AI and You

Transcript

Guest: Risto Uuk, part 1

Episode 139

First Aired: Monday, February 13, 2023

Hello, and welcome to episode 139! Today my guest is Risto Uuk, calling from Brussels, Belgium, where he is a Policy Researcher at the Future of Life Institute and is focused primarily on researching policy-making on AI to maximize the societal benefits of increasingly powerful AI systems.

Risto is in Brussels to be at the nerve center of the European Union, or EU, and is an expert on the EU's AI Act. The EU is perhaps the first governing body in the world to contemplate regulation on AI at the multinational scale, and their AI Act is already making waves in that respect. I had questions about it that were piling up and I wanted an expert to help me get the answers, and so I turned to Risto, because he is that expert on the Act. Whether you live in Europe or not, any legislation they make with respect to AI will have global consequences and inform policymakers around the world, so this is a big deal.

Risto previously worked for the World Economic Forum on a project about positive AI economic futures, did research for the European Commission on trustworthy AI, and provided research support at Berkeley Existential Risk Initiative on European AI policy. So we couldn't be in better hands for understanding more about the EU's AI Act.

Let's get to the interview with Risto Uuk.

Risto Uuk, welcome to Artificial Intelligence and You.

Thank you for inviting me.

And you're joining us all the way from Brussels now because you are involved in the European Union's **AI Act** and their policymaking with respect to artificial intelligence, which we will get on and spend quite a bit of time on. And what I want to do, first of all, is get an idea of what it's like to be in that job, because that sounds fascinating. Maybe you could introduce us to how you got started. Say, you were at the Future of Life Institute; you started there. How did that come about that you got involved with them?

Yeah. Maybe the most relevant kind of introduction to this would be that I was part of the European Commission's AI high-level expert groups work in 2019, where they drafted the AI trustworthy AI ethics guidelines. And I was involved in a project around trustworthy AI assessment list, which was essentially a voluntary toolkit for companies to evaluate how trustworthy their AI is in various kinds of phases of the AI lifecycle, from the development phase to pretty much being able to implement or deploy it on the market. So I was involved in that kind of project. After that, I worked for the World Economic Forum together with UC Berkeley Computer Science Professor Stuart Russell and economist Daniel Susskind on a project around positive AI economic futures. So this was imagining a world where AI systems could automate away almost most jobs in the world. How could that kind of world still be positive? And we interviewed top economists in the world, like Daron Acemoglu from MIT, Erik Brynjolfsson from Stanford to come up with five positive scenarios. So I was involved in that kind of project, and then I had a good opportunity, I saw a job posting from the Future of Life Institute. That was one and a half years ago, roughly, when in 2021, the European Commission released their draft AI regulation called the AI Act. And the Future of Life Institute wanted to start working on that pretty much right after the draft came out. And so they hired two people in the European Union - my colleague Mark Brakel and myself to work on the AI Act. And I saw this as a really good opportunity because the EU AI Act essentially is the first ever major comprehensive AI regulation in the world. So I felt like it would be wonderful to work on this because of kind of the possibilities that AI systems can bring to the economy, to society, but also because of the risks that I could see. And yeah, I saw it as an amazing opportunity.

I think I got some of the timeline backwards there. So you are working on EU AI policy, and you're kind of seconded to FLI as a role of amplifying the interpretation and the outreach of that for their audience. So then we go back earlier, what was your background before you got into AI and how did the introduction to that come about?

Yeah. Before doing the AI stuff, let's go back before the European Commission's Trustworthy AI assessment list work. I became interested in AI stuff as a science editor, I worked at Estonian Public Broadcasting, and I wrote a couple of articles about AI. For example, I interviewed an AI safety expert in the US: Roman Yampolskiy, from Louisville University.

Roman's been on the show.

