RAW FILE

ITU

AI FOR GOOD GLOBAL SUMMIT

GENEVA, SWITZERLAND

DAY 2, 16 MAY 2018

16:00 CET

ROOM K

PANEL DISCUSSION

Services provided by:
Caption First, Inc.
P.O. Box 3066
Monument, CO 80132
1-877-825-5234
+001-719-481-9835
www.captionfirst.com

***

This text, document, or file is based on live transcription. Communication Access Realtime Translation (CART), captioning, and/or live transcription are provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings. This text, document, or file is not to be

***

>> Could I encourage you all to take your seats, please. We are about to commence the panel discussion.

We are to you delighted to have a one-hour panel discussion in which representatives from some of the many NGOs who are present at this wonderful meeting have an opportunity to tell us how they feel that the work of their organization intersects with this important theme of trust in AI.

So we have three speakers. With have a Hagit Messer-Yaron, a member of the Working Group on the World Commission on the Ethics of Scientific Knowledge and Technology of UNESCO. We have Elena Tomuta, Chief of Software Applications Section, Comprehensive Nuclear-Test-Ban Treaty Organization, CTBTO, and Joe Westby, a researcher on technology and human rights with amnesty international. They are each going to speak for ten minutes, then we will have a discussion with all of you at the end.

Hagit, welcome.

(Applause)

>> HAGIT MESSER-YARON: Thank you all. It's a great pleasure to

be here, and today was a fascinating day. I hope we will stand up to the level that you set already in this very interesting day.

As I was introduced, I am hear as a member of COMEST, the Commission on the Ethics of Scientific Knowledge and Technology in UNESCO, and I want to talk about trust in AI by educating engineers to ethically aligned design.

I should do it; right? That direction, this direction? No? What did I do wrong? This one? Technology.

I have the right slides here.

So okay. That's the right one.

Thank you.

I put together my different hats, as I said, being a COMEST member, but I am also a Professor of Electrical Engineering for, I don't know, 40 years doing research and teaching for engineers, electrical engineers. Actually, I am teaching signal systems. I never called it AI, but that's what I do. And I am also a member of -- okay, I am a Fellow member of the IEEE. There are only about 200 out of 400,000 member in the IEEE. And I am a member of the Executive Committee of the Global Initiative on Ethics in Autonomous Intelligent Systems.

So putting all together, I want to talk, as I said, about -- now it's not me. Ah, it works. Okay. I want to talk about educating engineers to ethically align design of autonomous and intelligent systems, and before starting, I want to say -- and I am talking about engineers. I really mean something much wider. I am talking about anyone who is involved in the design and development of autonomous intelligent systems, meaning not only engineers, electrical, robotics, et cetera, but also computer science and others. I use the term engineers, but it is wider. But it comes from the technical point of view and not from other aspects of design like policy, et cetera.

Also, I want to know that use the terminology AIS, autonomous intelligent systems, and not artificial intelligence. Because I think that's what we are talking about. Autonomous systems that are made of several components, and AI is only one of the components. Also, sensors, also mechanical parts, everything comes together into the system as building blocks, but the key features in terms of what we are talking about here in this conference is the system that we are talking about are autonomous and also are intelligent, so this is for my point of view a better terminology.

So we are talking today about trust in AI, and trust is very much related to ethics. Ethics is fundamental for fostering trust in AIS technologies, and it is crucial for current and future engineers to be educated on ethically aligned design. The need for ethics in autonomous intelligent system design has been emphasized at the international level by UNESCO, the World Commission on the Ethics of Scientific Knowledge and Technology, COMEST, and they actually, this committee already produced one report about ethics of robotics, and now the committee is working on another report on the ethics of IoT, which are very much related to what we talk here today.

