HOUSE OF LORDS

Artificial Intelligence in Weapons Systems Committee

Corrected oral evidence: Artificial intelligence in weapons systems

Thursday 23 March 2023

10.05 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.

Evidence Session No. 1 Heard in Public Questions 1 - 14

Witnesses

I: Professor Noam Lubell, Professor, University of Essex School of Law; Georgia Hinds, Legal Adviser, International Committee of the Red Cross; Daragh Murray, Senior Lecturer and IHSS Fellow, School of Law, Queen Mary University of London.

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

Professor Noam Lubell, Georgia Hinds and Daragh Murray.

Q1 **The Chair:** Good morning. It is very good to welcome Professor Noam Lubell, Daragh Murray and Georgia Hinds. We are looking at you remotely, Georgia; I take it you are probably in Geneva. Is that correct?

Georgia Hinds: Yes.

The Chair: We will move around the committee covering particular areas of interest, but members are free to pursue follow-ups. If you want to add to answers later on, you are free to do that. The session this morning is being webcast and a transcript will be taken, which you can check for factual accuracy. Our plan is to go on till about noon, but let us just see how we go.

This is our first public session, at which the House requires us to declare any interests we may have. Mine is quite remote, but one of my sons-in-law, as a consultant, has done work for the ICRC. I put that on the record.

Let me begin by asking you a $64,000 question. How do you define AWS and why is it important to do so?

Professor Noam Lubell: It is an interesting question. I am not entirely convinced that we should be defining them for the purposes of these debates. We can. There are lots of definitions, and sometimes allied states end up having different definitions themselves, which is part of the problem.

There are four reasons why we maybe should not rush to define them. First, it is very difficult to have a definition that would not include existing systems, including systems that have been in use for decades. All sorts of active radar homing or high-speed anti-radiation missiles that have been in use since the 1980s would fit a lot of the definitions that are being bandied about right now, which would complicate things.

Secondly, the definitions tend to place too much emphasis on autonomy rather than the artificial intelligence elements. Autonomy does not equate to an intelligent system. You could have dumb autonomy. You could argue that a land mine is a form of autonomous system, but our concerns are not so much with autonomy as they are with the AI that is driving a lot of these systems. Equally, you could have an AI-driven system that is not entirely autonomous and is supervised, but that raises all the same problems we want to discuss. The AI is more important than the autonomy.

Thirdly, autonomy is a characteristic we apply to just one function within a system. The system might not be classed as autonomous, but it may still have autonomous functions. Again, that makes it difficult to say a system is autonomous.

Lastly, we are talking about a very fast-paced field where things are changing all the time. If we tried to come up with a hard and fast definition now, we would have all sorts of systems that fall outside of it within a very short timeframe.

At the end of the day, we need to be focusing on the particular characteristics of systems and looking at how those apply to different functions of the system. I am not sure trying to define autonomous systems as a whole and applying everything to that is the right way to go forward.

The Chair: Is the logic of what you are saying that, if we have this rather leaky set of definitions of AWS and autonomy, in order to arrive at a working definition that will be reasonably robust against developments, we need a much more constraining concept of what AWS are?

Professor Noam Lubell: Rather than trying to have specific definitions, we need to come up with basic principles, guidelines and rules that apply to whatever systems we use, regardless of how we define them. That would be the way to go. Bear in mind that, while the definitions keep talking about weapons, some of the concerns might be with earlier stages in the targeting cycle of these systems, such as the identification of objects.

Again, I would much prefer to speak about the particular characteristics and problems that come up with AI and apply those across the targeting cycle rather than trying to talk about autonomous weapon systems as a thing.

Daragh Murray: I should not have let Noam go first. I agree very closely. My main concern with the definitions of autonomous systems I have seen is that they can be overly narrow. If we think of the targeting cycle as the identification, selection and engagement of a target, autonomous weapon systems often do all three components. You might think of a drone that is deployed to the battlefield to identify enemy tanks and engage them. That tends to be what we think of when we think of an autonomous weapon system.

As Noam was talking about, particularly in the future, the military is going to be an interconnected military with lots of different sensors drawing on a wide variety of datapoints and data sources. That is going to be characterised by pervasive AI across the entire system.

One of the things that will inevitably arise is the use of AI to identify a target, drawing on intelligence. That will then be passed to a targeting matrix and selected for engagement. The only autonomous component might be in the identification, which is entirely separate, and then it might be a normal fighter jet that attacks the target. Avoiding an overly narrow definition is really important.

Georgia Hinds: On this, the ICRC has put forward a definition, including in the CCW, that has been picked up by a number of states and aligns

quite closely with the proposals by the UK. The definition we have put forward is "a weapon system that selects and applies force to targets without human intervention".

I agree with the comments that have been made. It is definitely appropriate to take a functional approach. We are basing it on the process by which the force is applied rather than tying it to technological specifics. I agree with what has been said and the UK Government position that tying definitions to technology would absolutely risk them becoming redundant very quickly.

For us, our definition does extend to cover certain existing weapons, such as air defence systems on warships or active protection weapons on tanks. A mine would fit our definition as a very rudimentary autonomous weapon system. This demonstrates that, for us, the definition does not require complex technology. It is not a defining feature that an AWS incorporates AI. It might, but it does not necessarily.

The key point for us—this is what raises the most concerns from a legal and an ethical perspective—is that, after the weapons system is activated by a person, it is the system that self-initiates or triggers a strike or an application of force. It does this in response to environmental information. It receives this through its sensors and it has a generalised target profile that has been input at the activation stage.

The difficulty there for us, and what is at the root of most of our legal and ethical concerns, is that the user is not choosing and does not even absolutely know the specific target, the precise timing and the location of that force application. For us, this is a really fundamental difference between an AWS and some of these other types of AI-enabled systems. This might be an AI-enabled decision support system that is just providing intelligence information or feeding into the targeting cycle but does not displace those critical functions of selecting and attacking the target.

This is where I would differ. It is not problematic that our definition captures current examples. There are really two definitions that need to occur. First, you capture all autonomous systems that may raise humanitarian concerns and that need to be regulated. That is based on the idea of them selecting and attacking targets, which is very different from non-autonomous systems. Within that, you then have the definition of certain types of autonomous weapon systems that might require specific prohibitions.

We then move to the smaller category of unpredictable autonomous weapon systems, ones that do not allow humans to control and limit their effect. That would be a category of prohibition for us, along with autonomous weapon systems that are designed and used to apply force against persons.

All the other ones currently in use that do not fall within those smaller prohibited categories are subject to general regulations and constraints

on use that are already being input around duration, location and type of targets. By their nature, they are primarily being used against military objectives, such as incoming projectiles or enemy equipment, which are much more stable in their characterisation compared to civilian objects, such as a hospital that might be used temporarily as a base.

The Chair: Can I ask you to keep your answers fairly short?

Georgia Hinds: Yes, sorry, I got carried away.

The Chair: We have quite a lot of ground to cover. We can pick up that issue of regulation and constraint in a moment, because Lord Clement-Jones has questions on that.

Q2 **Lord Browne of Ladyton:** Since I have not spoken, I will go through my interests. I am the vice-chair of and a consultant to the Nuclear Threat Initiative; I am a director of both the European Leadership Network and the Asia-Pacific Leadership Network, which work on arms control matters; I am a member of what, to my embarrassment, is called the Group of Eminent Persons of the Comprehensive Nuclear Test-Ban Treaty Organization; I am a member of the Top Level Group of UK Parliamentarians for Multilateral Nuclear Disarmament and Non-proliferation; and I am very proud to be an ambassador of the HALO Trust.

