HOUSE OF LORDS

Artificial Intelligence in Weapons Systems Committee

Corrected oral evidence: Artificial Intelligence in Weapons Systems

Thursday 8 June 2023

10 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. 7 Heard in Public Questions 97 - 108

Witnesses

I: Lord Sedwill, Former National Security Advisor; Hugh Durrant-Whyte, Director of the Centre for Translation Data Science University of Sydney.

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

Lord Sedwill and Hugh Durrant-Whyte.

Q97 **The Chair:** Good morning, Lord Sedwill and Professor Durrant-Whyte. It is very good to have you with us. You know the format of the session, I expect. It is being broadcast and afterwards you will receive a transcript so that you can make any corrections of fact. Of course, you can add to your evidence at any stage.

Might I begin perhaps by asking you the general question about costs and benefits: what are the benefits and what are the hazards of proceeding down the AWS road?

Lord Sedwill: Good morning, Lord Lisvane, and to the committee. Thank you for the opportunity to be with you this morning. Just for the record, I know you have had a submission of evidence from BAE Systems. I should just declare an interest; it is on the register. I am a member of the BAE board, although I am not appearing in that capacity. As you know, I am appearing essentially as a national former security adviser.

The Chair: Yes, understood. Thank you.

Lord Sedwill: On your question, the short answer is that AWS—autonomous weapons systems—can enhance both the offensive and particularly the defensive combat power of our Armed Forces, just as previous technological innovations have over the past several centuries, I guess. It means that we will need to rethink doctrine and the concept of mass. The parallel I have in mind is the Six Day War in 1967, where Israel, although having much smaller forces than its adversaries, was, through better capability, better planning and doctrine, and better training, able to prevail very swiftly. AWS offers that opportunity to countries that adopt it and build these capabilities into their doctrines.

As you mention, the hazards—Professor Durrant-Whyte will know more about this than me—are essentially the same issues that we have seen quite a lot of reporting on over the past few days about autonomous systems generally, which is essentially maintaining control: how do we ensure that they act only for the purpose and under the control that is set for them? How do we ensure that the rules of engagement that we apply to our Armed Forces now apply equally effectively to autonomous systems, particularly if those systems have to operate, as they might, essentially in a defensive capacity without very much human intervention? I am sure we will explore those issues further.

The Chair: You used the phrase "under the control". That raises the question of the human in the loop and meaningful control, but I think we will indeed explore that in a bit more detail.

Professor Durrant-Whyte, I know you are joining us from Sydney. Although it is all black behind you, I hope it is not too inconvenient a time in the evening.

Professor Durrant-Whyte: No, it is only 7 pm, so we are good. To some degree, autonomy in weapons systems is a matter of degree. I used to give a presentation when I was in the UK showing that we effectively had an autonomous torpedo in 1897. Many of the components in that—the single processing and so on—were clearly done in an analogue fashion but nevertheless, once it had been released it was autonomous. We have had a whole series of those, all the way through.

If you look at today's modern missile systems, they are almost necessarily autonomous because things happen so fast, so quickly and in so many complex ways that you need to leave that weapon system to make its own decisions about what it thinks it has seen and what it thinks it should do about what it sees. Therefore, it is very important to grasp that, although we are talking about autonomous weapons systems now, they have been in existence and in use for a very long time, so it is not a new problem at all.

The second thing that is probably worth pointing out before I head to costs and benefits, which is what you were talking about, is to try to make the point that one of the problems in the current age is that AI or artificial intelligence—I am pleased we have not used that yet—is to a large degree just enhanced use of algorithms and data. It is not pixie dust. It is growth out of techniques that people such as BAE Systems have been using for 50 years in lots of different applications: the way that images are processed, that targets are discovered and identified, that control is actuated into the weapons system itself. All these things are to do with combined statistics, control and a lot of other technologies that come together to deliver that. Therefore, we are on a journey rather than a sudden leap into something completely different. It is important for the committee to appreciate that.

The benefits are clearly there. We now face threats that are genuinely really hard to engage, such as hypersonic missiles. With the speed with which incoming systems affect you, there is no way of having a human in the loop. You have to have weapon systems that can identify, make good decisions and combine all effects, such as electronic attack, and cyber and kinetic effects, to guard against that sort of thing.

Exactly as was said, there is a huge role in defence. In fact, you could not have a modern defence system without that level of autonomy. There is no way of protecting yourself other than using autonomous weapons systems in all these things such as attacks of thousands of drones at the same time. It is everywhere. Again, underwater warfare is just a huge area; the data that you get back, such as the acoustic information, is not even addressable by a human, typically. You have to have a machine understand that signature. You have to have continuous contact with the enemy, to be able to deploy, play games, the whole sort of thing, at a speed and a tempo that is just not possible with human engagement. That is important.

The other part of it, which is challenging, is the offensive component. When I think of the offensive component, there is this thing where you

are clearly launching a kinetic effect or whatever, but the biggest thing in this area at the moment, which is growing, is is not the kinetic effect but the cyber effect. It is the general electronic warfare domain, attack using radar, attack on communication, influence, and a range of other things. It is how you influence the enemy and the population in so many different ways.