Oh, great. I interviewed him. I wrote an article about that and wrote several other articles, and I really became more and more interested in the topic. And then there was this initiative called the AI Safety Camp which was like a research camp where people got together for a couple of months to work on research projects. And I decided to take part of that. I wrote an article about the AI policy-making or trying to kind of find ways how general political policy-making process could be used to evaluate AI policy-making, what kind of learnings could it provide. And it turned out people were quite interested in that. Right now I think he has thousands of views and lots of people have cited it. So I saw that I could contribute to this field and people were interested in what I was able to create. So I became more and more interested in that and I did a master's at the London School of Economics in Philosophy and Public Policy. I also studied a PhD in economics to study AI automation issues, kind of thinking about the world where AI systems could do many, many different economically useful tasks instead of humans. Like what kind of implications does that world have? And then moved closer to the European Commission, EU institutions, the EU policymaking stuff.

Now, when a technology becomes the target of policymaking, it's an indication of a milestone of a certain maturity that it's reached the point where it's getting the attention of people that are concerned with general public welfare. And so we have a lot of thoughts about policy-making. In Silicon Valley, they are generally - not always - resistant to policy because it can often inhibit innovation and their ethos of moving fast and breaking things. And sometimes the thing that gets broken is us, hence policy. And if I look at the state of policy in different countries and regions around the world with respect to AI, it seems that the only place that I hear about this to any significant degree of progress of putting words on paper is the European Union. Would you agree with that? Can you characterize the relative maturity of policy development in AI according to nationality?

Yeah, that’s a really good question. So regarding the approach of moving fast and breaking things, the European Union is well known to be different from that in terms of philosophical inclinations. The European Union tends to be characterized from the perspective of the Precautionary Principle or precautionary approach where quite cautious about new technologies and try to cover those technologies using different rules and laws. So that’s quite common in the European Union. And I do think that the EU AI Act is one of the first comprehensive laws on AI in the entire world but it’s not fair to say anymore that it’s the only one that is kind of trying to develop some types of legislation around this topic. There are examples, for example, from China. They recently had a law passed, which is definitely not very comprehensive AI regulation, but it does cover, from my rudimentary understanding, I don’t pay so close attention to China, but it focuses more on recommender systems. So a subset of AI systems rather than AI systems as a whole or comprehensively. But that’s still an example. Then it seems that the EU already has had some impact on other countries that are drafting legislation around AI. For example, Canada has their own AI Act in the making from what I understand. In the US, there was recently the AI Bill of Rights, which is definitely nothing very significant in comparison with the EU AI act, just four principles I think, or something like that. But still, that’s some kind of a start. And in the US, there’s also the NIST AI Risk Management Framework, which is just voluntary, but a lot of effort is going into that. And we could see that becoming some type of standard in the US where if companies don’t follow that, then maybe there are some legal repercussions as well. And also the EU has had an influence for example, on Brazil, from what I understand. Last year, I think Brazil’s Congress passed a bill for a legal framework for AI. So you tend to see these types of things popping up everywhere because AI systems are becoming a bigger thing in everybody’s lives. And potentially with some leadership from the EU, other countries are seeing the relevance and importance of that. And maybe sometimes some countries don’t want to also be left behind in terms of regulatory leadership because in that case, the EU has almost total control if the EU comes out with some kind of legislation and everybody just follows that entirely. Because many companies provide AI services and products to the EU market, which is really significant and they have to abide by these rules. So it oftentimes makes sense for them to abide by those rules elsewhere as well. So in that case, through the Brussels effect the EU can have a lot of power. So yeah, these are some thoughts around the global initiatives around AI regulation.

It’s interesting that you mentioned the National Institute of Standards and Technology because I think that does point to one of the differences between the US primarily and European democracies, that the US would rather have this voluntary regulation. And I don’t think of NIST as being regulation - because I don’t think it is - it’s more like a Good Housekeeping Seal of Approval or something that a company would follow because it was in their best interest, not because they were being forced to, like a UL or BSI label on some piece of equipment. And then eventually maybe that turns into some kind of building code and you can't install something if you want to keep your ticket unless it's got that symbol on it. So I want to get into now the EU Act, what it is and why it matters, and why we should pay attention to it. Because I'm looking at this now as basically innovation in the field of AI policy-making. They're leading the way there in the EU on that. And first of all, a really basic question that's going to expose my naiveté on this, which is, what is the phase of this legislation? Proposed, enacted, where is it in the pipeline?

