



## Artificial Intelligence: Managing the ethical challenges

Access the podcast at:

<https://doi.org/10.1787/54c51e69-en>

### Please cite this podcast as:

OECD (2021), "Artificial Intelligence: Managing the ethical challenges", *OECD Podcasts*, Duration: 23:05, OECD Publishing, Paris, <https://doi.org/10.1787/54c51e69-en>.

**Host:** Karina Piser

**Speakers:** Anna-Michelle Asimakopoulou, Lorena Jaume-Palasi

**Duration:** 23:05

**Date:** 27 July 2021

All podcasts express the opinions of the interview subjects and do not necessarily represent the official views of the OECD and/or OECD Member countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.

# **Artificial Intelligence: Managing the ethical challenges**

**Intro** [00:00:11] Welcome to OECD Podcasts, where policy meets people.

**Karina Piser** [00:00:17] If a self-driving car gets into an accident, who's to blame? The person in the car or the technology powering the vehicle? What about the manufacturer? If an automated job search tool discriminates, based on race or gender, who's at fault? The software or the company? As artificial intelligence becomes increasingly entrenched in our everyday lives, reckoning with these delicate ethical questions is more urgent than ever. A growing regulatory framework aims to set boundaries on just how far AI can intervene in human behaviour and what our legal system should do when things go awry. But troubling gaps remain. Welcome to OECD Podcasts. This episode is the third in a series on Artificial Intelligence brought to you by the OECD Global Parliamentary Network and the European Parliament's Panel for the Future of Science and Technology, also known as STOA. STOA brings together 27 members of the European Parliament from various committees to examine developments in science and technology. I'm Karina Piser, and today we'll tackle the complex ethical challenges central to A.I., looking at fundamental philosophical questions about humans relationship with technology and more practical concerns about how to best regulate these new tools. Later in this episode, we'll talk to Lorena Jaime Palasí, founder of The Ethical Tech Society, a non-profit research organisation focussed on automation and digitisation. But first, we'll hear from Anna-Michelle Asimakopoulou as a member of European Parliament. She's also the vice chair of the Committee on International Trade and a member of the Committee on the Internal Market and Consumer Protection Development and Petitions, as well as the STOA panel.

**Karina Piser** [00:02:00] Miss Asimakopoulou, thank you for joining us. It's difficult to keep up with the ever growing body of research on a high and how to best regulate new technologies. Can you give us a sense of what the landscape looks like in terms of regulations and what are the points of convergence across regions and nations?

**Anne-Michelle Asimakopoulou** [00:02:18] Well, first of all, let me just say that that A.I. is the flavour of the digital decade and these concerns that you mentioned are completely legitimate. This is a big discussion that's happening in all sorts of fora all over the world. The G7, the G20, the OECD, UNESCO is about to adopt its recommendation on the ethics of artificial intelligence. In fact, I was reading a report from the AI lab in Boston, which has created this tool box to track all these AI principles that are being discussed. And they actually identified 117 documents that list AI principles in the last five years. So that just gives you sort of an idea of the wealth and diversity of these principles as they are being expressed. But if I had to choose, let's say, the common trend that I think everything centres around, I would say that the keyword is: trust. So you have to trust in the use of AI and that in itself has a few core principles. There is transparency, being able to see and understand what AI is doing, explainability, fairness, safety, accountability and privacy. Those would be. I think the main things that that go through all of these bodies who are discussing this. Now having said that, you also want to know what the inconsistencies are, right? And what the difference is. I think the main difference is that some of them are a lot more detailed than others. And some of them take a broader view of what we should be looking at, like: they have environmental concerns, they have societal concerns more broadly defined. Probably the broadest definition is the EU High-Level Expert Group definition. But I don't think there are inconsistencies. It's mostly overlap and some of them are broader.

**Karina Piser** [00:04:27] So what I'm hearing is that there are overlapping sets of principles that are similar, but not necessarily identical. I guess I'm curious if this kind of patchwork of principles, of visions about the

way AI should exist, does that create inefficiencies and regulation? How are we supposed to understand this kind of broad framework that you describe?

