Sustainability Unwrapped, a conversational podcast about responsibility, ethics, inequalities, climate change, and other challenges of our times. Where science meets practise to think about the world and how to make our society more sustainable one podcast at a time.

Hello, everyone. And welcome to this episode of the *Sustainability Unwrapped* podcast. My name is Robert Ciuchita. And I am an assistant professor at the Hanken School of Economics. And I teach, and I do research on digital marketing.

Today, it is my pleasure to have a conversation around one of the most interesting topics of today, which is artificial intelligence. And for that, we've invited two guests that help us understand a bit more about what some of the opportunities and challenges around artificial intelligence are, and especially how they might relate to sustainability and ethics. So, I would like to directly give the floor to our guests. And I will start with Martina Ciac. Martina, maybe you would like to tell us a bit about yourself and especially-- so, who you are and what brought you in to AI.

Yes. Thank you, Robert. And hello, to all our listeners. I am Martina Caic. And I am an assistant professor in strategic service design at Aalto University School of Arts, Design, and Architecture here in Finland. But my background is in business. I did my PhD in marketing in Maastricht in the Netherlands. And I focused on-- my research focused on how disruptive, emerging, new technologies affect human values, for example privacy, safety, independence, self actualization, human well-being, in terms of different functional, psychosocial, cognitive health, and overall service experience.

In particular, my studies were mostly in elderly care, and I focused on social robots. These are those type of robots that can behave and look like humans or in human-like manners. They can understand facial expressions, or they can engage in conversations. And they're introduced to different elderly homes as a machine that can help prolong the independence of older adults. And I was really curious to see how elderly people are going to react to these robots and whether they will accept them, whether they will indeed perceive them as human or social entities. So, that's how I got into the topic of robotics first. And then that extended more into artificial intelligence a bit later on. But as an introduction, I think this is fine.

All right. Thank you, Martina. So, moving from Finland to the other side of the world, it's my pleasure to introduce Stefano Tempesta. So, Stefano, welcome. And maybe you would like to tell us a bit about yourself and what brought you into AI?

Yes. Hi, everybody. Thanks, Robert. Well, after Martina's introduction I can only shy in front of these. But if I can give my contribution-- yes. I'm based in Melbourne, Australia, although I'm originally from Italy. I'm a CTO at Connecting Software and Microsoft Regional Director, which is a programme of trusted advisors to Microsoft and the community. I do a lot of these advisory services. Not services, but really complete contribution to the community. I'm currently advising to the Department of Industry and Science here in Australia on the National Blockchain Road Map. And my current focus is on helping people around the world to gain and own their own digital identity with initial and specific focus on the education sector.

And I'm also technology advisor at a start-up here in Australia called the Carbon Offset Solutions, which is a climate, action, and sustainability network with a mission to slow carbon dioxide emissions and remove the excess atmospheric CO_2 by using a regenerative agricultural technology-- normative agricultural technologies. So, my passion for AI was driven by the desire of connecting the study, the academy, behind it to application to the business world and also to agriculture, to robotics, to industry in general.

So, I'm trying to-- I'm a person of technology. I'm a person of action in some way. So, I like to connect that aspect of machine learning to a real practical usage and get some business or some real world application out of it, but also with a special angle on sustainability and impact on their society, so trying to make the world a little bit better using technology and using the best of our capabilities as humans.

All right. Thank you, Stefano. So, before we delve into some of the key topics of today's discussion-- if this is your first *Sustainability Unwrapped* episode, the idea is that we tried to bring together a research perspective and a practise perspective. And we try to understand, what are some of the key topics and what are some of the key challenges that cutting-edge academic research is dealing with when it comes to the topic of AI? But also, what are people actually doing in practise? What are organisations actually doing? And what are they struggling with? What are the opportunities?

And I think maybe as a disclaimer, we will be only scratching the surface in our conversation today because I think it's probably an understatement to say that AI is a very complicated topic, which can be approached from multiple angles. So, that's just more-- this is the ABC of the introduction, if you will. And we have also compiled a set of resources that people who want to read more and who want to go a bit deeper into some of the things that we discussed today, the links to these resources will be available in the episode description.

So, let's maybe start with probably something that sounds simple to begin with, but it is actually quite complicated. Could you maybe describe, at least in your understanding, what artificial intelligence or AI is? And do you maybe have an example from either your own work or something from popular culture maybe that provides an idea of what you think AI is? So, maybe give the floor to Martina now.

Yeah. As you just said, it sounds simple, but it's not very simple at all. I'm actually currently working on a project with one of my PhD students where we are doing systematic literature review to try to understand the definition of AI. And we're still working on that. So, I don't have a simple answer.