On this issue of definition, as an observation, we have to have a starting point. From what we have been doing in preparing for this, the issues we are concerned about are pretty clear. They are highlighted by Ms Hinds' evidence to us. If we do not start with some definition—because we are all legislators—where do we start?

If we come into an environment in which the US Department of Defense, NATO, Human Rights Watch and the International Committee of the Red Cross all have definitions, we cannot just ignore them. We need some assistance as to where we start with this. I put this in the form of a question. Is it not sensible for us to start where everybody else is, which is struggling to find a definition?

Professor Noam Lubell: I am not sure there is a place where everybody else is. There are different definitions out there. At times we have even seen significant differences between the US and the UK, to the point that one is saying, "We use them but only with human supervision" and the other is saying, "We don't have them and we would never use them", when they are talking about the same systems. That has changed. The definitions, including those of the US and the UK, have developed with time.

One can certainly talk about the use of—again, I would not limit it to weapons—AI in the targeting cycle and in military operations. I would have something around that as my starting point, rather than defining particular weapon systems. This came out in all the comments. Artificial intelligence can play a critical role in who or what ends up being targeted even outside of a particular weapon. If you are concerned about the

dangers that could happen, there is a risk that using some of the definitions out there could limit your ability to address these concerns.

The Chair: Daragh Murray, are you in agreement with that?

Daragh Murray: I am. The danger with the current focus of the definition is that it draws our attention to something that is quite sensational. It is the Terminator-style autonomous weapon system that gets deployed, selects targets and engages on its own. That is only a very small subset of potential systems.

As Noam was saying, it is the use of AI across the military that is maybe of more concern. The danger with the current definition is that it just misses an awful lot of really important issues.

Q3 **The Lord Bishop of Coventry:** I cannot think of any relevant interests I have to declare. Does this more sophisticated concept of definition you are talking about provide a basis that international law can work with, in the way Georgia Hinds was suggesting will be needed at some point, in terms of regulation and even prohibition? Is some sort of fairly well-defined and agreed definition necessary for lawmakers?

Professor Noam Lubell: It depends on what you want to apply it to. If you are trying to create a new legal instrument, for example, that applies to a particular category of weapons, you absolutely have to define them. Any instrument dedicated to a particular set of weapons usually has some form of definition.

I would suggest a more practical route that would help us in the future. If we are trying to say, "These are the types of technologies that are now being used", we need to clarify certain aspects of international law, international humanitarian law in particular, that would apply to any autonomous weapon system, however you define it, as well as other weapon systems. These particular rules need to be clarified in terms of what level of human involvement or supervision you need. There is no problem with saying this would also apply to non-AI weapons, which sit outside this definition, because these should be things that apply across the board.

My starting point would be to say, "These are the technologies that are coming out. This is how they're going to be used. Let's clarify international law to make sure the existing principles and rules are understood in a way that means we can use them for these weapons and address the concerns that come up". That could apply across the board. It would not require us to come up with a specific narrow definition, right now, of autonomous weapon systems.

Georgia Hinds: There is definitely a risk with definitions that are too narrow. We need to be mindful of that in the sense that any prohibitions, regulations or even policies that are developed would not have a meaningful impact in addressing the very real humanitarian, legal and ethical concerns.

We have seen that with notions such as fully autonomous weapon systems or autonomous weapon systems completely outside human control. They definitely seem to run this risk of what was referred to as this fantastical Terminator scenario. We very much support the idea of a baseline definition from which we then build, recognising that we do call for new rules rather than simply a clarification of the existing IHL.

Q4 **Lord Fairfax of Cameron:** First of all, I have one interest. I am the co-owner of a private security company called Hawk-i Worldwide Ltd.

I have quite a narrow question. It was picking up on your point about targeting, and the scenario where the targeting is done by AI and then there is a human in the loop afterwards. In that scenario, the AI says, "We're good to go", but it is then for a human to decide whether to accept that recommendation. Would that fall out of the definition? You were talking about targeting, and I was just thinking about that situation, where, after the recommendation from the AI, you have the human making the actual decision.

Professor Noam Lubell: I do not know whether you are planning to run through all the questions, but this also comes up later in one of the questions around the problems that may arise. Just briefly, there is a real issue about understanding what we mean by "human involvement". There are different phrases used, such as "human in the loop" or "human on the loop". By the way, if we are thinking about future technologies, in the UK and the US people are working on brain-computer interfaces that place a human in the loop physically plugged into the system. These things are happening. Is that enough or not? Where do you place the human?

The MoD has spoken about context-appropriate human involvement. At what stage does that come? Do you need a human at every stage or not? This theme is probably going to come up quite a bit in our next discussion, but it is absolutely something that one needs to be aware of. Where is the human placed? Will the human automatically trust the system? There are deep sociological, psychological and other aspects around that as well.

Lord Fairfax of Cameron: I have a follow-up question. It is a bit of an odd one, perhaps. I am just thinking of the self-driving situation. It is a big issue at the moment, whether we will have robo-cars. I wonder whether this is applicable to the situation we are considering.

It is thought that robo-cars may in fact be much safer than cars driven by humans. They may reduce the accident rate to—I do not know—a fraction of 1% rather than whatever it is at the moment. The fact they are driven by artificial intelligence means that we humans expect them to be 100% safe, so you understand what I am trying to say.

I am just wondering whether that analogy is at all applicable to our scenario. The fact it is an AI deciding in the end whether to push the button is unacceptable, whereas in fact the casualty rate or the risk of bad decisions may in fact be much smaller.

Daragh Murray: The fact it is an AI is probably the crux or the heart of the issue. Maybe that is the ICRC's objection to the idea that the AI targets individuals. From a legal standpoint, there is not necessarily a distinction. If somebody is targeted legally, it is about whether it is compliant with IHL. There is a possibility that AI could produce far better results. It can take a lot more information into account. It does not suffer from fatigue, adrenaline or revenge. If it is designed properly, I do not see why it could not be better in some instances.

For me, the big stumbling block is that we tend to approach an AI system from a one-size-fits-all perspective, where we expect it to do everything. If we break it down, in certain situations, such as identifying an enemy tank or responding to an incoming rocket, an AI system might be much better.

Georgia Hinds: The potential for autonomous weapons systems to improve compliance with IHL comes up a lot. I definitely understand that autonomous weapons can offer potential military benefits over direct control or remotely controlled weapons, such as increased speed or operation in communication-denied environments.

We would caution against conflating potential military benefits with IHL or humanitarian compliance. Speed can pose a real risk for compliance with IHL. If human operators do not have the ability to monitor and intervene in processes because they are accelerated beyond human cognition, that means they would not be able to prevent an unlawful or unnecessary attack. That is an IHL requirement.

There is then this point about artificial intelligence and autonomous systems not being subject to rage, revenge and fatigue. Like with self-driving cars at the moment, this argument lacks empirical evidence. Instead, we are engaging in hypotheticals where we compare a bad decision by a human operator against a hypothetically good outcome that results from a machine process.

There are many assumptions made in this argument, not least of which is that humans necessarily make bad decisions. It ultimately ignores the fact that humans are vested with the responsibility for complying with IHL. They are the ones who can be held accountable if they do not, and there are processes for this. There are many assumptions made in that.

On the incorporation of AI into other decision support systems, feeding into but not replacing the targeting process, we agree this could bring greater information and a greater situational awareness. As has been said, under the right conditions and accounting for things such as automation bias, it could produce good results. Again, I would make a separation between those systems that are feeding in but not replacing the ultimate human decision in targeting and autonomous weapon systems, because they fundamentally raise different legal considerations.

