Series name: The OII Podcast (Oxford Internet Institute)
Link to series: https://podcasts.ox.ac.uk/series/oii-podcast-oxford-internet-institute
Episode name: How Can AI Be Deployed Ethically in the Defence Sector? With Professor Mariarosaria Taddeo and Sir Chris Deverell
People: Mariarosaria Taddeo, Chris Deverell

This transcript was auto-generated using TurboScribe.ai. It has been reviewed but may still contain errors.

Transcript

Hello and welcome to the Oxford Internet Institute podcast, part of the University of Oxford. In each episode we look at issues and developments in the digital world that matter to us all. Today we're joined by Professor Rosariamaria Taddeo and Sir Chris Deverell.

Rosaria is Professor of Digital Ethics and Defence Technologies at the OII. Her research focusses on the ethics and governance of digital technologies and ranges from designing governance measures to leveraging artificial intelligence to address the ethical challenges of using defence technology. She's also a member of the UK Ministry of Defence Ethics Panel on the use of AI in defence.

Chris is a retired four-star general who is now advising many start-ups in the defence sector. He's also given evidence to the House of Lords Committee on AI and weapon systems. I'm Veena McCoole, Media and Communications Manager at the OIAI.

Welcome to the podcast Rosaria and Chris. Thank you, it's great to be here. Indeed, thank you.

So, for those of us who are less familiar with the world of defence and warfare, Chris can you share a brief overview of the history of AI's impact on the sector?

Yes, I mean I think it's much less than you would think, which is you know not perhaps the answer you were expecting. The penetration of AI in defence, European defence at least, is pretty minimal. There's more of it in Ukraine I think and also in the US perhaps than elsewhere but even in those countries it's far from ubiquitous and you know this is not to say that AI is not coming to defence but it's not yet here on any meaningful scale.

There is research and there is a fair amount of what you might call innovation theatre going on and some useful demonstrations and trials and proofs of concept but hardly any AI has made it through into the core programme and there are all kinds of reasons for this. Partly it is that defence procurement is not as agile as the private sector. You know Aside from when there is an urgent operational requirement it's innately risk averse.

It's a super tanker that takes a long time to turn around. Partly it is that even in the private sector the application of AI is less than you would think given all the hype. and partly it is that the operating model in defence procurement isn't exactly optimised for making bets on small companies which are the sort of startups that are leading on the application of AI as opposed to the large tech companies which are focused on building models.

And I think another reason is uncertainty surrounding regulation.

Am I allowed to buy this stuff? What hoops do I need to jump through to do so? There are high level policy statements but that hasn't yet translated into buying confidence at lower levels. So I think perhaps rather surprisingly I would say that the impact hasn't been enormous so far but it certainly could be and I think that's why it's a very relevant topic to have a conversation around. Yeah indeed I agree.

I mean it's certainly the trend that we should be looking at is not what we've seen yesterday or today but where things are going and the Ukraine war has been a watershed in this sense because we have understood that all this data that we have produced within the military or in defence and outside and this technology AI can be a great support for the entire spectrum of defence operation whether it's intelligence, situational awareness, cyber or kinetic.

The trend is interesting because indeed there are questions which are outside of as you were the immediate theatre of operation but the things that Chris was mentioning for example the difficulties of start-ups and SMEs assessing credit and funding for their activities is one key element that is emerging in defence as hindering an injury for the development and adoption of AI. But the trend is there if you think that you know now Google, OpenAI, all major companies are either already working or announcing or opening up to the opportunity of working with defence organisations.

That tells you know if these people move in that direction that means that there is at least a list of market opening up and it is true that one of the factors that is at the moment I think creating an healthy place because if we were adopting AI too quickly I would be perhaps even more worried. It's not just the procurement that is an issue but it's also the individual responsibility of the members of these organisations. because it’s actually what you can buy, what responsibility liability you take once you adopt something that we haven't yet tested as much or there are not so much regulations

about it. We don't have yet standards in terms of assessing or using this technology so there are a lot of questions which are at the crossroads at the border with innovation and marketing but also governance of these technologies that are determining the pace of this adoption which I think is happening anyway.