**Anne-Michelle Asimakopoulou** [00:04:47] Well, there's two parts. I mean, overlap can be a good thing in the sense that we all agree and we're all converging, let's say, around certain basic principles. And that's a good thing. That's something like, for example, I hope this Trade and Technology Council with the United States is going to be doing, which is let's find the basic principles we all agree on. So overlap is good. Overlap when you when you're talking about legislation or regulation is not good because then you don't have certainty, and legal certainty is very important. And when we struggle with the level of regulation that AI actually needs, is what we're trying to find a balance between having a regulatory framework that on the one hand is lax enough that it encourages innovation and allows especially start-ups and SMEs to be involved in these technologies. And that's a big concern. And, on the other hand, though, you want regulation that makes you feel that AI is trustworthy and that AI is safe and that your basic rights are being protected and you're getting the benefits of these technologies. So overlap is bad when you have lots of laws. Overlap on principles and a basic understanding of what you're trying to do is good.

**Karina Piser** [00:05:59] We focussed mostly on the role of governments in setting ethical standards and this tension you identified between enacting solid regulations without compromising technological innovation. But it's not governments who are creating these technologies. So how much of this is up to companies and how much can governments really do without corporate accountability? What is the private sector doing in terms of setting its own standards but also abiding by the standards governments set?

**Anne-Michelle Asimakopoulou** [00:06:25] Well, look, I mean, I think that you'll find when we're trying to legislate in a future proof manner, which with technology is very, very difficult because there's just so many advancements all the time that technology gets ahead of legislation, legislation gets outdated. The private sector theoretically should be able to follow this more closely. On the other side, I mean, when you let the private sector set its own rules, not everybody is a good guy. So and not everybody adheres to it. And I think that the way technology has now permeated through every aspect of our existence and operation and work and personal life, I think that people expect a framework that will protect them and safeguard their rights that's enforceable and that's the job of government. So that's not to say that the private sector shouldn't set its own standards. In fact, there should be interaction as much as possible. I mean, when we legislate, we have to take industry's perspective on board. Otherwise, there is a danger that we stifle innovation and that's not something we want to do. So it should be a collaborative process. But in the end, someone's responsible for making the rules and that is state.

**Karina Piser** [00:07:48] I'd love to hear some examples of, you know, maybe in the past decade how certain technologies have advanced ahead of the law and how legislators have worked to keep up with new technology.

**Anne-Michelle Asimakopoulou** [00:08:01] Well, AI is a good example of that because, I mean, some people say, well, there's really nothing new, and yet the people who are not in favour of regulation say, well, there's nothing really new. I mean, when you're talking about a self-driving car, you're really talking about product liability. So what's new? OK, you have a new technology, but the basic principles are fine. So maybe you just need to tweak the legislative framework. We have GDPR, for example, that covers a lot of stuff. We have a product liability legislation. Tort law covers a lot of that. That's fine. But we do need

to address these issues when there is a risk inherent in the technology that makes it dangerous. You want a self-driving car, maybe because you don't want to own a car, because maybe you want to have less cars, it's better for the environment, or because it's easier as you don't have to learn how to drive. Sometimes it might actually be safer because you don't have, for example, the problem of drunk driving. At the same time, you need to then address the question, well, when something goes wrong, whose fault is it? Is it your fault? Because you didn't take over? Is it the manufacturer's fault? Is it the software producer's fault? That's why AI needs its own framework, because it's a lot more complicated. The principle so that there should be human oversight or that there should be transparency. These are things that can be very future proof, regardless of how the technology evolves.

**Karina Piser** [00:09:30] Right. So these two examples you raise about self-driving cars and bias in hiring processes, these are really important. I was just reading about a case where a boy was killed in an accident with a self-driving car. And it presents us with this tricky question of this is automated technology, but there's a human using it. So how has the regulatory framework evolved to tackle these grey areas that seem really important?

**Anne-Michelle Asimakopoulou** [00:09:53] Well, frankly, it really hasn't yet that and that's part of the problem. Self-driving cars are not something that's the norm. Far from it. But it will be. And that's why as these cases come up, we really have to start thinking of them in a different light. And just the fact that we're discussing it like this shows that there are limits to the existing legal framework surrounding these questions. So we do have to dive deeper. And that means that policymakers have to have a much better understanding of the technologies and of the issues and of how all this works, because otherwise you can't you simply can't regulate if you don't understand.