But I will maybe start by saying that AI is not a new term. AI has been coined in the mid 1950s. And I think that the father of AI, John McCarthy, defines it as the science and engineering of making intelligent machines. And in that, it would be something about machines doing tasks that would require intelligence if they would be done by humans. And at least that's my understanding of AI. But then there are these different typologies or levels in AI. We might talk about narrow or weak AI, general or strong AI, and maybe the super artificial intelligence.

And when you asked for examples, I think most of today's examples are on this narrow or weak level of AI, which is the type of AI that can only do type of tasks that it's designed to do. So, it doesn't learn, doesn't go beyond what it's programmed to do. Maybe here I could give an example of Alexa or Siri or chat bots that we might have on different websites. So, this is how I would say that the current status of AI is.

But then there are some predictions that, maybe within the next 10 or 50 years, AI will get to the level to be able to do some of the tasks that are a bit more on a human level, such as sensing or empathy, consciousness. But these are just speculations. But, yeah. You said a simple definition. So, I'm going to give some space to Stefano as well.

All right. So, Stefano, what is your simple definition of AI?

No, exactly. I don't think there is a simple definition. But we can try. We can try to put some boundaries into it. Because, like what Martina said, AI is really about mirroring or mimicking the behaviour of people with machines, so machines that look like humans but without-- yet-- all the aspect over fairness, inclusiveness, and empathy, which are probably not even connected-- they're only now imaginary. They're not connected to the brain. And we associate artificial intelligence with brain capability. So, then when we call the computers to do something intelligent, what we are saying is, you, machine. Try to do what a person-- a real person-- can do using their brain, which is recognise images, recognise voice, speak [INAUDIBLE]

So, all these are tasks that machines now can do pretty well and sometimes even better than a real people-- translating from one language to another one, recognising patterns or anomalies in big data sets that a person will struggle. But that is part of it. That is not the brain part of it. But there are a lot of other attributes that we typically associate with a soul or with heart, being empathetic, being fair, being transparent and understandable. And these machines are not there yet. There are principles that we apply to make sure that machines, again, can remove by us-- and I'm sure that we will talk about it in a moment-- from the data that they analyse.

But this is still a long way to go. So, yes. So, machine learning or deep learning or cognitive learning-- these are all different flavours of artificial intelligence. They are all about mirroring the brain capability of users, of people, at scale. As an engineer, I like to use this expression that we do this with machines to reach a scale that otherwise people will not be able to achieve individually.

All right. Thank you. I think it's extremely interesting. And also in my role as moderator, I'm going to try to somehow maybe summarise some of the things that you were saying. I really love this idea of distinction between brain and soul or a heart. I think that's something that is very useful to consider and how-- yes, indeed. When it comes to more mechanical processes or tasks that are clearly defined and can be done at scale, that machines can be much better than people. So, artificial intelligence can translate texts from one language to the other, but sometimes the results can be kind of strange because maybe artificial intelligence doesn't understand things such as irony and things like that.

So, there are always nuances, but the reality is that nowadays on our phone, we can take a photo of a text and the text, which is [INAUDIBLE] language, like in our case in Finnish, and we would get an approximate translation in English, which we simply did not have 10 years ago. And that's where we are now. Another thing that I found kind of interesting in your definitions was that Martina was very much talking about-- and I'm guessing this is maybe a bias on your focus on users and consumers, but you were talking about technologies that have very much a consumer, or an end user, focus. So, you were talking about the type of applications that you would have in your pocket, on your smartphone. But so, that would be almost like AI as an outcome or as some sort of other application that people are aware of.

But in what you were saying, Stefano, there was also AI more as a process or something that is happening in the background that can help applications. I'm not sure that everybody understands that, let's say, image recognition that is used whenever people have to go on these websites and they're asked, which one of these is a bus stop and so on? I'm not sure they understand that they're

actually training AI, and those applications can be used for, let's say, things like self-driving cars. So, again outcome versus process and the fact that we have consumer-facing applications, but we also have background applications that are being used to empower other types of applications. I think we're going to talk about that a bit more, but.

I was just looking this morning before starting this conversation-- I was looking for some recent statistics on AI. And I found a report on Statista that was talking about the revenue worldwide market of AI around $10 billion US in 2018. And something that I found interesting was the AI contributions to China GDP in 2030 around 26%. So, clearly there is a lot of buzz and a lot of talk about AI. And people are putting a lot of numbers and values and investments in AI and so on.