Again, this is an enormously complex area, yet in a sense it is the least impactful. If I can go out there and spread this information and attack using that content, disrupt communications and all of these sorts of things—exactly the kind of thing, in fact, that we have seen in Ukraine and, indeed, that the MoD had war-gamed in Ukraine—it is exactly the effect that you wish to achieve. Again, it is hard to get away from the degree of autonomy that is necessary.

Having said that, and coming back to the earlier comment, there absolutely is a supervisory role. You do not just go in and disrupt someone's communication, jam GPS or fly in some kinetic effect. There is genuinely a supervisory role in that process. At some level, there needs to be the right authority to make that happen. Certainly, the challenges when I was in the MoD were a lot to do with the fact that the regulatory systems for that type of effect, including autonomy, are so very different from what they are for guns, kinetic effect or anything like that. The committee has to deal with that disconnect in a significant way.

The Chair: That is very helpful. Thank you very much.

Lord Hamilton of Epsom: At the risk of being pedantic, surely "autonomous" means that there is no human intervention. Therefore, to talk in terms of having the human in the loop and that sort of thing is a contradiction in terms. Presumably somebody within the military hierarchy has to be responsible for this system when it comes to a supervisory role, but the bottom line on that is that their responsibility is one of pulling it if it is not doing the function for which it was originally designed.

Lord Sedwill: That is right, and Professor Durrant-Whyte has essentially addressed part of that already. The human intervention is, as I mentioned, essentially in setting the rules of engagement and determining when systems should be applied and in what circumstances. Once engaged, the systems will be autonomous. As Professor Durrant-Whyte says, there are degrees of autonomy. They could be completely autonomous. As he said, it would be simply impossible for humans to be directly engaged in dealing with, for example, a swarm attack by multiple hypersonic drones, so one would essentially have a monitoring system that was capable of responding. The issue there, as is so often the case, is ensuring that the system can distinguish with complete confidence and reliability between friend and foe. One might have an entirely autonomous system doing that. For the offensive, one would want to have much closer human control, making choices about the systems that are deployed.

The other thing we should keep in mind is that not all military operations are peer-to-peer combat. These systems will be of most use in those circumstances, but if you think of the counterinsurgency campaigns of the past 20 years or other interventions engaging the military, in those cases there will be systems with some autonomous capabilities supporting the troops on the ground, identifying threats and so on, but in the end we will still need people to engage with other people to achieve our objectives. It is easy to get caught up in the "Terminator"/"Matrix" world when one thinks about this, but it is going to be a much broader spectrum of usage.

The Chair: Setting the criteria is presumably a really important element in terms of the application of international humanitarian law.

Lord Sedwill: Indeed. If one looks at AI generally—of course, the Prime Minister is currently in Washington talking about the broader questions of AI—on can look back. Professor Durrant-Whyte mentioned that autonomy goes back to the 19th century. Essentially, the rules around the use of AI were probably best defined by Isaac Asimov in 1942 with his three laws of robotics: law one, a robot—of course, one now applies that word to a much broader range of capabilities—cannot injure or, through inaction, allow injury to come to a human; law two, they must follow orders, as long as that does not conflict with law one; law three, they can protect themselves as long as that does not conflict with laws one and two.

Those laws would form a pretty good basis if they could be designed in and audited for AI more generally to deal with many of the risks that have been identified in some of the coverage recently, including those raised by some of the tech companies involved in this. Of course, those laws do not really work in exactly that way if one is dealing with autonomous weapons systems, because they are designed to have lethal effect. Therefore, we would need to alter their order and be sure that we could build into any system, both defensive and offensive, that we deploy, in a reliable way that could not be amended and could be audited, maybe even by another AI system, that it follows precisely the rules of engagement intended by the human command.

Professor Durrant-Whyte: I largely agree with that. Again, I emphasise that it is a matter of degree for autonomous weapons systems. We have them already; you have to follow procedures and so on to launch a missile, a torpedo and all these sorts of things, so they are there. It is really an issue of the degree to which you make decisions, to which you intervene in them, and to which you allow things to progress according to a computer-based decision rather than a human-based decision.

Again, I emphasise that it is there already. It is not pixie dust; it is very much the process of thinking logically through what is the right thing to do. For example, if you needed genuinely to suppress an air defence system—let us take that as an example—you have a range of options to do that already in lots of different ways. Whether you choose to apply those autonomously or in a more civilised way, in the end, the

commander makes those decisions now but he himself is supervised by political decisions. It is just a flow-on from that, in my view—new tool sets, I would agree.

Q98 **Lord Browne of Ladyton:** Professor Durrant-Whyte, I just want to explore a couple of the things that you said. I intended to ask a different question but you have encouraged me to think about this one. Of course, we have had autonomous weapons for a long period. The very use of the word "ballistic" implies that we are no longer in control of a weapon on re-entry from space or wherever.