Now, the challenge that I see in the committees I am part of sees is the curriculum of most programs around the world do not include developing tools for raising awareness to ethical consideration in AIS. Actually, it was mentioned yesterday by Wendell Wallach, what I just say. Of course, there are examples, but very seldom. Generally speaking, engineer is very specific, they take time to talk about more physics, more mathematics. But I think now we have an opportunity in joint effort of international organization can facilitate and accelerate a change in this. I am talking about a bottom-up initiative that can come together with policy and other top-down initiatives, but I think the bottom-up approach is most important.

I want to say that we are not the only ones who are dealing with this. Take, for example, the national academia of sciences, engineering, and medicine in the United States. Two years ago they published an interesting report. The title is "infusing ethics into development of engineers." And they actually reviewed I think 10 or 12 programs in different American institutions that teach ethics to engineers. This is an important initiative. First I like the title, "infusing ethics into engineers." Probably they are not conscious and you need to infuse them with ethics because they are dealing with other things. But this is important. But again, this is only a U.S. example, and we are talking globally. As mentioned during this session also, there are cultural gaps, and we need to think differently than only in the U.S. example.

Also, I want to emphasize that educating, not infusing, but educating engineers to ethical thinking needs to bridge over cultural gap between technology and different language, humanities. You can call it art, but I like the terminology art as it is used in this area. Sometimes we use art as anchoring for accountability, responsibility, and transparency. So we have art in double meaning here.

The American report, when they were talking about talking about ethics, it was quite broad. We have ethical guidelines, Code of ethics. We have professional ethics, which is the behavior of the engineer, if it is ethical or nonethical behavior. But now we are talking about different kind of ethics. Now we are talking about a technology ethics, meaning the effect of the technology of ethical consideration of the technology, and usually even if you have already classes for ethics in engineering, they are not for this dimension. So it is not sufficient.

Okay. So here. Trust in AI in the IEEE global initiative on ethics of AIS is part of the mission of this global initiative, and I put here the mission, I can read it to you. It's very short, and I think it's very important. The mission of the IEEE initiative is to ensure that every stakeholder involved in the design and development of autonomous and intelligent systems, so-called engineers, is educated, trained, and empowered to prioritize ethical consideration so the technologies are advanced for benefit of humanities. Very clear mission, and I want to emphasize the education part, as I said in my presentation. As you probably heard -- not probably -- you heard today about this initiative and the IEEE global initiative produced a very interesting report, ethically aligned design report, which is the -- version 2 is about to be finalized. And then the version 3 will be the last one. And the 244 pages are very interesting, but I put here the five what we call general principles that are recommended by this report. The five principles are human rights, well-being, accountability, transparency, and the wellness of (?). Sometimes when I am talking about teaching engineers into EAD, into ethically aligned design, people tell me but when you work for the industry, it's a structure, and the engineer is just engineer. There are other people that make decisions.

I want to finish my talk by giving an example how the end of the line engineer can actually infuse ethical consideration into the system. And I take only one example, I want to take the last one, awareness of misuse, and here is the example I want to show you. Let's see.

So the example is about a project that I could give to my students, third-year students, for one engineer, I can ask the engineer to construct an algorithm to count people in high-density crowd. Okay? This is a project. And if the student is lazy, he or she can follow a paper which was published in 2015, which actually described algorithm, and then the student needs only to implement this algorithm. So if you see this picture, you see that there is an image of the crowd, and then the project is to get -- to tell how many people are in this crowd. And in the right side you can see the flow chart of the algorithm. So you take one page of the image, and within this one page, what you need to do is to count heads. Okay? And then you do all kind of marking, analysis, et cetera, et cetera. At the end of the day from each page you get a number, and then you get the total number.

