HOUSE OF COMMONS

Foreign Affairs Committee

Oral evidence: Artificial Intelligence and diplomacy, HC 1927

Tuesday 24 October 2023

Ordered by the House of Commons to be published on 24 October 2023.

Watch the meeting

Members present: Liam Byrne (Chair); Brendan O'Hara; Bob Seely; Henry Smith.

Questions 1-28

Witnesses

I: Professor Michael Ambühl, former Swiss State Secretary for Foreign Affairs, and trained mathematician; Dr Pia Hüsch, Research Analyst for Cyber, Technology and National Security, Royal United Services Institute; Professor Corneliu Bjola, Associate Professor of Diplomatic Studies, Department of International Development, University of Oxford; and Professor Nick Boström, Professor and Director, Future of Humanity Institute, University of Oxford.

Examination of witnesses

Witnesses: Professor Ambühl, Dr Hüsch, Professor Bjola and Professor Boström.

Chair: Welcome to this afternoon's meeting of the Foreign Affairs Committee on artificial intelligence and diplomacy. Thank you so much to our guests and witnesses for joining us this afternoon for what should be quite an interesting session. We have merged our two panels and we have lots of questions, both deep and simple, for you. We are very much looking forward to your advice and counsel.

We will start off with some introductions. Michael, let me turn to you first.

Professor Ambühl: Thank you very much, Chairman, and thank you for the invitation to this hearing. I am Michael Ambühl. Until recently, I was a professor of negotiation and conflict management at ETH Zürich. Before that, I was a career diplomat in the Swiss foreign service, where I negotiated a couple of agreements with the European Union and mediated in some international conflicts. A year ago, my newly established consulting company was asked by the Swiss foreign ministry to make an analysis of the possibilities of digitalisations of negotiations.

Dr Hüsch: Hello. My name is Dr Pia Hüsch. I am a researcher in RUSI's cyber team. RUSI is the Royal United Services Institute, a security and defence think-tank here in London. Before I joined RUSI, I did a PhD in international law and cyber operations, so I am an international lawyer by training, but I have focused on the regulation of cyberspace and interstate cyber operations.

Professor Bjola: Hello. My name is Corneliu Bjola. I am an associate professor of diplomatic studies at the University of Oxford. My field of expertise is digital diplomacy, by which I mean the use of digital technology by ministries of foreign affairs and international organisations in their work. I have researched and published on how digital diplomacy works, especially with respect to public diplomacy, crisis communication and international negotiations, as well as the dark side, which means countering disinformation. At the moment, my area of research has moved in the direction of the use of artificial intelligence in diplomacy and foreign policy. Thank you very much for having me.

Q1 **Chair:** Perfect, thank you. We have merged our panels this afternoon, so the first few questions were originally aimed at Michael and Corneliu but, Pia, please chip in where you see an opportunity to make a contribution.

Michael, perhaps I could start with you. I guess our opening question is to ask you for a sense of the uses to which you think artificial intelligence could be put that would be beneficial to our Foreign Office in improving its performance here in the UK.

Professor Ambühl: I can answer this question only from the Swiss point of view, and specifically from the point of view of the analysis that we were just asked to do for our ministries in order to give an overview to my former colleagues in Bern. We made a categorisation of the methods and a

categorisation of the negotiations. I believe you have received a handout from me.

Chair: We have.

Professor Ambühl: You may be interested to have a look at the first point while I mention categorisations A and B. It might sound a bit technical, but I think it is good to have an overview of what we did.

We said we would look at the digitalisation question of negotiations and two methods. There is the classical, qualitative method of algorithms and software, which are tools without self-learning capacities—the so-called non-AI methods. On the other hand, there are new tools with self-learning capacity and machine learning—the AI method.

We tried to compare these two methods on different negotiations, which we also tried to categorise according to three axes, to have a sort of three-dimensional view. The first categorisation is the complexity of a negotiation; if a negotiation is rather simple, that means it is repetitive and has well-established standards, for example free trade agreements or double taxation agreements. The second axis or categorisation would be technicality; for the purpose of our analysis, we would define a negotiation as technical if it can be isolated from a value-based context and if it would be accessible to technical, factual, legal argumentation. The third is structure, meaning the number of players involved. Is it a two-person negotiation or is it multiple, meaning multilateral institutions?

On this basis, we made an analysis of different concrete cases. If you wish, I could give you a short overview of these examples, but maybe I should wait for your feedback.

Chair: Let me come back to you to follow that up, but maybe Corneliu could add to that answer from his perspective.

Professor Bjola: Absolutely. I will be using two criteria to understand better the potential impact of artificial intelligence on diplomacy. One refers to the notion of risk. In which areas is it safer to use AI, compared with others? The second is about value added. What kind of contribution could it make that would be better than what we have now?

I will start with the first category: the area in which the risk is lower if you make a mistake, and where it can make a good contribution. At the moment I see a lot of efforts in the area of consular affairs, because you have repetitive work that requires a lot of human resources and that can be done better. Even if you make a mistake, you can fix it.

The second area is where there is medium risk, but where you still require a lot of human resource to complete the task, for instance in public diplomacy. Public diplomacy is a way to understand perceptions about the country. We live now in a very geopolitically tense context, and it is better to understand things now as opposed to understanding what is going on in a week or month. It is a question of having AI deployed to absorb information about you as a country—the FCO or the UK in various

countries in various places in the world—and reacting appropriately. This is an area that, yes, can be done with your team of people right now, but it can be done more effectively if you have these kind of tools deployed.

