HOUSE OF LORDS

Artificial Intelligence in Weapons Systems Committee

Corrected oral evidence: Artificial intelligence in weapons systems

Thursday 20 April 2023

10 am

Watch the meeting

Members present: Lord Lisvane (The Chair); Baroness Anderson of Stoke-on-Trent; 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 Sarfraz.

Evidence Session No. 3 Heard in Public Questions 24 - 42

Witnesses

I: Courtney Bowman, Global Director of Privacy and Civil Liberties Engineering, Palantir; Professor Kenneth Payne, Professor of Strategy, King's College London; James Black, Assistant Director of the Defence and Security Research Group, RAND Europe; Dr Keith Dear, Managing Director, Centre for Cognitive and Advanced Technologies, Fujitsu.

Examination of Witnesses

Professor Kenneth Payne, Dr Keith Dear, James Black and Courtney Bowman.

Q24 **The Chair:** Good morning and welcome to our witnesses. It is very good to have you here with us—Professor Payne for the second time. This session is being broadcast and will be transcribed, so you will have an opportunity to correct any factual errors afterwards. We will go around the committee pursuing particular lines of interest and inquiry.

There is one thing I am particularly keen that we should do this morning, and we had a brief chat about it beforehand. It is inevitable that, in a subject such as this, we get an almost inextricable mixing of strands—political, military, strategic, ethical and so on. We will try, if possible, to compartment our questions a bit more than in previous sessions. If you could complement that effort by keeping your answers pretty much within those compartments, that would be a great help, and shorter is always better. Would you like to introduce yourselves briefly to the committee?

Courtney Bowman: I am global director of privacy and civil liberties at Palantir Technologies.

Professor Kenneth Payne: I am professor of strategy at King's College London.

James Black: I am assistant director of the defence and security group at RAND Europe, which is the UK and European arm of the RAND Corporation.

Dr Keith Dear: I am a former 18-year intelligence officer, former expert adviser to the PM on science and tech, and now managing director at the Centre for Cognitive and Advanced Technologies at Fujitsu. I should stress that I speak in a personal capacity today.

The Chair: Understood, thank you very much. Let me begin by going back to those twisted strands, as it were. Taking the first, there are lots of ways in which AWS might impact on events in combat or in preparation for combat. We understand that there may be a preliminary stage where AI is informing a subsequent engagement. It would be very helpful to have your impression of the likely impact of either AWS or AI-powered systems on speed of escalation, if we can take that issue to begin with.

Dr Keith Dear: My reference point would be a 2019 article in *Nature* called "Physics in Finance", which described the way in which we have automated decisions in the financial sector, where trades now take place at 0.004 seconds, faster than a camera's flash. In order to overcome human error, we have a speed bump in place, which is also overseen by machines. It runs at a similar speed, so the alignment and the protection against errors is machines guarding machines.

I would extrapolate from that and note a paper published in December by Anthropic, an AI company, which talked about reinforcement learning with AI feedback. Their argument was that, increasingly, they are getting AI alignment around what we would want it to do by using another AI in order to inform and train it.

We are heading in a direction where the escalatory barrier to AI will be other AI guarding the guards, as it were. I do not think that it is a simple question of whether that means that we are likely to get fewer or a greater number of flash crashes in the military. With those guard-rails in place, we will probably see it broadly flat, but the way in which we guard against escalatory risk will change.

The Chair: That, of course, immediately raises the question of where you set the speed bumps and the criteria that you employ.

Dr Keith Dear: Yes, and the trade-offs that you are willing to make, because those speed bumps induce risk. In the financial sector, a delay of one millisecond is said to equate to the loss of hundreds of millions of dollars a year in competitive disadvantage to somebody who does not have such a delay. In the financial sector, you can regulate and mandate it, but you cannot do that in war. The trade-off that you make in your system versus somebody else's will be a profoundly difficult decision, and we have to think through those implications and adjust according to circumstance.

James Black: I agree. To build on that, it is not the case that the escalation risk is simply a function of AI's greater capacity for speed of action or speed of decision; it is also a question of what those safeguards are. Related to that, when we think about escalation, we have to consider our level of understanding of the adversary or competitor with whom there would be said escalation. There are areas in which AI can improve and inform our understanding of adversaries—their strategic culture, how they think, how they make decisions, and how they are likely to respond to certain actions or signals that we may be advertently or inadvertently sending.

There are some benefits there in our understanding of how our actions might be perceived and, therefore, how we might manage accidental risk of escalation, which is already a great issue when we are talking about two humans, let alone machines. The added challenge that we are encountering now is that we are not talking just about understanding human decision-makers in different national capitals around the world, how they will respond to our actions and how things may escalate; we are also trying to understand how their own approach to and integration of AI within their own decision-making will inform their escalation ladder and, therefore, how we can control moving up or down that escalation ladder.

The problem is that these very quickly become quite complex, recursive questions. A lot of the theory around deterrence and managing crisis stability and strategic stability is very Cold War era, humancentric and

bipolar, but what we are talking about now is a world that is multipolar, with multiple major actors such as Russia, China, the US and Europe, where it is not just a human component in decision-making but also a machine component.

The Chair: Presumably, if you have an intelligence-led picture of how capable a potential adversary is in AI and AWS terms, that could also inhibit escalation.

Professor Kenneth Payne: It could. When I was last here, I pointed to a RAND USA study of a war-game that pitted human-machine teams on one side against human-machine teams on the other side, and studied the escalation spiral. It turned out to be quite a rapid escalation spiral in this particular tabletop exercise, because the opposite of what you have just described, which was uncertainty about how much the adversary had outsourced to automatic decision-makers, meant that you had to jump the gun and get your retaliation in first, so up the escalation spiral you went.

The broader point, to build on both these answers, is that deterrence, escalation and coercion are psychological as well as material factors. We have a decent understanding of how humans and human groups go about thinking about that. We do not have a similar level of understanding about how machines go about that.

If you had asked me this question a year ago, I would have said, "Well, machines are very different and they're calculating the odds in a very different way". I would have pointed to AlphaGo and the way in which that DeepMind computer played the board game of Go. It made some radically different moves from those a human would have made.

I have been given some pause for thought by the arrival of large language models, which you are probably very well aware of. I point the committee in particular to a paper that came out from Stanford a few weeks ago, where the latest variant of OpenAI's language model, GPT-4, was put through a battery of standard psychological tests of the sort that you would give to adolescents and small children to see whether it grasped theory of mind. If you buy into this idea that deterrence and compellence is about trying to gauge what other people are thinking, it is quite interesting to look at what GPT-4 is doing. It sailed through the tests. It had a 10 year-old human's level of understanding. Another person could have a series of false beliefs about a situation that it could profitably capitalise on.

At least on the surface of it, that one paper looks to me like there is something going on here. There is some way in which machines are thinking differently, but there is some qualitative overlap in how they are trying to gauge what another mind or another agent is doing. More broadly, this study of how algorithms relate to other agents is absolutely vital for thinking not just about escalation and deterrence, but more broadly about AI in society.

Courtney Bowman: Recent advances in generative AI are certainly impressive and point to a lot of potential development in applications, but

we are still at a place where these technologies operate within boundaries. The framing of the broader question is right in terms of evaluating time as one of the most essential and valuable resources for war fighters and decision-makers.

From a tactical perspective, this translates to an understanding of how we get to accelerated frameworks where human decision-makers who can operate within the contexts and boundaries of international humanitarian law can still exercise their expertise to make sound decisions. By using, for example, models that provide computer vision capabilities to evaluate satellite imagery, identify potential targets and then surface those for decision-makers to make sound decisions against those identified potential targets, we get to a place where we can still make rapid advancements and deal with a changing war-fighting landscape but ultimately make sound decisions that are defensible. That is at the tactical level.

At the strategic level, there are still considerable opportunities in war-gaming exercises to play out different scenarios, to examine potentialities and to be prepared when those potentialities come to fruition. There is a lot that can be done here. We are at a place where we still need to bound the technology and understand what the constraints and limitations are before we go too far down the line in assuming that certain types of agency are being propagated or produced through advanced states of technology.

The Chair: We have been talking up to now in terms of recognisable and organised power structures. I was going to ask you about the potential role of non-state actors, but, if I were to do that, I would be trespassing on an area that Lord Clement-Jones is about to explore.

Q25 **Lord Clement-Jones:** I want to come back to the point about what is being done to safeguard the use of AI in weapons systems from potential outside interference, but also the issue of inherent problems and malfunctions, and what safeguards there are. You started to deal with that, Dr Dear, which really raises the question as to whether there is any acceptable level of risk in that respect. I wonder whether you could deal with both of those.

Dr Keith Dear: Is there ever an acceptable level of risk? The answer is always yes. One of the challenges that we have at the moment in the deployment of lethal autonomous weapons systems is thinking of them as if we needed a whole new framework. We have a framework for the deployment of lethal force. It is international humanitarian law and, within the military, the law of armed conflict. That has certain principles that you have to meet—for example, distinction, proportionality and humanity. These are all things that you can test the ability of a machine to meet.

My contention for a long time, while still serving as a regular, was that we ought to begin baselining the performance of humans in how they perform in particular scenarios, and then comparing it to the performance

of machines. If those machines produce a lower false positive and false negative rate than humans, particularly under pressure, it would be unethical not to delegate authority to those machines. It would be more likely, for example, that, in the future, the MoD might find itself sued for not having delegated authority to a machine by a grieving parent of a soldier or of a civilian killed in combat, when, statistically, you can prove that that individual would have been much less likely to be killed by a machine with the delegated authority.

We already do that with humans. If I delegate authority to somebody in the military, and then they go rogue, act poorly or do something unexpected, we go back and audit: "Did you train that individual correctly? Did you select the right individual? Why did you not pick that up in recruitment? Why was it not noticed?" It would be largely the same thing. "Before you delegated authority to that system, Keith, did you check its training data? Did you understand the parameters and the orders that you had given it?" If all those things are satisfied, there is then a decision around accountability. In the end, we do not need new frameworks; we just need to apply the old ones more rigorously.

