

Professor Yuval Noah Harari In conversation with Lord Hague of Richmond

RUSI, 13th November 2018

Lord Hague of Richmond (WH): As chairman of RUSI it is a great pleasure to welcome you here tonight to the Royal United Services Institute and to the Duke of Wellington Hall specifically. The Institute was founded in 1831, with the Duke of Wellington our first chairman and we claim to be the world's oldest think tank. We have here tonight one of the world's newest and deepest thinkers, to do some 'thinking' with us at this think tank; it is a great pleasure to have Professor Yuval Noah Harari with us tonight.

We will range more widely than the normal subjects of conflict, defence and security on which we specialise – although no doubt these will feature in our discussion – because a discussion with Yuval is a discussion about *everything* that human beings have done, or are doing, or might do in the future. He has gone from specialising in world history, medieval history and military history to writing three acclaimed books about most of the things that we human beings get up to.

In *Sapiens*, his first book, he described our history and how we as humans drove other species to extinction, and much else besides, and having passed through a cognitive revolution, an agricultural revolution and a scientific revolution, we are now on the brink of the end of Homo Sapiens, as we move from natural selection to intelligent design.

In *Homo Deus*, he described what our future might look like. How we might become unaccountable gods wreaking havoc on our own eco-system – which we have already become in many ways – and that we might divide as a society, as some become super human beings. Now, in his latest book, *21 Lessons for the 21st Century*, he turns to today's most urgent issues.

So, Yuval, welcome to RUSI and I want to start with a point that is a very familiar subject to everyone here but that you articulate very clearly and mercilessly at the very beginning of your latest book, which is what's happening to liberal democracy. You write that liberal elites are in a state of shock and disorientation. Here we are dealing with Brexit, with President Trump, with the Italian government – we are definitely in a state of disorientation.

You write that in 1938 humans had three stories to choose from, and by 1968 they had two, then in 1998 there seemed to be only one, liberalism, and from 2008 we are down to zero. So let us start with expanding on that idea and for those of us still attached to liberal democracy, what hope is there?

Professor Yuval Noah Harari (YNH): There is still a lot of hope, that should be said from the beginning. We talk a lot about stories and maybe the first thing to say is that humans, homo sapiens, are story telling animals and think and talk about the world and our lives in terms of stories; very few people think about the world in terms of numbers, equations, or statistics. The basis for all human cooperation is belief in shared stories.

Stories are extremely important, at least in my opinion, and when we look at the broad history of the last hundred years we see a shift from the situation where politics was a battleground between huge visions about the future of human kind; huge stories that claimed to explain everything that happened in the past, everything that is happening in the present and what will surely take place in the future.

You had three such stories – the fascist story, the communist story, and the liberal story – and they got knocked out one after the other, until at the end of the century you had just one story standing, the liberal story. That caused a lot of people all over the world to think 'okay, that's it, that's the end of history'. Not in the sense that nothing else will happen anymore, that time will stop, but in the sense that we know it all; we now perfectly understand the past; we understand where we are and we know we are heading towards a world of

globalisation and liberalisation, democracy as the dominant political system and the free market as the dominant economic system; borders and walls will get lower and even be erased completely, until all of humankind becomes a single global community.

Lots of people believed that, not only in the core western countries but all over the world. But quite rapidly over the last ten years people lost faith in this story, not only in the developing world but above all in those core countries of Western Europe and North America. Some years ago some western governments were still trying to force this vision of history on other people around the globe, even at the price of war. Now they are not even sure of this story in their own countries.

It is the most shocking and terrifying thing to switch from having one story to having zero stories. When you have two or three stories competing with each other you still do not know who is going to win; you still have doubt. But to have just one story you have certainty; there is just one thing that explains everything and there is just one thing that can happen. This is the point of maximum certainty but makes switching to having zero stories much more frightening because you do not have any explanation about what is happening in the world and you do not have any vision for the future.

I think that what we are seeing now in the world, above all else, is a vacuum of stories and visions. This is partially being filled by old stories which seem to be making a comeback, like nationalism or religion, but they do not really offer a vision for the future. What characterised fascism, communism and liberalism is that they really had a vision for the whole world and humankind; maybe not a good vision, but they had a vision.

If you look at the rise of nationalism today it simply has no idea how to organise the world as a whole. Nationalists can have very good ideas sometimes, about how to run a particular country, but how should the world as a whole be managed? There is still a huge vacuum there and the most positive I can see is that they think in terms of networks of friendly fortresses; that you will have each country building walls and digging moats around themselves to defend their unique culture, identity, and ethnicity, but countries will still be able to cooperate peacefully to some extent and will not be a return to the jungle of everyone against everyone.