However, we are particularly interested in AI-enabled weapons systems rather than just autonomy in itself. We are particularly interested in weapon systems that are enabled by a technology that, to a degree, learns as it goes and which may not give you the same answer two days in a row to the same question because it is non-deterministic. That is what we are interested in. Are you saying that that is just a linear progression of what we have been doing all along the line in terms of autonomy?

The second point I want to make is that you described to us, I think very graphically, what you thought modern warfare is like in terms of technology. We are living through the extension of the invasion of Ukraine by Russia, which happened in February last year. We were promised by a lot of experts that we were going to get the first cybergeddon. That just did not manifest at all, so our most immediate experience of the most significant conflict that many of us have seen in our lives is that this is not how we conduct warfare in the 21st century and that artificially enabled weapons will be a significant shift from what we have seen up until now. Do I have this wrong? Have I misunderstood you, or is that what you were telling us? Have I mischaracterised it?

Professor Durrant-Whyte: No, but let me address your questions. The first thing is around AI. One of the problems with AI, as I said at the beginning, is that we call it "AI", because if you write code—and I have been working on AI for 40 years—essentially AI is applied statistics. What you do is take data in, process that data and convert it into decisions, whatever they may be. Most of those decisions are what I think of as standard statistical techniques; they are basic methods. For example, neural networks were invented in 1957. These are not new technologies. We have been using them for a long time. We process images, we bring that data together and we use that data to make decisions. That is essentially what AI is. Even for things such as large language models, which people are talking about at the moment, if you dig into how they work, all that is really happening is the next letter. The choice of the next letter is a probability of which letter is most commonly seen on the internet associated with the phrase you just put in. It is statistics.

I sometimes worry that people think that AI is completely different from everything we did in the past. It is not. The big thing that has changed is that we have lots of data and lots of computer power, and we can do hugely different things than we have ever done before—design

molecules, predict disasters, et cetera. But it is because of that; it is not because there is something fundamentally different about AI from anything else.

Therefore, yes, it is a progression. It might be a steep line in the sense of our computer power, our ability to get data and everything else, but nevertheless, it is a straight line. I genuinely do not feel that what we are seeing today is anything different from when I first worked on AI when I was 20 back in the 1980s. We can do much more but it is otherwise fundamentally the same. That does not mean we cannot do lots and lots with it. Do not get me wrong: I am not saying it is not powerful, but it is not fundamentally different. That is important.

To come back to Ukraine and other things like that, on the contrary: we are seeing drone technology, guided missile technology, guided shelling and targeting in lots of different ways, so we are seeing a lot of instances of autonomy applied to warfare. I have to say, I am as surprised as you at the cyber side of things. I come back the fact that we war-gamed the Ukraine situation when I was at the MoD, because it was one of the things we were worried about even then, and we believed that that would happen. We also believed that Russia would have air dominance, but it does not. No one really appreciated, I guess, that things were perhaps not as much as we thought they were going to be. But for other jurisdictions—for example, Israel or China—cyberwarfare is here and present in spades. It is likely to have a huge difference.

The Chair: Lord Sedwill, do you want to add to that?

Lord Sedwill: There is one point I would make in response to Lord Browne's question. In this kind of discussion, it is probably helpful for us to try to make a distinction between essentially the tactical and the system. A lot of what we are talking about here has been individual autonomous weapons, and essentially the use of autonomy at the tactical level. I think what Lord Browne is driving at is: what happens if and when we reach the conclusion that, because of the speed of response and the need to have a comprehensive response, whether defensive or offensive, we essentially hand over the command and control system to an algorithm, as opposed to that being the individual decisions being made by commanders?

There is a distinction between autonomous weapons, of which, as we have said, there is a long, linear history and the step change that there would be if, for example, we said that to deal with the ballistic missile defence issue in the UK—we do not have very many at the moment, as you know—we were going to have an essentially entirely automated system to identify friend or foe, that would be able to launch, et cetera, and the political decision was to develop it and allow the whole thing to be automated. That is a qualitatively different thing from individual autonomous weapons. It is probably useful just to make that distinction.

The Chair: I have an eye on the clock. We might need perhaps to come back to some of those issues a little later.

Q99 **Lord Houghton of Richmond:** I have a question that primarily draws on Lord Sedwill's experience as a National Security Adviser in setting national risks; Professor Durrant-Whyte might want to comment on what he comes up with. The committee has been very much bounded in looking at artificial intelligence within weapon systems, when these are weapons systems that, if you like, exist above the threshold of formalised warfare. My argument consistently is that these can be dealt with purely by extensions to our systems of regulation of the battlefield, in which, for some time to come, the introduction of meaningful human control will be an element.

However, from the perspective of national risks, there are an awful lot of artificial intelligence enabled vectors of attack between countries, if I can pompously call them that, that exist below the threshold of formalised warfare and get at things like alternate truths, fixing elections, and the denial of food, energy, infrastructure and power systems. All these represent far greater risks to the country and are below the threshold of formalised warfare. Lord Sedwill, I know you are not currently NSA, but going back to your time in attempting to set these things, do you think that artificial intelligence poses far greater risks to countries in its non-kinetic sense, and that we need to know this to put our study in context, as opposed to things that are no more than an extension of current frameworks that ban, treaty limit, or regulate their employment?