In 10 years' time we will be talking about this. Yes I mean I think the Ukraine example is worth just a brief discussion. I mean so one could be forgiven for thinking that if you look at all these drones that are being used in their thousands in Ukraine that you know they were essentially lethal autonomous weapon systems or at least had some significant degree of AI in somewhere in the process but the reality is for a lot of this period the biggest use of drones has been first person view drones which are piloted by an operator to a target.

A response to that was in the form of electronic warfare which jammed the signal and so what have Russians in particular done? They have come up with a drone that trails a fibre optic cable behind it so again it's still not an autonomous system it's controlled by an operator and a lot of the drones that are fired particularly by Russia they get given a set of target coordinates and fired at them. You know they're not clever they're fairly dumb munitions they just navigate their way to a position and attack it whatever is there. So right now AI wouldn't be the dominant factor in the drone war that's going on in Ukraine.

That is not to say that people aren't experimenting and indeed in some cases using systems that have an AI component but it's far from everywhere. The drone is everywhere but AI isn't. I think that there is also another element to consider related to the drones or the Ukraine element here which is that again looking at this in perspective, all these technologies drones, but even other kind of weaponry that is used in Ukraine as in everywhere else, is now collecting huge amounts of data.

The reason why we're collecting this data is because we're going to use it to develop AI. going on so whatever let's say delay whatever slow pace we're using now we're having now for adoption of AI in defence is going to be completely overthrown or accelerated going forward because we are now collecting a lot of data with testing technology with battlefield data which we didn't have in Europe before. and that is going to create a huge push forward for the development of these technologies and so for its adoption. And just to you know show how also these elements have to do with governance. There is a governance issue now about who's collecting this data, which are battlefield data and whether these are in the hands of the Ukrainian government, or the tech companies who are providing these technologies and where this data is going to be stored and who's going to access it because if you are a US company then perhaps you're taking battlefield data back to your own country and what does that mean for the sovereignty the digital sovereignty of the Ukrainian government over data that are produced effectively by a war they are engaged in.

So there are all these questions that needs to be asked now because they are telling of again a path we have embarked on already, but also of aspects which would be crucial to make sure that the adoption of AI in defence for all the positive it will bring will also be managed, and governed and ruled in a way that doesn't breach values and aspects of defence that we actually all want to be respected going forward.

Rosaria and Chris you both raised a lot of question marks there that kind of still exist and I'm wondering what is something that most people either don't know or understand about how AI is currently used in warfare governance and military operations.

I mean Chris is more qualified than I am on this point but one thing I'd like to say is exactly what has already emerged from the first part of this discussion there is a lot of hype and a lot of sci-fi inspired narrative about AI in defence and both things are detrimental. The sci-fi aspect as in any other discussion about AI it's a huge distraction from the very you know concrete and problematic aspect that we have must focus on.

Take the governance of data, take the attribution reliability, take the standards for the adoption of this technology in defence. So the hype is to sorry the sci-fi element has to be really kept under strong control because it's just a huge distraction from serious discussions we as society need to have. And the other is this idea that with the hype with the idea of hype in defence it comes the tacit assumption that because it's already there you cannot do anything anymore about it.

And so you know it's already it's already AI based defence so whatever rules are not there we cannot do anything anymore or whatever mistakes have been made it's too late to fix them. Whereas there is a huge potential in the fact that we are in the early days of this transformation and so it means we can intervene to steer it the way we think is more appropriate for the defence forces or liberal democracies. But then Chris knows more about it.

Well I mean I totally agree that kind of killer robot hype is unhelpful but I think another way of saying that is there tends to be a lot of conversations around the risks of acting but less perhaps much less around the risks of not acting. And so what I mean by that is that in this in the area of defence that there's a lot of noise about lethal autonomous weapon systems but I don't often read cogent analysis against or regarding the other real risks that we might fail to grasp if we don't think about this deeply. And I think that unequal distribution of commentary on risks is unlike, for example, in the health sector, where you do see a lot of recognition of the benefits of AI alongside commentary about the risks that we need to manage.