Q5 **Lord Houghton of Richmond:** Just in terms of my interests, I have a general interest in being an ex-Chief of the Defence Staff, and then I

have multiple interests in being either the chair of or an adviser to a range of defence and security businesses that use AI in their products. I will just list them: Draken, SecureCloud+, Thales UK, Tadaweb, Whitespace and Rebellion Defense.

My question is perhaps a bit convoluted. Does humanitarian law always proceed from the assumption that a human decision is better than a machine decision or the fact that there always has to be the accountability of a human in the chain of delegation of a decision? Might it be a plausible scenario that a human being could delegate to an autonomous machine the decision-ability and still remain liable in the event of that turning out to have been a bad choice?

Can you unpick that? In my experience, the battlefield is full of flawed decisions. Many of them precede arrival on the battlefield, I have to say. To pursue some element of perfection in all this could be a bit like dancing on a pin. For some of the reasons you have absolutely said, when there is a time imperative and no collateral risk, delegation to an autonomous system is a far better option. Otherwise we forgo all the benefits of technological advancement in this particular area. Could you perhaps comment on any of that?

Professor Noam Lubell: There is a lot in there. On better decisions or not better decisions, it is important to acknowledge that I agree with what our colleague from the ICRC said: "Better for whom?" The military side and the humanitarian side might not always, in this sense, see the same thing as being better.

Speed was mentioned, but we could also think about accuracy. On both sides of the equation, the military and the humanitarian, you can make an argument that accuracy is a good thing. If these systems are more accurate, there may be a humanitarian reason to use them.

We have seen some of these debates play out, with similar questions to the ones being asked now, about the precision weapons that emerged over the last few decades. You can see that, on the one hand, there is the argument being made, "There will be less collateral damage so it is better to use them". At the same time, one could also argue that that has led to military strikes being carried out in situations where previously they would have been unlawful because there would have been too much collateral damage.

You can now carry out a strike because you feel you have a precision weapon and, while there is some collateral damage, it is lawful. Had you not had that weapon, you would not have carried out the strike at all. There are two sides to this coin.

On the delegation to these systems, part of the problem is how we discuss and debate this. I am not a technology expert myself, but I probably know enough to say there is a difference between what we call general artificial intelligence and narrow artificial intelligence. With narrow artificial intelligence, we are talking about systems that that are created to perform a particular task. They are a tool. It is not the

Terminator scenario of general artificial intelligence, something that sets its own tasks and goes about achieving them.

The systems we are talking about do not decide, in that sense. We are using human language to talk about a tool. It executes a function, but it does not make a decision. I am personally not comfortable with the idea that we are even delegating anything to it. This is a tool, just like any other tool. All weapons are tools, and we are using a tool.

I do not know whether you want us to get into the whole accountability question now or later.

Chair: We will get into that later.

Professor Noam Lubell: Then I will not get into it now, but there are solutions to the accountability problem based on the understanding that these are tools rather than agents.

Daragh Murray: I agree. That is really important. I would be also hesitant to use the word "delegation" in this context. We have to remember that humans set the parameters for deployment. The tool analogy is really important. We should be able to set the types of activities a tool will undertake. That is where the bounds of responsibility and facilitating IHL compliance come in.

Thinking of it as general AI is—this is what we were saying earlier—maybe a little bit distracting from the more likely deployments we are going to face in the short to medium term.

Georgia Hinds: I would add a point on the precision argument. Again, autonomous weapon systems are often equated with being more precise or more accurate. We struggle a little with this argument because the use of an autonomous weapon, by its definition, reduces precision. The user is not choosing a specific target. They are launching a weapon that is designed to be triggered based on a generalised target profile or category of object.

The reference to precision here generally relates to the ability to home in on a target better and maybe to use a smaller payload, but that is not tied specifically to the autonomous function of the weapon. We spoke about precision-guided munitions. There is nothing in our definition that is capturing those. That is to do with incorporating better guidance systems such as GPS or laser guidance. That does not mean you have autonomy in the function of selecting and attacking the target, necessarily. A remotely piloted system can have exactly those same precision benefits.

We should be careful not to confuse those kinds of precision systems with autonomous weapon systems. As I said, that could incorporate a land mine, which is definitely not a more precise weapon.

On the IHL obligations and the question of what they require of individuals, there are certain IHL obligations that are specifically directed

to persons. It is difficult to get around that. Even more broadly, fundamental assessments in IHL such as distinction and proportionality rely very much on value judgments and context. When you recognise someone is surrendering or you have to calculate proportionality, it is not a numbers game; it is about the anticipated military advantage.

Algorithms are not good at evaluating context. They are not good at rapidly changing circumstances. They can be, as it is called, brittle. In those circumstances, I would query whether we can say there would be a better outcome for IHL compliance when you are trying to codify qualitative assessments into quantitative code, which does not respond well to these elements.

Q6 **Lord Clement-Jones:** I declare an interest as chair of the council of Queen Mary University of London.

I am going to ask about the regulations or limitations on use that may be needed, but we have already seen a bit of a divergence, in a sense, on what actual problems or major concerns we are talking about. There is more of an emphasis from Professor Lubell and Dr Murray on AI, and more of an emphasis from Georgia Hinds on autonomy or AWS, basically. In a sense, I need to ask a two-part question even in that respect.

Georgia Hinds, could you tell me what your major concerns are in respect to this technology?

Georgia Hinds: I have already touched on the concerns from a legal perspective, but, because you have these context-specific judgments and IHL is directed at humans, we are concerned about the ability of an autonomous weapon system to facilitate a user's compliance with their IHL obligations.

IHL is about processes, not just results. You might have an attack that in fact causes excessive incidental civilian harm, but, if the commander or the decision-maker went through the proportionality assessment in good faith, they used the information that they had and a reasonable commander would have made the same decision, the attack could still be lawful, even if the circumstances changed and affected the end result. This is about human judgment and a reasoning process, which cannot be outsourced.

We also have ethical concerns. There is a concern about effectively substituting life and death decisions with a machine process. We held an expert meeting on this in 2017 with independent ethicists and experts from other fields. We published a report that confirmed clear challenges around the loss of human agency in the use of force, moral responsibility and human dignity. These concerns are complementary but additional to the legal concerns.

They are often ignored because the experts, like me, are not ethicists and are perhaps a little uncomfortable talking about these things. It is worth noting that ethics often drives the development of international

humanitarian law and international law more broadly. We are not seeing a unique case here.

Professor Noam Lubell: From the perspective of legal concerns, sometimes we are conflating technological challenges and legal ones. The example of proportionality was mentioned. Certainly right now we cannot envisage any AI system that can do the proportionality balance.

In proportionality, as I am sure some of you know, on the one side you are putting the concrete and direct military advantage you anticipate, and on the other side you are putting the collateral harm to civilians and civilian objects that can be expected. You then balance the two.

As has been said, that balancing is, at least for the time being, something we cannot imagine a machine doing. We have AI systems that can calculate one side of it. They do that already. There are all sorts of collateral damage estimation tools and so on where we use these systems, but balancing the two we leave to humans. That is a technological challenge. I do not see it as a legal problem. The technology cannot do it so clearly it would be unlawful to use it.

Sometimes we are raising technological problems and saying they are legal problems. That is not a legal problem. The technology simply cannot do it; therefore, you cannot use it for that purpose, although you could possibly use it for other purposes. Generally speaking, the concerns I have right now are about trying to clarify how technology is going to be used at least in foreseeable future and how the existing rules apply. To me, that is the way to do it.

I am not particularly worried about the UK or likeminded states rushing to release systems that select and lethally engage their own targets, in that sense, but other states might. Non-state actors will be making great strides in this field as well. Ultimately, we will see—this is part of the concern—some form of an AI arms race partially because you may need AI to defend against AI. Any arms race can be a race to the bottom.