But the problems with this vision of the world as a network of friendly fortresses are, firstly, that fortresses do not tend to be friendly, as each fortress naturally wants to increase its security, territory and prosperity, even at the price of what other fortresses want and without some global values and organisations you cannot reach an agreement and the idea of friendly fortresses rapidly becomes warlike fortresses. The other problem is a new problem, that the world now has three major challenges which are global in nature and simply cannot be solved within the walls of single fortresses.

These are nuclear war, climate change, and technological change, particularly Artificial Intelligence (AI) and bioengineering. You cannot stop nuclear war just within the walls of one fortress. You cannot prevent climate change just within the walls of one fortress. And you cannot regulate AI and bioengineering within just a single nation because you do not have control over scientists and engineers in other countries and nobody will want to stay behind and just restrict their own development if the other fortresses are not doing the same thing. This is why I think there is no real vision there, just something that fills the vacuum.

WH: Your point is that liberalism could have coped with the first of those things, but it does not have an answer to climate change or technological disruption and so, even this cherished liberalism that has been the mainstay of our politics for so long, when faced with these is in a bad state?

YNH: Yes. You can make some slight corrections to the liberal recipe, such as people not liking too many immigrants, so you can have stronger regulations on immigration and go back to 'the good old days' of the 1990s or early 2000s, but in the long term the deep meaning of the ecological crisis and even more so the disruptive technologies mean that we need to rethink the foundations of the liberal order.

In terms of ecology, the main problem is that liberalism is premised on the promise of economic growth and that you can satisfy everybody because the pie will just keep growing. It is not impossible but it is difficult to continue economic growth if you want to simultaneously fight climate change and ecological disruption. It is not impossible, but it demands a lot of effort.

When it comes to disruptive technologies, clearly democracy is in a much worse position because the implication of the new technologies, especially AI and bioengineering, is that they undermine the most basic assumptions of the liberal order, about human free-will, about individualism, about these basic slogans that ‘the customer is always right’ and ‘the voter knows best’.

I still think that, of the options available in 2018, liberal democracy remains the best option because it enables us to rethink and question our foundational values and norms and try to construct a new system in the most open and accountable way. I do not think we can maintain the order as it is, but it remains the best order out of which to try and take something.

WH: You give us really good advice in this book, which is that we should be aware that we are bewildered, but not panic – which is what I often felt like saying to colleagues in government – but you go on to almost make us panic, with this point that the merger of information technology and bio technology is quite likely fatal to democracy.

This is quite a hard point for people to grasp because we can see lots of things that are damaging to democracy and to a former politician’s eyes that includes the way in which social media has developed, which has really fractured how people deliberate, which is no longer done together in a democracy. But you are introducing a much bigger and more worrying argument, that this technological change, which few people want to stop, could be fatal to democracies. Why do you think that is the case?

YNH: It will be fatal to democracy as we have known it. The big advantage of liberal democracy above all other systems is that it is adaptable and can reinvent itself. But for the way we have thought about liberal democracy for the last two centuries the merger of info and bio tech is fatal because of what happens when an outside system, an algorithm, knows you better than you know yourself, how you feel, can predict emotions and manipulate them, as well as predicting your decisions and make them on your behalf, and this is true of the marketplace where corporations know your preferences better than you and can predict or manipulate your choices. More and more of the crucial decisions in people’s lives will be taken for them, like what to study, where to work, whom to marry, or for whom to vote.

When it comes to politics, the government could know you better than your mother. The idea that somebody else knows you better than you know yourself is something that we all experience as children, such as being told that you are not angry but hungry and should eat something. This is very good when you are three, but what happens when you are thirty and it is the government, and all kinds of government not just liberal democracy?

People say this could never happen because people have mysterious things like souls, spirits and free will. But I think this is Eighteenth Century mythology, which held on for two hundred years because there was not technology to do it, but very soon we will have the technology to do it and it will force us to rethink the fundamentals of things like the free market and democratic politics.

WH: It is quite bleak some of what you are saying, to put it mildly, but one of the great things about your writing is that you have very memorable phrases and metaphors that stick in the mind. The one that stuck in my mind in particular, was the ‘data cows’; that the cows we currently see in a farmer’s field are the descendants of much more adventurous animals and we have reduced them to this state, just standing in a field and processing food for us. Now we are doing this to humans, you say, and are turning us into biochemical algorithms that process data, taking in emails and tweets, before putting more out again, and as such are just

part of the data flow and no longer autonomous individuals, but becoming the ‘data cows’. What do we do about that because we cannot stop technological development, so how do people looking for policies and regulations try to deal with these things?