But I would like us to maybe take a step back and discuss, OK, we kind of talked a bit about what AI is. But maybe also, what isn't AI? So, I'm guessing, maybe for many of the listeners, it's still a bit of a fuzzy concept. And to further delineate it, maybe it's a good idea to also give some examples of, what isn't AI or what we shouldn't call AI although maybe in the media it's referred to as that? Any thoughts on that?

Well, I can maybe piggyback on Stefano was talking about-- the brain and the soul or the heart. And in my research with older adults, we would actually [INAUDIBLE] whether they perceive robots to be competent and warm, whether they would perceive the warmth coming from them. But one thing that they were not aware of is that the robots were controlled through the Wizard of Oz method, which means that there are behind the scenes researchers who are controlling the robots during the experiment. So, the robot isn't actually on the level to really show empathy or to really engage in these nice conversations with the elderly.

But they were not aware of that. So, they did perceive the robots to be friendly, to understand their troubles. But they were not aware that they were actually talking to humans behind the robots. So, in this case, I would say we need to make a distinction that the robots are still not on that level. And that would be a case from my own experience, where people think, based on the experiment, that they are actually interacting with AI when that's not the case. That would be the first thought.

All right. Stefano, any examples from your side of what does not constitute AI?

Yes. Actually, there is-- I do a lot of public speaking in the community or [INAUDIBLE] communities. And a lot of times people ask me, but I have this problem. And can I resolve it with some artificial intelligence or some machine learning? And I first try to ask, do you need know your machines, your algorithm, to improve accuracy over time to make some prediction based on current conditions over

time? That will probably be a good candidate for some form of machine learning. But if you're just automating tasks-- so repeat a task over and over again-- and based on some variable, you decide to take one action or the other one, that's not machine learning. That's not AI. That's just a business process automation with some conditions and then you decide, OK, I'm taking this action or I'm taking this other action depending on some variables.

Or another sample-- if you're analysing some data-- identify patterns in a small, medium, large, whatever data set-- and then [INAUDIBLE] identify a draw some lines, draw some chart to identify some patterns or some data into your data. Again, that is not AI. That is a statistics. So, nothing fancy in that. This has been around for years and years. And it is mathematics. The value of application of machine learning is in the ability first of all, of working without data set being labelled, so without actually knowing what you're trying to find and helping humans in finding things where they don't know where to look at. So, having that extra capability to identify patterns in data that you don't know up front, one aspect.

But the second and more important aspect in my opinion, is the ability to learn over time and improve. So, you have a data set. And then you make some prediction. And that can be purely statistics and not based on something that happened in the past. That's the likelihood that something happens again in the future. Again, that's not machine learning, that's just statistics. But then with the new data coming in, the algorithm can learn more and more and become smarter and more accurate over time. There is the power of machine learning, which is a blurred soft edge with statistics.

But that's the difference between applying a mathematical algorithm that is predictable-- and then you get a consistent result of a time just apply a formula to the prediction that machine learning can do, which may change-- and likely will change-- as data is collected and more data are retrained over time.

Also from my side, what I notice a lot, especially within the past five years or s-- Martina started by telling us that AI is not new. And the discussion has been there since the 50s, or it started in the 50s. And I actually went this morning to Google Trends to look for, how has the change in-- or what is the trend in looking for artificial intelligence on Google since 2004? And actually, the discussion has been fairly flat. So, it's not that there has been a lot of change in recent years. So, these topics have been discussed and have been in the attention of media and researchers and so on.

But interesting to link to what you are seeing, I noticed that there was a bit of a shift between machine learning and AI. So, in the early 2000s, AI was a topic that was researched, or at least that

was searched, on Google quite a bit. And then machine learning kind of overtook-- at least since 2010 and so on, machine learning actually overtook artificial intelligence as the key search term. So, I was wondering since you brought it up, Stefano, can you explain, maybe for a lay audience, what machine learning is and how it works for someone that has never really heard-- and because you also seem to talk about machine learning and AI almost interchangeably. And I just want for someone that's never heard of artificial intelligence and machine learning before-- what's a simple way for them to think about this?

Yeah. You're right. They're not interchangeable. So, artificial intelligence is a broader category. And machine learning is part of it, as well as deep learning or cognitive learning. So, they are all sort of branches of a broader artificial intelligence. I tend to speak more in favour of machine learning because that's my area. I specialise in that section. I do a lot of machine learning, regression, classification. That is the area where I find the more tangible application to technology and to business.