An area in which things could be risky is, of course, crisis communication, crisis management—think about situations in the middle east or Ukraine—and trying to understand patterns that can lead to conflict. Interestingly enough, I know from my interaction with various ministries of foreign affairs at the moment that this is an area in which there is a lot of interest, understandably so. They want to understand whether a particular tension is likely to explode, and it is better to understand that now as opposed to in a week, because you need to prepare for that.

The problem is what to do about that. This is where we get into the technical aspect of AI—explainability and black box paradoxes—in the sense that while AI may provide some ideas about what happens, decision making requires an understanding of the process by which AI has reached that decision. In other words, when you decide a solution, if you go back to policymakers and say, "Why did you pursue that option?", your answer cannot be, "Because the machine told me to do that."

It is a question of accountability and explainability. From this point of view, I will be looking at using different levels of risk and understanding better the type of contribution it can make. Consular affairs is a low-risk area, and it is being deployed. In Canada and the UK, we have been using AI from this point of view already, partly for communication and on the dark side, meaning countering disinformation, which is an area in which I see these kinds of tools performing to a certain extent and providing value added. Where the situation is complicated is, of course, in managing crises—not necessarily to understand what is going on, but to better understand how to react to them. I am going to stop there for now.

Q2 **Chair:** Interesting. How well do we understand which diplomatic functions fit into which of your risk/value added categories? Do you think we are beginning to see any consensus on what kind of functions fit into what kind of bracket yet?

Professor Bjola: I think that is also a question about the evolution of AI. Why do I say that? Because AI is not static. What I mean by that is that what we fear about AI today, if we come back in two years, will be common sense. There is a question about perception. What is the risk at the moment, given the technology that exists—ChatGPT, let us say, which is being deployed nowadays to create content, despite the risk? The risk, for instance, is that when you create content for public diplomacy social media, the moment you use ChatGPT, the information that you put in goes somewhere in the cloud, which you do not control. From that point, it goes into a server that does not belong to you. It belongs to OpenAI; it belongs to some other organisation. You have to be very careful about what content you create using these kinds of tools, of course.

I think there is a pressure to use some of the tools in public diplomacy because the risk has been accepted. There is interest in using it for other

areas like crisis management, because of the geopolitical pressure, but there is also reluctance because of the potential political cost that may result from not performing well. If we come back in five years, I don't think the perception that we have now about this risk will have stayed the same, because technology is going to be different. We will probably get used to that, and to some of the intricate aspects of AI, in the same way that we got used to social media or Zoom, in terms of confidentiality.

Q3 **Chair:** Michael, would you agree that that is a good categorisation of the different functions and of the way of looking at where AI is appropriate and where it isn't? Where do you see particular diplomatic functions as being performable using artificial intelligence?

Professor Ambühl: I think it is applicable, on the one hand, in negotiations and, on the other, in conflict management. If I may, I could give you just a couple of cases we examined regarding negotiations and then I can give you our findings, which are on your handout. Would that be okay?

Chair: Sure.

Professor Ambühl: We have analysed a couple of cases, which are real cases, and also a constructed one.

The first example is an AI example: the creation of a resolution for the Human Rights Council, where we said, "Why don't we have a human rights and environment resolution?" The result with GPT-4 was very well done. It had one rogue reference to an ILO convention, which was useful for checking.

Just for fun, we also did a dummy resolution on human rights and railway workers. There, it also produced a nice text, although of course it did not really make much sense because railway workers do not need special human rights protection.

That was the first example. The second was where we made a comparison between a real text from a real negotiation and one with AI. The real negotiation was about drafting a co-operation charter—a real one in the middle east, with the six Arab Gulf states and Iraq and Iran. We compared the two methods—the AI one and the real, negotiated one—and came to the conclusion that the AI one gives an interesting start into a negotiation process. The negotiated text had a bit more creative elements and the AI text was more detailed.

We also did this, as a cross-check, with a nuclear co-operation agreement in the same area—the six Gulf states plus Iran and Iraq—based on the Beijing declaration. You might remember that in March this year there was an agreement between Saudi Arabia and Iran regarding nuclear co-operation, and we fed this in. According to our experts in the ministry—we gave it to them to make the assessment—they came to the conclusion that the draft nuclear memorandum was a very good text without mistakes. However, it was not very creative or specific.

Can I go on with the third and the fourth examples?

Chair: Yes, very briefly. They are very illuminating examples.

Professor Ambühl: Thank you. The third text was in my former area of expertise: Swiss-EU negotiations. As you know, we are not a member of the European Union; unlike you in the UK, we never were. We need bilateral agreements, and we are in the process of a so-called bilateral III package. There is an ongoing and very difficult question of dispute settlement. You also have this in your negotiation of the post-Brexit settlement. There, the AI result was rather disappointing, because it did not give any new elements unless you really put them in or said, "I wrote an article, why don't you use this?" Then it nicely re-copied those ideas, but that was not very convincing. It then said rather banal things like, "One has to be careful, because at the end of the process the Swiss voter will have to say yes or no in a referendum."

The last example, if you will allow me, is non-AI; it was a classical software algorithm. I think this is also important when we talk about digitalisation in diplomacy. We made an algorithm for multilateral WTO negotiations. It was a real case on e-commerce and digital trade. The UK is now, after Brexit, a full member again of the WTO; before, it was always in the hands of Brussels.

We checked whether, if in a meeting of 15 member states or countries they found a result to different issues in this area, it would be so-called Pareto optimal or whether it would be Pareto inferior, meaning that it would be possible to find another solution that would be an improvement for all the participating countries—that is the definition of Pareto improvement. The algorithm could calculate such a Pareto front that normal human intelligence, with just a piece of paper and a pen, could not. That was a very good experience.