**Karina Piser** [00:10:32] And again, on the problem of bias or discrimination on racial and gender lines, is there also the fundamental issue of first creating change at the societal level? Because, I mean, obviously, these are technologies that are supposed to do in some ways replace human action, but the algorithms that drive them are created by humans. And that is, from what I understand, the reason they replicate some of these biases. So how do we how do we kind of thread the needle with that?

**Anne-Michelle Asimakopoulou** [00:11:00] If there were no societal bias, then the algorithms wouldn't be biased. But there is societal bias and there may be more or less. That's one piece of the puzzle. The other piece of the puzzle is how is this incorporated in the algorithm? Because there's ways to hide the bias or maybe it gets magnified. So we still need to look at the technology and regulate things that will allow us to check if this bias is causing damage.

**Karina Piser** [00:11:42] Now we'll turn to Lorena Jaume Palasí, founder of the Ethical Tech Society, to hear more about regulations, but also get a sense of some of the philosophical questions essential to understanding these new technologies. So when it comes to artificial intelligence, we find a number of ethical and regulatory issues that sometimes overlap and sometimes don't. And accordingly, a kind of slew of regulations set by different companies and countries commonly known as AI, principles that seek to match the regulations to these ethical questions that keep popping up. So what are these principles and how do they show up in our daily lives?

**Lorena Jaume Palasí** [00:12:19] There's a lot of companies that in the last two to three years have been creating different treaties or ethical principles or guidelines or norms and protocols, it's sort of like trying to standardise and create some sort of global principles or universal principles for a specific region or for specific context. These type of principles and ideas about creating global standards are very important and are part of an old tradition of standardising processes, both on a normative level and also on a technical level. So this seems like unconnected or sometimes to a certain extent, we look at regulation and ethics as sort of the antidote to how we create technology in our everyday lives. With those principles, we decide who has access to technology. How does this access look like and how are people that are interacting with those accesses for our consumers or users of these technologies discriminated? Because in the very end with these technologies do is discriminate. It's the job of algorithmic systems to decide to what sort of category a person needs to be ordered or needs to be sort of assigned to.

**Karina Piser** [00:13:39] If the goal, like you say, is standardisation, should we be concerned that regulations vary so much by region and country? I guess, in other words, what are the pitfalls of this fragmented regulatory framework on artificial intelligence?

**Lorena Jaume Palasí** [00:13:54] The conversations that we're having right now are the consequence of a specific idea of how to structure the world. Actually, the origins of both algorithms and laws and ethical principles have something in common that is quite old, and it's based on old enlightenment ideas with the card, with Leibnitz, those scientists that thought, OK, you can on the first place, the idea that there's something like rationality that sort of can be detached from who is observing something, you can create objective facts, you can create objective observations that are neutral and then your body doesn't matter. Your point of view and your social position doesn't matter, because if you follow this rules, that is the case. Now, we know from the scientists, from marginalised positions for the last hundred years, that this is not the case, that this is impossible, that our bodies matter, that it does matter, and it does play a role. If it's women observing something or it's a person of colour observing something or it's a white person or it's a men and so on. But this belief still holds on. When we create technology, we think that we created neutral and objective rules, which is not the case in the first place. And the same applies, though, for law, for regulation and for ethics. And to a certain extent, this is part of the problem and said we have both at the regulatory side and also at the technological said because we're trying to achieve something that we think is rational, neutral, universally applicable.

**Karina Piser** [00:15:41] In April, the EU started discussing some proposed rules for AI, and it seems like things are really starting to shift globally toward maybe more regulation. The United States even is kind of hinting at a firmer position. I am curious what has been accomplished in this space and what are the gaps that remain and kind of bringing all of these theoretical issues that you've discussed into practise?