I guess in order to answer your question around the trend of AI shifting to machine learning, I guess that most of the buzz around AI a few years ago was especially pushed by Hollywood in some way, so creating an expectation that the machine-- I keep on saying machine learning-- that artificial intelligence is the super powers that make you travel time or do some prediction on conspiracy or on crime. Think in *Minority Report,* a movie a few years ago, and all these sort of-- or also *Star Wars* and these sort of things. Seems like everything is artificial intelligence.

But then what happened is that actually, a very specific application of this artificial intelligence capability-- machine learning, which is an analysis of data to identify in patterns in historical data set to make prediction that something is likely to happen again or a quantity is obtained again. So, I think of the Amazon model. The people that have bought these also bought this, this, and that. So, it can make a recommendation in that aspect.

Or given some different conditions, large data set that contains-- think of people that want to apply for a loan in a bank-- then, what is the risk of lending to people that are in different categories? So, that is what machine learning is all about. It's about analysing data set, identifying data sets that are too large for humans to make something out of it in a reasonable time, also able to identifying something that people don't know. Because we as engineer, data scientist, we know, and we don't know what we don't know. Simple, as silly, as that.

But this is where the ability of machine learning is-- to identify some key components, key features,

that we can anticipate. Help us finding something that we don't upfront and then learn, over time, and improve. So, what happened, then, is that this technology actually works. And it can be applied now to some very tangible use cases-- risk assessment, anomaly detection, network intrusion, object detection, voice to text, or text transcription, text voice synthesis. So, at the end, what happened is, hey, actually we can apply this technology to do something tangible, to get some business out of it.

So, the shift from Hollywood, which speaks about the beauty or the big potential of artificial intelligence, into something that can be applied to make money, to make society better, to create application that people actually use. And Martina mentioned Alexa, Siri, and so on. Then, the shifter is also in terms of interest from people. People don't look now into talking to HAL-9000, but actually about building a chat bot they could understand and could provide customer service. This is an application, which in my opinion is more interesting because it provides a result that I can use now and not just watch a movie about it.

Thank you. Thank you, Stefano. And I think that we can probably continue talking for another hour about the distinction and about machine learning applications. But I think it's important to be aware of those distinctions and to try to understand. And I'm sure that you will have some resources for our listeners in case they want to see more and understand more. But it's impossible to talk about AI and not breach this Hollywood perspective. And I would like to turn now a bit to Martina.

So, you do work on social robots. So, are we close to the machine uprising or not there yet? If we listen to Hollywood right, we had the *Terminator* already at the end of the 80s. And then we had *The Matrix* at the end of the 90s, where machines are taking over and turning us into their slaves. So, what do you think? Are we moving in that direction?

Well, I would say that we are definitely not there yet. But I think it's good to think about the dystopian futures and to kind of understand what kind of futures robots can maybe usher us towards and to understand in which ways they can distract or disturb currently held values or well-being of consumers. Again, coming from more of a consumer perspective and not the AI-- I'm not an AI person. But I remember maybe a couple of years ago, there was an article about recruitment robots that they were developing somewhere in Sweden. And I got really excited about it because they were writing how it's going to be completely free of human biases and how it's going to improve the recruitment process.

And after this very short term of enthusiasm, I started thinking, but really can we say that it's going to be free of human biases? And is it really going to be better than actually having humans interview

candidates? And then I started getting more and more into this area of ethics and maybe algorithm bias. And that's sparked my new interest with some of my master's thesis students. And I don't know whether you would like me to maybe elaborate a little bit more on what I mean with algorithm bias-- no?

Well, we're going to go to that in one second. And I think that's a perfect segue. It's just that, one thing that I wanted to maybe summarise from this part of the discussion before we move to the AI challenges particularly in relation to ethics and sustainability, I think it's extremely interesting that you are bringing these two perspectives. And one is this very process-oriented, which is, how does AI work and that's machine learning? And at the end of the day, for machine learning, you don't need to have some sort of anthropomorphized robot-like thing. Machine learning can happen in the cloud. Or most of the time, actually, it does happen in the cloud.

So, probably a very good example of machine learning, at least in my understanding, would be the AI from the movie *Her.* Would you agree with that? Because it's just this idea of something that happens in the cloud. It doesn't necessarily have some sort of manifestation whereas, I think for a lot of people, it is easy to relate to the idea of a robot being AI because that's also the 60, 70 years of sci-fi has been teaching us and telling us to expect. But I think interestingly speaking also for our listeners, it's good to be aware that AI doesn't necessarily need to be a robot.

A robot can have AI. But nowadays, at least if I understand from the conversation, we are still developing AI as software, so maybe at one point it's also going to be connected to hardware in a different way. But I think it's probably safe to say that we are not close to what the sci-fi visioneers were envisioning, which also makes sense because they were not aware of the technology developments that we will have now.