YNH: There are two different questions; one about articulating an alternative vision and the other about how you get enough people and countries convinced of your vision to make it work. In terms of the first step, I do not think that it is impossible, but you need to acknowledge the biological realities of human beings, how they make decisions and where human desires and choices really come from. But also the enormous potential, for both good and bad, of new technologies to really ‘hack’ human beings; the same technology can be used to give people the best healthcare in history, from some system that monitors your body twenty-four hours per day and identify all sorts of illnesses from cancer to flu at the very beginning, when it is easy to take care of them. It is a system which can help people make better decisions in their lives; again, when we talk about making decisions about what to study, who to marry, or who to vote for, people often make terrible decisions.

That can be a good thing if the system works *for* you and helps you to make decisions like what to study in college; very often people make bad decisions and waste years, lots of money and energy, because they have distorted views of themselves and they do not understand the field into which they are entering, whether ballet dancing or being a lawyer, or do not understand what it will take later on in their career. So, enabling better decision making could be very good.

Of course, it could also be used for very terrible things, such as that same system recognising early signs of cancer or flu could also mean that citizens in some countries like North Korea may have to walk around with a biometric bracelet which constantly monitors how they feel and if they happen to see a speech given by Kim Jong Un and the bracelet recognises signs of anger, that could be the end of them. That is technologically feasible within five to ten years and there are all sorts of governments around the world experimenting with the aim of creating a total surveillance regime.

This would be far more extreme than George Orwell’s 1984 because it is a system that can look *inside* your body. In 1984 the system monitors only the outside, as I recall, such as where you go, who you meet, and what you say, but what was going on inside the brain was off-limits. However, very soon we will have the technology to really achieve a genuinely total surveillance apparatus, in which you can surveil the whole population down to the level of what’s happening to blood pressure and brain activity every minute of the day.

The big question is how we use this for good and not for evil. It is not just about monitoring individuals in the service of government, but you can flip that and use it to monitor the government in service of citizens. The technology is neutral in this sense and can go either way, but the question is how we formulate the good vision and then convince enough governments to cooperate.

WH: So, our lives *will* be governed by external algorithms, but that could be under the control of an authoritarian government, big corporation or even AI where no one is quite sure where it lies any more. You go on to tell the story about what happened on 7th December 2017 and most people have not noticed yet because they were going about their business, or watching the news about Brexit, so please remind us what happened when the AlphaGo program started to play chess.

YNH: Almost every month we hear about another achievement of AI, so one of the latest headlines was when some new software for playing chess defeated the previous best software for playing chess. That does not sound like it should be news because it has been twenty years since a computer beat the human chess champion of the world Garry Kasparov, but what was amazing about the new software is that it taught itself how to play chess.

It started from zero, with just the basic rules of the game and was not taught anything about moves or strategy and had no access to previous games (i.e. the centuries old bank of knowledge about playing chess), it just

played against itself, teaching itself chess and then going on to beat the world champion. But the really interesting thing is how long it took it to go from zero knowledge to complete mastery and top of the world – four hours.

Centuries of human's playing chess and passing knowledge and it took only four hours to be defeated. It is still a long way from playing chess to taking over far more complicated tasks in the real world, but the writing is really on the wall. As you say, I do not think we will reach a point when we have computer overlords, but the more realistic and scarier scenario is that we might soon reach a point when all the powerful positions are occupied by humans. You may still have a Prime Minister and CEOs, but they choose from a menu written by AI. For instance, the financial system is likely to become so complicated within the next twenty to thirty years, and so fast moving, that no human being is really able to understand the global financial market.

You will have a situation when an algorithm goes to the Prime Minister and says 'sir or madam, we have a problem, we are facing financial meltdown and I cannot really explain to you why because you are a human and will not really understand it, but you have three options of what to do, but neither can I explain them to you because it is too complicated and you need to decide now because it works in milliseconds'.

You would still have a human 'figurehead', but in more and more fields, from finance to terrorism identification on the basis of more and more complicated patterns, you will reach the point when an algorithm tells a defence minister that 'this is a very dangerous terrorist but I cannot really explain to you why I have reached this conclusion, just trust me. This is not some bizarre science fiction scenario because we are in some fields very close to that point.

WH: You tell a story in your book about a Palestinian who was arrested after being identified erroneously in this way.

YNH: The Occupied Territories represent, in many respects, a laboratory for such things and this is why Israel is one of the leading countries in the world for the development of surveillance technology. There are glitches, like this case, when somebody posted on a Facebook page and the algorithm mistranslated the message 'good morning' as 'let's kill them', which are two words that are very close to each other, and they went and arrested the guy. This is a tiny, tiny example of what is happening right now, but you can really run with your imagination of where this can lead us in twenty or thirty years.