So my philosophy here is that we need balance and that characterising the problem in an extreme way is really unhelpful. Yeah I think that's a very good point. I mean as an activist I always get this kind of objection.

As the ethicist who enters the room I often joke and say you're equal to this very strict parent who tells you all the things you must not do and never tells you the thing that you should be doing. But actually being an ethicist is much more like being a wise parent who tells you well don't do drugs, don't drink and drive but also make friends, go enjoy life, read books. It's a trivial analogy but to say that the point of ethics especially of ethics of technologies and AI in defence is exactly to try to strike the balance between the risks, understanding the boundaries, the risks that as our society we don't want to take but also the opportunities.

And we have to make sure that that happens now while we are still embedding this technology.

Think about privacy for example. It's not much to do with defence but it's a good analogy.

Privacy is a fundamental right but it's an absolute right. We don't enjoy privacy totally for nothing. We can modulate privacy.

So during the pandemic for example some of us we gave away a little bit of privacy in the name of public health and security. Imagine genomics and imagining having to explain to your grandchildren in 30, 40, 60 years that you didn't cure cancer because you wanted to protect privacy. You didn't cure Alzheimer's.

73 million people in Europe have been affected by it in a few years' time because you wanted to protect privacy. That's the wrong ethics. You need to find the balance.

And so it's in the same in defence. We have the principle of necessity which we cannot maintain in absolute terms because that means that we trample on other principles like proportionality distinctions but we have to find the balance. And this is where ethics becomes crucial to steer the debate on this balance.

And going back to autonomous weapon systems this is a good example because the debate on this topic which is important because it's important how we decide to take life in war for societies has been so much polarised over the past 12 years. We started discussing about autonomous weapon systems in 2012 so it's not yesterday, but we never managed to get to any conclusion because it's so polarised between those who think it's going to save the life of soldiers and those who think it's going to breach human dignity, that we never even managed to find a common ground for this discussion. Then you fast forward to 2021 and we have autonomous weapon systems more or less fully autonomous being deployed without adding any regulation in place.

So it is true that these extreme tones, this polarisation, too much ideology as it were, is not helpful when you're trying to define governance for technologies that are shaping our societies., so effectively trying to shape our societies. Yeah and I think another thing that's unhelpful, I don't know what you think about this Rosaria, but quite often we see opinion dressed up as fact. and what I mean by that is that you often cannot verify the counterfactual. So you just assert it as your position as a truth and actually it's not, I don't mean you personally, I mean there is a tendency to do this rather than to actually interrogate the evidence.

We have a tendency to describe our position as fact aligning with our view of the world and I think an example of that would be around the risks of AI being escalatory. I think we do need to attend to that risk, it exists, but not to ignore the real possibility that it could equally have the opposite deterrent effect., you know states resisting going to war because they cannot tolerate the consequences as seems to be the case with nuclear weapons. So I think it's really important to have a nuanced debate about this rather than just be contentious.

So you're both clearly advocates of a balanced approach to this. Rosaria, I wonder how this ties into the topics that you cover in your newly published book, The Ethics of Artificial Intelligence in Defence. I'm curious why you decided to write this book as well.

Well, several reasons. The book really puts together about 16 years of work in this area. So we started this conversation by saying that AI was only about to being implemented in defence 16 years ago.

It was really like an extremely long shot. And so academically speaking, I wrote this book because it was nice to see how all the pieces of my thinking about this, they actually can be consistently put together. So there is this academic element.

But there is also something that seemed to me more relevant than just my own personal ego and academic satisfaction. is that while talking with defence practitioners, whether it's the MOD, whether it's DSTL or NATO, there was always a genuine interest in understanding ethical questions and finding an ethically sound approach to deal with this transformation, but also the will of not do this in silos. So the policymaker who works on cyber and the policymaker who works on procurement, they know that there are different corners of the fence, and they're very worried that ethics becomes something that is done here, not there, or here in one way and then in another way. So there was a need for a systemic approach, as he went to defence and AI and ethics.

And so I felt that I perhaps could try to fill this kind of gap amongst this need. And also there is another element. In no other place, I think, if not in a book, you can outline all the nuances of this debate.

