Interview #2

**Jan:** So to start I'd like to ask if you could broadly describe your area of work.

**Interviewee:** Yes. So my work combined three main elements. Is a technology policy and geopolitical dialogue. My main focus is assistive technologies, healthcare and education. It encompasses effective, expressive, social AI and robotics for such areas as autism dyslexia, cognitive sensory and physical impairments.

In the past, I was a founder of companies in this field, but now I'm also working with cohorts of technologies funded by the government to participate in frameworks adoption and ethical frameworks. The geography typically focuses on the EU plus the US and also the Middle East. Two years ago, I was a curator of the global AI summit in Riyhad which was actually quite inspired not only by the idea of, you know, AI ethics, but also diversity in historical complexity of data sets behind the machine.

And we brought many technologies focused on healthcare education. So I continue to be kind of an ambassador of assistive technology. Happy to answer your questions.

**Jan:** Thank you. So in your area of work, I'm guessing you'd say there are ethical considerations in the development, of AI systems and in their deployment.

**Interviewee:** Yes.

**Jan:** Could you elaborate maybe a bit, you know, if you know of any existing frameworks or guidelines in that field of work that might ensure or facilitate that there's this ethical development and deployment?

**Interviewee:** Yes, so first of all there are three levels on how ethics can be let's say implemented and how it's facilitated.

So, first of all is a policy and in December, the European Commission finally introduced the AI Act and we have these amazing policies in emerging what we call regulatory sandboxes. So it's a special authority at the level of member states, which helps companies’ technologies specifically related to high and acceptable risks to be better compliant.

So they will test and provide the guidance for them is the first thing. But even before it existed at the level of the European Commission and Horizon Europe. As evaluators, we were provided by not only evaluation of a visibility and sustainability of startups but also ethical considerations. It includes data privacy and alignment with the GDPR.

Second, is a mitigation of potential risks and risks include physical or non-physical harm to well-being, safety, environmental impact, and also economic impact. Also now this era become much more let's say diverse. You also can find corporate ethics frameworks. One of the most famous is the Deloitte ethics framework or BMW.

So companies also try to be proactive about this. You also can find the ethical manifesto created by Google or Microsoft. What actually brings even more impact is intergovernmental frameworks and such examples includes UNESCO, UNICEF, the United Nations, World Health Organization, and on one hand, there are more kind of general frameworks, such as a UNESCO AI ethics recommendation framework.

And by the way, two years ago, when I curated the global summit in Saudi Arabia, I worked for the government and during the summit, they actually signed the compliance with the UNESCO framework. So it was a kind of a political statement. Then the government investors, and corporate world in this country will be aligned with the ethical frameworks.

So UNESCO actually have a lot of influence, but also there are more specific frameworks. For instance, there is a World Health Organization, AI generative AI for neuro conditions or neuro disabilities. There is a UNICEF work around AI in children or girls and digital solutions for people with disabilities.

So there are more specific frameworks. Finally, there's a big impact on the university academia and non-governmental organizations. So you can, for instance, I'm personally part of many ethical communities, such as the Montreal AI Ethics Institute. All tech is human is in New York, United States Alan Turing Institute, UK.

So typically, they are university-based or community-based. Or a kind of open-source movement. They connect independent ethics or AI professionals who create open source documents or open documents, which are available for everyone. Also, they trying to be kind of an evangelist of ethical AI and actually, even before the AI Act existed or a UNESCO ethics framework existed, these community initiatives had a lot of impact.

So, in short, now, ethics is actually quite complex, big world in industry. So there is a policy-driven ethics and kind of specific groups. There is intergovernmental ethical frameworks. There's universities and institutions working on it and finally, corporations creating their own ethics frameworks.

But what is important is that they all basically discuss similar things. They’re talking about fairness, transparency accountability, agency, stakeholder involvement and so on and so on and so on. So, at least, the basic statements and elements of these frameworks are similar, but there are some details dependent on the area, country, geography, or geopolitical differences.

**Jan:** All right. Thank you. There are a lot of these upcoming developing sort of guidelines because obviously in the past couple of years, AI has I mean, it's been around, it's been developing for quite a while, but in the past few years, there has been this sort of boom where everybody now is sort of in on it.