Lord Sedwill: I think that is right. If I were writing this into the national risk register now I probably would not have it as a single risk for the reasons Lord Houghton set out, because in some areas it is essentially an enhancement of the existing risk. We know that there is a risk of cyberattacks that could disrupt our mobile communication systems and so on. Applying more sophisticated techniques, through machine learning, quantum, et cetera, to those threats means that we probably have to raise the level of our defences against that risk. In the end, risk is essentially a pretty simple equation: threat minus capability. If the threat goes up you have to improve your capability, otherwise your risk goes up too.

If we are worried about that, and we would be right to be, it would be essentially: what are the enhancements to some of the risks that already exist—the disruption to mobile communications, power supplies, et cetera, that you mentioned—that new technologies, including machine learning, could present? There would be some areas, like deep fakes in elections and communications, which probably have not made it through the threshold to the national risk register in the past because the technology has not really existed to enable that to happen. Therefore, there are some risks of that kind that probably should now break through to the national register, because the technology means that those risks are real. I would tend to want to look at it risk by risk and ask: are they new risks and/or does this enhance the potency of some of the existing ones?

Professor Durrant-Whyte: I agree with Lord Houghton very much. I suspect that what we think of as the kinetic weapon area or this sort of

thing can largely be dealt with through an extension of the rules-based order that we currently have. I agree that there are issues to consider.

Having said that, where we have no real understanding of the potential risks is in these areas around influence, attack of civilian structures and these sorts of things. AI has a very large role to play in that. Despite what I said about it being applied statistics, you have only to look to see what is happening in terms of influence, fakes and a whole range of different things and, to be honest, the impact it does have on the public to see that it would be a real concern. I also have to say that our experience when I was at the MoD was that our adversaries are much better at this than we are, because there are different ways that we have dealt with it in the past. It is a challenge and a risk that is worthy of real consideration, in terms of autonomy and AI.

The Chair: Our adversaries are better at it than we are? What are the fields and the methods that are giving them the edge?

Professor Durrant-Whyte: Some of the advantage they have is that they are not regulated in any way. They are more than happy to put up groups that can go out and attack, organisations, companies and the public at will, influence elections and so on. I do not think the Government here would ever allow us to influence someone else's election.

The Chair: That is a sobering answer. Thank you.

Lord Sedwill: To give you an example, which is not particularly AI-enabled but will give you a sense of what could happen, after the Salisbury attack we had a single narrative, very largely based on investigating the truth of the fact that it was a GRU operation. The Russians did not counter that with a single counternarrative. We are used to narrative, counternarrative. They did not do that, partly because they did not have a credible one, so they sought to confuse. I think at one point we reckoned that about two dozen alternative narratives were being promulgated on social media and through the mainstream media by the Russians, not because they believed that any one of them was going to essentially defeat our narrative but just simply to create enough noise that our narrative was lost, particularly in the areas they were seeking to influence, which was not our population but the populations of much of the non-aligned world.

One could imagine that happening in a much more sophisticated, targeted, industrial-scale way if enhanced by modern AI techniques. That sort of subversion has been on the national risk register for ever, long before there was even a national risk register. There are qualitative differences of that kind with counternarratives, deep fakes and so on that we will have to address. We counter risk by doing the same ourselves: we have to find a way of sticking to our values while making sure that we are effective in promulgating our narratives, values and principles.

Q100 **Lord Triesman:** I suspect this is probably principally a question for Lord

Sedwill, but I would like to try to peel back another layer on the potential and multiple additions to the national risk register. In a previous Select Committee, which I had the privilege of being on, that dealt with existential risk, among the existential risks that were considered, especially by Lord Mair, who is an eminent engineer, was the accidental risk that occurred when a dam overtopped. I am not choosing that example because of events of the last 72 hours but it was one of the risks. The secrecy that surrounds the national risk register is understood; one does not want to advantage people who have malign intent, but it is also true that people must respond to a risk that may occur and then occurs. They must be forearmed to be able to do that effectively. If there are now some other things that should be on the national risk register—maybe not just one but several things—who should be in on the secret? Who needs to know what those risks are to try to deal with them? Being locked in a room in Westminster may not be a solution to anything.

Lord Sedwill: I agree with the premise that you put forward. When I was National Security Adviser, the procedure for the national risk register was that there is a national security risk assessment, which is confidential, and then a version of that is published every couple of years or so as an open national risk register for first responders throughout the country. I think another Lords committee reported on resilience a year or so ago. I rather agree with its view, which is that the primary document should be the open, unclassified one—the national risk register—with a classified annexe for those things that are genuinely classified national security risks, and we should turn that around.