One of the frustrations you get if you work in this area for a long time is that the tendency is to go for the extreme opinion, for the unchecked fact, for the things that can polarise because that attracts more attention. But then actually it is the nuances, it is the trade off that we have to reach if we want to get forward with it. And this requires a little bit of arguing and elaborating.

So a book allows you that privilege, I guess. Yeah, if I may say, I think Rosaria's book is a very valuable contribution to the debate, which I commend to anyone interested in this subject. I think it's very important to consider the risks of new technologies, as long as we don't ignore its benefits.

And with groundbreaking technology, which AI definitely is, the risk of unintended consequences is higher than with things which we're well used to using, including, of course, adverse unintended consequences. And we need to understand what they are, which is important to do in a way that is not overly simplistic or emotional. And I really enjoyed reading Rosarioa's book for that reason.

If we delve deeper into some of these key ethical risks of AI in the defence industry, which we've begun to cover, which ones really stand out to you both in terms of needing really urgent regulation and management? Chris, maybe we can start with you.

I think I would start from the predictability problem, which is made famous by AlphaGo's move 37 against Lee Sedol, when the machine did something that the human had never attained, and probably never could have. And I would say that the lack of predictability is at some level inherent in AI systems.

And some have, therefore, gone from saying this to saying that we must not use AI in a military context, or we must not use AI at all. But I think it's, you know, this is where Rosaria's book comes in helpful, because she actually talks about contexts, and she describes three scenarios, sustainment and support, adversarial and non-kinetic, adversarial and kinetic. And, you know, for reasons she lays out in her book, and which are sort of obvious, the risk level of getting it wrong increases, the impact of getting it wrong increases as you move up that ladder.

So, for example, I think it makes little sense to prevent, say, a military logistic application of AI, which is allowed in a civilian context. You know, and that will certainly be the case. A lot of back office systems are already legal, and many would say ethical, in the civilian sector.

And there's absolutely no reason to my mind why they can't be used in a military context. But as you become adversarial, and you start to be having a direct impact on opponents, either non-kinetic or kinetic, you clearly have to think about that predictability problem more. What I would say is I think it's getting easier over time with one proviso.

I mean, it's definitely not an issue that's been solved, but better interpretability tools, better, smarter guardrails, testing, validation, and, you know, advances in theoretical research are all helping to address the problem. The proviso being that the models are getting more powerful, which pushes in the opposite direction. And I would say, you know, we do need to characterise and understand this risk and apply appropriate measures to mitigate it, by which I do not mean that we should envisage a world where the risk is removed altogether.

Usually, in my experience, if you pursue the removal of a risk altogether 100%, you induce new risks. And I think, you know, the example of driverless cars is quite interesting here. So, if you go to parts of California or Arizona today, you can book a driverless taxi.

And, you know, clearly the people in those regions have concluded that they are not without the possibility of accidents. This risk is not so severe as to prevent the use of AI altogether, not least because it may ultimately be the case that driverless cars are safer than human control. So, I think we need to really examine the predictability problem and understand it.

That would be where I would start. I'm not sure it really meets your criteria of an ethical risk, but it's a factor in the thinking about AI. I agree completely.

The predictability problem is the core question, because what we're saying is that not all sorts of AI, but statistically based AI by this very nature is something that can produce outcomes that the developers or the users never intended. And this could be good outcomes, alpha or beta outcomes. You might remember the Thai bot on Twitter a few years ago, which learned how to regurgitate Nazi language and all sorts of violent content there.

So, predictability is not something that happens in extreme context. It happens every day when we use some forms of AI. Hallucination of LLMs, those are the result of unpredictable behaviour.

Unpredictability is the other side of the coin of control, because if it's not predictable 100%, you don't control it 100%. In high-risk domain, limited control is problematic. Now, I agree with Chris.

The question is not how do we make it controllable, because we know that this technology is going to be to some extent unpredictable. This is something that Wiener was mentioning in this late 50s, Wiener and Summers. So, it's not something that we have never known.