So people who may be in the past were not looking at it with, any sort of concern are now starting to consider these ethical implications, which has resulted in a much larger discussion and global debate on it. And from these guidelines and frameworks that you've mentioned, have you noticed any emerging trends surrounding the ethics of these different frameworks?

**Interviewee:** Yes, and there are both positive and negative trends. So, first of all, on one hand, ethical frameworks become more specialized. On one hand, is the area of specialization. That's what I mentioned, like AI ethics and healthcare, AI ethics and education. There's more focus, let's say, in literature, for instance, there is AI Ethics by UNESCO, but there is also the document called AI Ethics and AI Literacy for students and teachers.

So more specialized frameworks focus on education, or a kind of literacy in particular areas or frameworks focus on economic development or building, authorities overseeing such activities is the one thing. Second thing, there is more focus on diversity. Now you can find frameworks focused, let's say on women or solutions addressing women or what we call designated groups, let's say people with disabilities, or impairments.

The third, there are more country-specific frameworks. Let's say Arabic countries. They working on nature language pro testing for the Arabic language. There are companies in Saudi Arabia and UAE, that create their own companies and their own generative AI and they need specialized frameworks driven both by the government and the community.

So it's become more like a diverse, more complex is one thing. But the second thing, unfortunately, and is it and this issue I'm actually working on just right now since we have an emerging kind of a multipolar world, but it's just the beginning. There's a lot of collisions. There are a lot of wars, and military conflicts and countries are not in good relationships anymore.

There is a much more kind of a movement for sovereignty and countries actually try to create the walls. So if just a few years ago, there was more like an exchange, open world, we're kind of a, we're all coalition or alliances. Now there's more like a national AI ethics. Let's say in the EU. The recent AI ethics frameworks. There is more focus on disinformation involvement in election political influence, even the recent UNESCO and UN public letter and initiative signed by many, many countries. Was focused on not only children's safety in online safety but also the aspect of political influence from from beyond.

It's quite unique, it is also similar for the United States, partially for the UK and I expect for Canada. At the same time in the Middle East, they're not focused on it or in China, they're much, much more focused on investments. So they, their AI ethics focus on more kind of on the facilitation of AI.

So from, let's say this point, we're starting to see differences. So some countries trying to use AI ethics actually, for some kind of censorship or even building walls. So to protect their own citizens or their own values, their own culture. Some countries use AI ethics in a way, which actually may be punished by some violations, but still facilitates the investment. So it should not punish national companies and one of the examples is the recent EU Council of Europe treaty on AI. It was a very controversial document because on one hand this document actually had external stakeholders, not only from EU but also from the United States and the UK and these countries actively lobbied for some exemptions for private companies. So some big companies should be excluded from the list or from the punishment or for law enforcement. So there is more of this political influence and since, let's say, for instance, the EU would love to build up National AI unicorns. They need to balance.

Yes. They would love to be ethical, but at the same time, they would love to build big AI companies like Mistral in France and they need American investors. So they should not be actually against private companies in the US. So there is more of this political element and it's actually become very active just like a one, two years ago. It's specifically right now it has continued to evolve.

**Jan:** Alright, thank you. And so as you're saying, there's a bit of, of a, a back and forth sort of in the, in the sort of ethical aspect. And it comes sort of to the next question where obviously there are certain issues that you, that you brought up, accountability, fairness, transparency, and other things like, like privacy.

How would you say that, the trade-off in these different ethical principles is assessed in your field?

**Interviewee:** Yes. So first of all, I think it really depends on the country and there are different opinions from different stakeholders. So, for instance, some experts believe that from a bigger perspective, the antitrust policy in the United States is more efficient than the policy in Europe.

However, in my practice, I believe the situation is quite the opposite and I see, for instance, the cases, in which the governments in the US actually use the data from social media without any permission to confirm or not confirm pensions for people with disabilities. So they scan their personal data from social media.