Lord Clement-Jones: That is very clear. Then there was the foreign interference point.

Dr Keith Dear: Sorry; would you mind repeating it?

Lord Clement-Jones: I asked two questions. The second was about the inherent problems, malfunctions, risk and so on. The first was about what is being done to safeguard weapons systems from potential outside interference—the foreign state actor point that the Chair made.

Dr Keith Dear: There are efforts to protect weapons systems against catastrophic degradation—i.e. if somebody interferes with the training data. Adversarial AI, which some of my colleagues work on, is designed to detect when somebody is trying to interfere with it or cause a system to malfunction.

To the point that both James and Ken made, we can think about this in the way that you would try to influence a human who is operating a weapons system. As you might try to deceive a person through camouflage or through misinformation, so too with AI, and defence is still building in the safeguards to those systems. Indeed, we at Fujitsu are building in some of those defences that can be applied.

Lord Clement-Jones: Do you agree that this does not raise new issues and that it is all pretty much covered by how we would deal with human actors?

Dr Keith Dear: That is my contention, yes.

Lord Clement-Jones: That is exactly the question that I am asking, Mr Black.

James Black: I agree that the processes and structures are in place, as Keith has said, both the legal structures at the international level and, within defence, the approaches to signing off new weapons systems, AI-related or not, as appropriate and falling within the UK's legal, ethical and policy obligations. Of course, it falls within the structures, as mentioned, of training and then evaluating human soldiers or commanders and saying that they are fit to deploy.

As was mentioned, there are these broader questions about how safe and secure is safe and secure enough. That is really a societal debate at the moment. It is not specific to the military, although, clearly, there are inherent features of military conflict that make it the pointy end of that debate.

If we look at driverless vehicles in broader society as an example, there has already been a real debate about how good a driverless vehicle needs to be. Does it need to be as good as your average human driver or does it need to be held to some sort of higher standard? That is not an objective measure but a political and cultural choice, and some of this similarly applies in a military context. As Keith has said, over time, we might see broader societal levels of comfort change around risk-taking with a human decision-maker versus risk-taking with a machine decision-maker.

I would raise a couple of other points. In many ways, this is forcing into the open some grey areas that have existed within our legal and ethical systems for ages, but, due the black box nature of the human mind, we have said that we are willing at a certain point, organisationally or as a nation, to take a risk that a human, if deployed in a certain situation, will respond in a reasonable way. When we start to reduce that to ones and zeros, as a metaphor, we force ourselves to come to much sharper, clearer and more binary—and I use that word pointedly—choices.

Building on what has been said about wider problems and external interference, it is not just about the inherent features of the technical system you are talking about and the external interference; it is also how that technical system fits within the wider defence enterprise. Are there ways in which introducing a certain AI-enabled capability might change the organisational culture or incentives within defence, for example, in unanticipated ways, or are there broader dependencies? That is probably another angle that I would add that we have to consider: it is that integration challenge.

Q26 **Lord Hamilton of Epsom:** Dr Dear, I am just wondering whether, before we bring these new systems into the inventory, they can pass some test that they would do the job much more efficiently than the stuff we already have. The worry is that, if this gets spun wrong, you could see the whole thing being completely rubbished as being extremely dangerous and everybody saying, "Well, let's not have it at all". We would then be extraordinarily disadvantaged if we ever ended up in war against somebody who had this equipment and we did not.

Dr Keith Dear: The baselining exercise that I talked about is perfectly possible as a form of validation and verification, whereby we can test how a human performs in a simulator, such as how often they make an error in either direction, and how the machine performs. We can test all of that and we can begin to do it in the real world.

The last part is more of a political question on what the likely response of the public would be to an error, and we are probably right to imagine that there would be a more extreme reaction to a machine error than a human one. I am not too sure how much defence and government can do to mitigate that. That wider political question is much more challenging than the technical one.

Lord Browne of Ladyton: Both Dr Dear and Mr Black have raised this issue. The public tolerance for this capability being deployed in their interests has to enter this through the public policy process—people such as me. I am all too familiar, from my experience, with the criticism, "Had you done this in a war-fighting situation, my son would have lived", and discussions about technology. We have all lived through the last two decades of that being at the forefront of the public mind.

Where is the current ongoing consultation with the public taking place? Are we going to do this retrospectively? Yes, okay, I feared that, so we should consider that. It is important that this consultation with the public takes place as we move along this environment, so that we are clear that those who take responsibility for this at the end of the day have the public with them.

Dr Keith Dear: That is about leading the debate and not following public opinion and reacting.

Lord Clement-Jones: Professor Payne, you talked about GPT-4 earlier. Do you agree with what you have heard from the other two witnesses?

Professor Kenneth Payne: It is interesting that you frame the question like that. There will be another GPT along shortly, and that points to the speed of the discussion. When I was here last time, I said that I find that a lot of these discussions are anchored in the present, rather than even six months or a year out. The technology is moving so quickly that that is an issue.

Just reflecting on those two answers to your question, my suspicion is—and I think other people here share it—that the public will hold machines to a higher standard when it comes to targeting discrimination, proportionality and so on. That certainly seems to be the evidence in other arenas, for example robotic surgery or autonomous cars. The problem is that pushing against that is a security dilemma, and that might shape the need to take risk.

Lord Clement-Jones: That is interesting. That then distinguishes your answer, because you think that there are greater security issues in those circumstances.

Professor Kenneth Payne: I do, and I suspect that that will lead to less exacting standards than one might find in the regulation of autonomous transport or healthcare.

As another aspect of this, it is not just that machines will be replacing what human operators or soldiers do, but that they will be doing qualitatively different things. There is no human equivalent of a 10,000-strong aerial swarm or of a submersible shoal that can stay at sea indefinitely, so the technology permits a qualitative shift.

As a last reflection on your question, it is not simply a matter of benchmarking against international humanitarian law. International humanitarian law captures some tensions between fundamentally different philosophical logics. We often fudge the issue between a deontological logic that says everybody has a right to life and a consequentialist one that says we sometimes have to weigh numbers of deaths in the balance. How do you code those tensions to a machine? I do not think that there is a straightforward way that you can capture a British ethics of warfare and upload it to autonomous machines.

Courtney Bowman: In terms of the broader societal implications, it is worth considering how a movement towards machine-facilitated or machine-augmented decision-making is shifting the balance from a procedurally oriented notion that is consistent with the just war tradition towards something more focused on outcomes, or numerical qualifications or evaluations of how a machine has performed, versus the process of working through distinction and proportionality considerations.

Where that leads us is a question that we need to decide as a society, but, to go back into the weeds of the discussion and answer an earlier part of your question around safeguards for these new capabilities, there is a lot of space for innovation that needs to be built out in thinking about systems of systems and how different components of artificial intelligence and autonomous weapon capabilities fit into a broader construct of carrying out a chain of activity.

This opens up new opportunities for design parameters that can reinforce principles such as failing safely, so figuring out the appropriate moment for a human to step into the loop to provide the appropriate level of insight and accountability. This is a point of technological innovation that is often glossed over when we go too high in the discussion, so it is worth getting into the specifics and understanding what we are talking about. If we are operating at this ethereal level, we will lose sight of some of the critical considerations that technologists such as I are well adept to address.

Lord Clement-Jones: Do you have a different risk appetite, though, between human and AI systems, so to speak?

Courtney Bowman: My sense is that both still have their place. As promising as new technologies suggest they are, they need to prove themselves in the field, so there is much work to be done in fielding the

technologies and testing them in near-live situations to be able to prove that, for example, LLMs are capable of delivering on the promises that they seem to provide in a laboratory setting.

Before we go too far down the road in making assumptions about what these new technologies can do, we are still bound to a certain set of normative, value-oriented assumptions about how we should operate and how we should pin our hopes for these technologies to the existing structures of international and national law.

Q27 **Lord Grocott:** I want to pick up the question that you began to ask, Chair. Mr Black talked about it being a much more complicated world now, with lots more participants, if you like, in these kinds of issues. Where are we as far as non-state actors are concerned? I still do not necessarily have the systems that we are dealing with very clearly in my own mind. How close might we be to non-state actors getting hold of and developing these kinds of systems and being a threat to everyone else via them?

James Black: Maybe I will talk in general terms, and then some of the technologists can be more specific. Generally speaking, the locus of innovation within defence, specifically in this area, has shifted away from the public to the private sector. When we talk about non-state actors, it conjures images of violent extremist organisations, but it should include large multinational corporations, which are the types of organisations that are at the forefront of developing this technology. This is not something like early computing or the jet engine, which came out of government-funded labs. The private sector is leading the way here, not in every area, but in a lot of areas, and is, therefore, shaping the debate about tech governance as well as the capabilities that are available. The starting point is that non-state actors are controlling this.

Of course, it is different if we start talking about hostile non-state actors, be they private security companies or violent extremist organisations. We are starting to see the ability of these organisations to use robotics; we have seen ISIS using commercially available drones either for reconnaissance and targeting reasons, or as crude improvised explosive devices or delivery devices for them.

Moving forward, a lot of this stuff is going to be difficult to control from a counter and non-proliferation perspective, due to its inherent software-based nature. A lot of our export controls and counter-proliferation or non-proliferation regimes that exist are focused on old-school, traditional hardware such as missiles, engines or nuclear materials. This sort of thing is a different proposition and, clearly, a challenge.

I add one further challenge as someone interested in deterrence theory, to the question earlier about escalation. In general, we do not have particularly good theory for understanding how to deter non-state actors. All the deterrence theory has evolved out of Cold War nuclear deterrence—the USSR, the USA and the West. It is not really configured in the same way to think about non-state actors, particularly those that

have very decentralised, loose, non-hierarchical networked command structures, which do not lend themselves to influencing in the same way as a traditional top-down military adversary. Perhaps others can talk more on some of the latest developments.

Dr Keith Dear: I completely agree with James's final point on how we deter non-state actors by imposing or threatening to impose costs on them. Deterrence in conventional conflict is about what cost you might impose. "What weapons do you have that you could respond with? How would you make us hurt if we hurt you?" That is all true, and I strongly agree that we do not have a good way of deterring violent non-state actors in that domain.