We know that this is part of the nature of AI. The question is, what kind of risks as societies are we willing to accept? What kind of ethical trade-off are we going to make? This is important, because in society, we always make trade-off for many things, the environment, as I mentioned before, privacy, defence, security, surveillance. The point is where the thresholds are.

The thresholds here they concern, are we happy to attribute or not attribute more responsibility for the actions that this machine do, when this may involve an act of force or an act that implies harming another human? It's a matter of control. How much leverage, how much gap we want to allow given that something might go wrong in a context of war? So, these are questions that require an answer that implies the balance we were mentioning before. I think at the moment, we're lacking the debate that takes us to those answers.

The answers are not going to be produced tomorrow. If they were, I would be very suspicious. But we need to have a public debate on, basically, how our defence organisations are going to be shaped and informed by this transformation.

It's crucial, because I found that in liberal democracies, defence is a topic we don't talk as much as we do with finance or healthcare. Because, honestly, we all think that at some point, you know, you might need to go to a bank, and having a good healthcare system helps. Defence doesn't seem so central.

So, why don't we should discuss about this? And it's also a topic we don't feel so comfortable discussing. A defence organisation is there in our liberal democracies to protect the values of democracies, human rights, plurality, and justice. Those are the first organisations that need to be held accountable for those values.

We cannot allow this transformation to breach them. I always say that, you know, democracies will have to defend themselves by reaching these values. We might as well join sides with the other forces because we won't be different from them.

So, I think that the biggest risk is overlooking these old problems, all these problems, because we mark them as extreme or useless or, you know, possible solutions are detrimental. And then we need practical solution. Just to give concrete examples, we need standards.

How much predictability are we willing to accept? How much transparency? What type off caring, sorry, handling of data and levelling of data we're going to have? How do we collect and share this data within our defence organisations? These are concrete answers to this big high-level question. So, this is the whole process we need to create. And the other technical problem that I see emerging more and more, but I'm not original in saying this, is the interoperability.

Defence forces work in alliances. Without governance, without standards, there will be different AIs, different standards, and that will hinder the ability to have a technology that we exchange, but also being able to rely on the result of that technology to do coordinated actions. I'll just to give an example., and then I'll stop.

Imagine Italian intelligence as some sort of information elaborated on some sort of AI system that uses different security standards than the one that the UK has. And imagine that information will be useful for a UK-led operation. Would the UK forces be able to use that information or not, given that it's been produced with different technologies and different standards? So, the ethical questions, which are very high-level and they refer to justice and they refer to the values of our societies, then end up informing very specific aspects of the governance.

And this process, this translation, takes a long time. So, this is quite important that we start discussing these questions now. Yes.

But I think, you know, what I would also add is that there is a danger that that conversation takes us the next 50 years. And if it does, then effectively what we have done is put ourselves in an extremely vulnerable position, because we're going to talk about, I hope, what our adversaries are doing in this area in a minute and paralysed, you know, our ability to exploit the opportunities.

So, you know, I completely agree with the fundamental point that in liberal democracies, the armed forces need to defend and pursue the values of the societies that they represent, or that they defend. I could totally get that. But equally well, we need to be very careful that we don't spend an impractical amount of time addressing these issues.

And therefore, we have to take some risk. I think a thought experiment is quite useful here. Let us say that you could devise a system that had a 99% probability of correctly identifying and locating a target, a 99% probability of then selecting the appropriate weapon system to use against that target, and a 99% probability of then hitting the target in a way that is discriminant, proportional, necessary, and so on.

In other words, compliant with the law of armed conflict. That, I think, is a conceivable state. I'm not saying that's the state we have today, but that is a conceivable state.

What we're trying to deal with is a situation in which we don't have that level of confidence today. So we have to move quite quickly to working out what level of confidence is enough for what kind of applications. And there will be different levels of confidence, I think, required for different kinds of applications.

But if we could get to a situation where we had 90%, or 99.9%, or .99, or whatever the number is, if we could get to a position where we had that level of confidence, why

should such weapon systems not be used? They are, by the way I have defined them, better than humans.

