification? It sounds as if you would be assured by what we have recently been told by the MoD in a witness statement: "There remains a need for a conscious, accountable human actor in order to be satisfied that the relevant legal requirements of that decision are being made". That is built into policy, so no doubt you will have some reassurance from that but not be convinced by it because of lack of transparency, as Lord Browne was indicating. Would that be fair?

Richard Moyes: I think those are needs. Where I would differ from the UK civil servants on these issues is on the need to move this to the level of international rule-making on the matter. I cannot, of course, interpret their position exactly, but I do not feel conceptually that I am a million miles away from UK officials on these matters. It is more about how we position ourselves politically in the international landscape.

Q116 **Lord Clement-Jones:** I do not think the answers to this question will take very long. To what extent are the UK Government a leading voice in conversations on international regulation, and where could they improve?

Professor Noel Sharkey: I think Richard Moyes has a very good answer to this on regulation, but I would say that they could be a leading voice if they were to step up and take a lead. At the moment, they do not. They

tend to just go with what the United States says. I often see them huddled in a corner talking together. Maybe I should not have said that; I am not sure that that is a permissible thing to say.

Lord Clement-Jones: You have probably damaged US-UK relationships irretrievably now.

Professor Noel Sharkey: Yes, I might have.

The Chair: But you are entirely covered by parliamentary privilege.

Professor Noel Sharkey: I will just make one very quick comment, because others have more important things to say about it, I think. One of the problems with the UK is that all the time since the beginning, since there has been Parliamentary Questions, it is always, "We will not be developing autonomous weapons. Therefore, we don't need to have a prohibition or support a prohibition". I would call that extremely blinkered. It is like there is no other community out there: "We'll develop them and we'll be very responsible and very safe". We will always be very responsible and very safe, and that worries me, because we need to stand up and stop the others who are not responsible and safe from developing their weapons. I will stop there.

Richard Moyes: I think the UK could do significantly more in international leadership on the issue. I do not know to what extent it will be apparent from the evidence you have heard already, but there is quite a substantially developed international policy conversation on this issue. There is a broad centre of opinion as to where a legal regulation on this issue should go. In the UK's 'Integrated Review on global Britain' and the 'Defence in a Competitive Age,' the UK talks quite a bit about being at the forefront of norm-setting and rule-setting and that this is important to our position in the international landscape.

Frankly, perhaps taking off my Article 36 and Stop Killer Robots hats here to say something a bit more at a national political level, I broadly agree with that posture from a national perspective about the UK's role in the world. We just do not see it in practice in terms of actually building the partnerships and driving the conversations forward. Although the civil servants on this issue have certainly been good to work with and there is a lot of thoughtful work going on there, I definitely do not see political leadership on that side.

Our assessment would be that there will probably be a resolution at the UN General Assembly later this year in October. That is a resolution that will probably invite the UN Secretary-General to consult states internationally on this issue. This is getting out of the rather narrow conversations there have been in Geneva and inviting all states to make their views known on the matter. It is likely that that resolution will not contain detailed formulations about what the outcome of this should be, so it is not a prejudicial formulation.

This is something that the UK would straightforwardly be able to support, from my perspective. It is a key test of whether it wants to be a dynamic actor on this issue, whether it takes a leadership role on that, whether it encourages other states to join that resolution, or whether it holds back and tries to keep this conversation in a rather stifled pool in Geneva, where it has been for the last 10 years. I think there is a concrete test of UK leadership that could come in the next few months.

Lord Clement-Jones: Dr Walker, do you agree?

Dr Paddy Walker: I would agree with that. I would be a little bit more outspoken, which I can be. I think the UK should understand that it is not in our interests to leave international law and norm-setting to others. The CCW, where this happens in the United Nations, is a case in point. I first went there after Human Rights Watch published its rather pejoratively named killer robots report in 2012, so that is 11 years ago. In this institution, which is where all the discussion that Richard has articulated takes place, it is Russia that is deciding on the agenda's ambition, Russia deciding on the scope, Russia deciding on the pace.

My view would be: how can this be strategically valuable to the UK, whether it is, as Noel suggests, huddling in a corner with the US or not putting our best foot forward? My recommendation would certainly be to adopt the October initiative and to step up leadership on the matter and perhaps move that negotiation of the statutory instruments from the CCW to the United Nations General Assembly.

The Chair: Thank you very much. I am now looking to Lord Triesman and Lord Sarfraz to take us into the world of future-proofing.

Q117 **Lord Triesman:** I will do my best, Lord Chair, to do this very briefly. Therefore, I will apologise in advance that the questions will come cruder in order to make them brief.