**Lorena Jaume-Palasi** [00:16:05] Well, I think there are a few major frictions that we need to consider that are still not addressed regarding groups, and they're fundamental. On the one side, our legal European culture is based on centring the individual on everything we do. So fundamental rights are individual rights. And our law system and the way it's created, it's individualistic law system from the methodology to the object, it sort of treats it's about individuals and on the other side. And that's a major difference. These technologies are not about individuals. They are about implementing statistical models. And statistical models are more about understanding collectives, are based on average. It means that you're stripping off individuals from their particularities and trying to create a sort of model of an individual in a very fine, granular way because of the new technologies that we have. But it's set up a very different thing. So those

technologies are really very much about understanding what's the shape of a society, what's the shape of a group. It's about understanding the woods, so to say. So on the one side, we have technologies that are about woods and on the other side we have regulations that are about regulating the trees. And with this idea, it's like trying to regulate the woods by regulating tree by tree. If there's a problem, it will not affect a single individual, but it's going to affect probably a collective, at least one collective. I think it's about balancing in the same way that the woods consists more than just trees. It's the same with societies. Societies are more than the sum of individual interests, but a good society sort of doesn't put one of those two aspects in the main point, but tries to balance both of them. So I think the gap here is trying to balance on the one side individual interest and the protection of human rights and on the other side, protect collective rights and understand that there's that space there that needs to be regulated. And right now, it's not really that.

**Karina Piser** [00:18:22] Yeah, that that tension between the individual and the collective, which might seem a bit abstract, I think really gets to the heart of a lot of these ethical concerns. For example, when it comes to bias and discrimination in automated technologies. At this point, we've all read about how facial recognition shows a lot of built in biases that perpetuate the racial inequities we're trying to fight elsewhere in society. So I'm curious if you could give some concrete examples of what principles we could apply in practise to make these tools more inclusive and follow the direction that at least a lot of people are hoping to move toward at the societal level .

**Lorena Jaime Palasí** [00:19:01] Discrimination is not a statistical question. So you cannot create a perfect system that's impossible in the same way that you cannot create perfect human beings. Human beings fail as well, and they discriminate. So we cannot ask as a democracy to have moral citizens. This is what precisely autocracies demand from their citizens, that they behave morally to have a good society and democracies. We try to create rules that make inclusion more possible, but not demanding morality from our citizens, but trying to create more opportunity and more spaces. Now, the question is how do you do that? One of the main problems in this type of technologies regarding discrimination is the fact that most technologies are essentially lactic and universal meaning. With that, they first try to create human categories and the solutions that they create. And those same categories are supposed to be universal things like fundamental decisions about, for instance, what type of gender categories that we want to have. There is going to be female and male only, or do we want to add different additional categories. It goes farther on ideas of ethnicity, which how many types of ethnicity do we want to put in there? And depending on the format, how do you want to reflect the different colours of skin? And because we create them like that, because we don't have enough categories to define the complexity of life and the ambiguity of human life, and because mathematics is unable to mathematize ambiguity, those discriminatory problems happen.

**Karina Piser** [00:20:44] So what can regulators do practically to make these AI technologies better, to make them less discriminatory, to make them less vulnerable to human bias?

**Lorena Jaime Palasí** [00:20:57] We could concentrate more on processes. We could think about how these processes could be more inclusive and can be, so to say, creating in such a way that they have a governance of human beings and machines and think about it more organically that allow potentially marginalised and discriminated communities to ask for redress and to be able to address where those technologies are discriminatory. I think there is also this sort of idea that there implies digitisation implies using artificial intelligence. And to be honest, digitisation and new technologies is not an objective in itself. It's an instrument to achieve specific societal goals. And those goals need to be defined first and not the

other way around. When we decide that there's a part of a human process that we want to automatize, this is a moment which is an opportunity to think about how do we do this type of processes and where they write and what is it that we want to enhance. Those things have to do with cultural conversations, that have to do with ethical conversations as well, and the need not only to find a place within the regulator, but as a society. And I think that the polarisation that we're seeing nowadays sort of shows that we're at that point where we are discussing the fundamentals of our society.

**Karina Piser** [00:22:33] Thank you for listening to OECD podcasts to learn more about the EU Parliament's panel for the future of science and technology work on AI, go to STOA at the European Parliament's website. To learn more about the OECD work on AI, go to [OECD.AI](https://oecd.ai).

**Outro** [00:22:53] To listen to other OECD Podcasts, find us on iTunes, Spotify, Google podcast and [SoundCloud.com/OECD](https://SoundCloud.com/OECD).