The concern right now is less about a system you are going to release into the wild that is going to choose who or what to attack. I would go back to earlier comments about the way AI is used in the earlier stages of the targeting cycle, for example in detecting and identifying objects of interest or military objects. We know about this already from the law enforcement area. My colleague here has worked quite a lot on the problems of bias and other issues that have come up in using AI for law enforcement purposes.

If those problems repeat themselves when we are using AI in the earlier intelligence analysis and target and object identification stages, that is where, at least in the very near future, problems can arise from the use of AI, which we really need to keep our eye on.

Lord Clement-Jones: Daragh Murray, do your concerns mirror those of Professor Lubell?

Daragh Murray: To a large degree, yes, but I have a couple of others. I would very much endorse the point that we should interrogate how AI applies across the entire lifecycle rather than just focusing on one element. I would also agree that the problems are not necessarily with the law itself; they are about ensuring legal compliance and understanding the technology, how it works and how that can match the legal requirements.

One of my main concerns is about how we acknowledge the characteristics of an AI system or an autonomous system in terms of how it works. It typically works on the basis of correlation and not causation and a probability score. It is an inference that an object is a tank or that an individual is a member of an armed group based on a variety of sources. It is really about understanding how those work so we can apply them to IHL rules. That will be key, but it is very value heavy and context dependent.

One thing to acknowledge is that that decision is made by humans at the moment, but it is an imperfect decision. We have seen a lot of issues where errors have been made, particularly with the classification of an individual as a member of an armed group or not. It is about better understanding how we can acknowledge the characteristics of an AI system or tool and ensure legal compliance on that basis.

Lord Clement-Jones: This is one of the big questions we have to try to get to grips with, but, Georgia, would you like to make start on the regulations or limitations on use that may be needed to address your concerns and possibly those expressed by the other members of the panel?

Georgia Hinds: I would love to. The ICRC always loves to have a go at assisting with laws and regulations. We have recommended, in the CCW and elsewhere, that states adopt a new law at the international level to prohibit certain categories of autonomous weapon systems.

The first is unpredictable autonomous weapon systems. By that I mean systems that are designed or used so their effects cannot be sufficiently understood—we have already heard about the need for understanding—predicted or explained. This is most likely to arise with those systems that are incorporating AI, particularly machine learning.

They can exhibit what is called the black box effect—I am sure some of you are familiar with it—where their functioning is opaque. Because these systems do not allow the user to control or limit their effect, which is a requirement of IHL, they should be prohibited.

The next category is a prohibition against those systems that are designed or used to target humans. This is both a legal and an ethical concern. Coming back to the complex judgment that is needed for a person in armed conflict as to whether they are protected, whether they are surrendering, whether they are hors de combat or out of the fight in some way, or whether they are wounded or sick, these things are very

context-dependent. They rely on causation and intention, which are not well recognised by machine algorithms. They are also something that changes very rapidly. We really struggle to see how an autonomous weapon system could be used to allow a user to comply with the IHL requirement of distinction.

Picking up on Noam's point about technological changes, the other basis for prohibition is ethical. As I said before, that is something additional to the law. It is not something that can be fixed technologically because the idea of removing human agency in a life and death decision remains regardless of whether you are targeting a legitimate combatant or a protected person. Those are the two prohibited categories.

For those that are not prohibited, we recommend a combination of limitations on use, which reflect the existing constraints that militaries are using. These are the specific limitations: limits on type of target—as I said, they are mostly used on military objectives by nature, such as incoming projectiles or enemy equipment, things that are stable in their characterisation—limits on duration, geographical scope and scale of use, to enable a level of control over the effects; limits on situations of use, to constrain them to areas where civilians are not present in order to reduce proportionality difficulties; and requirements for human-machine interaction, by which I mean a level of human supervision, with the ability to intervene and deactivate. Something like self-deactivation of the machine is not enough. There needs to be a real ability for the human to step in if circumstances change.

This is our proposal to reinforce and strengthen existing IHL in light of the particular risks of autonomous weapons.

Lord Clement-Jones: I am sure colleagues will want to unpack that later, but thank you.

Daragh Murray: I fully agree that any tool that cannot comply with IHL should be prohibited. The rules of IHL are very clear on that, and that has to be the litmus test.

The other element that is relatively important is focusing on the decision-making process surrounding the development of the tools. We do not have a lot of guidance in that regard. What are the different factors we should take into account when we are considering the tools?

Earlier, my colleague mentioned that a lot of the focus is on the process. Ensuring there is an adequate process from the outset is really important. That is something we have not paid sufficient attention to so far. Drawing on a more interdisciplinary approach, to understand the technological constraints, how we ensure legal compliance and what other issues should be taken into account, will be a really key component.

Lord Clement-Jones: We need a mixture of prohibition and regulation, basically.

Daragh Murray: I guess so, yes, as well as guidance.

Lord Clement-Jones: Do you agree with that, Professor?

Professor Noam Lubell: My starting point, not just in relation to this question but whenever we are dealing with new technologies or new situations, is to try to avoid saying, "We need new law" unless it is absolutely imperative.

One needs to keep in mind that, as soon as we say the existing rules are insufficient, we are effectively admitting to a legal lacuna that could be exploited, whether by us or by others, for years. It takes years to develop new law. I am always very uneasy when we say we need new law because of what it means about what we have right now.

I would also tend to agree with the UK Government's position that, by and large, existing IHL contains the basic rules we need here. That does not mean we do not need to do a lot of work to unpack and guide how it applies to particular situations.

In terms of the categories that have been mentioned that might need new regulation, sometimes we talk about systems being "unpredictable", but I am not sure whether that is the test. Do we need predictability or reliability?

If we think of existing weapon systems, loitering systems would be one example. You might release a system that will loiter in a particular area and look for certain signals, if we are talking about attacking radar systems. You can have a system that you know is going to attack a certain type of thing, but you do not know which one. You also do not know whether it might not find it and then self-destruct. Does that mean the system is unlawful?

I can tell you, "I don't know exactly where it's going to attack. I don't know the minute it's going to attack. I don't know whether it might decide to self-destruct. But I know that the only thing it can do is something lawful and, if not, it will self-destruct". You could argue that it is not predictable, as I do not know exactly what it is going to do, but it is reliable.

I am not sure what we are supposed to be looking for. Are we talking about unpredictability in the sense of something going off and targeting whatever it likes? I am not sure we need new laws for that. Under a host of IHL rules, it would be unlawful to use something like that.

As for systems that are used against humans, again, for me this is a technological problem. I agree. The systems we have now cannot recognise hostile intent. They cannot tell the difference between a short soldier with a real gun and a child with a toy gun—most systems are still not capable of doing that—or between a wounded soldier lying slumped over a rifle and a sniper ready to shoot with a sniper rifle. They cannot do that yet, but that is a technological problem. We should not use them because they cannot be used lawfully. I am not sure it requires a new law.

The ethical issue is completely separate. There are absolutely ethical questions there. If we want to create new laws because of ethical concerns, that is certainly valid, but it is not because of a legal problem. It is a choice.

Q7 **Lord Browne of Ladyton:** There is a certain irony in us having this conversation about a lacuna in the law in relation to technology in this Parliament at this time. We are about to employ almost all of our resource for about six months in trying to catch up with risks we did not engage with when we allowed technologies to be deployed into our lives. The Online Safety Bill, which is a large piece of legislation, is only one of them. There are a number of them in different areas.

That way of dealing with technology using regulation has not proved to be efficient or successful and it has caused a lot of risk. With respect, Professor Lubell, this is one of my major concerns about the approach you encourage on us. We are now living in a world in which there are lethal autonomous weapons out there. They are being used for sentry purposes by certain people. If the evidence of the United Nations is correct, UAVs that use them have been deployed.