We also tested this in a seminar with our foreign economic ministry. That was a nice tool which can now be applicable, for example in multilateral institutions, which should have an interest in negotiations in this multilateral area of WTO being Pareto optimal, meaning that it is not possible to have an improvement for all. That is what we did; there are many other examples.

Chair: That is really useful.

Professor Ambühl: On that basis, we then drew conclusions. Maybe I should stop there and come back to this later.

Chair: That would be perfect. Thank you.

Q4 **Henry Smith:** Professor Ambühl, what are the experiences of other foreign ministries in using AI?

Professor Ambühl: Thank you for the question. I am not sure whether I am in a position to answer that, because I do not know. I believe that, in the context of the mandate we got from our foreign ministry, they are all

up to using artificial intelligence, probably more in conflict management questions than in negotiations. In negotiation, it is not really applicable in a way that you can, in a bilateral or trilateral negotiation, use openly. You can just use it to inform preparation back home to know a bit better what the possible outcomes could be. Otherwise, I am sorry, but I cannot really answer that question, because we do not have comparisons with the UK, Germany or France. I can just give you our findings on what we believe could be, maybe from a general point of view, applicable to other foreign ministries.

Q5 **Henry Smith:** Thank you. That was certainly very useful. How does machine learning technology of this kind differ from other algorithmic tools?

Professor Ambühl: Very good: this takes me to our conclusions in the comparison of AI and non-AI algorithms. We believe that the less complex a negotiation is and the more repetitions it has, the better AI could give support. Of course, that is support on your side; you are not using it in an open area or telling people, "I use AI in order to make an analysis of whether you are talking rubbish or not." That would not be very positive. I believe it is clearly the case that if it is repetitious, you can use AI quite nicely.

If the negotiations are complex, however, and if creativity is asked for, I believe that the added value of AI is limited. Of course, it is not excluded, but if you need, as is normally the case, pre-existing data to make suggestions, then creativity is probably the most likely issue.

The third conclusion was that in multilateral institutions, AI methods are limited because of the lack of visibility and transparency. If you are in the UN Secretariat or other international organisations, it is difficult to say, "We have a proposal that was made by AI." People would like to know what the basis was. My colleague has already mentioned the words "black box", and you would not know how the result given by AI has been established, so there are certain needs.

My fourth point is that the more technical the negotiations are and the more quantifiable certain questions in negotiations are, the more I believe that classical quantitative methods can be used. Why do I say that? Because I believe that they are not really used enough. I believe that the potential is underestimated, probably because diplomats might have a certain reluctance to use such methods even though they have an advantage compared with AI. They are deterministic and if you repeat the inputs, they always give you the same outputs, so there is no black box. You know exactly that when you put in A, you get B—

Chair: Michael, if I may, I will just stop you there, because we have so much ground to cover and the clock is against us. I will bring in Brendan at this stage.

Q6 **Brendan O'Hara:** I have some general questions about predicting trends and crisis management. An important part of what the FCDO does is looking to the future and identifying where, when and how problems will

manifest themselves. How effectively could AI be used to predict a conflict and to horizon-scan generally?

Subsequent to that, is there a role for AI in crisis management in real time? Can it be trusted to advise Governments in a time of crisis? Is there a danger that AI could take over during a crisis and get it disastrously wrong? Politically, is there not a question that Governments might almost abdicate their responsibility to AI in a time of crisis?

Professor Bjola: Thank you very much for the question. In answering that, let me clarify one important aspect of AI. AI has two components, basically: one is about data and one is about algorithms. In order to understand whether we are able to predict something, we have to understand what kind of data we put into the system and what kind of algorithm we use, and that is where things get complicated.

I was in Rome last week. On a similar topic, the ministry of foreign affairs in Italy is developing a tool exactly on the lines that you suggest. For data, it is going to use news. It has two databases that it can access, with news from 194 countries. To give an example, you can imagine that if you use news only in English, you will get only part of the story, so what kind of data you put into the system is very important. It is good if you put in local stories. I asked a question about what kind of stories are used in China, because press is controlled—can you get an accurate picture of what is happening in China if you use only official stories? What I am saying is that in order to get an accurate prediction, you first have to understand the data that is being used and fed into the system.

The second point is about what kind of algorithms are being used to predict probabilities of potential conflicts. Again, that is where things get complicated, because it is about weights. If a protest happens in Paris, people will not necessarily be alarmed, because it happens frequently and we know that usually the Government control the situation well. If a similar protest happens in Kosovo, people will be alarmed because the context is different. So we have a question about weights, and those are subjective—you have to agree on them. It is not necessarily that the machine will give you the weights; they are something that you get with expert knowledge from the people on the ground to understand what is going on. Based on that, you get a result, and the result may be good in the sense that it predicts with 90% probability that something bad is going to happen in country X next week. Then, to go back to your question, how are you going to react to that? Are you going to deploy more resources to manage it? Possibly. If you have nationals in that country, are you going to get into a particular type of diplomatic engagement—alerting, or trying to build an alliance or coalition to deal with the potential crisis? I think those are answers that you are going to receive not from a machine—from AI—but from humans.

This is why I said that the third category is about risk, and to what extent you let the machine inform your decisions, as opposed to driving your decisions. Informing is about putting things on the table and letting decision makers, with support from other experts who know the context,

formally make an answer. Driving is eliminating the human from the loop, which I don't think is likely to happen in this kind of situation. It may happen with a low-risk situation, but in a big situation like this, I don't think so. But again, this is policy. It is about policy makers deciding how to deal with the results. But in order to deal with the results, you have to understand the quality of the data, how the algorithm works and the type of adjustments that you add to the algorithm.