Therefore, deterrence by denial tends to be the primary strategy. We now have barriers to stop cars coming up on to the pavement and driving into us as we walk down Whitehall. When we begin to think in those terms, we also need to be careful of not overstating the risks of violent extremists getting their hands on these weapons.

The mental model for a lot of us is the "Slaughterbots" video from Stuart Russell and Human Rights Watch, which I expect most of you have seen. That is propagandistic and overblown, and it is important to state why. If you think of that in terms of deterrence by denial, first of all, you would have things like jamming for any of those low-cost drones. Yes, it is autonomous, but it still has to fly. Then you have the potential to fry the circuits within it, to jam the GPS and other sensors or to have all sorts of camouflage and different systems that would stop that. As Paul Scharre says, chicken wire or a net would probably be enough to stop the drones in those videos.

When you think about how we do it right now, deterrence by denial includes things like air intercept and layered air defences. All of those things are possible at various scales to defend against the kinds of threats that are posed in that "Slaughterbots" video. It is very difficult for non-state actors to get their hands on high-end capabilities, whether that be cruise missiles or chemical weapons, which is why we do not see them using them. It is similar in this domain. While we should be concerned, we should also be wary of overstating the threat.

Lord Fairfax of Cameron: I do not know whether this question will be ruled out of order by the Chair, but I am going to ask it anyway.

The Chair: It will be too late then.

Q28 **Lord Fairfax of Cameron:** I am trying to link it to something that Professor Payne said about the speed of development in this area at the moment, which seems to be moving incredibly quickly each week, and the fact that all this stuff is dual use, military and civil. What do each of you think about the recent letter signed by 1,000 scientists asking for a six-month minimum pause?

Professor Kenneth Payne: They made their point and garnered some headlines. It is, of course, wholly unrealistic with an activity that is so

heterogeneous. We cannot even agree on a definition of what constitutes artificial intelligence research, so it was never going to be more than the headlines in the day's papers.

It is a reflection of the degree of societal unease about the rapid pace of change that people feel is coming down the tracks towards them, and perhaps also a reflection of Lord Browne's comments earlier about the extent to which they do not have a voice in shaping the way that our society responds to these technological developments. It captures that quite nicely, but I do not think that it is a proposition. I would just mention again the security dilemma that I referred to earlier. It just means that it will not happen.

The Chair: In an earlier evidence session, we were urged to embrace slow AI, and I do not think that I am misrepresenting the committee in saying that we were generally quite sceptical about the idea.

Q29 **Lord Browne of Ladyton:** This is a crucial point. We just have to accept that we will never get in front of this technology, so we are always going to be trying to catch up. If our consistent experience of public policy development is sustained, which it will be, that will go at the speed of light and we will go at the speed of a tortoise. That is the world we are living in.

Professor Kenneth Payne: Instinctively, I am reluctant to say that that is the case.

Lord Browne of Ladyton: It is what is being implied by it.

Professor Kenneth Payne: Yes, and I am loath to agree with an argument that an academic would sum up by saying, "That's technological determinism; you're ignoring all sorts of institutional and cultural factors that go into shaping how individual societies develop their AI", but it is certainly going to be challenging. I do not think that the existing institutional arrangements are adequate for those sorts of discussions to take place, notwithstanding the work of committees such as this.

Lord Browne of Ladyton: That implies that, from the point of view of regulating and being in control of this, we should bring it in house and into government control. The Government can sit beside it and understand it.

Professor Kenneth Payne: You see something similar to that in Italy's response to GPT-4. Unless I am mistaken, they have banned it, but I am not sure that that is tenable.

Lord Browne of Ladyton: It was with nuclear weapons.

Professor Kenneth Payne: You need rare metals, big industrial plants, centrifuges and so forth. You need a VPN to get around Italy's ban on language models.

Dr Keith Dear: On the first point, nothing would slow down AI progress more quickly than bringing it into government, and it would be hugely to our detriment. If you think as I do—and this was a guiding principle in writing the integrated review—automation, robotics and AI will completely transform our economy, for better and for worse. The idea that we might slow down progress in the UK would see our relative decline in the international system being precipitate, and one reason for that is that advances in artificial intelligence build on each other, so you get a flywheel of constant progress. You cannot afford to fall behind.

On the other hand, there are some really unpalatable questions being asked here. I agree with Ken that that letter was for the birds; nobody was ever going to act on it and, indeed, it could not have been acted upon. On the other hand, the big issue is whether we can learn from failure in artificial intelligence. If we can learn from failure and can have a catastrophic success as we potentially approach artificial general intelligence or artificial super intelligence, that is the moment when the world might find the political will to act collectively to get in front of the technology.

If, as Max Tegmark and other MIT physicists have argued, that is not possible, and what you get when you get to artificial super intelligence is a runaway, we have a real concern. I thought this coming into today's discussion on lethal autonomous weapons systems. An analogy that Tegmark uses, which is really illuminating, is that the reason why the tiger is in the cage is not that it is not stronger than we are or does not have has better weapons than we do, but because it is not as smart as we are.

For me, in all of this debate, the thing that we should be most afraid of is not the bombs but the intelligence. How we get out in front of that depends completely on whether there is the ability to fail and learn, then impose restrictions and get out in front of the debate. Without that, the people writing that letter were right to be concerned, but wrong in their proposed remedy.

Lord Clement-Jones: We have gone pretty broad here, Chair. I sense, therefore, that there is not a denial that it is possible to apply some form of risk-based regulation.

James Black: As has been alluded to, fundamentally, we are trying to shift from industrial-age government and governance structures to information-age ones. We talk about AI and integrating it as a sociotechnical system and it is very easy to focus on the technical side of that equation. There are lots of very smart people working on developing AI, so I am not dismissing it as an easy thing to do, but I would that argue the "socio" side of the sociotechnical system is the hardest bit. How do you, as has been mentioned, gather the political will to make decisions collectively, at a multinational level, quickly enough to implement updated regulations that reflect the latest developments, et cetera?

Then how do you then work much more closely between government and industry? As was said, you cannot simply put the genie back in the bottle

and bring it into government, so you are going to have to find a new way of working with industry that is much more collaborative and less transactional than what we are used to in other areas of defence procurement. How do you do all those things quickly enough to keep pace with the sorts of developments that we have been talking about? That is the real challenge.

We would need to come up with not just a new type of regulation, but an entirely new model of regulation that would update much more quickly, solve all those collaboration and co-ordination challenges, and be far more informed not just by risk but by uncertainty. When we talk about risk, it is very easy to think that we can quantify the probability or the impact of some of these things, but we often do not know what we do not know. We are talking about uncertainty, but we do not know how likely it is that we will have catastrophic events or how big an impact they would have.

What we need instead is regulation that is robust across as broad a range of bad futures as possible, while shaping us towards as many positive futures as possible, and that is a very big change in our approach to thinking about technology governance.

Q30 **Lord Sarfraz:** I declare an interest as a member of the advisory board of C3.ai and as a venture capital investor. Mr Bowman, these weapons systems will be trained on a large number of datasets and there are certain sorts of datasets—satellite data, weather data, sensor data, et cetera—that will be accurate and unbiased, but where is the risk for inaccurate and biased datasets?

Courtney Bowman: As a starting point, it is worth acknowledging that no data is unbiased. All data comes from a certain set of sensor capabilities that have some inherent focuses or implications that cannot be avoided because they provide a particular perspective or because they look at the world through a particular part of the visual or electromagnetic spectrum. They are always going to reflect some perspective on the world, and that is a fundamental bias.

The question becomes how you build that information into a system and broader capacity in a way that reflects an understanding of that bias and of the trade-offs that are implied in evaluating that bias. For example, a lot of the discourse and literature focuses on the questions of fairness in algorithmic design. Fairness is a concept that we think about in the world in looser terms, but we attempt to translate it into quantifiable metrics. Those metrics may not all be things that we can optimise against collectively, and we have to make choices. When we make those choices, we are making socially determined choices that need to be part of a public discourse.

The same thing applies in the application of information for the development of these technologies, but one of the broader questions that comes into play in the landscape of information that is used to build, train and deploy these types of systems is whether we are creating the full

infrastructure to support the pipeline that needs to be in place. It is not just a matter of collecting data, but a matter of being able to validate the quality of the data, and that is a procedural set of considerations. It goes to the point of collection and to the point of testing and validation, but also to a point that figures into the last part of the conversation, which is the prospect for regulation. How do we think about model maintenance over time? How do we make sure that these capabilities have enduring value and are persistent against these inherent challenges of what we call brittleness? These are real issues that are underaddressed and underdeveloped in some of the thinking around regulation.

Lord Sarfraz: Is there a mechanism in place right now to audit data, or should there be?

Professor Kenneth Payne: I do not know whether there is. Should there be? I am sceptical, because of some of the points that were just raised in that last answer. There is no single definition of fairness, of what sort of data you should be collecting and using or of what variables you should be measuring. Coming to agreement on those issues is problematic, and I refer back to the answer I gave earlier about the tensions inherent in just war theory. Those are questions about fairness as well. We reserve our right to change our minds about what values we ascribe to things.

As a second part to the answer, war is often the search for novelty, surprise and advantage, so it would be surprising indeed if you could anticipate up front what sort of data you might need and how robustly you were going to go about testing it, if the advantage comes from finding new ways, new data and new information that you can use in an asymmetric struggle.

To the question of whether you could ever have adequate data, the answer is probably no. I note that large language models are trained on almost the entire corpus of online prose in the English language and in computer code as well. They are running out of data on which to train large language models, so there is always more data to be had than you might think.

I am meandering slightly and giving a slightly unhelpful answer, but I suspect that we will struggle to cohere around what data we should collect, what data is useful and how we go about ensuring that it is fair when we cannot agree on a definition of fairness.

Lord Sarfraz: Are there any sorts of datasets that we should be particularly concerned about when it comes to these sorts of weapons systems specifically?