And this case was in the United States. We couldn't imagine something like that in the EU, for instance. Bye. So it really depends on the area. So in short for my area of health care education, I believe this trade-off of privacy and safety is more focused on in the EU, at least now there are some bigger, issues for like, big antitrust cases.

Why? Because the EU are ready to create l like these guardrails for smaller violations, but they still may be not fully ready to fight big tech because they're afraid to lose these companies. So I think yeah, you would be kind of better but it's a question of time. How long?

**Jan:** Okay and so this, this next question is kind of difficult then I think. So what I was going to ask is in this sector, how can policymakers and regularity body regularity bodies effectively address these ethical challenges? I mean, I don't think there's one simple solution, but what would you say are some ways they could address them?

**Interviewee:** Yes, so since the beginning, I have liked the idea of the European AI Act and many of the emergent UK, and United States, they are somewhat eco and similar in different at the same time, but they basically articulate the few main things. The first is a category of risks, called the risk-based approach. So we identify the type of the systems and we create some categories for EU these are unacceptable, high, low and no risk.

So, let's say the first thing, the second thing, we identify some completely prohibited cases, for instance, for EU AI Act is a lifetime, a real-time biometrics system or social scoring is the second thing. First thing we address some more unique systems, which require more attention. So let's say generative AI, and there are specific articles about generative AI, and also we introduce specific measures towards this system.

There are articles focused on special transparency requirements and the necessity to declare the sources of the training data. Confirm that you're compliant with the copyright. So you don't use videos from YouTube or something which is protected by the copyright intellectual property law is the third thing and the fourth thing, something which is continuing to emerge.

And what I personally advocate for. Is more kind of a scenario and group-specific regulation when we more specifically address the area of healthcare, education, labour, a group with disabilities, children, and there is an area where actually, let's say, the US is better because. The Biden Blue Bill and AI Direction actually addressed not only the categories of risks but influences of AI on the economy. Let's say the cases in which AI potentially could replace the jobs. How we could address this threat? From an economics, or social perspective, or if we talk about education, we should talk not only about the ethical frameworks but also about the literacy about the preparation of a teacher's students to be better prepared for the use of AI understanding of this.

So I think in this way the US actually improve the vision of the regulation with a more area economy, social class or perspective. Is also important. And one more thing is the children's protection. In this area, the UK is slightly better. So just following the Digital Services Act in the EU, they introduced the Online Safety Act which actually has more measures focused on online protection for children and minors. So it is another thing. So in short, I think we're all kind of coming in parallel. Some countries are better at risk categorization. Some countries in addressing like ethics in industries or particular areas. Some, let's say, on protecting designated groups or children.

But together we can connect this vision. Hopefully cooperate with intergovernmental organizations such as UNESCO, and the World Health Organization, and I think it could be really like an efficient framework.

**Jan:** Thank you. So you are just mentioning now some of the issues and how some of the countries tackle them.

The next question is, what would the potential consequences of failing to tackle them? For example, you mentioned that there's the UK who have an act to take care of minor safety or so on. In these sorts of instances, if they were not addressed, what do you think the consequences could be?

**Interviewee:** The consequences could be terrible.

So I would love to share a few examples. One from the UK. I'm starting from another area, but also very serious. There is an autonomous system focused on organ transplantation or decisions about surgery. And if AI makes a mistake or if we're not able to provide ethical and efficient oversight for the AI system, it can lead to dramatic outcomes for patients.

In one of the most well funded in the most like a well developed area of AI and health care are the areas of oncology, DNA sequencing, and drug research. If we are not capable of introducing ethical AI, mistakes in these areas could lead to dramatic consequences to the whole population. Inefficient vaccines, Inefficient surgery appointments or organ transportation or drug research have very dramatic consequences, this is the first thing.

The second thing is what you mentioned regarding the minors. So, generative AI has a Phenomenal potential for creating deep fakes, creating, let's say, pornography adult content. It can be used to attack and target minors. It can specifically target people. Let's say on the spectrum. Let's say autism, and dyslexia who are not capable of recognizing what we call addictive design or manipulative design or manipulative scenario. So they could be involved, not only in some kind of harmful practices. They actually can, let's say, lose money or money of their parents or get some financial or physical harm or even pushed to to suicide.