People talk about this as an arms race and there is a fear of losing out on capabilities. There is this constant argument that comes up in arms control. "Why should the bad guys have all these weapon systems?" This may be a forlorn attempt to do this, but those who were keen to have this Select Committee are trying to find a way in which we can at least get in step with this or get ahead. Making the mistake of banning weapons prematurely would be the worst of the options.

We are looking for recommendations as to how we can not make that mistake again, where we entertain the risk, then find out, when it manifests, that it is much worse than we thought it would be. From a legal perspective we should not be showing people where the lacunae in our laws are, but they know where they are. What is your advice?

The Chair: Please do not treat this as a three-hour essay question!

Professor Noam Lubell: Part of what I was trying to say is that the existing law is there. A lot of the problems and concerns that have been raised are covered by existing prohibitions. By and large, even the futuristic scenarios that are brought up are covered by the existing rules.

As I said, if we want to try to make progress, we need to unpack these rules. They are very general. When we try to apply the rules we are talking about to the detail, things break down.

Even if we talk about a new instrument, I have concerns. The statement made by the UK, the US and others just a few weeks ago included draft articles on this subject. Let us say that became the law. If you had a new instrument, the phrasing would be similar; it would be at that level. It is at that higher level: "You can do this. You can't do that". It ends up using phrases such as—this phrase has been used by UK—"context-appropriate human involvement in the use of weapons".

They are the correct phrases, but what does that mean? How do you use them in this context? How do you use them in that context? What does "context-appropriate human involvement" mean in the identification of targets by an AI system? Systems are already being looked at, such as SPOTTER and SQUINTER, to help detect objects. What does "context-appropriate human involvement" mean there? What does it mean for the commander deploying a system? What will it mean when we have systems that engage swarm technology?

I can lay out 20 different contexts, and in each one "appropriate human involvement" might mean something else. If we had a new instrument, it would not spell all of that out. It would stay at the relatively generalised level of what the law is in terms of whether human involvement is needed, but ultimately we would still need guidance on what that means.

The high-level things we would end up with in a new law can be found in existing IHL. I would urge us to spend more time on unpacking this and thinking about how these things would be used in practice, in the real world. That does not necessarily require new law, but it does require a lot more work.

Daragh Murray: Yes, I agree. The danger of looking at a specific type of weapon is that we miss the bigger picture, which is the more pervasive role of AI across the military.

The danger is that a lot of regulatory effort is spent addressing a particular type of weapon, the use of which for the most part is probably regulated already under IHL, and then we make the same mistake we have made in other areas of technology: we pay insufficient attention to it and we do not think through the future consequences by looking at the guidance, working it out and thinking through the decision-making process.

For me, that is the big concern. If we focus only on a specific type of autonomous weapon system, we are not thinking through how intelligence is generated using AI systems, which will affect targeting or detention issues further down the line. That is the big risk: that we get distracted.

Georgia Hinds: Just quickly on this issue of why we are calling for new law, we agree that existing IHL rules apply to autonomous weapons. They already constrain their use. They would prohibit certain types, particularly weapons that are inherently indiscriminate.

The ICRC does not often call for new international rules for precisely the reason Noam has identified. We do not do it lightly. We only have to look at the debates in the CCW, which have been going on for nearly 10 years. We can see that states can and do hold different views about precisely what the limits and requirements are for the design and use of autonomous weapons that flow from existing IHL.

This is not surprising. In every IHL expert discussion we have, there are always points of difference about not only the application of different treaty obligations across states, because you have that, but the precise interpretation of commonly applicable key obligations. This is not something we can hope to solve in a forum such as the CCW. In fact, we are wary of trying to do that because attempts to restate or reinterpret existing rules risk leading to the lowest common denominator of less protective formulations.

The fact that we do not have consensus on how the rules apply specifically to autonomous weapons is what makes it clear to us that we need something specific, shared and codified to provide that understanding and clarity. That is our rationale there.

Q8 **Lord Hamilton of Epsom:** I am a shareholder in the Herald Investment Trust, which apparently invests in technology companies—I have no idea what they are—and I am a vice-patron of the Defence Forum.

My question is really for Georgia Hinds. We are talking here about new laws that will be signed up to by western civilised countries. Professor Lubell has already referred to hostile actors. Let us put a name to a hostile actor: Russia. Vladimir Putin does not seem to lie awake at night worrying too much about international treaties and whether he is abiding by them. If we were to go to war with Russia, surely we would be at a distinct disadvantage because he would be using all these illegal systems against us and we would have nothing to play back to him.

Does it not worry you that we would be putting troops on the ground at a distinct disadvantage to the enemy we were fighting?

Georgia Hinds: This argument comes up a lot when we talk about international humanitarian law, but, at the same time, it does not stop countries such as the UK from very clearly committing to abiding by and following IHL. It is not based on reciprocity. It is based on the fact that they abide by international humanitarian law because that is who they are and that is the standard that they want to set. As a general point about international humanitarian law, it is not based on what your enemy does.

On the point about the majority of states, in the CCW two weeks ago we saw that the majority of states were in favour of a legally binding instrument, so it was not just a select group. It is more representative and broader than perhaps might be implied with the question there.

The adoption of national measures and constraints that is already happening is fantastic, but this is precisely our argument for elevating it to a shared, codified constraint in order to have that commonality and the more widely binding aspect. That creates norms, which is what happens with international law. Even if you ultimately have certain actors who do not sign up to certain treaties, this does not prevent them being transformed into what is called customary international humanitarian law, which is then binding regardless of whether you have a few who do not attend.

So there are a number of ways to counter that argument. This is not unique to this issue and has not stopped the progressive development of international law and the setting of stronger protections across the board.

Q9 **Lord Fairfax of Cameron:** My question may have been answered or discussed already, but I am going to ask it in any event. It is a question about the enforceability of regulation, particularly in view of the comment Professor Lubell made. You said everything is so fast moving in this area and that, in your view, there is likely to be a race to the bottom and an AI arms race. In view of that, and given the difficulties we have also discussed about appropriate definition, I just wanted to ask about the realistic prospect of effective enforceability of regulation. Georgia has touched on that already, but I just wanted to home in on it. If the genie is out of the bottle and some of these things are out there already, as has been suggested by Lord Browne, I am concerned about realistic prospects for effective enforceability.

Professor Noam Lubell: I suppose it depends whether your concern is with enforceability once particular systems are being used and how they are being used. That would be the same as the problems that we have with all weapons systems and actions in conflict, and potential problems that we see in current conflicts as well around enforceability, but that does not mean that we do not try.

If you are talking about enforceability in the sense of us trying to regulate the development of certain weapons while others go ahead and use them, that is a realistic concern. Over the years and over the centuries, there have been attempts to prohibit and ban certain weapons completely. Sometimes they have been successful. It is interesting to look at when they have been successful and when they have not.

Some of this is linked to the ethical, moral and social level. If you think about gas and chemical weapons, they seem so utterly abhorrent that it is easier to get that prohibition. It is a lot trickier when it comes to technologies that militaries see as being especially useful. They are usually more reluctant to agree to a total ban, so then you try to have regulations on particular types of uses. That might be what could be achieved here.

Again, I will just come back to what I was saying earlier in that I do not find it disturbing, other than the problem of admitting that there may right now be something that is not regulated, which, as I said, strategically makes me very nervous. I know that in the ICRC, as they have said, they usually do not like to say that either, so it means something that, in this case, they are concerned that the law is not there.