Indeed, I think Martina just started talking about the issue of ethics when it comes to AI. And of course from her perspective, it was about how a robot can make a less biassed decision. And of course connecting a bit to what Stefano was saying before, there are all these algorithms-- which is again another buzzword around AI-- an algorithm that optimises some sort of decision. And of course we can consider that the robot from Sweden was programmed to make those decisions based on some sort of algorithm.

So, what are the challenges of artificial intelligence when it comes to ethics and/or sustainability? So, maybe you started with ethics, Martina. So, you can continue with that. And then Stefano-- you were already discussing sustainability in the beginning, so maybe you have an example for that. So, what

did you learn about ethics in AI, Martina?

Yes. Thank you. Well, for the algorithm bias, which was a term that I got introduced to maybe two years ago-- and it can simply be explained as a systematically unfair treatment of a certain group of people in comparison to others based on the algorithm. And you can see that Stefano was already mentioning some bank loans. That can happen definitely in that scenario, also maybe in detecting shoplifters with the facial recognition in supermarkets, or the recruitment AI that I was just mentioning with this particular robot. Or now with COVID-19, there were many examples coming from education in which some universities didn't couldn't organise [INAUDIBLE] so they use some algorithms to decide which students will get the possibility to study at their schools.

One famous example is also Apple card that was found to be discriminating against women and giving women less credit than equally qualified men. So, what we see a lot in these algorithms is that somehow they tend to be biassed towards women and minorities. That was something that I found out in my research. And then we need to think about, how do we get there? So, whether it's because of the data-- because the data is, again, produced by humans, and we are having our own biases, and the data is never neutral, and algorithms are learning feeding from that data. But also, who is designing those algorithms and whether we have the diversity of teams involved in designing AI?

These became very interesting topics for me for further research. This is still in a nascent field of my interest.

Stefano, any thoughts on-- so, how do you see this idea of algorithm bias? And what are some challenges? And of course we-- as researchers, we stand on the sides and try to understand how these things work. But you actually work with developing algorithms and with the data. So, how does this phenomena sound from your side?

Look, you touch a critical point. But the reality is that it's all about people. It's not about AI itself. Martina made it very clear. It's about the data and about the algorithm that you write. So, AI is not something that has a brain on it's own. It's a computer that runs an algorithm. If this algorithm, which is a collection of instructions, is coded to act in some way, it's because the programmers have coded it to run in that way.

But also, the power of AI is it works because it learns from existing data. But if we feed them data that is unbiased at its source, then we obtain biassed result. There is no magic in there. And you cannot code the fairness. You cannot code the inclusiveness. You have to provide data that is fair and is inclusive. I have an example.

I was working some time ago on an algorithm for a detection of skin conditions and then using the capability to identify a face. It identifies skin conditions, like acne, like [? crow ?] feet, wrinkles, these sort of thing, then provide some recommendation on lifestyle improvement or products to apply and so on. But then we felt that not our machine learning system, our algorithmic, we data all the white, middle-aged women. So, obviously as soon as you start applying any other person, and also [INAUDIBLE] to any other person, the algorithm was very, very poor in terms of performance and even failed to identify those conditions.

Now, who can we blame for that if not just ourselves because we didn't provide enough data from a variety of different, inclusive variety of people from different cultural backgrounds? But also on top of that, there is a problem of culture, which is different from one country to a country. One aspect of the principles of AI is that you should protect and be respectful of your privacy. And also you mentioned the shoplifter. But I'm thinking of something broader to be honest, like cameras in the cities that identifies people moving from one area to another one.

The intention is to optimise traffic congestion, to provide public transportation that is on demand, so all good intentions. But who is using those picture? Is anybody detecting faces of people? Can it be used for crime detection? Or what if someone has an alibi and say, you committed a crime. No, I wasn't because I was 100 kilometres away, and there is camera evidence on that. So, where's the protection from a privacy and security perspective? Who is using this data? Can this data be manipulized?

And the reality is that this is a cultural thing because our cultures, especially in the Western countries-- they are extremely protective of our own privacy. And we don't want our own movement to be tracked-- location, phase, and whatever we do. There are some other cultures more on the Asian side that are exposed to this sort of experience that actually don't mind it because it's accepted and acceptable to have this sort of system that tracks people movements.

So then, whoever code this algorithm-- to get back to the data and algorithm for AI-- what cultural backgrounds should they introduce? And that is an open question. There is no answer. It depends. And it is the current way, the current challenge, for artificial intelligence.