WH: Thinking of RUSI's normal work, this can apply to military technology. In a democratic government it is normally a human decision to use force, but soon we will find that the next generation of war planes does not need to have a pilot, or a ship could be attacked by a swarm of a thousand networked objectives, which there is no way a human reaction could defend against that and you have to have AI controlled defences. You say in your book that autonomous weapons are a disaster waiting to happen. Is there anything we can do about that because we would have to stop almost the whole development of military technology to prevent that?

YNH: Not necessarily, but certainly we are entering an AI arms race which is leading us very fast towards a world of autonomous weapon systems. I think it could be a disaster for humanity, especially if these weapons are not in the hands of responsible democratic governments, but either dictators or terrorist organisations, criminal groups and so forth. The only real way to stop it is through strong global cooperation. This is one of the examples I mentioned earlier, in which we need strong global cooperation to deal with technological disruption.

The easiest example to give is 'killer robots', or autonomous weapon systems; it is very clear that you cannot regulate these kinds of technologies on the level of a single country or even a few countries. If you have a treaty against killer robots and autonomous weapon systems only between European countries, then very soon they will break their own ban because nobody will be willing to stay behind if the Chinese and Russians are doing it – we would be crazy not to be doing it ourselves.

It is not enough to sign a treaty banning autonomous weapon systems, like treaties on nuclear weapons, because with nuclear weapons it is still quite difficult to develop a serious nuclear programme that is secret; you cannot develop nuclear weapons without the world knowing about it. But with AI and autonomous weapons systems it is much easier to have a completely secret programme and you can also develop it in civilian contexts, claiming it is not a killer robot but a self-driving vehicle.

So, for this we need not only a treaty but real trust; it is not impossible and you can have such trust even between former enemies, such as France and Germany today, one hundred years after the First World War. The French and Germans trust each other enough that if the Germans tell the French 'trust us, we do not have a secret laboratory developing killer robots to invade France' then the French will trust them and have a good reason to trust them.

It is very difficult to see China and the US have the same level of trust and they are certainly not heading in that direction at the moment, but unless we have this kind of trust it will be almost impossible to prevent an arms race in autonomous weapons systems, which could be a disaster for humankind.

Another one of the issues with autonomous weapons is that, unlike nuclear weapons when unless there is an all-out nuclear war there is not much that you can do with them, autonomous weapon systems are not just waiting to be used in a doomsday scenario, they can be used in a lot of other kinds of scenarios. The kind of 'mutually assured destruction' that prevented the use of nuclear weapons since 1945 is unlikely to be relevant for autonomous weapon systems and once we have developed them they are much more likely to be used in a variety of contexts, unlike nuclear weapons.

WH: Overall you also tell us to get war and terrorism in perspective and that the proportion of people in the world who die from violence is now smaller than it has ever been and far more people are killed by sugar than gunpowder. You also ask us to put terrorism in perspective and another one of your metaphors that sticks in my mind is that 'terrorism is like a fly in a china shop, unable on its own to push over a single piece of crockery, but it can get into the ear of the bull that is sitting in the china shop and make that bull so angry that it stands up and wrecks the china shop'. This is quite good advice for governments and media organisations to have such things in perspective, so would you like to expand?

YNH: There is a big difference between terrorism and warfare. Terrorism, in the shape that we have seen it so far (as autonomous weapons could also prove a gamechanger in the field of terrorism), has demonstrated that the ability of terrorists to actually kill people, wage war and take over states has actually been extremely limited. The number of people killed is minuscule compared to air pollution, car accidents, or obesity. In western Europe more people have died of nut allergies in the last ten years than terrorism; you should fight the nuts before you fight terrorism.

Part of this is because of the immense effort put into fighting terrorism, but there is also the reverse logic there because the power of terrorism comes from the response of militaries which are like the bull in the china shop. If you look at the history of the Middle East in the last twenty years then you can see that it was not the terrorists who 'wrecked the china shop', it was the bulls who became enraged by the terrorists.

As regarding war, despite the increased levels of tension in the international system over the last five years, we are still living in the most peaceful era in history. As a military historian specialising in the middle ages, what amazes me about things like Brexit and the Scottish referendum is how peaceful everything is so far. In previous centuries, to make a decision like whether Scotland should be led from London, or whether Britain should be part of a huge European political system, it could only be decided by a major war, with thousands or maybe millions of people killed, wounded, or displaced. But now, so far, it is decided by very peaceful referendums and elections, in which a few people may be killed by fanatics, but not millions and you do not need to wage battles.