So, you know, I do hear and agree with the importance of considering these issues, but not to the point where we are so paralysed that we don't make some use, at least, of the technology. I agree, but I don't think we will ever get to that point, to be honest.

I think that's right, in the sense that, you know, if you look at UK government policy in this area, you know, it's quite a balanced position. Some would say it doesn't go far enough, others would say it goes too far, but it has got a policy and it has published it. And so it is not waiting for overwhelming consensus on the issues, which I think is right.

The point is exactly this, that what we've seen so far is that adoption has gone a bit faster than the public debate or the regulation. And taken to an extreme, this is also very problematic, because we find ourselves, we find our defence forces doing things that perhaps we don't consider legitimate. So the tendency seems to go in the other direction.

It is a little bit problematic, or is it problematic for one element, that a policy is not a law, and there is a different level of legitimacy, so to speak. And we have to keep this in mind, that it is fine to have a policy, it's good that there is transparency, and it's good that there is, you know, with transparency comes accountability. But those are not regulations.

And this is what we're lacking at the moment. We don't have a law, we don't have, not even, you know, the only regulations, the only governance that we have internationally about AI is the European Act, AI Act, which says clearly that they don't deal with AI in defence. So what is missing is a debate that includes AI , I agree, we don't have 20 years, we don't even have 10 years to make this discussion, and we shouldn't take that long.

But we need something that has a democratic legitimacy in terms of defining and guiding... I think, I mean, I think the government would disagree in the sense that, about the absence of a law, and it would say its actions were governed by international humanitarian law, and always will be. And, you know, in that context, it has said, the UK does not possess fully autonomous weapon systems and has no intention of developing them. So... Indeed, there are two questions there.

One is the international humanitarian laws, they are valid, but their application becomes a little bit problematic when you put AI into the equation. There are some aspects which are unclear on how to apply them. So for example, the impact on AI in on the principle of necessity.

You mentioned, you know, AI having an impact on escalation or in decisions of countries or states to engage in war. We did some research showing that actually, when you think about ad bellum, proportionality, that is not accessible. You cannot really say whether AI will play, particularly autonomous weapon system, will play a role in that.

So the debate is very open on that front in terms of how it applies, not whether it applies, but how it applies. And then you had another point, sorry Chris, taking over. So, the UK government would say that... Well, so I'd like to make I make two points.

Firstly, that the government would say that it is not going to do anything which runs counter to international humanitarian law. And then also the government has said as a matter of, as a specific point, that it has no lethal autonomous weapon systems, sorry, autonomous weapon systems, fully autonomous and hasn't any intention to develop them. So this is also very interesting for a few reasons.

One is that, as we discussed so far, the question is not just about autonomous weapon systems, although this is, you know, one of the biggest aspects to consider. The second, and it's not just the UK government, is that so far there have been about 12 definitions provided by governments around the world about autonomous weapon systems. They're all very different.

And if you look at the way they define autonomous weapon systems, sometimes it's also laughable. There is one, I think it's, I can't remember where it is. China.

We should talk about the Chinese one. Machines able to understand the intention of the user. So we're talking about magic.

Of course, if you put this threshold, you know, autonomous weapon systems are machines that are able to understand the intention of the users. We're never going to develop them because they're never going to have an understanding of what the users want. So of course, then those machines were never going to be developed.

Whatever is below that threshold might be developed. So it is true that there is, let's say, a pledge of the government, the UK government, about not developing autonomous, lethal autonomous weapon systems or fully autonomous weapon systems. The definition, however, remains a little bit vague and allows room.

And you know, it's understandable in the context of international relations. But we should remember that these are the points that cannot be addressed by a policy. They require proper laws.

And those laws need to have, of course, democratic legitimacy, but also, in my mind, they need to be substantiated by a public debate, because there is a strong connection, and here is the ethicist speaking. There is a strong connection in the way we wage war and the societies we are. I take war to be war itself, not just defence.

War itself is a benchmark of how serious we are about our values and principles. I do agree with that. But I do think we should just dwell for a moment on the 2018 Chinese definition, because as you say, Rosaria, it is so extreme that it allows the Chinese to do absolutely anything, almost anything they would want in this area.