Beyond that, even if we came up with some kind of new treaty or instrument, I do not see that it would actually solve the problem of how we use these things in practice. We would still need the guidance that we need now on spelling out what that means in particular contexts. What does human involvement mean? What does human control mean? It is a

phrase that is bandied about and is coming into all the various draft instruments. I am not sure it would even solve the problem.

Q10 **Lord Houghton of Richmond:** I do not know the degree to which any or all of you have been exposed to the emerging UK MoD policy on AI and weapons systems, but from everything you have said thus far, as I glance down its headlines, it seems to be fully compliant with all your concerns. It would be ambitious, safe and responsible. There is absolutely no intention to deploy fully autonomous weapons. There will always be the requirement, while technology demands it, to have a human in the system to delegate to.

This is contrary to what Georgia was saying about the need for new law, but to what degree do you think that the emerging UK MoD strategy on the use of AI is accommodating the concerns you have about IHL already, and that the residual challenge is therefore more one of how we continue to engage the human dimension of decision-making at the appropriate levels?

Professor Noam Lubell: You are entirely correct. From the perspective of the UK as a state with a very active military presence on the world stage and, at the same time, a public commitment to IHL, its position and the direction that it is taking is entirely reasonable. The right things are being said. That is not the problem from my perspective. I am sure colleagues on the panel here might see it differently, but I do not see that as being the issue.

For me, again, it is about what this means in practice. We have general statements about things. Obviously, there is a lot going on to which I am not privy, but I am not entirely convinced that enough work is being done on taking this down to the various contexts. I know there are people within the system who are interested in doing that. I have spoken to plenty of them over the last couple of years. The problem is spelling out what it means in detail rather than the strategy or the position.

There are also very specific problems that we need to look at. We keep focusing on the use rather than the development. How does the weapons review process, for example, apply to AI systems? How do you use a weapons review process for software? This is not just about AI; it is about cyber as well. How do you use it when you are examining whether something will be lawful but the thing you are examining will change constantly afterwards, sometimes even of its own accord? At what stage do you need to re-examine its legality? There are complicated questions as to how we manage this, but I find the headline strategy to be reasonable.

The Chair: Lord Grocott has been immensely patient. I am going to go to him now.

Q11 **Lord Grocott:** I have been patient, Chair, not least because it has been extremely interesting, but also because a lot of the area covered by the question that I am about to ask has been touched on quite frequently. By the way, I have no relevant interests to declare on this. In listening

carefully to everything that has been said, I am slightly alarmed in terms of our report, as and when it is produced, because we have heard that it is very difficult, if not impossible or maybe unnecessary, to define what AWS is, which is a bit worrying if that is what we are looking into.

I am also aware that, as Georgia Hinds said, so much of this is about hypotheticals. One thing that politicians cannot spend too much time discussing is hypotheticals. You actually have to make laws or participate in the making of laws. Are there circumstances in which AWS could improve compliance with international humanitarian law? That has already been answered in part, but because of my desire to get my feet, if no one else's, back on the ground and away from the ether, could I ask for specific examples where AWS could improve compliance with international humanitarian law?

To Georgia Hinds, I found what you had to say about possible new laws very interesting. Would the potential new laws that you spelled out prohibit any activities that are going on at present? Leave aside for a moment how you would define AWS. Can we just try to get it down in terms of specifics? Those are my two questions.

Georgia Hinds: In terms of specific examples providing greater IHL compliance, I honestly cannot give you any within our definition. I can definitely give you many that are currently being used that improve military advantage. With these air defence systems that operate beyond the speed of humans, I can absolutely see the military advantage there, and we would say they are currently being used within IHL constraints, based on those limits around the types of targets, the limited scope of time and duration, and the geographic scope. With those existing systems, we would say they are, by and large, from what we know based on open-source information, being used in compliance with IHL. That is where we have drawn the recommendations for our limitations on use from.

As for whether our recommendations would then prohibit any current activities, again, we are a little limited in our information because, even with the ICRC, states do not always want to share the exact extent of their capabilities. Where we are really running into problems with this is that there are technologies that operate on a continuum. They are remotely piloted drone technologies with the capability to operate in an autonomous mode, but we do not know whether they are being operated in that autonomous mode.

In current conflicts, I am thinking particularly about loitering munitions. They look like drones and are sometimes called suicide drones. Essentially, they will hover over a target and then dive and self-destruct. We are seeing potential that these are being used against human targets. It is unclear whether they are being used in an autonomous mode or based on existing remotely controlled technologies. For us, this is a key consideration, so it makes it difficult to answer your question as to whether our recommendations would extend to that, but certainly, if

these were being used as autonomous weapons and against humans, that would be captured by our prohibition recommendation.

The prohibition on unpredictable systems is designed more to capture the future technologies where there is greater integration of machine learning. These are systems that change their functioning after use, as Noam was talking about, putting them beyond the human ability to predict the effects. Swarm technologies are another one where, at the moment, they are being run as remotely piloted. As they become more autonomous and start to produce emergent behaviours, they could run into becoming an unpredictable system within our definition. We have tried to base it on existing use, so that it does not constrain or prohibit things that are already in use and that are already being used in compliance with IHL.

Daragh Murray: Hypotheticals are the problem. That is why focusing on the decision-making process itself is really useful, because it gives you a concrete way to think about the issues that arise. One of the issues is maybe that we are operating under a slightly different definition. When I am thinking of the issue, I am thinking more about how AI is used rather than a fully autonomous system that exists on its own. If an AI system is used, there are a number of ways in which it could contribute to better compliance with IHL. The ability to model predicted consequences is a really good example. You can use a system to model the likely effects of an attack, drawing on past history and the contextual environment, such as the make-up of a building, bringing a lot more intelligence into play in order to understand what the consequences would be.

Lord Grocott: Is anything being developed along those lines? You are in hypotheticals now, but is there a manufacturer somewhere in Britain working on the kind of equipment you are envisaging?

Professor Noam Lubell: There are projects being worked on with regard to identification of objects, for example, in a military context. Absolutely, there are projects and there is public knowledge of some of these projects being done. These systems could be better than the current systems we have at identifying what is military and what is civilian. I am not saying that they are but, if we are talking about possibilities, it is possible that you could have systems that would have advantages also on the humanitarian side.

Again, as I mentioned earlier, there is a flipside to it. For example, when you can be more accurate and target only the military, because IHL allows some proportionate damage, you could end up targeting something that beforehand you would not have targeted because there would be too much collateral harm. Now you can target it with a small amount of collateral harm and, therefore, those people who are in that small amount of collateral harm have now suffered when beforehand the thing next to them would not have been targeted at all because we did not have anything accurate enough. There is a flipside to this, but I cannot say there will never be an advantage. Absolutely, there could be.

It also depends on the definitions. If your definitional starting point is that these systems are the systems that do all these unlawful things, of course you are not going to have an example of a positive use of these systems, because you have defined them as something negative. It comes back to how we define them in the first place.

I hesitate to remind the committee of the title of this committee, but it is important to note that it is not "autonomous weapons systems"; it is "artificial intelligence in weapons systems". In that sense, I would presume from your own title that you have the mandate to think about things a little differently, to look not just at how to define autonomous systems in the way it is being discussed at the CCW, but more widely at the use of artificial intelligence across weapons systems and the targeting cycle. Someone chose that title for a reason, I hope, and not just randomly.

The Chair: That is advice very well taken, Professor.

Q12 **Lord Hamilton of Epsom:** It strikes me that we have an enormous problem if we bring the law into this, because we cannot define what we are trying to ban. The great advantage of all this is that we are talking about military structures. They are the ideal pyramid. They have General Houghton at the top and everybody else all the way down. Surely the responsibility for what happens with a weapons system, whether it is AI or not, should lie with the military commanders. To the extent that you place the responsibility within the military structure, you bring the human element into it, rather than trying to ban certain activities. Would you agree with that?