Yeah, really good questions. And some of it we have addressed to some extent already but yeah, essentially, I see it as one of the major comprehensive AI regulations in the world, one of the first ones. And based on the initial proposal I mentioned, that the European Commission came out with that in April, 2021, the draft version, that initial proposal assigns applications of AI to three risk categories. So the first category would be unacceptable risk. These systems would be prohibited from being used or rather it would be more precise to say that these use cases rather than systems. And for example, in the unacceptable risk category, there are things like government-run social scoring that we kind of think about are implemented in China, for example. That's in our mind when we think about these types of use cases. And the second category is high-risk applications. These are not prohibited, but these have to follow legal requirements, and few, for example, CV scanning tools that rank job applicants would be part of this category. And finally, there's the category of low-risk or no-risk AI applications. And here they basically are unregulated or have just merely some transparency requirements. And why this is important, I alluded to it already, but people think that this is similar to the EU's data protection regulation that came out in 2018, GDPR, where the AI Act could also become a global standard, like GDPR seems to have been. And essentially this could determine to what extent AI in principle has a positive rather than negative effect on people's lives across the entire globe if it goes beyond the EU borders, but also in the EU. The EU is a huge market, a lot of people, hundreds and hundreds of millions of people. Even if it only affects the European Union, then it could be really important for those citizens. But yeah, there's also potential to have an influence beyond the European Union through what is called the Brussels effect that I also mentioned earlier. And yeah, I think we ought to pay attention because of that reason already.

Right. So bear with me while I stay at a really high level here to make sure I've got the fundamentals right of this. So is it accurate to say that this is fundamentally a product regulation act? It's not talking about the companies that make it, it's not trying to certify companies, it's talking about products themselves. Is that the sole focus of this act?

Yes. This is correct.

And so what are the consequences then for that in terms of how it takes effect? If we were talking about, say, the production of electrical equipment, then you could say whether it gets approved for meeting some standard for use in a market and then there would be some sort of testing process and so forth. This doesn't sound that amenable to that sort of formal stamping. How does the rubber meet the road with an act like this?

Yeah. Good question. So first, I think you also asked about the timeline, so maybe I'll comment on that briefly. So as I said, in 2021, in December at the end of last year, the council of the European Union, which is one of the EU institutions that negotiates the final draft, basically, they together with 27 member states, developed a new version of that initial draft with all the member states' input. And in December last year, they essentially adopted the council's position. And now the European Parliament is continuing to develop their position and one of the predictions I've heard is that they will be finished with that perhaps in March this year or April. And then some stakeholders are predicting that by the end of this year, the final version might be passed. So that is kind of on the timeline. Regarding the question around testing and the product legislation aspects, I would say that currently, it looks like mostly this act will rely on company's self-assessment of their AI systems. They essentially go roughly on Annex III, there's this list of high-risk use cases of AI systems. They essentially go to that list. There are eight categories of risk use or high-risk use cases. They go through that; they assess whether their AI system belongs there or not. For example, I mentioned these CV scanning tools. Some AI company that is developing these types of AI applications, they see, "Hey, yes, our system is almost definitely covered by this. Okay, so what kind of requirements do we have to comply with?" They go to chapter two of the AI Act; there are requirements around risk management, human oversight requirements, robustness, cybersecurity, accuracy requirements, quality management. They go through that list and they, to their best understanding, try to comply with that, try to make sure that they have technical documentation in place, risk management procedures, all kinds of things like that. Of course, there's still discussions continuing over the involvement of third-party auditors - somebody else auditing whether these AI systems actually comply with regulation. But from what I understand, actually, the claim is that most of the products in the EU market are self-assessed. And this is kind of an interesting aspect. A lot of people, of course, don't see that as a good thing. They would want third-party auditors to check, especially the more risky AI systems. But we will have to see where the negotiations lead. But yeah, that's kind of a rough understanding I have of it right now.