Thank you, Stefano. And I think those are some wonderful examples to ponder. And I think probably nowadays, if we look into the media, we hear a lot of talks. And we've had this Cambridge Analytica scandal a couple of years back that actually turned out to be very misunderstood in terms of big data and analytics and machine learning capabilities. But there is this idea that we leave more and more

data traces that can be connected and that are collected by private, but also public, organisations and that somehow, these data can be used to make some inferences and to train algorithms and machine learning. So, the fact that we are leaving these data traces is clear.

But you raise an important point, which is, it is our responsibility as programmers of AI to be aware of the biases that we might introduce and to also be aware of the biases of the data. But now I want to go back to Martina and ask, how do we design ethical AI? Or, what are some principles that we-- so, aside from the data and the algorithm, which seems to almost somehow, and in some cases, be very, very separate from the actual end user of the consumer. But if we keep the end user in mind, is there anything from your research that you can-- or from your experience that you can think about in terms of designing ethical AI?

Yeah. From my research-- and I did talk a lot with elderly people and their formal and informal caregivers about some ethical ramifications of introducing social robots in their lives. And mostly they would talk about, as Stefano mentioned, privacy. There is also safety. There is also data management. Maybe there is this fear of, especially older adults, that they will not be genuinely cared for and that it's all going to be fake, and it's not real emotions that are going to be involved in this.

But the topic is really difficult. And we don't have any clear standards or principles on how to design ethical AI. I think that there are many initiatives around the world. Different countries started working on separate documents on how to really do that. Just this year, there was one white paper published by Berkman Klein Centre at Harvard University. And it's entitled, Principled AI. And they detected, across all these different initiatives around the world, that there are eight key areas that we need to work on. And I'm just going to enumerate them now. So, that is our privacy, accountability, safety and security, transparency, and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values.

So, they've detected that these are the things that now are being discussed around the world in, how can we design more ethical AI and robots? But I think we're still far away from having very clear standards that are going to be used within Europe already, but not even across the world. What are your thoughts Stefano?

No, [? Luka. ?] Totally spot on I know you have a huge list of these principles. But the challenge is, according to whom are these principles correct? Which cultural background? So, I want to be-- if you'll allow me. It's likely more provocative and introduce a bit of humour. I have three principles that

I apply to my algorithm. And these three principles-- I invite the listeners of this podcast to think for a moment if they ever have before-- and also you guys here.

So, the first principle is that AI should not harm any human. The second one is that AI should obey an instruction given to it by a human. And the third one is that AI should avoid any actions that could cause it to come to a harm itself. Now, digest this for a moment. And then, these are the three laws of robotics by Isaac Asimov, who is a writer. He's not a scientist. He didn't write any data algorithm and then machine learning. He wrote books. And *I, Robot* is one of my favourite books. And for me, this is a manifesto of AI. Well, he said that robot. He didn't say an AI.

But if a robot, an AI, doesn't harm a human, it doesn't do anything to harm a human and itself, then this is everything. Because there it is fair, it's transparent, it's inclusive, it's the safety and reliability-- all the other principles. Obviously, there is a long journey to get there and to actually implement this in complete entireness.

I love that you actually brought that because I use Asimov's, also, laws whenever I'm sharing my research on robotics. And I agree. It's actually very simply stated there but difficult. So--

Here you go. That's it.

It's showing people. Yeah. Yeah.

It is that simple, but not that simple after all.

And I think this also raises one of the most interesting questions about AI, especially when it comes-- because I think, again, you hit the nail on the head when you said, OK, ethical from who's perspective? Because we've already discussed that there are maybe some different ways that looking at it. And one is from the perspective of the algorithm and designing the algorithm and the data, so more or less the "how" part. And then we have the consumer, end user, part in which people don't really understand what's in the black box. They don't really understand what's behind machine learning and how data is collected, aggregated, modelled, used. And how does the machine actually learn? I think this learning part is still something that is very tricky for the average user or consumer to grasp, myself included.

And I think people do see these-- or they have these consequences which relate to privacy and personalization and the privacy personalization paradox and things like that. But when you talk about the laws of robotics, and of course the laws of AI by extension, it also brings this problem of-- and I think it's Nick Bostrom in his Super Intelligent book that talks about the-- what is it? The paperclip

paradox.

So, you have this AI that is supposed to-- its job is to maximise the production of paper clips. And basically, everything that it does is around optimising this one goal of maximising paperclip production. And in doing so, the AI is doing what it's supposed to be doing. But in doing so, it actually basically destroys the universe because at the end of the day, it is maximising an objective, and it's learning a lot in doing so. But this whole part of, let's not harm others in doing so, is put aside.

It's missing.