James Black: Again, you can ask whether this should be looked at through a consequentialist lens or other lenses, but the one that is obviously the source of a lot of controversy and is starting to cut through to the public discourse is the use of personal data or datasets about human and societal questions rather than physics-based questions. There

are obvious sensitivities around discrimination on race, gender, profession and all sorts of things, where biases have been shown to be baked into earlier versions of various AI systems.

Again, it is all context specific. We have talked about different comfort levels with risk in general, and the same thing applies with data. We were having a discussion earlier about malfunctions and misinterpretations of input data from sensors. Perhaps, in the future, we could expect greater comfort about employing AI or autonomous weapons systems in relatively uncluttered, relatively straightforward environments.

I used the word "relatively" there. An example would be a maritime engagement between a military vessel and an incoming missile. There is relatively limited scope for problematic biases from a political controversy perspective in the way that there would be if you were talking about deploying autonomous weapons systems in a crowded megacity with lots of civilians around, and they had to look at people's faces and determine whether they are a threat. Again, it becomes very context specific and comes down to a question of risk appetite.

It is not just about the inherent limitations of the dataset or of the availability of data, and then the limitations of the AI or whatever approach is being used to process and make sense of that data. It is also, of course, the transparency and explainability of how it has come to whatever decision it has come to, and then it is our level of assurance that that dataset, even if it is unbiased, is secure. We might have an ethereal, hypothetical, perfectly unbiased dataset, but, if our adversaries can poison that data and interfere with it in various ways, that is going to present other challenges.

It becomes a very difficult question, and this is why it is right that the defence AI strategy focuses so much on the data side of things and the transformation that needs to happen within defence from that perspective, which is a broader problem than just AI and is a broader challenge for government.

Q31 **Lord Hamilton of Epsom:** You have been talking about regulation. Do you have any faith in regulation? What we are talking about here is very fast-developing technology. Almost by the nature of it, regulation would be miles behind the curve and you would never catch up, so why does regulation have any relevance here at all?

Professor Kenneth Payne: A lot of the attention when it comes to discussing autonomous weapons systems is focused on arms control. I mentioned to you last time that I am a slight arms control sceptic when it comes to AI, for the reason that the technology is changing quickly and it is hard to agree what constitutes AI and what does not. It is used throughout a broad swathe of military activities, some of which confer fighting power, because they deliver lethal force, and some of which do so tangentially, for example sorting the pensions. That makes a contribution to military efficiency, but it is not exactly at the front line, so which bit are you regulating?

Then there is another series of questions. If you have arms control, you

need to have processes of validation and monitoring, but the signature for developing AI is quite small; you do not need those uranium enrichment facilities. A lot of it is dual use. You are talking about warehouses with computers and scientists. How can you monitor potential defection from any arms control regime? Lastly, there is a huge incentive to cheat on any regulation, if you believe that these technologies confer profound military advantage.

It is another topic that makes me uncomfortable, because I find myself arguing against what I want to argue, but I am slightly sceptical about the prospects for regulation.

Q32 **Lord Houghton of Richmond:** The generic question set that two or three of us from the committee want to come in on is in relation to what you understand by the terms of context-appropriate human involvement and meaningful human control in relation to AI weapons systems and how they might be translated into operational reality. In a way, we could challenge the premise of that particular question, because it comes from the context of human involvement being a good thing and autonomous weapons or the machine being a bad thing. You have already hinted that that might not always be the case.

In the context of certainly UK operations and the authorisation of lethal force, they operate within a set of delegations, all the way from royal prerogative through Ministers down to battlefield commanders, through such things as targeting directives and rules of engagement. Those delegations sit with an understanding of the need to be compliant with international humanitarian law, the law of armed conflict and all those things.

But the delegation is implicitly one of legal authority to use lethal force, but also the responsibility to use judgment in its application. This is where the problem with machines comes in, in most people's mindsets, because you cannot delegate authority or judgment to a machine in the way that we normally understand it.

If my own logic follows here, you are not exercising judgment or delegation in any of this. You are authorising autonomy. You are retaining a legal responsibility, but you are exercising it in the knowledge of the capabilities, risks and context in which you are granting that autonomy.

If you have followed me thus far, what we need from the panel is an analysis of the risk factors in play at the point at which the human authorisation of autonomous action is initiated. You have already mentioned that there are relative competencies at play between a human and a machine, where the machine may be more competent. That is brilliant, thank you, but I would like someone to expand on that.

You have mentioned the ability to build in machine fail-safes, or how a machine can fail safely, so that there is a sort of override that will soft-land it before these rather more dramatic scenarios. In terms of proportionality, discrimination and those sorts of things, the machine can be more helpful or the better option, because it then would be a matter of training the human to do the best conceivable delegations. Some of

the weapons systems, rather than these delegations being quite low, might be held at quite high level, particularly if you are on to thousands of swarming independent bits of machines.

In the analysis of the risk factors in play at the point of human authorisation of autonomous actions, what ones are we missing so far, or what else would you stress as being the major ones that are in play?

Courtney Bowman: There is a lot to unpack in the last part of your question, so I apologise if I do not hit all of it or certainly most of it. It is worth going back to a principled evaluation here and understanding what humans and machines are capable of and not capable of.

In terms of distinction and proportionality considerations, there are fundamental limitations in what machines, in their current and foreseeable forms, are able to do. They might be very good at making quantitative assessments over situations, providing a parameterised view of the battlefield and making recommendations, but, when it comes to making value judgments that build in procedural and legal considerations, those are functions of moral human agents that we still have not figured out any way to understand in our own persons, let alone programme into systems.

To go back to the crux of your question, around what other risk factors we need to be building into these systems, you touched upon something very important, which is training the human operators and individuals who are interacting with these systems. This goes back to an earlier point around fail-safes and user interfaces in the systems.

Building the capabilities so that individuals who are interacting with the recommendations or decisions promoted by a system can put those in context and know when to step in and exercise their control over that process is critical. If the end-users or operators do not understand what those systems limitations are, and that there are risks and confidence scores that apply to certain evaluations that may not be well understood or well surfaced, they are ultimately going to be challenged in their capacity to make sound decisions against those recommendations. One of the core considerations in assigning a risk profile for the use of these systems is the necessary level of training an operator has to have in order to step up and interact with the system.

Professor Kenneth Payne: This is another meaty issue and a meaty question. The basis of mission command, as carried out by the British and other militaries, is that you delegate responsibility to the lowest possible level, on the understanding that that lowest possible level is thinking something like you or that you can at least anticipate that they are broadly thinking along the same lines. It helps that they are human and that they have been encultured in the ways that the Armed Forces work. The Armed Forces put a lot of effort into inculcating that culture and that understanding of intent in mission command. Of course, that does not really work as well when you are talking about machines.

On the other hand, this term "meaningful human control", which has

come into the discourse really only in the last year and a half or so, is the last trench-line that is being defended in the long retreat from decision-making in the loop, to on the loop, to meaningful human control. The technology has driven that retreat from one position to the next. That is where we are now.

With meaningful human control, it is the boundary cases that are hard. You can quite clearly say that there is going to be scope for human control in the launch of strategic nuclear weapons, because there is still some time—20 minutes or so for an ICBM in flight—and the consequences of that decision-making are vast. The two experts are looking at each other and going, "Where did you get that from?" It is relatively straightforward to say, "Well, we want meaningful human control in this".

At the other end of the spectrum, you cannot see meaningful human control where speed is of the essence and it is a tactical decision defending a ship from incoming missiles with a system such as Phalanx. It is the huge swathe of territory in between those two decisions where you ask, "Where does meaningful human control come into it? Where are you making that judgment about how much you're prepared to outsource to the machines?" The technology will drive some of that with qualitatively different changes—shoals, swarms and what have you—and you will not simply be able to control each decision taken by each node in that network.

There are risks you are accepting with that, such as automation bias, where people just accept the judgment of the machine without exercising enough critical faculty, and training is an important part of mitigating against that sort of risk. You are also accepting risks about justice for when things go wrong. You can have perhaps punitive justice as a result of decision-making going wrong. Perhaps there is a responsibility at a corporate level or at a government level for developing, purchasing and fielding these bits of equipment, but what about restorative justice? How do you have restorative justice if a machine system commits a war crime? That is a risk that you have to accept if you are going to use these systems for anything other than close-in defence.

James Black: I will try to give as orthogonal an answer as possible, because a lot of really good points have already been raised. In your question, you mentioned that we have had some discussion around the different competence levels of human versus machine decision-makers. Another risk factor to consider on that perspective is what the possible cognitive impairments might be. We have already talked about some of those categories on the machine side, such external interference and the inherent limitations or malfunctions that are a risk with machine agents.

Of course, there are all sorts of potential impairments and problems on the human side. Humans get fatigued. They get angry. They get traumatised on the battlefield. They commit war crimes and rape, in very small numbers, of course, but these are things that humans can do, so we have to recognise that there are ways in which humans can be compromised actors. There are certain areas in which artificial

intelligence and autonomous weapons systems provide opportunities and perhaps avoid some of those problems.

Clearly, though, as we have discussed, a lot of it comes back to the complexity of the task that is being undertaken, the environment in which it is occurring, and then the scope, risk and impact associated with that task happening with a machine and going wrong, that task happening with a machine and going right, or that task being undertaken by a human or some alternative capability.

You start getting into quite specific analogies. We have talked already about defending a ship with Phalanx or a nuclear encounter. If you go to a really rudimentary level and look at a landmine, no human is making the decision there. In a very rudimentary sense, something has stood on it or a tank has driven over it and it has gone bang. Within the constraints that exist on landmines legally already, within international law, that is something that we are broadly ethically content with in certain bounded situations. The scope for that machine to move is zero. The scope for it to make decisions outside of asking, "Has something stood on me and should I explode?" is zero. Therefore we are comfortable with it.