Why would they have such a definition? I mean, I think it's reasonable to conclude that the reason for having such a definition is because they are developing lots of things just short of that definition. And that should worry us. You know, I mean, that is a real risk of adverse consequences, a real risk that in the end, we have to accept the will of the Chinese government, because the prospects of defeat are so great that we can't afford to go to war over them.

We can sort of see that trend as a result of that definition. It's really concerning. It is, I should say, also French as a definition.

It's close to, not as out there, but at least as big. And I think the UK changed its definition as a result of a debate within Parliament in the House of Lords. So, I mean, it's an attempt that all governments have done.

I think we can take this discussion about the definition of autonomous weapon system as indicative of another debate, which is like, okay, why should we regulate AI in defence as a capability when the other side of the world or the other actors in the world are not going to do it?

Yeah, I'm not opposed to the regulation of AI in defence. I just think it needs to be done carefully. And I think we also need to be really thoughtful about what more regulation is needed beyond that that is needed to regulate AI in the civil sector.

You know, a lot of AI regulation that applies in the civil sector will be equally appropriate in defence and doesn't necessarily need separate policy. We need to think hard about situations which are peculiar to defence, for sure. But I think there's a distinction I would draw between regulation and law.

I think we might disagree on whether new laws are required. But I'm not opposed to regulation done well. Yeah, no, I don't think so.

I agree with you. It's not a matter of having new law, necessarily new law. Although in some context in cyber, for example, there is a gap.

It's a matter of understanding how do we implement existing international humanitarian laws when we have AI. Nuremberg trial, one of the phrases that are like the most of the notes of the trial is that they say, well, we need to be able to ascribe individual responsibilities for the crime of war, because otherwise the morality of war is not upheld anymore. Take autonomous weapon systems.

And we know that attributing responsibilities to humans, because those machines are not responsible for the actions of the machine is very problematic. We need to find a solution. It's not a new law.

The law is there. But the application of the law requires some thinking. So this is where we need to work.

And I think that, you know, we could be smart and say, well, you know, we just need to apply whatever regulations we have or learn from lessons that we have in other high-risk domains, whether it's administrational justice or healthcare to AI and import those in defence. It's fine. But there is a delta.

Defence is one of the two aspects of sectors in liberal democracies where the state can use force. That's the details we don't find covered when we think about healthcare or finance. Here we need to think.

These are the trade-offs that we were mentioning before that I think we need to understand. How much, you know, we want to, the use of AI to change the way a state can apply force and to understand. I think we need a frank discussion.

I mentioned the principle of necessity before. There is a tendency, at least that is what I've seen in my experience, of not wanting to bring that at the centre of the debate. If you look at the AI ethics principles that have been produced by UK, US, NATO, those are principles that you could apply to a bank, to a hospital, to a school.

And this is no good because they miss the point about the application of force. We need to talk openly about that the fact that we have defence forces. If we attack, we're going to defend ourselves by using force and AI is going to play a role in it.

The question is, what role? If we were able to move out of this stereotype about the use of force that we don't want to consider, perhaps the discussion would also be a bit quicker and more frank. We would be able as a society to have a stronger position rather than a fragmented, polarised debate that leads us nowhere because if you take the autonomous weapon system, we have a UN group that's been working for 12 years on this topic. Not even a definition, as we discussed, has been produced of autonomous weapon systems.

That has not stopped anyone nobody from starting starting to develop, testing, possibly using these things, for example in Ukraine. So to me, this is the key lesson that we should learn from the debate on AI ethics and governance in the past decade. I do agree that there is devil in the detail.

I think that I kind of take your point about the principles. I mean, the defence AI strategy sets out some principles for the use of AI and defence and I take your point that when you just read the words on the page, they could be applied to a bank or a healthcare institution. But actually, in the detail, there could be a lot of complexity around the use of AI.

So there is a requirement to look at the detail, but it doesn't make the principles wrong. I think those principles are sound. Principles are sound.