So there are a lot of cases of cyberbullying when a group of people attack someone who is different and now they can use tools for that to automate this process which is quite terrible as well. The second thing and third thing there are many areas. I'm just like pointing out the most, like, critical and dramatic case is a military AI.

So, even though a few months ago, there was a meeting about kind of not-so-weak developments and military AI between the United States and China. The situation is quite the opposite. If you check the official AI strategy of France, the national defence is one of the 4 priorities and they actually even introduced this specific agency, a governmental agency focused on military AI and this area continues to grow.

And it's also very dangerous because these systems for instance sometimes can't properly recognize people with assistive devices disabilities. They can attack civilians. They can attack someone by mistake. It's very, very dangerous. So it's not ethical use of AI in law enforcement. Police and the military also lead not only to illegal detention of someone or arrest of someone but even sometimes attack or death.

So it's very dangerous.

**Jan:** So there's going to be now. Well, we'll see in the, in the coming months now or in the coming weeks, I imagine as well. The AI act being actually implemented across the EU block is sort of coming into effect and companies and governments having to sort of follow that.

How would you see the relationship between ethics and AI evolving in the coming years within the sort of industries that will be affected?

**Interviewee:** Yes, so if we talk about the EU, I would love to reflect a few things just after the AI Act was introduced. The SME, Social and Medium Enterprises Association of the European Union openly welcomed the AI act, and there was even a statement. We believe it would actually benefit businesses because the more transparent, the more accountable business will benefit society and citizens. So generally, there is a kind of conclusion and consensus that companies would love to see the AI act and they will try to be compliant and there's a lot of effort to better understand and they’re proactively working on it.

Also, there is a similar opinion from a venture capital perspective, from a startup perspective just in February I was part of a panel with one lawyer who was behind the AI Act and one investor. So, almost everyone was agreed that the AI Act is, is a necessary measurement, but at the same time, there is some criticism.

So what is this criticism about? One of the issues is too much politics about the AI Act, people believe that the AI Act could be more focused, not on, let's say, disinformation politics elections, but maybe slightly more on businesses and our citizens. So they believe that the AI Act has become more kind of an element of a Brussels effect and an opportunity to use the AI Act as some kind of asset for global influence or as a political tool.

In a similar way, digital services or digital market act also can be used as some kind of political tool. So, in short, companies will continue to be compliant and they will do their best to be compliant to act but I believe will be potential growing criticism of this political aspect. Maybe it will be reflected in some suggestions on how to improve it, or maybe to make AI Act more practical or more citizen-centric and more specific. But also it can lead maybe even to some political coalition who actually can be against it.

So, I think the majority would be will be neutral. So based on my community and people I know, even though they do not fully agree with some statements of the AI act or the way how it was introduced, they will continue to be. Compliant and one more concern they have is that in comparison to other processes, which were initiated by the European Commission, the AI Act was more lawyer-driven.

So, for instance, myself. I'm working on evaluating startups for the European Commission-funded projects and it was always driven by a technology community. So let's say you're a health tech entrepreneur, you evaluate health tech startups and they are like a community. It was more democratic.

The case of the AI Act, it was always behind closed doors created by politicians, and a group of some selected lawyers. So that's why society felt slightly isolated from this process. That's why there are some particular concerns on the purpose, even though everyone agrees that we need this regulation.

**Jan:** So it's actually really interesting the point you just mentioned that it felt a bit like on the lawyer side, the AI Act in particular. I had another interview earlier and the person actually one of their key focuses was that they felt the AI act was they didn't consult with the tech. People sort of enough on it and it felt a bit too much like they were playing politics and they were sort of listening to the lawyers.

That was their opinion and they sort of said that they would rather develop the technology and see where it would lead them They sort and they used America and China as examples in terms of that the development of their systems has been facilitated by maybe the lack of such strict or the word they used, I believe was over hard or overly hard regulations where they have softer regulations or not that many.

Do you have sort of an opinion on this or, or what you think the middle ground might be?