It is finding the grey areas between that example and the nuclear example that we have to deal with. The challenge therefore is also considering what the alternative capabilities are to deliver a similar effect. If we are using an autonomous weapons system, the alternative to defending a ship from a hypersonic missile situation is that the ship gets hit. There is not an alternative option there, perhaps other than not deploying the ship in the first place and inhibiting our own freedom of action. For other areas, it may be that we decide that we ethically recognise that there are significant risks with certain capabilities, but we can find alternative ways of delivering similar effects through other means. We are therefore likely to invest in those.

The final point to sum up my position on this would be that, in many ways, the biggest risk is our lack of understanding of risk. As you say, we are delegating a lot of authority. The people we are asking to make those decisions, inherently, have been promoted into positions not based on their understanding of AI but rather based on their historical ability to fly an aircraft, pilot a ship, storm a hill et cetera, and then to exhibit leadership potential and therefore move up in the military rank structures and so on. We need to continue investing in the AI literacy and understanding of our leaders at all levels, from junior to senior. It is then that you can make the most informed decisions about which risks to take.

Dr Keith Dear: At the point of delegation, there are some questions that are new. You should ask what tools have been applied. There are tools for checking for data bias. They are procedural, but of course, when you clean and transform data, there are inevitably, on occasions, values built into those decisions of what gets discarded. You ought to ask whether those things have been applied.

My colleague, Darminder Ghataoura, works on a thing called "Faircheck", which is a data quality assessment. It aims to make sure that your data is representative. I think that IBM has a similar tool, AI Fairness 360. There is an open-source tool called FairLens. There are tools that can help. They are not a substitute for judgment, at this stage, but can help you understand what you are using as input data to train your system.

You should be asking at the point of delegation today, "Have we used all the available tools? Have we applied adversarial AI tools to make sure that this particular capability degrades gracefully or flags when somebody might be trying to influence it to do things we did not want to do?" You would currently train a human operator of a system to do the same thing. For example, if you were fighting an adversary that began to place anti-aircraft weapons on top of hospitals, now you might say, "Hospitals are always off the target list, but we now need to ask some deeper questions", because you are being deliberately misled. You need to begin to train your system to understand the various ways in which it might be manipulated. They are new.

The only possible potential disagreement with other panellists is that I do not think most of the time we do trust when we delegate. Mostly, we train, parametrise, validate and verify. Before you delegate authority to your subordinate, you make sure that that person has had the right training. Do they understand the rules of engagement? Have they had sufficient experience of operating the weapons system that they operate under particular circumstances? Do we know how they respond under pressure in comparison to peers? The ability of an individual to do that determines where they end up and the level of delegated authority that they get.

Then we set parameters that we ask them to operate within. Then usually we validate and verify that, nine times out of 10, they operate within those parameters. Of course, sometimes they do not, after you deploy them. Then we notice that the outcomes we are getting are not the outcomes that we wanted, and we either go back and adjust the training or take that particular individual offline. It is broadly the same for machines and a lot of this is overblown.

We used the word "discrimination" today and it comes freighted with, in essence, negative connotations for obvious reasons in the modern environment. Of course, as a targeteer, discrimination is exactly what we want; we want systems that can be as discriminating as possible. You live in a world where there are on average 5,200 gigabytes of data on every person on the planet. Really crudely, if you equate that to Kindle books, it is tens of thousands; I think it was 80,000 last time I did the calculation. That gives you unprecedented insight and foresight, so you are going to need AI to be as discriminate as possible.

Similarly, when we think about DNA data, psychometric data, behavioural data, quite often the most catastrophic ethical incidents in combat are caused by information overload, where we cannot process all the information because there is too much of it, so we miss some crucial fact.

You should be asking at the point of delegation, "Is AI going to enable me to make a more effective decision than I would have been able to make otherwise?" There are some technical tools, there is the application of existing frameworks and there are probably some new questions you should ask before you delegate.

Q33 **Lord Clement-Jones:** We have described circumstances where there is no meaningful human control in certain autonomous weapons already, yet we have talked about meaningful human control in certain circumstances. Are we ever going to get to a position where we can have almost a ready reckoner to see where meaningful human control is appropriate and is not appropriate? Are we moving along a spectrum already? I think Professor Payne mentioned that we have already moved from human in the loop to meaningful human control, so we are already moving away from the degree of human involvement. How are we going to define where it is appropriate and where it is not?

Dr Kenneth Payne: It is going to be really difficult, because you reserve the right to change your mind in conflict. What you think in the calm of Westminster today might not be what you think when the odds are dramatically different in combat, so you can change your mind.

For me, there has been slight phoney-war feeling to this discussion for the last couple of years. There is a disjuncture between what I am describing with thousand-strong aerial swarms and the reality of a small RAF experimental squadron firing its first missile from a drone. There is a feeling of a gap between the rhetoric or the intellectual discussion and the practicality on the ground. It is much easier to see the radical nature of AI when you look at the basic research that companies such as OpenAI and DeepMind are doing than it is when you look at what is on the inventory of the RAF or the Army today.

That is why it is crucial to have those sorts of discussions, because the technology is inevitably coming down the track. We have a brief window while we can still have that discussion, but I am slightly sceptical about the idea of a ready reckoner.

Q34 **Lord Fairfax of Cameron:** Can I quickly ask about a very specific example? Reference is quite often made to the Harpy loitering munition in this regard. Do any of you think that that is in a grey area, or is it totally in accordance with current law of armed conflict, international humanitarian law and so on? Is that example problematic?

Dr Kenneth Payne: A couple of years ago there was a report about the use of a free autonomous drone for the first time to kill somebody in Libya. It attracted a lot of headlines—perhaps you saw it—and no end of follow-up about it. Was it actually fully autonomous and under what circumstance? What sort of drone was it? Who was using it? Who had delegated authority?

I mention that because it points to a real problem with autonomous weapons systems: from the outside, as third-party observers, it is really hard to know what the code is permitting, what its rules of engagement

are and even what its technical capabilities are. Harpy is a grey area in the sense of how far it can conduct fully autonomous missions.

Q35 **Baroness Hodgson of Abinger:** I would like to quickly come back to this security of these systems, in this context. We are talking from the point of view of people utilising them who have human empathy and are trying to conform to international humanitarian law. Today, we have people operating in this world who have medieval concepts of right and wrong. One looks at ISIS and some of the non-state actors, who have been extraordinarily effective in the face of today's very modern weapons systems.

When you think that this is software-based, there has to be a risk at some point that they will think that it is worth getting hold of this. With corruption and all sorts of things, there has to be that possibility. Do you build into any of these systems the ability to disable them completely?

James Black: As we discussed earlier in relation to counter and non-proliferation, it is much easier in regard to the hardware elements than the software elements, for obvious reasons. Physical stuff has to physically move round the world, cross borders, undergo legal checks and all these sorts of things; on the software side, it is much harder.

As has just been discussed, even once you have a deployed system, it is very hard to actually understand what rules of engagement it is operating under. What are its policy, legal and ethical constraints? Those of course can be changed on the fly. Even if we understand them at one point in time, an adversary might adopt rules of engagement that we are comfortable with and would view as fair, but then perhaps their perception of the threat or their value system changes, there is a regime change or whatever. They flick a switch and of course they can change that. There is an inherent challenge there from that black box angle that the professor was just talking about. That gets us into that discussion around the security dilemma and this race to the bottom on standards.

What is important therefore is not the UK seeking to join that race to the bottom on ethical standards. It is a multipronged approach of the UK seeking to shape the broader global debate on governance of these systems, which does not eliminate the risk that non-state actors can defer from whatever normative and legal frameworks are agreed, but puts some safeguards in place by limiting their access to working with certain companies, buying certain products or services or working with certain states.

That is something that is in the defence AI strategy. It is a big ambition for the UK to try to meaningfully influence those sorts of debates. It also speaks to the need for us to develop asymmetric responses, so that, when an adversary is employing AI-powered systems or autonomous weapons systems in a way that we ourselves are not ethically comfortable in doing, it does not mean that we have to respond with tit for tat and lower our own ethical standards; it means that, instead, we have alternative capabilities that we can use to counter that.

An example is the swarms that we have talked about already. If we came to a position that, ethically, we were not comfortable with using swarms, for example, we would really need to invest in counter-swarm technology such that others could not gain an advantage over us. It is a very complex issue and it speaks to some of the challenges that we had earlier about how you deter, influence and degrade non-state actors. We need to find those ways of learning to live with that as tolerably as possible, rather than thinking that we can eliminate that risk. As you say, we certainly cannot reduce it to zero.

The Chair: This may segue very neatly into the next area we would like to look at, which is specifically the role of the private sector.

Q36 **The Lord Bishop of Coventry:** Thank you, panellists. It is an extraordinarily interesting engagement. James, you referred to what is obviously underlying the discussion of the way that the locus of development has very much shifted to the private sector here. That clearly brings down a massive weight of technological and moral responsibility on the private sector.

I was very interested in a comment in an MoD policy statement that said that working with the private sector is a rather good moral check on what we are doing, because it will not want to work with us if we are doing things of which it does not approve. I would be very interested to know whether you agree with that assessment and whether there is that potential for moral check.

As a bit of a sideline to that, I am fascinated by your reference to the human mind as a black box. I wondered what sort of formation, if I may put it like this, the black box of the technologists' mind is. We are receiving evidence at the moment and an interesting comment that made me think was when someone asked, "What about a digital Hippocratic oath?" I wondered what you thought about that, but, more generally, whether you agree with that MoD assessment.

Then I have a second question. There has been scepticism over whether we need new international frameworks or even whether regulations work. I am interested in whether you think there is a role, and evidence of a role, of the private sector in shaping moral norms, as it were. I suppose that that follows on from the first question. Is there a shaping of moral norms? I think that we have seen evidence of that in your responses today.

A bit of topicality that relates to all this is the article yesterday or the day before of the Google chief executive officer saying, "We don't quite know what's going on here even in our own systems and we need some sorts of frameworks". Is that evidence, in a sense, of the moral check?