Also, if you look at the rise of right-wing populist parties – like in Hungary, Poland, and Italy – so far they are still far, far more peaceful than their predecessors a century ago. If you listen to Viktor Orbán you need to become very worried if you hear things like ‘the wicked Romanians are holding the sacred Hungarian land of Transylvania and we need to go to war to reclaim them’ and if you see hundreds of thousands of Hungarians volunteering to die in the trenches in order to capture those Hungarian lands. But so far that is not happening, although maybe in five years we could see something like that, but it is similar to Italy and their involvement in the First World War.

They entered the war in 1915, mainly in order to conquer or regain the lost Italian territories in Trento and Trieste. That was quite a small piece of land, today partly in Italy and Slovenia, but the Italians lost 500,000 soldiers killed and more than 1,000,000 injured. Do you think that Italians today would be willing to have 50,000 soldiers killed to capture a few dozen kilometers from Slovenia? So far it does not seem like it, so with all the frightening rise of the nationalist demons they are still nothing like what they were a century ago. It can become like that very quickly, but so far we are still in a much better situation.

WH: So, the danger is not so much that the world will breakdown into violence, but that in living longer and often *more* secure lives we will become imprisoned in this ‘matrix’, as part of the data flow, and possibly divided in a different kind of way. We may see the upgrading of some humans, through biotechnology, who then have less need of all the other humans in the age of AI and that a new divide opens up within society between those wealthy enough to be upgraded and those who are no longer relevant?

YNH: This is one of the long-term dangers, that technology will, on the one hand, make it possible to start enhancing and upgrading humans, and on the other hand, especially with the rise of AI, will make more and more humans economically unnecessary, effectively ‘useless’ and therefore also politically powerless. Different parts of humanity might have different futures and we might see a process of speciation within Homo Sapiens.

This is a long-term danger, but the immediate danger is that we will get embroiled in these nationalist battles and we could see a return to the extreme nationalism of the Nineteenth and early Twentieth centuries, and that involves two big problems. First, that we will have all the same problems of the First World War, but in addition we will have the problem of not being able to do much about the new technologies.

The next twenty to thirty years are crucial for our ability to regulate the new technologies and prevent the worst outcomes, whether it is autonomous weapons systems, massive unemployment and the rise of a ‘useless class’, or the rise of bioengineering and enhancing humans in all kinds of frightening ways. We will not succeed in preventing these scenarios if we spend the next twenty years fighting about the borders between different nations and levels of immigration.

WH: So we should be focused on the ownership of data, as the big upcoming political question, along with the transparency of algorithms? Understanding these problems may allow us to get a grip on them before it is too late?

YNH: Yes. One of the biggest questions is the ownership of data, as data becomes the most important asset in the world. If in ancient times it was land, so a lot of political conflicts were about the ownership of land. Then machines replaced land as the most important asset in the last two hundred years, so a lot of politics was a struggle about who controls the machines, factories, mines, rail system, and electricity system. Now data is replacing machines as the most important asset and so one of the biggest political decisions now is ‘who owns the data?’.

Yet most people, parties, and governments are hardly even aware that this is the question. So, data is being harvested and monopolised by a few corporations and governments, whereas most governments, political

parties, and citizens are hardly aware that this is happening or that it is the most important thing that is happening right now.

WH: Before we open it up to the audience I want to ask you about animals, as you write about there having been at least five other species of humans and we managed to get rid of them all, then we got rid of a lot of other things. You also write a lot about how we mistreat other animals today, particularly farm animals, and you argue very convincingly that in other mammals, as in humans, there has to be a great bond between mother and infant and that is something we rupture every time we undertake modern farming methods, tens of billions of times over. Do you think people will look back in a future age, as we look back on the middle ages, as this being a barbarous and ignorant time when we did not really understand what atrocities we were committing?

YNH: It could be, but again I do not know who will dominate the world in a century, but certainly if you think in terms of what kinds of atrocities are committed today in the world and you then look back to previous atrocities like slavery and you think 'how could people do it?'. The people who wrote the American Declaration of Independence, many of them had slaves. How could they just not see the gap there? You had very intelligent people who did not see the gap.

What is happening today, that is like the way in which we treat other animals, so there is a chance that in a century we, or other entities, will look back and ask why good, compassionate, intelligent and otherwise wise people just could not see that what they were doing on a daily basis to billions of animals because it is not isolated cases but an entire system. This is especially the case as the scientific discussion is basically over and while you can still hear some people who are not scientists claiming things like 'cows don't mind' because they do not have consciousness, feel pain, or even if they do they do not have emotions like bonds between cows and calves, which is just the humanisation of cows.