Q7 **Chair:** Thanks. We have been joined now by Nick Boström. Nick, thanks so much for joining us slightly earlier than planned. Would you like to say a word about who you are and where you are from, and we will bring you into the conversation?

Professor Boström: I am a professor at Oxford University, where I run the Future of Humanity Institute. We look at the impact of emerging technologies. AI in particular has been a big focus of ours for the last decade and a half—you could almost say an obsession. We are interested in both the technical aspects of AI safety and AI alignment and the broader questions of how this might reshape the world, and the governance and strategic challenges that the current rapid developments that we are seeing in machine learning will pose for our civilisation. I don't know whether you want me to elaborate further on my childhood—

Chair: No, that is great. I am going to turn to Bob Seely for our next question.

Q8 **Bob Seely:** Nick, like most normal human beings—pretty much everyone but panel experts—I don't know what I don't know on AI, and I don't know much about AI to begin with. You are from the Future of Humanity Institute. Does humanity have a future with AI? I have seen "The Terminator", with Skynet: all this horror movie stuff. What is the danger that AI will become so powerful that it ends up controlling us or giving human beings the means to control other human beings in a very malign and dark future for humanity?

Professor Boström: Right—let's jump in at the deep end. The first thing I would stress is that AI is not one fixed thing that we are dealing with, but a rapidly developing field. When we are speaking about AI systems' capabilities, the capabilities they have today and the capabilities they may have even just 18 or 24 months from now—let alone further on—are very different questions.

I think that as we develop increasingly powerful forms of general artificial intelligence, significant risks will be posed in many ways—including some existential risks to the survival of the human species. People have begun to discuss those more recently, and they come in various forms.

There is what we might describe as the more technical problem of how to align a particular AI system to make sure that it executes the intentions of whoever is operating or building it. Figuring out how to do that will require some technical advances, but even assuming that that alignment problem is solved, there is then the question of what ends these human users will put these AI systems to. We know that many other general purpose

technologies throughout history have been put to all kinds of uses, including very negative ones such as waging war.

There is also the possibility of a scenario where multiple different AI systems are used by multiple different agents, which might create new social equilibria. We know from the past, for example, how developments in information and communication technology have up-ended political systems—going back to the emergence of writing, which enabled states to keep tax records and so on, and onward to the printing press, with the hundred years' war, religious prosecutions and turmoil. More recently, we see various challenges coming from the internet and social media, along with the enormous potential for positives.

Q9 **Bob Seely:** Before I go on to slightly narrower questions, can I ask about the two bigger elements that you talk about? As I understand it, parts of the Russian system to analyse whether Russia is facing a nuclear attack are automated. In an AI age, is it potentially very dangerous if existential weapons systems such a nuclear missile response are automated using AI? If you apply that to game theory, it is not impossible in the future that an AI program would recommend the launch of those weapons if it reckoned that it had worked out how to destroy its opponents in a first strike. What is your thought on that?

Professor Boström: I think it would seem very dangerous to put AIs in control of nuclear weapons, but I also think that that is not where the bulk of the risk is coming from, because it is kind of obvious that that would be a dangerous thing to do. I think policymakers would be reluctant to do it.

Q10 **Bob Seely:** So I come to the second question; you may say that it covers the bulk of the risk or that actually I am still missing the point.

Big Brother—an Orwellian future. AI will soon have the power to monitor individuals, arguably even more so than it does at the moment. You can have closed societies such as China and Russia that monitor people to make free individual thought almost impossible by analysing how people move or think, or the sort of things they post. AI is also a threat in open societies when people are put under it: commercial AI might be used to understand people better and better. That is a potential threat—a benign threat, but nevertheless a threat—from AI in open societies as well. Is that anything to do with the sort of things that you are thinking about as well, or am I missing the point slightly?

Professor Boström: That is definitely one aspect of the challenge that will be very important. For example, imagine running sentiment analysis at scale. If you had access to everyone's social communications, whether because you are an intelligence service eavesdropping or because you are one of the tech companies providing the service, you would be able to determine with quite a lot of precision what everyone's views are on various different topics—political persons, a company in question or a product—and you would be able to train operators to sift those. Or if you are a dictator, say, you could pick out—

Q11 **Bob Seely:** I have been watching "The West Wing", which I am slightly

ashamed to say I have never watched before. I have been bingeing on it slightly in the past month. For the State of the Union addresses, they had their focus groups, turning the dials up and down. When do you think that we will have AI sentiment at scale in US elections or in other electoral systems in the democratic world, or is it already happening?

Professor Boström: I imagine that this is happening—if you have the data—even just to target ads. If you have Facebook and want to help advertisers to reach the particular customers who might actually buy something as a result of an ad, that already involves this basic technology. It will become increasingly powerful, and it could be used for totalitarian purposes, say. If you are a dictator, it might become easier and easier to detect pockets of resistance early on, if you have access to communications.

Q12 **Bob Seely:** I am assuming that there will not be a "big bang" moment, but will this happen in the next five years or the next 20 years?

Professor Boström: I would say more five years, or perhaps fewer than that. If you have access to the data feed, even current AI technologies would be quite up to the task.

Q13 **Bob Seely:** But you still have to analyse the data. If you are an American presidential candidate, or you are in Brazil, India, Colombia or anywhere, you can buy lots of data, but this is about using AI or having an AI program that helps you understand that data, helping to shape your words, image, behaviour, body language posture and everything about you in order to maximise your ability to manipulate target audiences. At what point does AI become a critical help in that, over and above just being able to process lots of data now, which is what computers have enabled us to do for the past 20 years?