Then, of course, the issue is how you follow up, as you have indicated. If we conclude that there is a major biosecurity risk or there are risks, as you suggested, to dams, our power systems or communications and so on, then we simply must find a way of engaging the experts, including, in many areas, the private sector, as well as local councils and local responders and so on, in the contingency planning to deal with that. That is not as complex as we sometimes think. We do not necessarily have to reveal the underlying intelligence sources to be able to engage people in thinking through contingency planning and identifying the capabilities that we would need. Sometimes the intelligence will tell you that a risk that, in a sense, is self-evident is more or less potent than we might have thought, but the risk itself is probably something that any reasonably well-informed expert would be able to identify.

The Chair: Professor, does that accord with your view?

Professor Hugh Durrant-Whyte: Yes, and I have nothing more useful than that.

The Chair: I will go to Lord Browne very briefly before we move on to the UK on the global stage.

Lord Browne of Ladyton: This is not a question; it is just a fact. The last risk register, which was published in 2020, has one mention of AI in the context of new technologies, which it describes as probably causing

"shifts" in "the nature of how we work in the future". It warns that we in the UK "must be aware of these changes and work through the implications on people and businesses in the coming years". That is what we currently have in the context of AI. A piece of journalism in a national newspaper says that, apparently, the Cabinet Office told journalists on 6 June that the Government will publish the updated national risk register soon, so we may know what presence it has very quickly.

The Chair: "Soon" has a number of interpretations in the parliamentary and political vocabulary, like "shortly" and "in due course".

Q101 **The Lord Bishop of Coventry:** As you said, to continue the theme of the UK on the global stage, this is a question about capabilities, whether we are using those capabilities well and responsibly, and whether we are exerting good influence. Lord Sedwill referred to the Prime Minister being in Washington. He is reported to have said these words, I think on the plane to Washington: "You would be hard-pressed to find many other countries other than the US in the western world with more expertise and talent in AI. We are the natural place to lead the conversation". Do you agree with that generally, perhaps, but also in terms of its application to AI in weapons systems? If you do, do you feel that we are using our capabilities well—by that I mean responsibly? How far can we influence others to use AI in weapon systems well and responsibly? Do you see evidence of that which has been claimed in written evidence we were given on Monday from the MoD, of the UK promoting "values-based norms and standards for military AI", particularly in the light of Professor Durrant-Whyte's comment about some other jurisdictions being less sensitive to rules and regulations?

Lord Sedwill: I might leave Professor Durrant-Whyte to comment more on the skills and talent in the UK. I believe we are in pretty good shape, but he will know that better than me. He is in Australia, of course, which is pretty good at this as well. He will have more to say on that than me.

The general point you make is essentially about doctrine in this area. We must look at doctrine; Lord Houghton will have views on this as well. There is something about setting rules and standards for the use of all weapons systems, and certain weapons systems, such as chemical weapons, have been banned. By the way, we should not assume that AI will necessarily make all weapons more dangerous. In many cases, it will make them more precise. Therefore, we should be able to minimise combat casualties, at least on our side, and civilian casualties on our adversary's side, so there should be benefits in international humanitarian law to the use of AI, as well as enhancing combat power if done properly.

All the other elements of doctrine will have to be brought into this, because we will set rules and standards for ourselves, as we do in all other areas of defence and rules of engagement and so on, but we cannot be absolutely confident that every other country will do so, so there will need to be defence and deterrence. There will probably be a role for arms control agreements as well, with verification mechanisms that will be very

different from human inspectors walking around and looking at sites as they did in the past, counting missiles and so on. It will have to be verification of code. If a country says, "Yes, we have applied unamendable code sitting at the heart of our AI defence systems that means that they are not fully autonomous, they cannot get out of control, they will only do what we want for them", there will probably have to be some kind of AI audit of that code that will tell us whether that is the case.

These are concepts that, as I think both of us have been arguing, we are familiar with. We just have to apply them in the new contexts of defence, deterrence, arms control, and verification. All of that should enable us to have a framework that we can be confident of in this new era.

Professor Hugh Durrant-Whyte: I will deal with the skills part of it, but the first thing to recognise is that 98% of what is going on in computer science and AI is arguably not in defence; it is in the civilian sector, commercial products and things like that. Truthfully, defence is a latecomer to the game in terms of what is generally happening.

When you look at what is going on globally, without question, China is investing arguably twice as much as everyone else put together. We need to recognise that. It genuinely has gone to town: if you look at the patents, the number of publications and the companies involved, it is quite significant. Yet it is important to point out that the US is still dominant in the area by a long way. It has more innovation, it has fantastic companies, and the big breakthroughs, ChatGPT being the latest one, typically come from the US rather than China.

The UK punches above its weight in a lot of different areas than AI. There are some fantastic companies. DeepMind is a very good example, but I can name lots of others that are genuinely changing the game in AI—stuff that has never been done before anywhere. Arguably, their impact is much more than all the things that are going on in China at this point. The UK needs to be proud of what it does in that area. You could pick very specific niches, such as computer vision, where the UK leads the US, I would argue.