I think all our witnesses, and particularly Mr Moyes, have emphasised the need for either revised or new regulation in these areas, for all sorts of reasons. Rapid technological advances are one of them, as well as greater clarity on the guidelines we need, and some greater legal scaffolding because the current legal scaffolding, if I can put it this way, is inadequate.

I am aware that there have been a number of examples of people who have tried to future-proof by getting together in groups that have considered the variables in the future and shouldered it commercially and very successfully over a number of years. RUSI has done it; Chatham House has done it. We even had a unit that was set up in the late lamented Department for Innovation, Universities and Skills, led by Sir Keith O'Nions before he went on to Imperial College. In all those cases the future-proofing effort was abandoned as not of sufficient interest to anybody and too difficult. It all got put in a box called "too difficult". How can we get to a point where we try to future-proof successfully? I am sorry if I have put that rather crudely, but that is the question.

The Chair: We know exactly what you mean, Lord Triesman.

Richard Moyes: It is a key consideration to make sure that the rules you adopt are orientating to the future, because ultimately technologies will continue to develop in these areas. We have to have a rule set that is sufficiently flexible and sufficiently open to be workable in that context.

From our perspective, the key is to formulate rules that are based on human principles and human use requirements, the kinds of rules that I was talking about before to suggest that a user needs to sufficiently understand the system that they are using and what will trigger an application of force. That is a rule that is based on what the human needs to be able to do and it is not tied to specific technological configurations. That is something that we should avoid. We should not write rules that talk about specific forms of AI, specific computer configurations and the like. We want rules that specify what we as people want to see protected and retained in the use of force.

That, for me, is the understanding and —the sufficient ability to constrain the use of a system so that you are making meaningful human judgments about what will happen. Again, the antipersonnel prohibition is a prohibition based on systems targeting people. It is not based on systems functioning with some specific technological configuration. It would flow more from the human. That is the core of it for me: have rules that specify what we as people want to retain in the future in conflict, recognising that there may need to be guidelines and grey literature that provides more detail about how certain specific technical forms match up to that. The rules we adopt should be at a broad and human level.

Professor Noel Sharkey: I agree with Richard. I will just be very brief. Throughout the whole of AI governance in other areas other than this, the idea of future-proofing is to focus on the end users, on the people who will be subject to these systems, and do not focus on the technology. Focus on protecting people. That way you can future proof, I believe.

Dr Paddy Walker: I do not have much to add. I would go back up to that 40,000 feet that we started with at the beginning of our session: meaningful human control. I know that when you are sitting in the CCW and China and Russia are talking for days about what “meaningful” means and where the control is—. Nevertheless, it is meaningful human control and that targeting profiles piece. It is one target. It is not a rolling target. It is controlled by duration. It is controlled by space. I would agree with Noel that it is focusing on the end user. I would agree with Richard that it is those positive obligations on somebody who is actually using the kit to be able to understand what the effect of the weapon will be. If you have that in a bundle, all the technology noise that happens underneath is less important.

Q118 **Lord Sarfraz:** When we speak about regulation, it feels like we are focusing on international law primarily. Is there also a role for domestic regulation that you feel should be addressed? How do we make sure that we do not regulate ourselves out of business?

The Chair: As you will gather, there is a huge premium on brevity at this point.

Richard Moyes: It should be led by the international legal structure. There are challenges to regulating legally at a national level on these issues when you do not know where the international legal structure will land. This is an issue where I expect there to be an international legal instrument within the next three, four or five years. It will be quicker if the UK works on it. It will be slower, but it will happen if the UK does not work on it. It is a challenge to regulate nationally on this legally without knowing where that framework will settle at this stage.

Professor Noel Sharkey: I agree.

Dr Paddy Walker: I agree that it is an international forum matter. On regulating your way out of being commercially competitive in this space, that is less of an issue if you can establish the broad framework, which we must do. Lots can happen underneath. When I said that the technology is less important, I should have said that technology can develop alongside but within that framework. So I do not think that is necessarily a consideration. That would be my view.

The Chair: The final question comes from Baroness Hodgson and it has been specially designed to allow a one-sentence answer.

Q119 **Baroness Hodgson of Abinger:** Absolutely. What one recommendation would you make to the UK Government?

Dr Paddy Walker: I would listen to Richard Moyes.

Richard Moyes: I would recommend that they join the UN General Assembly resolution in October on autonomous weapons.

Professor Noel Sharkey: I agree with what Richard said, but I would recommend that they step up to the mark and show leadership by acknowledging that we need a prohibition on very dangerous autonomous weapons systems.

The Chair: Thank you. You have contributed greatly to our inquiry within cruel time constraints, so we are extremely grateful to all three of you. Thank you very much indeed.