And it's, of course-- it's a thought experiment. And then it's also this whole idea of how super intelligence can be benign as long as the input and the thing that needs to be optimised does not also take care of these other basic rules, which is you shouldn't hurt others, you shouldn't hurt the environment, et cetera, et cetera. But these are things that maybe a lot of people find easy to relate to because they are falling on their-- if not necessarily culture, you were saying, Stefano, culture-- but maybe something like common sense.

And I think at least in my understanding of AI, is that, at least right now, AI doesn't really have common sense. AI can always recognise the cat from the dog in 99.5 times. But it doesn't really know that the cat meows and the dog barks. That's not something that is--

Also, a lot of people don't have common sense, so.

Absolutely. But, where does this take us? Of course, we have a lot of fantastic books and thought experiments and a lot of interesting pop culture discussions if you think about *Black Mirror* and things like that. But what's in the future for us? And we've been talking a lot about ethics, and we clearly see there are issues that-- we have algorithm bias. We have unintended consequences of AI. So, people are-- develop engineers and people are trying to develop these amazing applications that can help people and provide value. But in doing so, because of intentional or unintentional biases, things happen along the way. And to the extreme, we get this extreme example of the paperclip maximising AI that basically destroys the universe for a very silly reason, which is producing more and more paperclips.

But what's in the future? So, think about the fact that we will have students and we will have people working in companies in different types of industries, in Europe but also beyond. We might have some policy makers. We might have some journalists. What would be some things that people should think about, should consider, when they think, when they write, when they work around AI? And

maybe if we can also relate a bit to the notion of sustainability because we've talked a lot about ethics. But how can AI also help us deal with some of the issues that we are seeing now and that are also very hot when it comes to politics and debate?

And I know that's not an easy task to do. But I just put it on the table, and I'm curious what you think.

Right. So, that's big question. And look, the short answer is that we need to get back to people and making sure that this AI principles that we've been talking and a lot of countries ever produced. I mentioned that I have this collaboration with the Department of Industry here in Australia. And the Department has recently published the AI principles for Australia, which are influenced by the cultural background over here. I'm pretty sure that in Finland, as well in China, in Russia, may be different. You never know.

So, my ideal future is that we actually have some global AI principles that are cross culture, sort of a common sense-- as you mentioned Robert-- to become the adopted principles where scientists, data engineers, software engineer actually comply to. But comply not in a sense of we are forced to, but because we believe this is the best way of producing software application and hardware applications that can have this sort of principles. So, this is one part of it-- making people talk to each other.

And then the second part is, when whoever is working on producing these algorithms, again, the inclusiveness should be reflected in the team itself. Diversity and inclusiveness. We've been talking a lot in different aspects-- in recruitment, so making sure that there is a recruitment for men and women that it is fair, that their salaries are paid equally between men and female or any other cultural background. It's not just a binary world. So, we want to encourage DNI.

But DNI applies also to producing AI software and hardware. Because if we have a team that is diverse, their output is diverse. Because if someone doesn't think about that aspect to consider, then you can't blame the team. The [INAUDIBLE] didn't know. And then the output is an algorithm, a software application, that didn't have that consideration in mind. So, we need to start making sure that whoever is behind the production or software, algorithm, application of AI has DNI at its core as part of [INAUDIBLE] a strong value within the organisation.

Just quickly-- by DNI you mean diversity and inclusiveness, just to be sure?

That's it. OK. Awesome.

That's the one.

Martina?

Yeah. I think it's spot on what Stefano was mentioning now. And I would like to go back to saying that, yeah. Bias comes from people. And we need to have more diversity in teams. But there is another thing that came to my mind, which is another buzz word that I've heard throughout the last couple of years. And that's algorithm aversion. And that a lot of employees, working with chat bots, are having some kind of a version towards [INAUDIBLE] not [INAUDIBLE] moment [? as ?] stupid and not contributing well, and then they don't want to interact with them. They don't want to teach them.

And that becomes a problem again. Because if there is no co-creation between humans and algorithms, then I don't see that we are going to develop them in the direction that we would like to, even with all these principles at hand. So, there also needs to be openness from older people working with algorithms to actually train them to be less biassed.

[INTERPOSING VOICES]

Maybe to push the conversation to the end-- we live in interesting times, whether we like to or not. And we are exposed to unprecedented developments in this information and communication technology when it comes to digital platforms, when it comes to data, when it comes to analytics and machine learning and applications. And we are still at the point where we are just, again, scratching the surface and trying to figure out how these things work and what's going to happen.