Professor Boström: It is already possible in ways that it has not been before not just to crunch big data if you have it neatly packaged. If you have a survey, for a long time we have been able to say what percentage of people answered A, B or C. Now, if you have access to the raw data of people's communications—their Facebook messages, their browsing history or any other sufficiently large aggregate of data—you will be able to form a much more complex picture of each individual. Then, for example, you can craft customised messages—one for every voter—that speak specifically to them and their concerns, worded in a way to resonate with that particular individual. That is possible right now, and it is limited only by access to enough contextual data of each person you want to reach and by the willingness of the campaign to risk sending out millions of emails, a few of which will probably contain something embarrassing from the AI that could backfire.

Q14 **Bob Seely:** Just as a round-up for anyone else who might want to come in on that: what are the risks and dangers for society of the widespread use of AI, and what are the risks and dangers of unregulated widespread use of AI? Does anyone else want to come in, preferably briefly if possible, please?

Dr Hüsch: I appreciate that a lot of recent discussions have very much focused on existential risks. We see the same in the preparation for the UK AI safety summit. Coming from a national security perspective, I appreciate those considerations, but I think that this narrative, which is very much fostered by big tech companies, comes at the risk of overlooking short-term current risks and implications. A lot of the things that you mentioned—manipulation, election interference and so on—are things that we are already seeing. Artificial intelligence technologies might make you reconsider the scope, scale and tempo at which these things are possible, but I would say that they are already here.

The risks I see from AI technologies, more in the short term, focus around things like increased bias and discrimination in decision making; privacy risks; data theft and so on; security risks; mis and disinformation, as you mentioned; manipulation and ethical concerns. Those are all aspects of risks of AI technologies that can already be addressed under current regulation, for example, through data protection.

Q15 **Bob Seely:** So your fear is more Big Brother rather than Skynet in "The Terminator", to put it crudely.

Chair: On that spectrum.

Dr Hüsch: Neither Big Brother nor the Terminator.

Bob Seely: A bit of both. Anyone else?

Professor Bjola: Just to connect to the last question that was asked—I think we are already there—about the idea of Cambridge Analytica on steroids. There was a paper published in February this year. It was not using AI; it was using ChatGPT.

The idea here is to create this kind of synthetic persona, as I think they are called. What does it mean? When you do a survey, you have a demographic profile: I want this proportion of males, this proportion of females, this kind of ethnicity, this kind of pay, this kind of level of education. Based on that, I am going to tailor my political message.

The synthetic persona, the paper showed, which created a lot of impact and attracted a lot of attention, is that you do not need that. You can do that, as we speak, with ChatGPT. How did they prove that? They created this kind of persona, and then applied it to US elections in the past. They found that by using that kind of demographic and techniques they were getting better results in terms of predicting the results of elections than using surveys.

The idea here is that it cuts down the costs and also democratises—let's say—the use of techniques to tailor your political communication in a particular way. You don't need sophisticated AI; you just need to understand how to programme a little bit.

Q16 **Bob Seely:** Tell me if I am wrong, but I see no evidence of AI being used in a meaningful way in democratic elections. You are saying that it is

being used because—

Professor Bjola: A paper has been published that shows that the potential is there to use it.

Bob Seely: Potentially so, but we are years away from it.

Q17 **Brendan O'Hara:** On the example you were talking about, six years ago I was on the DCMS Committee, which published a report on disinformation. We discovered there was an unregulated wild west out there. Democratic Governments were way behind the curve when it came to big tech and social media companies. We are where we are because of that situation that we found ourselves in, and the fact that the bad-faith actors were over the hill and far away before democratic Governments got their heads round this and what to do about it. Are we in danger of repeating that situation now with AI?

Dr Hüsch: I would say yes, in the sense that, if you look, for example, at regulation of big tech and what we have seen with social media, and a lack of early regulation, for example of social media platforms, there were a lot of very unregulated approaches. We see now the consequences of disinformation. We also see tech companies firing their integrity teams and so on—mechanisms that could at least try to help to counter disinformation spreading online. A lack of regulation has led to that. If we listen now to the same big tech companies to whom we have given a lot of unregulated power, voluntary standards and commitments to improve these situations, I think there is a certain irony in taking advice from them to keep hands off on further regulation.

Q18 **Chair:** I will just follow up on that, Pia: how desirable is it that we create mechanisms that might encourage the co-operation of lots of individual countries in some kind of global regime for AI regulation?

Dr Hüsch: I think it is pivotal that we do that, although I don't think it will necessarily be immediately fruitful. I don't think there is a chance that we will see an international treaty on the use of artificial intelligence technologies any time in the immediate future. But just because that might not be on the table right now, it does not mean that we should not try to have that dialogue with other countries. We saw last week, for example, that China has announced a global AI governance initiative. If we don't foster that conversation and dialogue, states like China and Russia have in the past been very successful at using international platforms such as the United Nations, when it comes to the regulation of cyberspace, as forums to pursue their geopolitical and strategic ambitions.

Chair: Nick, do you have a perspective on that?