But now there is scientific consensus, at least in the relevant fields of the life sciences, that certainly basic emotions like the bond between mother and child is common to all mammals and probably to many other species. The scientific discussion is over in this respect, but as in many other cases, what is obvious to scientists is still far from obvious to the general population, politicians and the market.

WH: If we agree with your diagnosis on all of these issues, about how we should live differently and think about what we are doing in the future, then really we have to do two things: we have to outrun those algorithms in our own minds, investing in human consciousness in whatever way, whether meditation or whatever we can as an individual task; but we must also globalise politics to find global solutions, which is why nationalism is not the answer.

But on that point, one has to admit that politics is not currently going in a constructive direction and is probably going in the opposite direction. What if political leaders refuse to globalise themselves, does technology give us all the chance to do something anyway? If five billion people, who are connected now technologically, all decide that they are going to do something about ecological collapse, the treatment of animals, whatever it may be, then is there any hope in that and this redefined liberalism may be able to call on that?

YNH: First, I want to emphasise that there is no contradiction between this global thinking and nationalism, it depends on how you understand being a good nationalist and patriot. If you think that being a good patriot means taking good care of your compatriots, their security and welfare, then I think good patriots today should be globalists because the only way today to really protect the security and prosperity of fellow citizens is through global cooperation. So, I think it is very dangerous to think in these either-or terms, that it is either about nationalism or globalism, when today really good nationalists should be globalists.

WH: But we are getting a bit less global in our unity and outlook aren't we, whether we are nationalists or whatever we are?

YNH: If the political system is going in one way, can the people just use some technological platform to unite and do something else? I do not have a strong belief in that, as I tend to be skeptical of technical utopias, that suggest we can just invent an algorithm or platform and circumvent the entire messy political process and system. It does not work like that and we have enough experience from the last twenty years, that many of the utopian visions of the 1990s about what the internet will do, in terms of connecting people, turned out to be completely wrong.

Even when you have partial successes, like when you look at the Arab Spring, you had something like Twitter and Facebook revolutions, which were used to get a 100,000 people into Tahrir Square and topple Mubarak, but the problem then was that old fashioned politics was needed to fill the vacuum and the new platforms were not up to it, so you got two very old fashioned organisations, the Muslim Brotherhood and the Army, filling the vacuum and it became a struggle between them. So, if you want to have a one-time demonstration, or even a revolution, then yes maybe you can have a technological shortcut, but the really difficult part of building a political mechanism does not have such shortcuts.

WH: I think I agree with you, but I am looking for some hope. The politician's mind says that there has got to be an answer and there is a great prize to the one who finds that answer.

YNH: Well, to finish this part on a hopeful note, I would just say two things. We just marked the hundredth anniversary for the armistice that finished the First World War and if people from back then could see the situation of Europe today, despite Brexit, despite the rise of populism, they would say that this is an absolute miracle. So, looking at the last hundred years of European history can give us a lot of hope that people can rise to the challenges.

Similarly, if you think about the last time that we faced major technological disruption, in the shape of nuclear weapons, in the 1950s and 1960s, maybe in the same think tank, sitting right here, you would have had somebody talking about the nuclear age and a lot of gloomy people in the room being convinced that 'we all know how it ends', with a nuclear Armageddon that destroys human civilisation. But it did not end like that and the Cold War ended peacefully. So, I think it is not inevitable, but it is possible that we will also manage to rise up to these new challenges.

Question: How linear are these projections? I am reminded, when you mentioned the First World War, of Norman Angell and his lovely book on *The Great Illusion*, which four years before the Great War predicted that there could not be any war because economies were so interrelated. He got a Nobel prize but not much else. You did define certain ways around it, in terms of collective action, but how robust are your projections and did you take into account that the linear projection may be the wrong one?

YNH: I do not believe in linear projections and history certainly does not move in a linear fashion. What I try to do in my writing is not to predict the future, which is both impossible and futile, as even if you predict the future accurately you cannot do anything about it; what is the point of recognising something that is inevitable but you can do nothing about? What I try to do is map different possibilities in the hope of influencing, even a little, the choices people make, in order to prevent the worst outcomes.

So, when I say something like AI might drive millions of people out of the job market and create a useless class and an immensely unequal society, this is not a prophecy about something that is definitely going to happen, but a very dangerous possibility which we can do something about today by adopting the right policies. It is the same when we talk about the threat of war; I do not think a new global war is inevitable, or impossible.

Norman Angell was correct in saying that it would be a terrible disaster for humankind to enter into a global war. He was absolutely correct about that. It is just very unfortunate that humans tend, very often to make very bad decisions, as individuals and as collectives.