That aspect of self-assessment in particular seems antithetical to the notion of regulation itself. It's like seeing someone committing a crime and hauling them over and saying, "You've done this. What do you think we should do to you?" It doesn't sound sustainable. So to get to the definition of the act here or what it is, I'm hearing two sections. There's one that is this set of definitions that establishes categories that then get used in the requirements to say, well, depending on which category your product is in, this is what you should do about it. Is that a good description at that very, very high level?

Yeah, exactly. I think the key aspect of this regulation is what is captured as high risk and what is not and also what is prohibited and what isn't. But the high-risk aspect is more relevant because that captures more use cases than the prohibitions part does. The European Commission's own assessment of how many AI systems would be captured by the high-risk requirement, they claim it would be like five to 15% of AI products' use cases. Although some companies and some startups recently have tried to do their own estimates, and they kind of tend to claim that maybe it's more than that, maybe it's 30% or something like that. But yeah, the high-risk list is probably the most relevant one for everybody because that decides what is captured on what isn't.

We'll get into how the Act defines AI in a minute; but I wanted to look at these risk categories because it was interesting that you mentioned in the unacceptable one essentially China's social credit is an example of that. And it points out to me a very different effect or a different scope of the effect of the product from things that we normally look at regulating. Like, is there a risk of it electrocuting someone, or did someone get poisoned by this pesticide from this apple, where there's a direct connection between the infringement and the person affected. But if you've got something like social media that is poisoning people's minds, you can't draw the direct connection and say, "this vote here was corrupted because of that." Does that difference in the traceability of the effect make a difference to how this act is constructed?

I agree completely with you, and you make a really good point and bring very relevant examples, because as we discussed earlier, this is seen as kind of a product legislation, but many disagree that AI systems are just regular products. Oftentimes they have a different view of AI systems. They're more dynamic and they will look more like networks and services and some other types of things than products. So yeah, usually when we think about the safety of products, then we think about risks to health, risks to physical health especially. Whereas in the case of AI, a lot of people are talking about risks to fundamental rights. And also in the case of the work at the Future of Life Institute, we have been very interested in societal-level risks. These are risks that, for example, an AI system might not even cause a particular person any damage, so to speak, but overall, for example, through changing how somebody votes, they could have effects on the entire society, on democracy, on rule of law. And these kinds of things are very hard to capture with a very strictly focused product legislation framework.

So perhaps the analogy is something like industrial pollution. You can't point to one person and say, "they got this sickness because of this power station 200 miles away," but statistically that's inevitable because of this science, and so we are going to regulate that kind of pollution. Is that a reasonable analogy?

Yeah. I think that analogy seems quite reasonable. In the case of the pollution example, maybe there's the issue of very small damages that we can't really capture, but overall, over the long term, the damage is very, very big. That might perhaps be the case with AI systems as well, in the case of misinformation or slowly, slowly reducing trust of institution and then slowly, slowly doing that, that could be similar to pollution as well.

Right. Of course, it's very easy to precisely quantify the quantity of nitric oxide emitted by something and far less so to measure disinformation, although at least we're trying now. So artificial intelligence has a famously fuzzy boundary of what it is. The definitions of it are really quite vague, even at the textbook level. And this is an act that is titled artificial intelligence so how does it draw the line and what is explicitly on the outside of that line that is nevertheless say, data processing?

Yeah, good question. So I think the kind of general way to look at it is the European Commission, initially, when they drafted this regulation, their aim was to kind of capture almost everything that could vaguely be related to AI. And because of that, they came up with kind of a broad AI definition. So right now, the definition in the AI Act is roughly that it's software that is developed with one or more of the techniques and approaches listed, and they have a specific list of AI techniques. And on that list, they have machine learning approaches, they have logic and knowledge-based approaches like symbolic reasoning and expert systems, and these kinds of things. But they also have statistical approaches and Bayesian Estimation and search and optimization methods. So they have very, very broad things listed under this definition.

Does that run the risk that someone comes up with something that's not on the list and it runs around the act? Like if there are neuromorphic chips being developed, there are brain cells on chips being developed where neurons, actual human neurons, are fused onto chips and there's this kind of biological silicon hybrid thing going on there. Are the definitions in the act broad enough to catch this sort of adventurousness?