James Black: I will pass to the private sector in a moment to outsource my answer. You raise a really good point. Certainly, we have seen in recent years some quite high-profile debates about push-back in major multinational tech companies by internal groups of employees who found out that there were Pentagon contracts or whatever that they did not

agree with, or that they were exporting to certain nations around the world that are controversial to work with.

There is clearly an inherent level of utopian idealism in some people who work in tech, because they are working to develop the tech that will change the world, improve things and so on. There are some for whom this is a very important moral consideration when they are developing things. For others, the bottom line, all the other things that exist within the private sector and the shareholder pressures continue to exert themselves.

We are seeing a shift, not just within tech but more broadly. There are debates around moving from shareholder to stakeholder capitalism and increasing interest in ESG considerations in investments, these sorts of things. There is a broader pressure on the private sector to think about these broader ethical, environmental and social dimensions to what they are doing. That applies in this area as much as elsewhere. There are positive and negative countervailing trends at play there.

The more fundamental question, though, is to what extent the state wants to accept and embrace this erosion of the Westphalian system and the monopoly that the nation state has on legitimate use of violence and force, and on setting and shaping those moral questions. There is absolutely a very important role for tech companies to play in those international discussions around setting up new frameworks and regulation, and shaping norms. It would be a very ill-informed debate if that was a purely state-centric one. It is notable that, at the UN, for example, a lot of the groups of government experts and working groups that exist on this have brought in tech companies, but also civil society, Red Cross, et cetera, in a more prominent way around these discussions of new tech than they have on other policy issues. That is a good thing, but that is a fundamental challenge to the state of how far it wants to push that. The debate we have had about whether to be led or to lead on technology is a really difficult and not straightforward one.

Whatever the UK decides in terms of its level of comfort with integrating private sector voices into governance, regulatory and normative discussions, it does not exist in a vacuum. It depends on what people such as the US, China and others do. We have to consider also how we can try to influence them and their approaches to these debates, as we do not have a monopoly on this by any great stretch of the imagination.

Dr Keith Dear: We probably ought to acquaint whoever said that in the MoD with the past 300 or 400 years of history of corporate capitalism, which would not give you much hope that the corporation was going to impose the right morals on the MoD and the deployment of AI. You should perhaps start with the East India Company and work forwards.

More recently, you might look to Milton Friedman's arguments that the moral responsibility of business is to its shareholders. For all that there is a rise of stakeholder capitalism and that businesses are beginning to take ESG seriously, they are doing so in order to better serve their

shareholders because, increasingly, consumers care about those things. It is the profit motive still.

The idea of outsourcing your morality and ethics to the private sector is probably unwise. It probably flows from this sense that, because a lot of this tech has come out of California, there is slightly woolly thinking around ethics, what matters and how the world works that informs an awful lot of decision-making in the current world leaders in AI, brought from their employees.

Demis Hassabis seems entirely well motivated in worrying about what it means if a corporation were to achieve an artificial general intelligence that could do everything from manipulate the stock market and onwards. Demis's lords and masters in Google appear slightly less concerned and would rather like it to be them, from what we can see from the outside looking in. There ought to be some real concerns about this idea that you can rely on the private sector. I hope that that answers the tech mind part of the question as well.

The Lord Bishop of Coventry: I do not mean "rely", but is there a part to play?

Dr Keith Dear: The private sector is going to have a huge role in deploying these technologies. One reason that the defence companies are not leading the debate on this at the moment is that the gap between what the MoD says it wants to do on AI and what it is actually doing on AI is so vast. You do not invest in these capabilities unless you can see the size of the prize on the other side.

If I go before an investment committee and say, "I want to build this capability because I think defence is going to buy it", they are going to ask me, "Where is the evidence that defence is going to buy this?" whatever it might be. There is not a lot of evidence to say that the MoD is. This debate continues to be led by companies that are looking at where they can make profit, which is still the private sector, not the public. There will be a role for the private sector, and an increasing one, once the MoD starts spending money on artificial intelligence and stops talking about it.

The Chair: I know that Baroness Doocey wants to raise the issue of the perception of the role of the MoD in relation to the private sector.

Dr Kenneth Payne: Keith's answer is more or less unimprovable. I wrote down here, "Do no evil", which of course was Google's motto for a period of time shortly before it stood up its ethics board; shortly after that, it stood it straight down again. I noticed the name of the leading language model company, OpenAI, which is not open about the algorithm that underpins it. I would be very wary along Keith's lines.

Notwithstanding the role for corporations to feed into an ethics discussion, that discussion needs to be as broadly constituted as possible.

It needs to be at the intergovernmental level, in this building and in society at large.

On the point of a lot of this activity being increasingly concentrated in the private sector, that is not inevitable. The story of the last decade or so in this country has been of computer science talent being hoovered out of the university sector to go and work for these corporations that have very deep profits. There are steps that government could take to address that issue. One concrete example is increasing the amount of compute that is available to researchers in a university setting. It is part of the new national computing strategy to do that. A sovereign foundation model AI capability would be a concrete step, as part of a wider project of democratising these activities again.

Courtney Bowman: Perhaps I could offer a concrete counterpoint to some of the other panellists. I agree on the broader consideration that morality should not be outsourced to corporate entities. I also agree that this is a collective societal enterprise to figure out the norms that should govern the future of war-fighting.

At the same time, I point to the fact that there are mission-oriented companies out there that are very focused and build into their core enterprises the notion of a moral responsibility. They are building effective technologies but also that effectiveness incorporates the concept of responsibility. I think that you will find that that mission orientation, in contradistinction to some of the consumer-facing entities that are coming out of California, is really focused on getting involved in defence applications with the right partners under the right circumstances. In many ways, they are trying to drive this conversation forward.

Q37 **Lord Sarfraz:** We have all seen these great videos of the Royal Marines wearing their jet packs, and flying back and forth between platforms. That was developed by an early-stage business and adopted, because it is great. In this space, it feels like you are going to have large tech companies being these enterprise solution providers for these platforms, companies such as yours and others. Is there a risk of us facing very high switching costs between these suppliers and getting locked in with large suppliers? How can we encourage the diversity of a private sector supply base?

Courtney Bowman: The emphasis on open technologies that enable transferability and provide APIs for different utilities to be plugged and played according to different environments and requirements is a critical consideration here. There is definitely space for creating these types of open standards and they are already broadly promulgated. Coming up with better ways of enforcing principles of common data standards is a useful way of ensuring that data can be ported from one system to another, alongside broad ways of minimising the risk of the vendor lock-in situation.

It is to everyone’s benefit to focus on a system that advocates for merits, so companies that want to win on their own merits should be open to

having platforms that enable some degree of transferability and allow customers to move from one system to the next. Ultimately, the systems that work best will be those that prevail regardless of any proprietary considerations in the structure of algorithms or of data.

Lord Browne of Ladyton: I do not think that the implication of the question about the role of the private sector in this area was that, as representatives of the public policy field, we were outsourcing responsibility to them. It is not unreasonable for us to ask that they should share the responsibility for this with us. This will probably very quickly lead us to the sort of response that my earlier questions got, which was that we need a meaningful partnership. There is a partnership in reality in which we learn to live with the development of these capabilities so that we understand the potential positive contribution that they can make to conflict, which may be inevitable, in order to reduce civilian casualties. That is where we should be, because the proportion of that to combatant casualties has got completely out of step. It is how we make recommendations that inform that process in the Government, not that we want to do the same.

I might say in passing that I do not agree with the idea that, if Governments control this, it will be a brake on development. Tell that to the Chinese. I am not suggesting that we control this, but we need to live with this more in the way in which we are training operators and developing.

We also need to bear in mind that, last time I looked at this—this was in the debates associated with the development of our DARPA—80% of the R&D in this space, generally, was public money. In the United States, it is that in spades. This is public money that is being spent.

How do we develop an infrastructure of the development of this so we come to have the levels of knowledge of how it works that enable us, at the end of the day, to train the operators of these systems, and those to whom we delegate the human control, to understand what they are dealing with? That is the challenge. How do we get that from the private sector, which, after all, is spending a significant proportion of our money on R&D? They are getting it indirectly as well by taking people from our public universities.

James Black: This is not a challenge unique to AI. There is a broader challenge of how public-private partnerships work in general and of industrial policy, strategy and so on. If I look at a defence-specific context, I am not saying that defence has perfect industrial policy, and acquisition and procurement approaches, because manifestly it does not, but that is a separate committee. Where it does have experience is working with traditional defence primes in more traditional, well-understood areas of hardware and, to a lesser extent, software. It has centuries of doing that, again, not necessarily doing it terribly well, but there are existing structures and processes in place and a culture built around that.

When we start talking about AI, we are not talking about the same type of relationship that you would have between, say, the MoD and BAE Systems, which is one of monopoly monopsony in some areas, say nuclear submarines. That is a very different type of market to trying to engage a multinational tech company based outside of the UK. Whatever the UK MoD could offer in terms of money to acquire something is going to be a tiny drop in the ocean of that company's revenues.

Our ability to shape certain markets is much more restricted than it is in other areas of the defence industrial base. That affects how we approach it. There are a few aspects to this. Can we be an intelligent partner and customer? Do we have the skills, knowledge and expertise within government at all levels to make informed decisions about what that industrial base looks like, who to work with, how to incentivise them, how to shape it, what kinds of products, outcomes and services we want to see et cetera?

Do we have the contacts and networks with those industries? Again, we do not have the mature, decades-long partnerships. Many of these companies have just been created in the last six or 12 months. We do not have that in the way that we do with big, well-established defence primes.

Do we have the carrots and sticks to influence their behaviours? On the carrot side, it is often money, and it is reforming our procurement processes and how we do R&D and innovation, such that, as Keith has alluded to, they can then make a business case within the private sector to invest in certain things that the UK or its allies would like them to invest in. There is also the stick. It is thinking about the regulation and the harder forms of intervention that we can also make to try to shape that.

It is good that this AI issue was addressed within the defence and security industrial strategy to an extent. It is good that the defence AI strategy signalled that this was a big area where change needed to happen. But we are certainly not at a place yet where there is a fully-formed governmental understanding about capacity to influence industry and its behaviours in this sector and how to do that most efficiently, at either a national or international level.