They are, I think, insufficient insofar as they could have addressed more openly the question of justice in war, basically. I think you're right that the devil is in the details. Indeed, the activities that, for example, the MoD is doing now, but also the DoD, is about how do we implement these principles and how do we deal then with balancing these principles with necessity in defence.

So it actually is a process, it's there, but it was brought into the debate much earlier. And also, I should say, let's say, there is no particular focus on it when you look carefully at some of the documents. If you take the DoD documents, the DIB, the Defence Innovation Board, produced this very nice, interesting document.

I was jumping on my chair when I read it, because at some point they say, we're not going to talk about justice or fairness of AI in defence, because defence is a competitive contest where, of course, you are looking to get the upper hand of the opponent. I don't believe they were, let's say, doing this in a naive way, because the fact that there has to be some form of justice in defence does not mean that you have to be fair, as we understand fairness in civil society. So just war theory tells us what we should be doing, so to get the upper hand, but not to commit atrocities, not to win in an unfair way.

And these are the things that, as a scholar, make you think, but you weren't there, you had all the legitimacy to address this question appropriately. By looking at the reality of things, you're going to use force via AI. Yeah, I mean, I do think that, to quote that example, if I may interrupt, there is a cultural statement in the US military in particular, which is that we don't want a fair fight, we want an unfair fight.

And what is meant by that is that we want to be certain that we're going to win. And so I think that's what was behind that language, a deep cultural belief that we shouldn't put our soldiers in harm's way, unless we're certain that we're going to win. I agree.

But there should be also the understanding that you cannot win using any possible means. And that's what just war theory is telling us. And I don't think that conflation is naive.

I think we're considering intentions, but at least it's a missed opportunity. Because if we had addressed at that point, the question of how AI inputs justice in warfare, we would be, as you mentioned before, already way ahead in this debate, which is happening anyway, because now we have to implement this technology and we have to address the issue.

You've raised a lot of ideas here about regulation, questions on how far that regulation should reach, the benefits of what currently exists and where the gaps are, and a word that's been coming up a lot is trade-offs.

And to wrap up this conversation, my final question is really around the consequences. So the other side of those trade-offs for the defence sector and beyond, if we don't get this form of regulation right. What are your thoughts on that?

Okay, I'll be brief and I'll go first.

The two extreme negative scenarios are this one. We lose the opportunity. I think we should be mindful of the fact that there is an opportunity to improve defence.

And especially in the current situation, this is crucial. So we can't really afford as societies to let this opportunity pass. So this is one huge risk.

The other risk is that we look the other way when this adoption becomes capillary. And we don't engage enough with the changes that this brings about to make sure that the changes do not override or violate the very principles of our societies. I'm saying this because I come to the work on digital ethics and particularly in defence technologies, because I am deeply convinced that digital technologies are reshaping the way we understand the world, but also our societies in a very fundamental way.

Now, the reshaping to me is not problematic. It's an opportunity to improve things. It's a way of going forward.

And I wouldn't want that the one way in which defence is being reshaped would be a missed opportunity to improve things under many circumstances, under many, let's say, categories, including the way we respect international human rights and laws, for example.

Yes, I totally agree with Rosaria's categorisation of the risk. At the two extremes, those are the risks that we miss the opportunity or that we behave in a way that's inconsistent with the values that we support.

I think our fear about those risks is likely to be driven by our experience. You know, in my case, I've spent my 40-year military career being taught, trained, educated, thinking about rules of engagement, about the vital necessity of behaving in a way that is consistent with the values of liberal democracy. I've seen that in practise throughout my career.

Now, that's not to say that mistakes don't get made. They absolutely do under extreme conditions. But that's where I come at it from.

And of course, I think somebody who is really concerned about ethics is more likely to be worried about the opposite risk. The truth is, we need to worry about both. And that's, I think, the answer to this.

If we keep that in mind, we'll get to the right answer.

Rosaria and Chris, thank you so much for your time today. It's been really fascinating to explore the world of AI and defence with both of you.

And thank you so much to our listeners for tuning into this episode of the OII podcast. If you've enjoyed it, please leave us a review and share the link to this episode with your network. We'll be back soon with more conversations.

Take care. Thank you.