What strikes me about many of these decisions is how unnecessary they are. If you look at the Second World War and think about what went through the heads of the Germans and the Japanese. One of the most amazing things about that war is that twenty years afterwards the losers were more prosperous than ever before, so why did they need the war? If you can be so prosperous after the greatest defeat in history it is obvious that you did not need the war in order to be prosperous. They just made the wrong decision because they thought that without conquering China or Russia they would never be able to prosper economically and they were a hundred per cent wrong.

Question: You do not talk much about the distribution of power. The digital age has democratised power, to some extent, inside societies. What people do collectively counts more. What do you say about education and would it not be possible to imagine a more optimistic scenario where decent education means that people take more responsibility and make better decisions? That does not lead to a doomsday scenario, but a better and more interesting society and I would have thought that the fourth industrial revolution is going to increase the economic and political power distribution at many levels of society instead of just at the top.

YNH: There will definitely be new jobs as a result of the AI revolution, but the big question is whether all the people who lose jobs will have the ability to educate and retrain themselves to fill those new jobs. Some people might, but not all people, and you could have a decision where some countries have both the economic and social capital necessary to retrain their populations to take advantage of all the enormous opportunities of this revolution.

But other countries might be left completely behind because the main thing that enabled countries to progress in the Twentieth Century was cheap manual labour and this will not work in the Twenty-First Century. So, you could have an even more extreme division of the world than what we saw in the Nineteenth Century with the industrial revolution.

The other point is about individuals taking responsibility for decisions in their lives. There is an optimistic scenario where AI and these monitoring systems, constantly tracking how we act and feel, actually empowers us to make better decisions. For example, again, when making the decision about what to study in college, maybe a system that has been following you from childhood could present you with a far more realistic view of your abilities, potential, opportunities in the job market and the difficulties of the different training courses and then you can make a much more knowledgeable decision about your future.

But that depends first on whether we develop the right kind of AI systems and second on whether people have the education and the motivation to use them. The danger is that people might just become lazy and learn to follow whatever recommendation the system gives them. Taking a real example, one of the most important abilities of humans is to look for information, but today a lot of people have completely renounced their personal abilities to search for information and just follow slavishly the algorithm, so 'the truth' is just the first three results of the Google search engine.

The only way they know how to search for information is just to ask Google and it does not have to be like that because the internet gives us immense abilities to look for information, but a lot of people do not utilise those abilities and their skills of looking for information are actually far lower than fifteen or twenty years ago.

Question: I particularly enjoyed how you started by saying that so much of history can be explained by the declining number of stories, but one of the stories in our field that remains very powerful is nuclear deterrence and it is one that is shared by communists and liberals, India and Pakistan, Israel, and North Korea. But I was reflecting on the impact of AI and decision support in that area and would be deeply pessimistic about the chances of international cooperation on killer robots, but also in relation to nuclear weapons.

It reminded me of the 1984 film *War Games*, when a teenage hacker gets into the Pentagon computer and thinks he is playing a computer game of global thermonuclear war but has actually started one. Yet the film ends with the computer running all the options multiple times and coming to the conclusion that nobody can win global thermonuclear war, so the only option is not to start. In a world in which future leaders are more reliant on machines perhaps such Armageddon scenarios are less likely?

YNH: There are many different scenarios and one is that hacking and AI are likely to destabilise the nuclear deterrence because you no longer know for sure if you have control of your own weapons or the other side has control of theirs. So, when a missile is launched from Russia at Great Britain, maybe it is the Russians but maybe it is North Korean hackers, or maybe it's a fat kid in Macedonia. So, what do you do then? Nuclear deterrence collapses if you do not know who is controlling the different weapons.

Another danger is that nuclear deterrence is based on the idea that nobody can win, but what happens if technology develops so rapidly that you gain some new wonder weapon and you think that now you can win, or the other side think that they are losing their last chance to win in ten to twenty years and that they will be so far behind that they have to act now.

The third scenario, to which I already referred, is that with AI you can have a lot of extremely dangerous and violent attacks which are not an all-out war; you can take down systems in a country and no one is sure who is doing it or why. So, if nuclear deterrence is a super-rational chess game between just two players, in which everything is clear and you know who is making every move, then the future scenario is a completely hazy situation in which lots of things are happening and nobody is sure about anything and that is a very pessimistic situation.

Question: The politicians of today are grappling with what they believe to be the biggest issues of our time. Whether we should be in Europe. Should we have immigration? Should we follow Hayek or Keynes? But AI seems to be a side issue, to be dealt with later, so if you were speaking directly to governments what would you be telling them to do now to prepare? Second, related, when you discussed the scenario where the democratic process is called into question because we have algorithms that can decide for us; democracy is the dominant political system in the west, with both its strengths and weaknesses, but can you imagine a scenario where such technology prompts a rethink about how we approach the political process?