The UK needs to have its capabilities and be world-class in them. World-class matters in this domain; being second best does not. On the other hand, like everything in security and defence, a lot of it is about our international collaborations. The fact that we collaborate with the US, Australia and other places, and the rest of Europe, is critical to the fact that we can maintain a significant lead in what we might do not just in weapons systems but in commercial systems. It is important to understand that. I worry a little bit that the UK is falling behind in terms of its investment in that type of technology, but nevertheless it holds its own in a lot of key areas, both corporately and in academia.

Q102 **Lord Hamilton of Epsom:** You mentioned international collaboration. Of course, the Prime Minister is talking to the United States about AI systems, presumably in the civilian sector, but there is a direct read-over

to defence systems. He is talking in terms of the EU, China and the United Kingdom. The notable exception is Russia, which I do not think will ever co-operate on anything like this. Where do you see these talks going and is there a read-over? If we are going to have a regulatory body dealing with the civilian sector, could it deal with defence as well?

Lord Sedwill: It is very difficult to see how any international regulatory body could be effective in regulating autonomous weapons systems, not least because it would require a range of capabilities around verification and so on that would essentially skew its entire work. I would need to give this some more thought; I have not thought about it a great deal so I would apply that health warning.

A body that tries to set a set of systems, norms and guarantees for the civilian world—whether based on Asimov's laws or something similar, but that would be a pretty good place to start, I would have thought—should focus on that. As Professor Durrant-Whyte has said, that is 98% of what is going on in this world, and a lot of that will be around questions such as whether AI systems should go through some kind of trial or pilot scheme before they are launched on an industrial scale. There is a suggestion about applying the same kind of approach that we apply to drug trials: should that be applied to AI, particularly if we are using AI techniques in terms of medical innovation and so on? Trying to include defence in a body that is helping to set standards and norms for all that, in the way that we would in many areas of activity, would probably skew its work. Weapons systems and defence and AI systems, as I mentioned earlier—the strategic as opposed to the tactical—are probably better regulated through an adaptation of the existing arms controls and other strategic mechanisms.

Professor Hugh Durrant-Whyte: I largely agree with that. Just picking the civilian area, there are two pieces. It is very hard to regulate the development of our algorithms and innovation in that sense. I cannot imagine how you would do it.

There is a lot to be said for regulating the way data is used. I can see someone coming in to say, "This needs to be tested against different communities or different outcomes", in the same way that you would run a clinical trial. That makes sense if you have some product assurance associated with it.

At the algorithmic level, you can always take something that is approved for one job and use it to the detriment of people in another way, as you can in many technology fields. That is very hard to regulate. I can have an algorithm for recognising friends and images and use it for targeting people, for example. It is very hard to see how you would control it at that fundamental level.

Exactly as Lord Sedwill said, medicines and things like that may be the best possible approach. Some regulation in the civilian domain is probably achievable. Regulating in the military domain would be much harder.

The Chair: Bishop, you set us going on this area of questioning and I rather cut you off. Do you have a supplementary before I go to Baroness Hodgson?

Q103 **The Lord Bishop of Coventry:** It is very interesting and reassuring to hear Professor Durrant-Whyte's confidence in our skills and capabilities. Could our witnesses say something more about the values approach and whether you see the excellence that you have described? Is that excellence something to do with a values-based approach? This is related to it, but can I come back also to that expression that Lord Sedwill used of unamendable controls, which I think is related to the values-based approach—you build in unamendable controls that reflect that? The thought that leaves me with is whether that concept of unamendable controls will stand the test of time as AI develops at pace. Is that what some of the scientists, manufacturers and inventors are saying: that that might not be possible in the future?

Lord Sedwill: It is a great question, and I am now beyond my area of expertise. I mentioned Asimov's laws because that is the earliest crystallisation of the kind of principles that should be built in to deal with autonomous artificial computer systems and so on. As I say, it comes from a short story written in 1942.

The point I was making, and this is where people with much greater expertise would have to work out exactly how we would do it, is that something of that kind is what we would be looking for as policymakers. Where we are talking about autonomous general intelligence, or AGI, as opposed to individual AI applications to unwind DNA and so on and to create new drugs—that is again the issue that we are touching on here: we must be quite clear about the difference between the tactical and the strategic—before any system of that kind were allowed out of the lab I would want to be assured that there was an auditable assurance that something like those three laws were built in: that it cannot harm through action or inaction, it would always follow instructions and so on.

As Professor Durrant-Whyte says, that may be phenomenally difficult to do. In a sense, that is the task we should say to the tech industry: "If you wish to continue to pursue all the potential for human welfare of these new technologies you must build in some safeguards of a kind in which we can have not just statistical confidence but complete confidence". That is a challenge we must put over to the people developing this: is there a way of ring-fencing some code or whatever it might be—as you can tell, I am way beyond my area of expertise here—that could not be amended and that would set the boundaries to how an autonomous system, essentially learning and evolving in its own sense, might operate?

Professor Hugh Durrant-Whyte: There are two issues here. First, let us pick self-driving cars as an example. Even this morning there was something about Tesla running over a dog. The problem is that the number of cases is infinite; no matter how much you test a piece of code, you cannot guarantee that it will work everywhere, every time in every

circumstance. It is just technically not a possible thing to do. You need to step away from that.