Professor Noam Lubell: Accountability is one of the biggest issues that keep coming up and give rise to the most concern. Pretty much everyone would agree that we cannot have systems in use for which there is no accountability. That is a basic starting point that nobody would dispute.

Part of the problem with the discussion around accountability goes back to something I mentioned earlier about the conflation of general AI and narrow AI, and these being tools and not agents. Nobody is talking about how we hold a machine accountable. We are not going to hold a machine accountable because this is not general AI. We may do in 20, 30, 40 or 50 years, but that is not the kind of speculation we are engaging with in these debates.

Then we are talking about a weapons system. If it is a weapons system, I would argue that the problem of accountability has perhaps been made to seem more complicated than it is, because accountability can be determined the same way it is with any other weapons system. There is no one size that fits all, because it depends on what went wrong. If you use an AI-driven weapons system and something happens that looks like it may have been a violation of IHL, you need to investigate what happened. That is the starting point. There is an obligation to investigate.

If it turns out that a commander deployed this system in a manner that was against the rules, such as to intentionally harm civilians or knowingly in violation of the operating instructions, the commander is responsible for that. If the weapon was deployed in accordance with its operating instructions with all the right intentions and yet something still happened, again, we do the investigation to figure out what that was. It could be a mechanical failure. We straight away go to a problem with the algorithms, but it could also be a loose wire in these systems. You need to look at that. These things happen with other weapons systems as well and we have mechanisms for dealing with that.

If we cannot find a mechanical problem, and we do not know what happened and it is something to do with the algorithms driving it, again, the investigation continues. The black box is a problem, but it is not a problem with every single system. It is particularly problematic with neural networks, deep learning and so on, but not always. You investigate. You also look at the computing side of it and, if you can figure out what went wrong, you look to see if it was a problem in the design of it. You look to see if it encountered something it had never encountered, and you can try to understand and rectify that.

If, after looking at it, you have this explainability problem in the black box and you cannot figure out what went wrong and why, you take the weapon out of use. You do not use a weapon that would be unpredictable in a way that causes problems. If a commander, or a military more generally, continues to use a weapon knowing that it malfunctions and we do not know why it malfunctions, there is accountability for that as well, as there is with knowingly using any weapon that could be malfunctioning.

We can find solutions to the accountability problem along those lines more generally. That is the focus on individual accountability. We also need to remember that international humanitarian law is a branch of public international law and there is always state responsibility, regardless of what individual accountability we find or do not find.

Daragh Murray: I agree with the distinction between general AI and narrow AI. AI tools are complex systems, but they are made up of different components and each component has a task. When we are thinking about accountability, we need to look at those different components. Have they been designed and developed such that they are capable of achieving that task? Are they deployed in a manner consistent with that task? Are they deployed in line with their intended circumstances of use or outside of them?

Once we break down the AI life cycle in that way, it becomes much easier to think about attributing accountability and responsibility, because everybody has a role in the system. In some instances, it will be within the military. If tools are used from outside, maybe as part of a corporation, it is just about understanding what the component is intended to do at each part of the life cycle.

Georgia Hinds: A lot of our concerns around accountability mirror our concerns at the front end. This is where our prohibition on unpredictable systems plays a role. Noam has brought up the black box effect, which does not appear with all systems, but, for those systems where it does, we have a fundamental issue because, for individual criminal responsibility, most legal systems require intent. They require a certain mens rea. If you have a system that is producing its own results or continuing to learn, and it produces a result that is beyond human intent—that might not be a malfunction; it might be that the system views it as an optimisation—you have a fundamental break with individual criminal responsibility.

You will hear in your next session from Vincent from SIPRI. They have done some very good work on accountability and responsibility, so they could speak much more eloquently on this. They have brought out issues around traceability, which is the ability to trace the operation, performance and effect of a weapon back to a human developer and user. There are specifics around autonomous weapons systems that do not arise with other weapons.

The other thing that they have suggested is, in contrast to other weapons, having a specific deemed scheme of responsibility for those who are involved in the development and use of autonomous weapons. This comes out a bit in annexe C in the ambitious, safe and responsible approach. It says that the UK will clearly establish authorities, hence accountability and responsibility, whenever UK forces deploy weapons systems with AI. There is a question there about whether that needs to be AWS specific, with deemed responsibility on certain operators where you cannot otherwise trace back intention. Whether that is fair and something military users would accept is a separate question, but that is a proposal that has been put forward as well.

Q13 **The Lord Bishop of Coventry:** My question on accountability has largely been addressed, but let me press it a little more and then ask a slightly wider related question. It is about a number of things.

Are these methods of accountability going to suffice for the future? That is such a difficult question, but Lord Browne makes a very important point, as did Professor Lubell at the beginning, about the pace of change and trying to think ahead, in a sense. That is one aspect. Are the accountability systems you are talking about going to last?

Maybe this is moving more into an observation, but an interesting tension has been evident in this fascinating session and is relevant, as Daragh Murray has pointed out, to our whole task. I hear Georgia Hinds talking about a specificity regarding autonomous weapons that moves things into a different realm, ethically and legally, than the AI that you are quite rightly saying we must attend to across the system.

If I may add in the supplementary to that, I hear from Georgia Hinds that the specificity of lethal autonomous weapons touches upon a fundamental ethical issue of killing not being franchised to machines. I just wondered

whether you had any reflections on that fundamental issue and whether that is something that needs to be addressed by our thinking. I am conscious of that expression, "Law is the hard edge of ethics", if I have that right. It is not unrelated to ethics. It is where it comes down to the realities.

The Chair: That is a challenging brief. Who is going to have a go at unpacking it?

Professor Noam Lubell: On accountability and keeping up, we will need to keep up, but that does not mean that it cannot be done, so long as we are sticking with narrow AI systems and we have not moved into the realm of systems that set their own tasks and then go about achieving them. The systems will get more sophisticated. We will need to keep updating our operating instructions and the understanding of everyone in the field who is using them, at every single stage, as to what that means, when and how they can and cannot be used, and what the risks are. We will need to keep doing that, but as long as we do the same basic accountability mechanisms should be in place.

In terms of the specificity of these systems, again, it is less about the autonomy and more about the AI. Daragh has been saying this as well. I am just as much, if not more, concerned by AI being used in the early intelligence analysis stages that then feeds into something that we think is completely done by humans, but is actually all based on information that was put together by AI.

The two of us, together with a colleague from the US, Ashley Deeks, published a paper a few years ago where we looked at the risks of having some of these AI systems being used in intelligence analysis for what we call the *ius ad bellum*, which is the decision to go to war. Ultimately, these are the grander decisions. We are sitting here talking about what might happen in a particular military attack that involved a total of 50 people, but war gaming has been done using computers for many years. It gets more and more sophisticated. What about when our decision to go to war in the first place is based, in large part, on intelligence analysis done by AI systems? We are not really talking about that, but AI being used in intelligence analysis and so on is just as worrying. For me, it is less about the autonomy; it is more about the AI.

As for the ethical questions, they are huge. I do not deny any of them, and I agree with some of the concerns raised by the ICRC and others around the ethical side of things. Absolutely, these are concerns. I separate whether we can use these things lawfully under existing law from questions of ethics, but the issue of AI determining life and death decisions is not unique to military contexts.