YNH: There are many things that can be done but the first essential step is to build trust, as you cannot regulate it on the level of a single country. You need strong international cooperation and trust in order to do anything meaningful about AI regulation, so the first step is to build trust because without this we can have a lot of very nice ideas about what kind of AI to develop and what not to develop, but we will not be able to enforce it because we will be in a kind of race to the bottom where everybody says 'yes, we do not want to do it but we cannot trust the other guys not to do it, so we must do it first'.

Potential AI-enabled political systems could be something better because any technology can be used in many different ways. If the AI systems are created to empower the individuals and to survey the government to fight corruption then this could really strengthen democracy in a new form, but we need to realise the immensity of the challenge and the fact we have very little time to confront it.

Question: How do you think the relationship between the US and China will pan out?

YNH: I do not know, but one of the most dangerous developments in the world is the beginning of an arms race between China and the US, especially in the field of AI. Three or four years ago hardly anybody talked about it, but now we are really in an arms race. Certainly, the US government is far less aware of this arms race and in a curious way much of it is between a government on one side and private corporations on the other side, but the US government is also starting to realise what is happening and coming on board.

This is extremely difficult because this almost guarantees the worst outcome and again it is not inevitable if we can reach a situation of cooperation, but for that you need to build trust and at present we are going in the opposite direction.

This is one of the most frightening developments in the world and the really sad thing is that it is not inevitable or even really necessary. It is not a law of nature that there must be an arms race and a clash between China and the US, but we are on that course and at present the situation only gets worse and worse.

WH: To finish, you say that to outrun the algorithms we need to know ourselves. You say that people wonder if you are out of reality spending two hours meditating, but you say that actually that *is* reality before you go back to all the tweets, emails, and funny puppy videos, so perhaps you could expand on this before closing.

YNH: The oldest advice in the book is to ‘know yourself’. For thousands of years we have had all these prophets, saints, and philosophers – from Socrates to Confucius, Jesus and Buddha – going around telling people to know themselves better and know who they are and why they make the choices they make and what is really in control of them. This has always been good advice, which most people never followed, but throughout history you did not really have that much competition.

If, in the days of Socrates, you did not really make the effort to get to know yourself, then you were still a ‘black box’ to the rest of humanity; maybe not to your mother or your spouse, but to the rest of humanity. But right now there are all sorts of governments and corporations that are very busy trying to ‘hack’ you and that means to know you better than you know yourself. Then you reach the point that they can predict your actions and they can manipulate you, without you even realising. The easiest people to manipulate are the people that believe in free will because they believe, therefore, that no one can manipulate them.

The belief in free will was good for centuries, but now it is becoming really dangerous because it makes us incurious about ourselves. If you really believe that every decision you make reflects the free choice of your soul or spirit then you think that you know yourself perfectly. That is really a barrier or curtain to really starting to explore your reality, but if you realise that actually no, you know very little about what is going on in your mind then the next thought or desire that pops up you will ask where it has come from, why you are thinking it, or why you want one thing and not another. It is only then that you start realising that this is the result of all sorts of biological or genetic mechanisms and outside influences, which makes you much more curious about yourself.

One way to understand meditation, is that it is just a systematic observation of what is happening inside you, in order to start having a much more realistic understanding of who you are and where all these thoughts, emotions and desires come from and whether it is really from you or the influence of other things.

For me one of the most shocking things when I started meditating while doing my PhD in Oxford eighteen years ago, was that I know almost nothing about myself. It is like when you have a big test or presentation in the morning then you go to bed saying to yourself you must have a good night’s rest, but suddenly all these thoughts and worries come into your mind and you will say ‘shut up, I need to get rest!', but you cannot.

In a way this is a little like meditation, as you do not want to do it and you just want to get to sleep, but you have this stream of thoughts and worries which you have no control over and cannot tell it to just stop, or direct where it is going. So, meditation is like that but much more systematic and deeper. On courses of ten days, not just for half an hour before you eventually fall asleep, you observe all the sensations and emotions that come up in you.

For me and anybody I know who has been to do such a meditation course it is a really shocking experience, for the first time to take a really direct look at what is actually happening inside of you and this is far more interesting and in touch with reality than all the tweets, emails and funny puppy videos on YouTube.

WH: Thank you so much for joining us tonight because there are many people here, like me, who have read hundreds of books, pamphlets, and blogs but despite the fact that we have read all of those things your writing really does make us think about new things and new ideas, maybe even reengineering our own brains and looking more deeply into ourselves, as you have just advocated. To do that for millions of people is a great achievement, so we pay tribute to you for that and thank you for joining us here at RUSI tonight.

END