However, there are two other ways that people try to approach it. The current way, even in things like ChatGPT, is open code: everyone gets to look at it, everybody gets to check it, and that code evolves. A good example would be Unix, which is the standard operating system people have now used for 50 years. Because it is open source it is almost indestructible, because people have worked on it for so long and checked it in so many ways. It is open. With UK companies particularly, and UK academics are good at this in AI, you will find that almost everything is open source. That is a fantastic way of implementing what Lord Sedwill has suggested at a pragmatic level.

However, there is an opportunity—there are groups working on this, and we tried to work on it when I was at the MoD—of formally, mathematically proving that an algorithm does only this and no more, and providing an absolute guarantee using a mathematical process rather than a "test until it breaks" process. In the future, that will be the only way that we will get these things right, but it is only an emerging discipline. We need to understand that that will not happen in the near future, but it is the way things will need to be done.

The Chair: Lord Sedwill, you are under the pressure of time, as are we. Are you both able to stay with us for another 10 or 15 minutes?

Professor Hugh Durrant-Whyte: I have to exit on the hour.

The Chair: All right. Take your leave of us at the time you must, but let us see whether Baroness Hodgson can squeeze her question in while we still have you.

Q104 **Baroness Hodgson of Abinger:** We have slightly touched on elements of this question, but there has obviously been a lot of publicity nationally and internationally about AI. Is the language of an arms race in AI development useful, and to what extent does the development of AI-powered weapons by hostile powers create a compulsion to develop and deploy such systems? In that context, how reliable is our information about what other states are developing?

Lord Sedwill: I can stay for another 10 minutes, so maybe we should let Professor Durrant-Whyte have a go before he must go.

Professor Hugh Durrant-Whyte: AI is hugely overhyped at present. Again, I go back to something I said at the very beginning: you need to be careful because we use the word AI. If we used "advanced statistics" instead, people would not get so hyped up about it. That is a challenge. If you go back historically, the reason the word "AI" was invented was because a group at MIT was trying to flog a project to Defense in the US. That was back in the 1950s. That is why it is there.

We must be careful. It is not quite what you are suggesting. Other countries are developing these things but we are as well. It is a

progression in lots of different areas and it is not something that has suddenly changed and is suddenly magic; it is just a progression of algorithms, data, larger computers, better sensing, all of these things—precision, as Lord Sedwill said. It can bring benefits, not just the negative things associated with it. That is important to understand as well. Please do not think that somehow AI is changing the game. It is not, and I worry about the hype that is going on in the press now.

The Chair: As the clock is about to strike the hour, not only in London but in Sydney, thank you very much indeed for joining us. It has been really useful, and we are most grateful.

Professor Hugh Durrant-Whyte: Thank you very much.

Lord Sedwill: I agree with Professor Durrant-Whyte there, but this is true of all developments in weapons systems: it is possible to compensate for inadequate capacity through better capability. I mentioned the example of the Six Day War in 1967, where the much smaller forces of Israel outfought their adversaries because they had more sophisticated capabilities, including human capability—they were better trained and so on.

Of course, whereas some countries can simply use mass, the UK, as one of the leading but below the top handful of powers in terms of sheer scale and mass, cannot afford to fall behind. We must continue to innovate in this area as we innovate in others. I argue that we should try to make that a source of competitive advantage in defence, compared with others. There has been an awful lot of speculation recently about the size of the army in Europe and so on. In the AI world, the number of troops may be much less relevant than how much combat enhancement is available for each troop, system or individual. We must stay competitive in this area.

The other point I would make echoes a point Professor Durrant-Whyte made. The big qualitative shift in defence doctrine in the past century was the development of nuclear weapons. I do not think we should see AI as being as significant as that, in that that completely changed doctrine because of the concept of mutually assured destruction, so wars could not be fought. AI may make that issue more potent as it enhances systems, but I do not see another step change of that kind coming as a result of it. It may just be that that particular step change becomes even more relevant than it was back in the 1940s.

Q105 **Lord Mitchell:** A statistic that came up, which surprised me, was that 98% of all AI is not in the defence sector, meaning that 2% is. I was surprised by that. Does the UK Government engage effectively with private industry in developing AI systems? You are perhaps in a very key position to comment on that. Is their approach to policy-making capable of keeping pace with the speed of AI development?

Lord Sedwill: Again, I am not an expert on this. The effort is clearly there. The MoD has just issued its paper*, Ambitious, Safe, Responsible*,

setting out its approach to this. You have had evidence from BAE Systems. Its research in this area is partly driven by demand.

The 98% statistic was new to me as well, just for the record; it was from Professor Durrant-Whyte. It probably is right, given the degree to which AI is used in all sorts of areas to enhance capability. "Machine learning" is probably a better way of describing it. If we think about defence, we have been thinking about this for several years. For example, the concept of autonomous mine hunting has been around for a few years and there are programmes to try to develop that kind of capability because, for obvious reasons, it is much better to get machines to do that work than to put our own people at risk. There are detection systems for threats, including tracking our adversaries' nuclear submarines. All those things will involve the kind of algorithms that we are talking about, and all of that development is live.