Self-driving cars were mentioned earlier. It is a little cliché in these debates to use the old-fashioned philosophical trolley problem, but you can apply it to a self-driving car that needs to choose. Now there are only two options left because of an oncoming truck. Do I go up on the pavement and run over these two people or stay on the road and let the

passenger be killed? The car is having to make that decision although, again, I hesitate to use the word "decision" in the context of AI. These are life and death situations that are going to involve AI systems. On a larger scale, AI in traffic management systems is going to determine, for many countries, a lot more life and death situations than any military AI, and that is without speaking about AI in health and other contexts.

AI is coming into critical life and death decisions across every realm of society, and I am concerned that we keep trying to debate it within the military context as if this is something separate and unique. We then end up having a different relationship to AI in the military context from others, when that same AI system or a different version of it will make life and death decisions, or execute life and death functions, regularly in other realms. We do not have enough joined-up thinking in that sense. The people debating it here need to be talking to the people debating it there.

This is a much bigger question about human-machine interaction and how we incorporate AI into our society across all realms, and that is where we need to be having these big debates. Then they can feed back down into the military side and everywhere else. We need to understand ethically what our relationship is going to be with AI and then think about what that means for the military, for law enforcement and in the health sector. Then we might think different things, but I am concerned that we are having these big AI ethics debates in narrow silos rather than with joined-up thinking.

Daragh Murray: For the foreseeable future, we will be talking about narrow AI, so systems that operate within human-defined parameters. The accountability processes we are talking about will be appropriate. My bigger concern is with the use of AI in intelligence decision-making because intelligence should, I hope, pervade everything that the military does, whereas the targeting or the actual engagement is a relatively small part. It has potentially bigger impacts on both the military and humanitarian components than just the specific engagement of targets. Ethics is way beyond my paygrade, but the point about healthcare in particular is really well made.

Georgia Hinds: On accountability, yes, we have existing systems for individual and state responsibility. These will continue to operate, but with autonomous systems, even if we are talking about existing systems, at the moment they are, for instance, particularly vulnerable to adversarial conditions, such as enemy interference or spoofing, as they call it. If this produces unforeseen results, who are we tracing that back to?

It brings in some different considerations for accountability. We can talk about programmers and say that, if there is a malfunction in the code, that is a programmer responsibility but, in reality, when we talk to programmers, it is actually very complex, even in narrow AI. There is not just one programmer responsible for the whole piece of code. It is far more complex and a process that involves many people who might see

only a narrow part of the development. It is very difficult to have that traceability, which is something that SIPRI has specifically said needs to be looked at for autonomous systems.

The Chair: Lord Browne has some questions on international fora, which we have been touching on over the course of the morning, unless all your questions have been answered, Lord Browne.

Q14 **Lord Browne of Ladyton:** They have all been touched on. We should at least tell our witnesses that we have got the message about having a broader look at this than only in relation to autonomous weapons systems. It is undoubtedly the case that AI is used in intelligence, reconnaissance and surveillance and, if the reporting of it is correct, it is being used in Ukraine to significant effect, so it would be a mistake if we did not touch on that. Defence and security is special and different because we train people to exercise lethal force on other beings. We do that to a limited extent in our domestic security, but it is at the heart of what we do when we create our military capability, so it is special.

I will not make any apology for finding myself in the same space as the international community on autonomous weapons systems and the issue of the interface between humans and machines in decision-making, because that is where the international community got itself. It went to this inappropriately named convention. Its full title is the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons—and this is important—which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects. We went there, and I think we went there because of 100 years of history of arms control.

This issue is a discussion about international humanitarian law, but de facto it is about arms control and whether we can, as we have done with many other weapons that were subject to and could be deployed consistently with international humanitarian law, come out of that discussion saying, "We'll not use them". As legislators, we are also being encouraged by many of the people who have been responsible for the development of potentially AI-enabling weapons systems to say, "Don't do it".

Are we in the right place in doing that or are there alternatives? Is that likely to be successful? If it is not likely to be successful, why is it not going to be successful with these weapons systems when it has been with a whole list of others that we are able to identify? In particular, I am interested in the United Kingdom's position in relation to this and your comments on that. This has been an extraordinarily rich session for us.

Georgia Hinds: Thanks very much for the very difficult question on where we go from here. Two weeks ago, I sat through the discussions in the CCW, but this has been going on for nearly 10 years without reaching consensus on the best ways to address many of the risks that we have been talking about today. In those discussions, we were encouraged to see that there have been constructive moves. We are talking about substance, and there are an increasing number of states that clearly recognise the risks, recognise what is being called this two-tier approach

of prohibitions on certain AWS with restrictions on the remainder, and recognise that there are pragmatic, principled ways to articulate those limits. Having said all that, of course there are still challenges within the CCW, and a lot will depend on the next few months as the chair works towards a report.

For us, the bigger picture is the momentum towards regulation, and that includes discussions occurring in other forums, including the UN General Assembly First Committee with the resolution of 70 states last year, which the UK joined. The Human Rights Council is discussing it with a new advisory committee looking at new technologies broadly, but with a strong focus on autonomous weapons. We think that the CCW remains a strong forum that we will continue to engage in. In terms of its history, it has a very good track record that can be drawn upon. For instance, the protocol on the use of mines and booby traps, the amended Protocol II, which ultimately then led to an anti-personnel mine ban convention as well.

Coming back to the earlier question that was raised about what the point is if we do not have all states joining this, not all states are party to the anti-personnel mine ban convention, for instance. That is not to say it is not effective. It is not to say that the UK does not say it is effective either. Similarly, the protocol on blinding laser weapons under the CCW was one of the first prohibitions we saw on the use of a weapon against persons. This provides us with a good precedent when we are talking about autonomous weapons.

I am afraid I cannot give you a good answer about why we were able to achieve it more quickly with these other ones than with autonomous weapons, but I hope that we can be optimistic. As I said, at least there are discussions occurring in other forums. It is good to be multidisciplinary, to have different considerations of the issue being looked at and maybe not to be so narrow.

Professor Noam Lubell: You can never rule anything out, but it is not particularly likely that we will see a new binding instrument come out of the CCW, for reasons already previously stated. Other forums may pick this up, and some are already. By the way, we have not mentioned regional forums. There are discussions and there may be things happening at that level, but what will we end up with?

Let us say we did end up with something, whether CCW or elsewhere. What would it actually say? We are not going to end up with a total ban on using AI because it is just too useful, and states are too concerned about others using it. We cannot equate it to some other things that have been completely banned, so we are talking about regulations on use that we might end up with, and I just do not see how anything in an instrument form would get into much more detail than what we are already seeing come out, whether in the statement by the UK and the US or the ICRC position. They are not as far apart as you might think. There is a lot of common ground there in terms of accepting the need for

controlled human involvement and accountability. Everyone is accepting all of this but, at the instrument level, it stays very generalised.

I would just remain with a concern that what is missing is the detailed guidance and the unpacking. Yes, we all agree that there needs to be accountability. How is that going to work? Spell it out. Give details of how accountability is going to work. What does context-appropriate human involvement mean in this context or in that context? What does it mean for intel? What does it mean for targeting? What does it mean for this technology or that technology? Without that, for me it is a question not so much of whether we can get an instrument, but of what this instrument is actually going to do unless we have the more in-depth discussion and guidance.

Daragh Murray: The in-depth guidance is key. It might also be useful to think through a framework about how we approach the decision-making. This would mean more practical guidance, potentially at an international level, setting out the types of issues that should be taken into consideration and the hurdles that might be overcome. It might be really useful in allowing us to go into more depth on specific issues.

The Chair: Thank you very much. I am going to draw proceedings to a close now. The three of you have had the great achievement this morning of, first, answering a dazzling array of questions in a way that we found hugely helpful, but secondly, in the process, adding substantially to the committee's agenda. On both accounts, we are extremely grateful to you. Thank you very much indeed.