Dr Keith Dear: I pick up on the point on R&D budgets, because that ought to be a thing that we can check as an objective fact outside of committee. It is still true that DARPA funds, and it is intended that ARIA will fund, an awful lot of invention. The budgets of large corporates in this space completely eclipse government budgets, in both the US and the UK. That slightly changes who has influence over what gets invented and to what end.

One of the challenges for the kind of partnership that we all want, whether in industry or government, is that defence gets the market that it incentivises for. The suppliers are shaped the way they are. The market in defence in the UK and globally is shaped the way it is by the way in

which the people who are buying buy. For example, defence companies tend to run as lean as they can until there is a clear invite to tender. Because the people writing those tenders are trying to keep up with the pace of change that all of us are struggling with, inevitably, though it is going to sound like a sharper criticism than I intend, what is being procured is yesterday's technology tomorrow. As a consequence of that, you do not really get the partnership. Corporations are holding back on the investment, waiting to see, "Are you going to buy this thing?" By the time you tender for it, the private sector, which is serving other, more agile customers, has moved on at a rate of knots.

To overcome that, yes, as Courtney said, the MoD ought to make sure it includes open standards, particularly data standards and interoperability, in all its tenders. It should fundamentally reverse the way it procures and clearly set what outcome it wants to achieve. What is the thing that you are buying for? Tell industry what the size of the prize is, if you do that more effectively than the others competing for it, and then set the rules in which we are going to compete, such as open standards, interoperability data standards, maybe even some particular ethical considerations you wish to build in. Set the rules and let us compete in the game.

Without that process and a continuation of the current procurement process, you will not get the kind of partnership you want, because you are looking to maximise competition between suppliers and to drive price down to the maximum possible extent. Those two things are not irresolvable but are, necessarily, in tension.

Q38 **Baroness Anderson of Stoke-on-Trent:** This will be very short, I hope, although I could be opening a whole moral dimension here, but that is usually your territory, rather than mine. One thing that we have not really touched on, where I would like clarification, is the global versus national nature of the technology that we are talking about. You have all touched on the ethical-moral framework that potentially can be built in or whether the national mission can be built in.

However, as you have just highlighted, Dr Dear, we are going to be a very small procurer, potentially, on the national playing field. Even with our allies—we talked about California—their world view is slightly different from what we would want our world view to be. How do we make sure that, in this type of technology, we can procure in a way and the technology is developing in a way that conforms with our national mission and our morality or ethical needs in terms of military goals?

The Chair: Without confusing matters, can I ask you to reflect on that for a moment, because I know that Baroness Hodgson wants to come in on precisely that point? Let me come to Baroness Doocey now and then return to the issue of the global stage, as it were.

Q39 **Baroness Doocey:** Dr Dear, you have already touched on this. Do you believe that the MoD engages effectively with private industry, which we are told is at the very cutting edge of developing AI systems? Will its

approach to promoting innovation and streamlining procurement processes allow it to keep pace with private industry and not lag behind?

Dr Keith Dear: They are two quite different questions. On the former, engagement, yes, it does it really well, all the time, in multiple forums, as fairly as it can so as not to bias future procurement. It has private sector representation on the ethics committee, for example. You can see the list of members. I think that Professor Nick Colosimo at BAE, who I expect you have called or probably will, is one of those. Yes, there is an awful lot of engagement.

The second question is much more challenging. I have probably already answered it in terms of procurement. Unless something changes, I do not see how defence will keep up in this area. Was it Senator Warner in the US who set this? You will know, Ken and James. For a while, the US had a target that 50% of its systems had to be fully autonomous or uncrewed by day X. They have since waived that but, for a while, it was the thing that overcame the inherent conservative nature of defence procurement, in both its procedures and the cultural norms that inform it, and forced the adoption of autonomous systems and experimentation in that space.

Thinking about what kinds of tools we might get to change the internal incentives such that the MoD changes the external incentives and moves to the kind of procurement that I outlined earlier, and I noticed a few people nodding along to, is essential. A more direct answer is that the engagement is great, but the procurement systems are not.

Baroness Doocey: You do not think that that is going to change, by the sound of what you are saying.

Dr Keith Dear: My own view is that the internal incentives and structures within the MoD make it extremely difficult to change the procurement process, at the moment. That is a much more expansive question and I am conscious of Members' time. It is fundamentally an operating model question for defence.

Baroness Doocey: Do you have a different view?

James Black: No; I agree. A lot of work goes on at RAND around how you improve defence acquisition or, more pointedly, how you make it less bad. That is probably how to frame it. As has been said, there are some fundamental issues around skills, incentives and the fact that military officers change post every two years, so there is an inherent short-termism to programme decision-making.

There are some challenges in cost estimation and modelling capabilities. There are challenges in the culture around risk, which is avoiding it rather than managing it in an informed way. There are issues around commercial practices, strategies that are employed and the KPIs put into contracts. You could go on and on about the sorts of issues around defence procurement.

To be clear, there are also green shoots. There are areas of positive reform and experimentation. There are always good examples in as broad and large a defence enterprise as the MoD is. It is one of the larger ones in the world. It can acquire things quite quickly and in a fairly innovative way; it is just how you scale that up to become the mainstream norm across the entire organisation that is a challenge. We have seen challenges, frankly, procuring large, fairly traditional bits of metal—ships, planes, tanks et cetera—in decade-type timeframes, let alone trying to procure something software-based within six weeks or six months.

Yes, it is positive that the defence AI strategy and the DSIS both signal an interest in moving much more towards agile, spiral, DevSecOps and all the jargon words that get thrown around about quicker ways of doing procurement. We need to be realistic. Yes, while I imagine there will be improvement, that will clearly never overtake the pace of change in industry.

On the broader question, I suppose, of how that kind of global, national angle comes in, there is some work that I am involved in at the moment around what that co-development, co-design, procurement and then use of AI within alliances or coalitions looks like. There are a lot of challenges there and it is not just the ones we have all talked about as to how you influence the market if you are a relatively small player, much of it exists overseas and you have relatively few levers as a state to exert influence over the incentives and industry.

It is also about the broader political risks. If we have allies and partners that are employing AI in ways that we ourselves are not comfortable with, there are some operational considerations there. There are also reputational and broader political considerations. It speaks to our need to be working very collaboratively and multinationally on this. There are obviously a lot of US-led initiatives in this space around AI. Existing institutions, such as NATO, are also thinking about it, quite clearly.

It is crucial for the UK to be as influential and meaningful a player within that as possible. That speaks to the UK needing to have a clear value proposition to its allies. On some programmes, that means a bunch of money to say, "We're going to buy our way into a joint programme and work with you on R&D". In other areas, it means we have to bring the skills, the unique S&T, the finance or something that enables us to assert some sort of unique value proposition to our allies. Without that, they will be making decisions that they would then largely dictate to us.

Dr Kenneth Payne: It is global in the sense that the marketplace is global and AI will affect societies around the world to greater or lesser extents. It is not global in the sense of providers or innovators in AI. It is effectively bipolar for the most part, China and the United States, with some tier 2 actors, the UK being prominent among those. The recent history of the industry has been one of centrifugal forces pulling towards those two poles. Small companies get acquired. Talented researchers want to go and work in San Francisco. Those tendencies to centralise are common features of the landscape.

Where does that leave the UK? We are a big player in AI. We are not in that league. We might want to think about the things that we would need to have sovereign capabilities in and not just leave to actors overseas, such as quantum computing, a national foundation model and specialist AI activities in intelligence. We might need to think, "Do we really want to rely on the market for those?"

Courtney Bowman: It is right that the UK should play to its strengths, but it also should not sell itself short. I represent a US-headquartered company, but our largest global office, with over 900 employees, is based in London. There is a tremendous wealth of talent and people are drawn to the UK to work and live. The question is one of how you channel all that talent to useful applications in the L&D space.

To go back to the broader point around the potential frictions in not just procurement but the application of technologies, there is a set of considerations around how you create the incentive environment for this technology to be developed and deployed effectively. Right now, there are a number of programmes that are very focused on admiring the problem and build capabilities in laboratory settings in this sort of bubbled-off framework. They are very interesting and useful, but, when it comes to operationalising that technology, making it meaningful in the field, it falls short.

We often think of a notion of fielding to learn. Programmes that are focused on providing the private sector with opportunities to engage directly in near real-life settings with MoD partners, to be able to build in time and adapt to real circumstances on the ground, prove to be most effective, not just in delivering efficacy with AI and automated weapons systems but in understanding the specific normative context of the application and the specific applied ethics challenges.

We can all get caught up in the academic trolleyology problems of figuring out whether the automated vehicle should go this direction or that direction. The actual ethical challenges of applying this technology come to the fore when the rubber hits the road, where you are actually exposed to the challenges on the ground and understand what the war-fighters are experiencing in context. Creating incentives for those type of interactions is an effective way to build them out between the public and private sector.

The Chair: Baroness Hodgson wants to explore this global standing issue a little further. Perhaps in responding to her you could pick up any unanswered elements of Baroness Anderson's questions earlier.

Q40 **Baroness Hodgson of Abinger:** What impact might the development of autonomous weapons systems by the UK have on its position on the global stage? How is the UK viewed through this development?

Dr Keith Dear: One challenge we face is that defence as a whole is not a market. It is not a market in the normal sense of the word where you are constantly competing equally for customers. It is heavily state-influenced,

and that determines where companies are, where they are headquartered and who they sell to. There is a huge amount that goes on behind the scenes to decide who gets what and where. Therefore, for countries such as the UK, and many others are similar—France is not completely dissimilar, nor are Korea and Japan—you have to build for export, if you are going to achieve the kind of scale that you want to.

There is a challenge that comes there around lethal autonomous weapons systems. One thing—and I am going to say the quiet part out loud here—that allows you to sell to almost anyone is that often, while you might sell them a highly capable piece of equipment, you know that they will not have the training to operate it at a level that makes it a threat that you would otherwise worry about. If you have a system that is truly fully autonomous, when you sell it, in theory at least, you sell a piece of equipment that is unaltered from one of your own.