**Interviewee:** Yes, I think the text of the AI act itself is it was quite good. It's a funny thing. I work on the open program about AI ethics and I brought the program in October I used the AI act as a reference and funny enough in December the text became different so I released the course, but at just the last minute, the text became quite different.

My main concern is that people don't understand the actual impact of these slight differences. So just a quick example in the initial version majority of effective computing systems or so-called emotion recognition belonged to a low risk and they had special transparency requirements. So just providing necessary let's say explanation to the user, just pointing out how this system is used, the sources of the training data and so on. But very quickly on the recent deal, the European commission actually prohibited four cases of effective computing for workplaces, for education, for law enforcement and for immigration and also added many cases to high risk. So, why I have concerns for sure. Sometimes maybe it's necessary, but my, concern is how they could do this so quickly in just one month because it affects hundreds and hundreds of systems related to healthcare, including disabilities.

So just one line of a lawyer actually could lead to the liquidation of many EU-funded companies and innovation in scientific labs. So they just don't understand how much impact it actually brings. So the main concern and it's not only my opinion, but the founder of a responsible AI institute, how easy they add something, remove something.

Let's say they added concerns about generative AI. Just the last moment, they added these articles about generative AI, then quickly added some articles about effective computing. So, I have concerns that from the beginning, they didn't have enough knowledge about the AI industry. I mean, they had like a very general legal framework and in the end, they started to add some additional articles. But It's such a huge industry. You can't do it so quickly without preliminary consultation with many, many stakeholders and I believe it could lead to very irresponsible policies actually. So my main concern is not about the text or initial text and not even about the EU approach, but rather of the working group who was behind this document.

So I believe the European approach could be as efficient as a British one or a Chinese one or an American one, but I believe it should be kind of a more kind of responsible approach. So they needed to add all of these articles from the beginning and gradually improve it without such quick, deep changes, just the last moment which actually drove a lot of anxiety for so many businesses for so many technologists and in the last moment, people just started to lose the hope in this document, because it didn't look professional because just in the last year they realized there is a generative AI. So you didn't know about generative AI before. You didn't know about effective computing. Why have you started to think about this just the last month or two last months? And it was a kind of a concern for me and it's still quite concerning. But my hope is that My hope the EU has sufficiently strong social and technological institutions and institutes, which help society to provide feedback to improve it.

Hopefully, my belief. Because, or instance, the European Innovation Institute in Horizon Europe, they incubated over 7000 startups and emerging technologies is a huge ecosystem. There are many professors and scientists, and I hope this community help to improve this framework. But initially, yes, I believe it was not professional.

It's my main concern.

**Jan:** Thank you. So the last question now on that I have this this to focus a bit on on the end result of this thesis, actually. So the thesis is undertaking a systematic review and then it's going to compare the results with the existing obviously it's going to sort of try and deduce what sort of exists already or the sort of things that are already out in the public from what has been published also making reference to the legislation.

So of different countries and of different regions to sort of compare it with what's what's being written was being done. And also then once it's complete, I'm going to develop a website to sort of host the, well, to host all of the articles that the systematic review cites, and to also have some of these key points defined quite clearly.

Do, do you know, are you aware of any existing resources, which, which would have a similar functionality that would group together literature and legislation regarding ethics and AI?

**Interviewee:** Yes, I would love to share a few examples. My favourite one is the centre of digital policy and AI in Washington.

So I can share the link. It's I think the number one institution. So they mostly cooperate with Georgetown University and Washington University, and basically, they publish the democratic value of an AI index. So, they compare different countries, they sign different public letters they kind of facilitate this cross-Atlantic dialogue and publish as articles are interviews, the involvement of the policymakers for US from EU from UK, but also for, for the rest of the world in some way as well as the first thing and a second thing is a group of Ethics organizations, which I already mentioned before they're not so systematic, but they also could be useful as a source is the AI Now Institute in New York, the All Tech is Human community, the Montreal AI Ethics Institute in Canada and there are some other communities as well, but they're less systematic and the first option would be like the ideal one and is exactly what you described.

**Jan:** Okay. Thank you very much. Those are all of the questions I have for today.