Professor Boström: Yes. I think it is going to be very important, but also difficult. The problem will probably have to be broken down into different areas and applications and kinds of AI, and then each of those will have to be worked on. For example, I think one particular application area where there would be broad consensus is on preventing the use of next-generation large language models to facilitate the production of biological

agents or other criminal activities such as cyber-crime, and to do some kind of pre-testing to ensure that before these next-level models are made broadly available, they can be shown to be resistant to those kinds of applications, or that there are some sort of safeguards in place. I think that would be in basically everybody's interest. The AI companies—the labs themselves—are also keen to try to do this. If you have many different competing developers of AI, the overall regime is only as strong as the least scrupulous of them, so having some standards that prop up the bottom of that competitive field would be one obvious area where policymakers could work.

It will probably have to start with something vague, like gesturing towards some principles, but I think there should be a statement that ultimately AI should be for the benefit of all humanity. It is too big to be the sole possession of any one company, or even one country. I think getting some in-principle statement in, and committed to early, might be valuable, even if it does not initially have specific regulatory teeth.

Chair: Michael and Corneliu, can I bring you in on this question?

Professor Ambühl: If I may go back to a previous question regarding what to do with mistakes and rogue statements coming from application of AI in diplomacy, I will bring in my simple way of looking at things as a former diplomat. Of course, AI can very much give the wrong references or make wrong statements or wrong probability predictions. The problem is that it is not easy to detect the wrong statements. As a rule of thumb, I would use AI only for topics in which I am an expert—in a way, where I know more than AI; where I have more experience—so that it would be easier to realise when the AI is in a way hallucinating. That is my first remark.

If I may come to another point that was made on how you can use AI in conflict management, I am not in the area of regulation; I am just in the area of normal application in a foreign service where you would like to have some conflict prevention. Of course, as has been said by a colleague in the room, you need a lot of data in a conflict to get a reasonable, sensible prediction. Most likely, you also need a pattern in a conflict, which can then be detected and made more transparent with an AI method.

There is a problem with delicate questions. The question of Russia was mentioned before, in the sense that you would like to know when Putin is launching a nuclear attack. That would be a nice question—one we would love to know the answer to. To give you an answer from my point of view, AI cannot produce a lot of things because there are not enough data. There has been only one nuclear attack so far, very fortunately, and from my point of view, there are too many nuclear power states, so it is very difficult to make a prognostic. Even if you get one, it will most likely not be useful. The same goes for the terrible Hamas attack three weeks ago. The question arises of whether it would have been possible to make a prognostic about it, but a machine could never precisely predict such a top-secret issue, which was dealt with in a very secret way. It would not have been possible to give a useful analysis there.

What is much more needed, I would plead, is human intelligence—good, plausible analysis, of which there are many examples. Not to create publicity for the *Financial Times,* but today the lead political commentator made an excellent comment about what could happen again in the middle east. I would contend that human intelligence is much better placed to give some reasonable assessment than AI in cases where there is a lack of data.

Chair: That is very useful. Thank you.

Q19 **Brendan O'Hara:** The potential power of AI is certainly beyond my comprehension, but how do democratic Governments work with regimes, knowing that there are bad-faith actors, around the regulation of AI? Should there be a two-tier system in which we work with our friends more than we work with our enemies? Do we work with the bad-faith actors we all know about? Is it possible or even desirable that we do such a thing?

Professor Bjola: Thank you very much for the question. In my view, the question of regulation, along the lines that you suggested, has three elements. You have to think about regulation at the level of conceptualisation: how and what exactly do you regulate? We have to balance risk versus innovation. There is a potential innovative aspect with important implications for the economy, security and education, but that has to be balanced with the question of risk. That is where I think the debate is going on—at the level of conception. It is the same with how granular—how precise—you want to go with regulation. Technology is difficult because it is dynamic, but it can also affect that balance.

The second element is the format. Yes, a number of countries want to create national legislation, and they think it's understandable why, but at the same time these things do not stop at the border, so there is a question about collaboration, how to collaborate and with whom to collaborate. Here, the notion of like-minded countries comes to the fore: who are the like-minded countries that share your values, your perspectives and your political understandings and that you would like to engage with in order to produce this? We now see collaboration through, for instance, the Trade and Technology Council between the EU and the United States. There is a G7 Hiroshima dimension as well. I think these are good formats that have to be pursued.

The third element of the regulation equation has to do with the implementation. Let's assume that you have a brilliant conceptualisation of what you want to do, and you have found the perfect format. Is it going to stay on paper, or do you have a useful instrument that can actually ensure that the measures you have designed produce effects? On the last aspect, we don't have a clear answer yet on how to implement, because the model that we have had in the past, like the UNFCCC, on climate change, based on voluntary agreements—yes, they might work, but not with the expected results.

To step back, for many the issue is to understand exactly the three components and not forget about the last one, which is about implementation. The risk of doing this is also something that has been written about: it is the fragmentation of the digital international order between more authoritarian persuasions and more democratic persuasions. That is also something to bear in mind—what this polarisation and fragmentation could mean and whether it can be stopped.

Q20 **Brendan O'Hara:** On that, if China, for example, is so far ahead of us that it is rolling out the rules, and the global south are accepting that they are now playing by the rules that have been set by China, how do the UK and other democratic countries react to that? Do they set up their own rules, in competition? Or do they accept the rules that have been set because they are already in place, and try to get agreement about their implementation? How do we react to what China is doing in rolling out its rules on AI?

Dr Hüsch: I wouldn't be so passive. Yes, China has a first-mover advantage, but the regulation that we have seen come out of China so far is on quite specific aspects of AI technologies—for example, on generative AI or on deep synthesis, what you use to make deepfakes and so on. China comes at it from a very specific angle of trying to secure social order and its own regime stability.

Interestingly, there are some similar narratives. For example, China also very much wants to balance risks and limit the risks, but at the same time still see that technological innovation and economic growth are possible. That is a statement that the UK would probably say in the same way.