There are arguments against this. You could say, "We always have export variants". Yes, you do, but you rely a lot on the fact that the operators will not be of the same calibre. That will be less true when you deploy a lethal autonomous weapons system and try to sell it overseas. The worry is how that shapes our market and whether it shapes it to the detriment of developing the capabilities we need for our own Armed Forces because we are so influenced by how we are going to sell this overseas. How do we get round things such as the ITAR regulations? That would be a big concern around anything that was fully autonomous.

Dr Kenneth Payne: The UK is pretty well placed and will be a key signaller of norms in this area of lethal autonomous weapons, AI more broadly and AI in defence. There are some dangers ahead geopolitically. If my hunch is right that AI has a big impact on fighting power, and if it is unevenly distributed and different states differ in their ability to instrumentalise it, there are implications for the shifting balance of power. We know that times of shifting power and great uncertainty are not a recipe for stability. That is the main risk that I see from lethal autonomous weapons.

There are some countries that currently have big industrial-era legacy-type militaries. They employ a lot of manpower, often as a means of societal control. Those are the countries that are most at risk from this sort of technology. The UK is definitely not in that bracket, but a lot of the countries to which we sell weapons systems are. That has some implications for British society as well. We will have to have a think about what sorts of weapons systems we export and to whom we export them, as these new technologies come on screen.

My final point is that Britain is currently a tier 2 power. Discuss. My sense is that the gap with America will only widen as a result of these technologies coming on stream, because of the disproportionate amount of R&D that goes on and because of the centre of gravity of AI research that happens to be in the United States.

Courtney Bowman: The UK is in a position to harness its traditions, its wealth of talent and its moral guidance, to not just adapt but to some degree drive the development of these advanced technologies and the thinking around their use and application. In many ways, the UK plays a disproportionate role with respect to the size of its defence force.

I would caution this committee to think through the distinct advantages that could be built out through potential regulation or other approaches to help steer the direction of development. I know that there was earlier discussion in previous witness panels around this concept of slow AI. There is a concept of moving quickly while applying some of the principles of slow AI to ensure the durability of these capabilities.

It is one thing to think about how we exercise the first mover advantage and provide the frameworks to be able to innovate rapidly, but there is another set of questions around how you ensure that those capabilities that you have innovated rapidly have an enduring impact on the war-fighter on the battlefield. There is an opportunity to think about that long tail of events in development, and incentives around creating these types of weapons systems, such that the UK may actually be able to exert an advantage over a longer timescale, even if it is currently in a position where it is a bit behind the race with respect to the US and China.

The Chair: Would you see those developments coming about organically, as it were, or having to be underpinned by regulation and international agreement?

Courtney Bowman: I do not know that regulation is necessary. It might be a useful tool, but more direct guidance and setting firmer expectations on capabilities deliveries for weapons systems that focus on this long-tail set of considerations is one way to incentivise and push the developers and programme managers towards factoring these in as hard requirements for the technology that they are building.

This also insulates against the risk of the hype cycle. We are very much steeped in a lot of hype around certain classes of artificial intelligence. Some of these capabilities are going to prove themselves to be extraordinarily valuable. Some of them, on the other hand, are going to wash out. Creating incentive structures that force entities to deliver not just current impact but sustainable impact may be one way to create that longer trail of value.

The Chair: This question of "me too" competition is one that Lord Hamilton wants to explore.

Q41 **Lord Hamilton of Epsom:** My question is about the arms race and the degree to which pressure is being put on us to develop AWS because other forces are too. When I mention that, can I go back to what Dr Dear said? I do not want to paraphrase him, but I want to suss it out a bit. He seemed to be saying that the public position of the Ministry of Defence is that we were having nothing to do with AWS, but in practice we were going ahead and doing it anyway. Is that right or wrong? Can we tease

that out a bit?

Dr Keith Dear: That is definitely not what I was saying, in fact perhaps almost the opposite. There is an awful lot of rhetoric around what the MoD might do in AI. I was speaking at an event recently. A friend of mine, Joe Robinson, who is known to others on the panel, is CEO of enterprise at Improbable. At the end of the talk, Joe said, "What you have said, once again, about where AI is taking us is really exciting, Keith, but can you name a single major artificial intelligence project within the Ministry of Defence, or UK defence more broadly, that has made a significant difference to the way we fight and has been applied at scale?" I was stumped. He said, "I don't know one either". He is ex-military as well. He said, "If you and I don't know of one, I'm going to guess that it doesn't exist". My contention is that there is a very big gap between the rhetoric and the reality, but in the opposite direction to that which you suggested.

Lord Fairfax of Cameron: I am very grateful for that clarification, because I took it to mean something completely different when you said it earlier.

Baroness Doocey: I did not.

Lord Hamilton of Epsom: We have now spelled it out.

The Chair: We do not want to embarrass you, but we have it clearly on the record.

Courtney Bowman: That sentiment is right. If it is useful to motivate the MoD to make strides in developing technology and creating the right incentive structure to work with the private sector, incredible value can come of that, but the arms race should not be about winning just in terms of rhetoric and hype. There is also a race towards ethical, responsible and trustworthy deployment of these technologies. The UK is well positioned to help drive that set of considerations.

Dr Kenneth Payne: There is a danger that people who are reluctant to call it an arms race feel that framing it that way will accelerate and exacerbate tensions between states when it comes to developing AI. I do not have a reluctance to call it that: there is an arms race afoot.

Other sceptics of the term "arms race" point to the dual-use and general purpose nature of the technology and say, "You do not talk about having an arms race for electricity as well". I see the logic, but that does not discount an arms race when you are talking about developing autonomous weapons systems or systems that can be used more broadly in defence. Yes, AI is dual use, but part of that use includes shaping fighting power.

One thing about the arms race that I think I said to you last time is that it is often posited to be an arms race between either America and China or the West and China, the other leading pole. But that may not necessarily be the only way to frame the concept of an arms race. It is a

multipolar world; other states exist. Other states and other groups will try to instrumentalise AI in the defence national security realm.

We kind of project our fears on to China a little bit. The view from the other side of the fence, evidently, is quite similar. China feels that it has shortcomings in AI, and in science and technology more broadly, relative to the West, so we should be cognisant of that. We have recently discovered, after the invasion of Ukraine, that the Russian military was not as mighty as we perhaps feared. We would do well to keep that in mind and guard against the temptation to demonise people on the other side of the arms race.

James Black: Some very good points have been made. There is perhaps a useful analogy here. You said that we do not talk about that with electricity as a dual-use technology, but there are some technologies where we are absolutely having that sort of debate.

One dual-use example would be space at the moment. There are some useful parallels and lessons there, perhaps, for the UK, in developing both its military space capability and its broader space sector. We have acknowledged that there is an arms race going on at the moment. We have also said that we are going to place a moratorium on development of certain capabilities ourselves. In that case it is direct ascent destructive anti-satellite missiles, because we do not think that they are responsible or ethical in their employment. At the same time, we have said that we are going to invest heavily in other types of capabilities to make ourselves credible.

Crucially, and this is where it is most useful to the AI debate, we have tried to position ourselves as this convening power between the US and China. If the US and China try to shape the normative debate or the international legal or regulatory debate, of course they are open to accusations that they are just superpowers trying to defend their existing hegemonic positions in those markets.

The UK does not have some of that baggage because it is not such a big player. At the same time, it is well enough informed and has enough capability to be credible. It can also draw on its broader soft power capabilities and lineage, legal expertise, financial expertise and these sorts of things to act as a broker for those middle nations that are not the US or China but are still going to vote in the UN ultimately. We have seen that through initiatives such as the FCDO-led push at the UN open-ended working group on space.

Crucially, the key lesson from that is always that, to get a seat at the table, you need to be credible. It is making sure that you have that suitable level of understanding within government to be an intelligent contributor to those discussions. It is making sure you have the effective public-private partnerships, the kind we were talking about, so you can mobilise the rest of the UK and have a Team UK voice within these kinds of fora. It is then making sure you are developing the S&T and the capabilities within the Armed Forces and so on that make you relevant to

those discussions. Without that leverage, you are not able to be that convening power and shape those norms.

Chair: Lord Fairfax has our final question. I suspect it may be one that can have a single-sentence answer.

Q42 **Lord Fairfax of Cameron:** If you could, what one recommendation would you make to the UK Government in this area?

Dr Keith Dear: I would argue that the MoD should put its money where its strategy is, moving 6% of its budget every year into emerging technologies by arbitrary rule. I could back that up with where that comes from, but that would be my contention as the best way to overcome the incentives that militate against the adoption of these technologies at the scale we need.

Chair: That was admirably concise. Thank you very much.

James Black: Define the UK's desired USP and value proposition in this space. We cannot do everything; we cannot be everywhere all at once. What are the two, three, four or five things we want to be really good at in the value chain? That is going to be the basis of how we engage with our allies and partners. We are going to lead in those areas and accept that we will not lead everywhere else.

Dr Kenneth Payne: I have a very narrow one, which is that the UK needs to put some thought into re-establishing a second leg of the nuclear triad. All our eggs are currently in a submarine boat and this sort of technology may have an impact on that. More broadly, I have mentioned several times before that I favour the establishment of some sort of democratically selected public commission, of the sort that happened in Ireland around the discussions on abortion, to discuss these issues in the round as a way of widening that public debate about the issue.

Courtney Bowman: I will re-emphasise that point I made earlier: invest in field to learn. By that I mean investing in programmes that advantage moving quickly from policy and innovation to real-world operational deployment of these technologies that actually exposes the challenges as they exist.

The Chair: Thank you very much indeed. You have been very generous with your time and expertise. I know that I speak on behalf of the whole committee when I say that you have moved our understanding and insight into these issues forward significantly this morning. I spotted in your evidence a number of epigrammatic phrases, which will be strong competitors to appear in the eventual report, so we will see about that. Thank you again very much indeed. We are all very grateful to you.