But in an ideal future, what would you [INAUDIBLE] how would you say-- if we don't manage to blow ourselves up by 2050, how do humans and AI live and work together in 2050? Again, you have absolute, full creative rights on this. And, what is the ideal future? Are we living in a world where machines do most of the heavy work for us, and we get to just enjoy following our passions and creative outputs? Or, how do you see it? What is your scenario for the future?

Martina, what do you think?

Thanks for giving me the stage. Yeah. I always like to think that I would-- I just don't want the society to be oppressed by robots and AI, that we somehow get to the future in which we can cohabit, we can learn from each other, we can co-create. And it's a better future. It's not that we're going to be walked [INAUDIBLE] robots and things. I think that's very simple. But that's what I hope for.

Exactly. Look, we can come out with a lot of scenarios that would be an incredible plot for some Hollywood movies. But I honestly don't want that. What I want is that we apply AI for good purposes--

for protect the environment, for resolving or minimising the risk connected to climate change, for preventing situation as we had this year where we are all able to anticipate risks connected to pandemics. And I can think of an application of AI for good that, in 20 years or 30 years from now, can give a sight to the blind, can give voice to people that can not speak, can improve the mobility of people, or make people that live in their areas of war or emerging countries access to education, identity, that we don't have now.

Not strictly connected to AI, but our report from WHO is that there are over 1 billion people that leave without an identity. And this means they don't have any form of access to any sort of services-- education, employment, and so on. And I'm not talking about digital identity or Facebook account. Here I'm talking your identity, your identity card, your passport. So, they are people that don't exist-- a billion. It's one in seven of the worldwide population.

So, this is the challenges that we need to address. And I want to see robots, AI, helping-- also protecting our cultural heritage. We are more and more global. We are easy to move around-- probably a bit less this year, but we move around very much, and we life everywhere. Where are our roots, our cultural background, going? So, again AI programme that can help preserve languages, preserve culture, preserve [INAUDIBLE] and help people that can not travel also have the possibility to actually immerse themselves into travelling.

A dream of mine, if I can really put the last word on a dream here, is the possibility to create an augmented reality or virtual reality where we can immerse ourself and travel to places and explore places without actually going there. Because yes, it's not the same. I agree. But our carbon footprint can be reduced. And we still enjoy the experience. And we meet people from everywhere and every site. That is a good application of AI, in my opinion.

All right. Listening to your conversation, it got me thinking of Garry Kasparov, of all people. And he has a very interesting Ted Talk and also a book. And so, probably Garry Kasparov can be considered one of the first "victims" of AI. And I put that under inverted commas because he was the grandmaster at the chess that was defeated by the machine. So, there was a machine developed by IBM that was optimised for playing chess. And it completely defeated the best chess player in the world. And that was already in the 90s. And of course nowadays, we also have the AlphaGo, which is a bit of a more complex story.

But still, Kasparov is the first "victim." Again, inverted commas. And very interestingly, in his Ted Talk and his book, he actually talks about what both of you guys are saying, which is that we shouldn't be

afraid of the fact that AI-- he basically learned from the fact that, yes. The machine can do those things better than him. And he was the best in the world at that. But he also understands that we don't need to be afraid of that.

And we need to learn how to work with that. And Martina used the word co-create. So, how can humans and machine get the best out of each other? Because yes, the machines can be much better at computation and optimization, but there are still things which relate to common sense, culture, et cetera, et cetera, that are still inherently human. And those things can feed into the machines and into the learning once technology evolves and we do that.

But we do see that, as people, we also have a responsibility that we understand that we need to contribute to that, and we need to work with the machine towards our own goals. Because as you, Stefano, were saying, AI can play a fantastic role in maintaining and cherishing culture. But if done in the wrong way, it can also rewrite history to the extent that some particular-- people prefer to do to provide a particular view of events and so on. So, that's something that has to do with engineers. But it also has to do with researchers and educators and policymakers and journalists. So, I think it's a really interesting time to be thinking and working and writing about AI machine learning and all sorts of applications.

So, thank you very much for taking the time to be with us today and for bringing a lot of very interesting points and things to consider. And if people would like to get in touch with you, how can they do that if they have some questions or if they want to react?

They can write an augmented reality software that will travel to Australia and come and find me.

That sounds very good. How about you, Martina?

It's much easier for me. They can just send me an email. And I guess we can leave our contact at the end.

That's what I wanted to say. We would provide our contact details. So, if someone wants to find out more, then at least they know where to do that. All right. So, thank you very much for taking the time for this and for your really, really insightful thoughts and comments. And I hope this was a pleasant experience for you, as well. And with that, I say goodbye. And look forward to the next episodes of the podcast.