The difficulty is that the differences lie in the detail and what we understand, for example, as a risk. China's definition of a risk is certainly not the same as the UK's definition of a risk. I think one way to counter Chinese proposals is to give more substance to generic principles—to generic language. The current language, which is used also by China, very much stresses sovereign equality and aspects that certainly appeal to other states. The idea of stabilising social order and so on is perhaps also something that appeals to other states, particularly those of a less democratic like-mindedness. For me, it is really about filling those vague principles with specific details in order to demonstrate where the differences lie.

Q21 **Henry Smith:** Professor Boström, what do you think are the dangers of fragmentation in the regulation of AI, particularly between different groupings of states whose philosophies are at odds with each other?

Professor Boström: As with other areas, fragmentation in some cases can be a good thing in providing a rich ecosystem where many different ways are tried, rather than there just being one uniform standard across the globe, meaning that if it is misconceived, there is no way to find out because it is the only game in town. However, for certain specific application areas there are huge externalities—that is, what happens in one place does not only affect that place but can affect all of humanity. In

those particular cases where there are strong global externalities, it can be important to take a more globally co-ordinated approach. That particularly holds when there are existential risks of one kind or another, either from the AI itself or from it enabling people to do new things with chemical or biological weapons. In those areas at least it would make a lot of sense to try to work towards minimal standards.

There might also be shared interests, including with regimes that otherwise have different views on many issues. For example, China would like to see the democratisation of the ability to produce weapons of mass destruction no more than the UK, so there might be possibilities for agreement on minimal standards—if anybody wants to make an AI model available to the public for general use, some minimal criteria and testing requirements that it would have to pass to show that it cannot lend itself to those uses—on which there could be wide agreement.

Dr Hüsch: I would like to add that the narrative of fragmentation is one that we are quite familiar with in a cyber context. For many years now we have been discussing the potential of the balkanisation, territorialisation or fragmentation of the global cyber space as we know it, and the internet, but we have not actually seen it manifest to that extreme degree. That is partially because in some ways we have seen it fulfilled—for example, the blocking of access to certain websites and so on behind the great firewall in China.

At the same time, there are the different regulatory approaches that we outline—in China, the US and the EU—and like to contrast with each other, but in reality they always have to be balanced. They are not as extreme as we perhaps might like to think in a discussion like this one. Just because the EU is framed as the human rights-focused one, that does not mean that they don't care about innovation. Just because the United States prioritises innovation, that does not mean that they don't care about human rights. In the end, part of the fragmentation discourse has not quite materialised to that effect, just because those approaches are not quite as extreme as we sometimes like to make them.

At the same time, something that we have not talked about a lot yet are technical standards that avoid some of that fragmentation and foster co-operation and compatibility between different systems. The UK has been active in that area, but it is vital that you also continue to push for international standards for AI technologies.

Q22 **Henry Smith:** You have anticipated my next question about standardisation and the UK's involvement in that. Can you be a bit more specific as to what you think the UK could bring to the issue of standardisation more globally?

Dr Hüsch: Yes. We are talking here about technical standards as opposed to, perhaps, political or legal frameworks. Technical standards are there to promote integration and compatibility between different products or devices—they are the reason why you can use your phone abroad and so on. They can be precise criteria. Their strength is that they are often

perceived as technical, and so less political and less politicised in negotiations.

At the same time, however, it is important to acknowledge their importance in the sense that technical standards are never free of values. It is important to get the balance right. The UK has, as part of the national AI strategy, launched an international standards hub, which has had very positive reviews—I think an Oxford study was published last week—so there are already positive assessments of the UK's role in that and in leading the diplomatic process and the technical side of the discourse.

I will also flag that we have seen in the past that other states—for example, China—recognise the importance of technical standards and setting examples in those processes. Paying attention to international bodies, such as the International Telecommunication Union and so on, and other standard-setting bodies and making sure that they are appropriately funded, staffed and resourced to participate in those international engagements is key.

Henry Smith: Thank you. Professor Boström, any comments on that in terms of standard setting?

Professor Boström: No, I have no particular point on that.

Q23 **Bob Seely:** May I come in on that? That really grabbed my interest. Looking at what the UK can do, there are tech issues and there are ethical principles. I understand, Pia, that you are saying that tech reinforces ethics and that technical issues are not value-free, but this is something in which the UK can potentially play a world-leading role, not only because we have the universities that have had and discussed ethical principles for centuries, but because we have a significant tech sector. We are one of only three countries that has a trillion or a billion—it must be trillion, mustn't it? We have a very large tech sector, whatever the figures are. We have one of the three leading tech sectors in the world, so we have a role here, and potentially a very valuable one. Does anyone want to develop those points? I would be fascinated to listen. Also, on the idea of two-tier co-operation, do we co-operate closely with like-minded countries and then aim to have a more general set of principles with countries like China and Russia that have self-interest but may absolutely not share our view of the current world order or the ethics around it?

Professor Bjola: On this issue, this kind of collaboration is already taking place. I would like to highlight the issue of data agreements. For AI to function, yes you need an infrastructure, 5G, semiconductors, the hardware and the algorithms, but you also need the data. The data is particularly important. How is it being used? How is it being collected, stored, transferred and analysed? Data agreements in the past few years have emerged as an important diplomatic tool, by which this kind of combination of engagement with the outside world is important.

Q24 **Bob Seely:** Can I just come in very briefly? Are you talking about stuff

like takeovers in mobile phones, which could potentially allow lots of data transfer to China? Are we talking about that sort of thing?