I think you're right. They might not capture if you come up with examples like you provided. You may end up finding some type of very novel AI approach that would not be captured by this list. From what I understand, this list is under Annex 1 in the AI Act, which means that the European Commission, through a mechanism called delegated acts, could potentially update this list and put new stuff there. And this wouldn't take as long of a time as the drafting of the entire AI Act, so it could be quicker. So in that way, the AI Act could be updated according to new developments. But I think the initial goal of the commission was to be quite broad so that ideally it would capture AI systems for quite some time. But regulation, generally speaking, is not super future-oriented, it is still focused on the present issues and what's presently possible. So, for example - we might discuss this later - but in the case of more general system like large language models, they became a thing only in roughly 2020 basically with GPT-2 from OpenAI becoming a thing actually used on the market and becoming a bigger thing. And when the European Commission worked on their AI Act, they didn't have these systems in mind. They were just focused very narrowly on AI systems that can be used for a very specific purpose, but not those ones that could be actually used for many different purposes. So you can see very quickly how the type of ideas that regulators have in mind could become outdated somewhat quickly. Maybe in the next couple of years, these very narrow specialized systems become a rarity on the market, and we only have things like Gato from DeepMind, which are able to do hundreds and hundreds of different kinds of tasks. Maybe those are the systems that we will have. It's very hard to foresee the future.

I think that's a really interesting point there. You're saying that they were initially focused on what are the risks of narrow AI systems. But now we have seen things like ChatGPT, which *everyone* is talking about. Not a day goes by when someone who is not even close to the field of technology, let alone artificial intelligence, asks me about it, where the risks seem to be more of the lack of our imagination of what it could be used for than any vertical application that is apparent to us.

That’s the end of the first half of the interview; this is broken into two halves because we’re already getting pretty dense and I want us to have time to digest this part. I always like asking people how they got into a field because I think that there may be people listening who could be thinking about what it would be like to be in that job and whether they want to get into that field. Maybe it’s a consequence of having two school-age daughters and constantly thinking about what to expose them to in terms of career and vocation possibilities so they can make the most informed choices I can possibly give them.

I thought that was a very good introduction to the basics of the Act and now we have a good idea of what sort of thing we’re dealing with. I’ll get into what we’ll talk about next week shortly. You know, I’ve said this before, but I just love talking with experts and learning more about these things and helping you understand them better too. It’s one of the things that puts me in the flow state, you know, maybe you’ve heard of the flow state, it was a term coined by Mihaly Csikszentmihalyi, the Hungarian psychologist who discovered this alpha brain wave state when we are caught up with something so completely that time disappears and we’re just present. That’s what this does for me. Maybe you can think about what puts you in the flow state.

In today’s news ripped from the headlines about AI, NASA has an AI called ExoMiner, which is able to find exoplanets, or planets in other star systems, some of which are Earth-like. No, it’s not an autonomous spacecraft traveling faster than light. It’s a deep learning network running on their Pleiades supercomputer that looks for very faint signals in data gathered by the Kepler space telescope to sort out real exoplanets from noise that accidentally resembles them. Jon Jenkins, exoplanet scientist at NASA's Ames Research Center in California's Silicon Valley, said that it comes with explainability built in: "We can easily explain which features in the data lead ExoMiner to reject or confirm a planet." "Now that we've trained ExoMiner using Kepler data, with a little fine-tuning, we can transfer that learning to other missions, including TESS," said Hamed Valizadegan. "There's room to grow." TESS is the successor to the Kepler mission and stands for the Transiting Exoplanet Survey Satellite, which tells you how they find the traces of these exoplanets, because it’s when they pass between their sun and us, which is called transiting. That little blip in the image that we get of the star when the planet goes in front of it, even though the planet is incredibly smaller than the star, is what gives us that data.

Next week, we’ll conclude the interview with Risto Uuk, when we’ll talk about the types of risk described in the act, types of company that could be affected and how, what it’s like to work in this field day to day and how you can get involved. That’s next week, on *AI and You.*

Until then, remember: no matter how much computers learn how to do, it’s how we come together as *humans* that matters.

http://aiandyou.net