Q106 **Lord Sarfraz:** Professor Durrant-Whyte talked about the hype around AI, and he talked about open-source systems. What has changed is that, because of the hype around AI, the open-source community and the number of individuals engaged with AI has grown a lot. That is a difference from a few years ago. That also means that some of the cleverest technologists in the world are perhaps those who would not pass security vetting to come into the MoD, let alone become defence contractors. When you were NSA, what were the ways to engage with the best and brightest who would not be the traditional folks who defence would engage with?

Lord Sedwill: That is a great question. One of the issues that we had was that some of the best and brightest in this area, for their own ethical reasons, were unwilling to work on defensive systems. They were prepared to work with DeepMind, for example, on the development of some of its systems in medical and health and so on, but were not willing to work on defence. The Ukraine invasion, as it has more generally, has reminded everyone that, properly managed, defence is a deeply ethical part of a democratic system. I suspect, therefore, that some of those attitudes have changed.

In the end, it is simply the case that we have to be more thoughtful about how we engage some of the best and brightest here. Of course there will always be secure areas where nationality, vetting and so on matter, but those are a small fraction, to be honest, of much of what we will see in this area. I suspect that we will also see that in defence, AI, machine learning and quantum computing, when that comes along, as with most other algorithms, will probably start out by being 90% based on a civilian algorithm with defence enhancements, rather than being based on something "defence-ive" from scratch. Pattern recognition, facial recognition and even targeting systems are based on algorithms used in completely different areas of civilian life, as I am sure you know. It is bound to be blended.

Q107 **Lord Fairfax of Cameron:** Quickly, because I know that you are short of time, if you could make one recommendation to the UK Government in

this whole area of AI-enabled weapons systems, what would it be? If you have time, I have an unrelated question. Both you and Professor Durrant-Whyte distinguish to some extent between offensive and defensive weapons. My question relates to defensive nuclear weapons. Would you have any comments in that regard?

Lord Sedwill: My one recommendation would be to invest. For this country to punch its weight in military terms over the next 25 to 50 years, we should be investing in this area. That also means being able to invest in a way that has a higher risk appetite than government investment traditionally does. We now have mechanisms to do that through ARIA and so on—some of these new government investment opportunities. The recommendation is to invest. This will be the future of defence capability and the UK needs to be at the forefront of that, just as the Prime Minister and, I should imagine, almost any Government would want us to be at the forefront of this in the civilian area.

On the nuclear deterrent, this goes to the question of the point at which you have control. I could imagine, as I think that we have both described, having an almost entirely autonomous ballistic missile defence system, particularly as hypersonic missiles become more of a threat because of the speed of response required. That would be the individual capabilities—things like the Patriot system, directed energy weapons or whatever it might be—as well as the whole system: if you like, the corralling organisation of that system.

I do not think you could ever contemplate that in the offensive area. We go to great efforts already to maintain positive political control within the deterrent. It is why every new Prime Minister, almost on their first day, has to write the letters of last resort to the nuclear submarine commanders, appoint nuclear deputies if they are incapacitated or incommunicado in a conflict, and so on. All those controls would have to be maintained.

That does not mean, of course, that AI algorithms would not have a role at the tactical level in improving targeting, developing the systems, making them more accurate and reliable and all that. Of course, since the test-ban treaty, we already test all our warheads in a virtual environment. That is AI as well. In terms of the deployment of nuclear weapons and the deterrent as a whole, maintaining positive political control, not just positive human control, is essential.

The Chair: Thank you very much. Lord Houghton has a very brief question to conclude.

Q108 **Lord Houghton of Richmond:** It is a quick and somewhat naughty one, because I know that you are not speaking on behalf of BAE Systems. In a world where the future discriminator of military capability is more likely to be based on software, not hardware, would you share the view that the UK's defence industrial base is too dominated by defence primes, whose primary source of income is the glacial replacement of generational platforms?

The Chair: "Time allowed: three hours".

Lord Sedwill: If I answer that question, I will either be reprimanded by the House of Lords for a conflict of interest or be sacked from the BAE board.

You have a very fair point and a challenge for defence generally. I do not think that one has to particularly get into the question of primes; it is a question of how we engage effectively and create the procurement systems, as you know, to engage effectively with some of the smaller, innovative software houses, particularly those spun out of our universities, where there is, as Professor Durrant-Whyte said, some very impressive work. Defence needs to get much more sophisticated about that, ensuring that the providers of the hardware platforms are open to plug and play of systems, not only AI software systems but other systems. If someone develops a more sophisticated targeting system or a more sophisticated version of something else, they should not be essentially suppressed by the primes. That would be true, by the way, of all defence sectors.

The Chair: Lord Sedwill, thank you so much for your help. You have got us thinking in whole new areas of our inquiry and we are extremely grateful to you. Thank you again.

Lord Sedwill: Thank you. I am glad to have participated.