Professor Bjola: Yes—what happens to the data in my phone when I am in country X? I use this data because I am going to access the local mobile phone company. It is about what happens to this data, because it could be used in different ways. The point I am making is that this kind of agreement allows you to export rules and standards, and that is happening already. For instance, the UK has an agreement with like-minded countries—with Canada and others—and the condition for that is to respect the data rules that are valid in the UK. Why is that important? Because you have already created an environment for an ecosystem that can actually develop. That is one element of that, aside from the 5G standards, the Huawei story and that type of thing—the data.

Q25 **Bob Seely:** Sorry to interrupt again. Just so that I understand, because I assume that I am the least knowledgeable person in the room on this, it is not just a question of doing a big treaty that says, "Let's be nice to each other and not use AI to learn how to attack each other"; it is actually the bread-and-butter issues of data transfer—

Professor Bjola: Absolutely.

Bob Seely: And it is about company ownership and all these things where data goes. It is these very basic issues, because without the data you cannot do AI.

Professor Bjola: You cannot do AI. It is also one step in terms of data services. It creates a lot of this bread and butter, as you put it, rightly—it is one of the constituent building blocks that can later influence the strategy towards AI.

Q26 **Bob Seely:** Are we being too lazy about that in the UK? Are we allowing all these companies from other countries like Russia and China to hoover up data?

Professor Bjola: It is the same. For instance, I think there have been some efforts in Asia. There is a battle in some of the Asian countries and between, to a certain extent, the United States and China, about how these kinds of standards are adopted. From what I have seen recently, there is no engagement with any African countries from either the EU, or the US, or the UK on this particular aspect—data agreements.

Q27 **Bob Seely:** Very briefly, when people say "Oh, the Chinese have got my data"—apologies: I know that is very sloppy way of putting it—the issue is not just that the Chinese are hoovering up data everywhere, but that they are doing it so they have the basic building blocks to then add AI and start making decisions on manipulating human beings, and how they do it. That is the issue: not just having the data now, but what you do with it in the decades to come, if you can keep getting it.

Professor Bjola: I cannot speak about the second part—about whether the Chinese do anything about the data. What I am saying is that once you create an environment with this kind of room for storing, analysing

and collecting data—once it is created and accepted—then of course the next step, later, about how to introduce AI tools is already set in place.

Bob Seely: Okay. So it is the raw materials for your recipes.

Chair: But it is enabling a kind of data colonialism. That is the risk we need to guard against.

Professor Bjola: Absolutely.

Q28 **Chair:** We have a vote coming, so I am going to move on to our last question. Do you think there should be a global hub for AI regulation, and could steps towards that be something we can hope for from the AI summit that we will see on 1 and 2 of November? Michael, can I come to you first? Any thoughts and comments on that will be very welcome. I then invite everybody to make some final comments and reflections.

Professor Ambühl: Thank you very much. Indeed, this Government questioning the digital space is very important. The problems have already been described. We have large tech companies playing an important role: they drive digital transformation at enormous speed, and at the same time they benefit from an inconsistent, or non-existent, regulatory framework. Governments, or states, should up come with concrete steps, and it is very good that this AI summit is coming. I believe that the like-minded Governments should try to play an important role and, if I may say so, not leave it only to the big guys like the US, China and the European Union.

Chair: Perfect. Professor Boström, anything to add to that?

Professor Boström: It is difficult. Some articulation of general principles that serve to stake out the general direction, even if vague, would be good on the one hand. On the other hand, a few particular, narrowly focused agreements—I keep bringing up this thing about biological weapons, and a few other closely related things—would make sense.

Then there would have to be processes whereby the much wider range of impacts from AI could be continuously discussed, and regulations developed over time. But it is such a fast-moving field that I don't think you could, right now, lay down a 200-page detailed plan for how to govern AI globally, because it is going to be a different beast one year from now. So even aside from the diplomatic challenge, the fast-moving technological field makes it very hard.

Dr Hüsch: I would appreciate an international body that addresses AI regulation, as well as knowledge change, regular meetings, dialogue and so on. However, whether that is now in the form of the "AI safety institute" that it has been rumoured will be launched or announced at the summit, or other international fora, I do not see as so decisive. There are existing mechanisms, groups and efforts. Rather than just adding to the list of potential meeting places, we should try to foster resources. We should make sure that existing mechanisms are well resourced and staffed, for example, rather than adding to already full calendars with more agenda points and more meetings, ultimately duplicating efforts.

Time is running out and resources are always limited, so fostering efforts and concentrating on specific initiatives is better than just adding to the list.

Professor Bjola: I think these kinds of summits are productive, because they link together people to develop a common language—a common language for what AI means, what the risks are and how to mitigate the risks. That is one element in which summits could make a solid contribution. I see the potential to connect that to the larger efforts. There is a UN digital compact, which is supposed to be finalised and presented by next year, so I think there is an opportunity, once we develop this kind of common language, to connect things further and make sure that the risk of fragmentation, which is probably difficult to avoid, does not set us back in the sense of connecting with the larger efforts of international co-operation. The example that I mentioned—the digital compact, which is being prepared for the Summit of the Future in September 2024—is one example of how these kinds of efforts could be co-ordinated or integrated.

Chair: I am afraid that the votes are almost upon us, so I must bring the sitting to a conclusion. Thank you so much, all of you, for what has been a hugely enlightening session. It is clear to us as members of the Foreign Affairs Committee that this is not only a novel field of foreign policy but something that the Committee will have to return to again and again. I am very grateful to you all for giving us such a flying start on this topic.