But if you look carefully this this picture, in this picture we have people. And we can actually recognize the picture. The resolution is very high, and you can recognize the people. So if you store this image after you finish counting people, there is potential misuse by taking these photos and (?) privacy of the people in these pictures. So what can the engineer do? And here is an algorithm for engineers to face such problem. First, recognize an ethical issue. And the ethical issue that any engineer can see that there is a potential violation of privacy within the algorithm if you keep the image as it is. Then get he the fix. You need to ask yourself do I really need all the details? The algorithm is based on counting heads. Do you really need the eyes, the ears, the mouth, the nose, et cetera? Then evaluate alternative options. Here, for example, the option that I would suggest is to add preprocessing part to the algorithm, and the preprocessing part can be edge identification, meaning you take the photo, but you only look at the contours, not at the faces. Okay? Then quantify the trade-offs, meaning this preprocessing, will it add processing time? On the other hand, it can actually save on memory use because you don't need to store the image. You can store only the contours. Then is the performance of the algorithm helped by using only the contours? But on the other hand, you prevent a potential misuse, you prevent bias because nobody cares about the dollar of the people anymore. You look only in the contour, et cetera, et cetera. Then at the end of the line, there are decisions.

This is the case where not the individual engineer makes the decision. It can be on higher level. But if it is initiated by an engineer who took an ethics course, then I am sure the decision will be right.

So to summarize my talk, trust in autonomous -- autonomous -- AIS is based on ethically aligned design. And regulation is not sufficient. The role of the individual engineer in building trust in AIS is extremely important. Education, and in particular education when the students are still in school, meaning if you are infusing ethics into the development of engineers, it's necessary to guarantee sustainable ethically aligned design of AIS. And if you will guarantee such sustainable ethically aligned design, then we can contribute to trust in AI.

So thank you, and I think I didn't pass my ten minutes. Thank you.

>> Thank you very much, Hagit.

(Applause)

And second speaker in this session is Elena Tomuta, who, as I said, is from the Comprehensive Nuclear-Test-Ban Treaty Organization, and Elena is going to be speaking with us on experience with establishing trust in an AI application for arms control. Elena, welcome.

>> ELENA TOMUTA: Thank you.

Okay. So I think this presentation is different than most you've heard today in two ways. One is that it's somewhat different the way in which we apply AI is a different application domain than most of the SDGs which have been the focus of discussion so far. We are concerned with arms control. And second, we've been actually developing this application for several years now, and we have just released an operational version earlier this year. So this is more of a case of describing the challenges and how we address challenges in establishing trust in a system throughout its development and operationalization.

So a little bit about the CTBTO, which is not an NGO. It's an international organization. So the CTBT is the Comprehensive Nuclear-Test-Ban Treaty, and the CTBT establishes a ban on all nuclear explosions on the earth's surface in the atmosphere, underwater, and underground. And the Comprehensive Nuclear-Test-Ban Treaty Organization for which I work is tasked for building up the verification regime that can detect undeclared nuclear tests. So we are based in Vienna, and the cornerstone of this verification regime, which is nearly complete, is an international monitoring system which, at the moment, has around 294 stations all over the earth. And these stations continuously transmit data to the CTBTO in Vienna, and we then process this data and produce data analysis products, which we make available to our Member States.

So a little bit about, just in very simple terms, what kind of processing we do. The largest part of this processing deals with seismic, hydroacoustic, and infrasound data. The idea here is if a nuclear test were to be performed underground or on the surface of the earth, that would cause a small magnitude earthquake that might not necessarily be felt by humans but will be detected on the stations of the international monitoring system. And the first processing stage, once we have acquired this data, we automatically detect anomalous energy levels in the signals from these stations, and we characterize these signals. And then there is a second automatic processing stage where we determine the events that led to these signals to be observed on the monitoring stations. And this step, this building event step, results in an automatic bulleting which then human analysts review and correct. Finally, we make this available to our Member States. And if there was a nuclear test, then an event would be present in this bulleting that would signal the fact that this test was conducted.

So the Treaty is not in force at the moment. After entering into force, we have timeliness requirement that we produce this review event bulletin within 48 hours after real-time, and at the moment, with the number of analysts we have and with the quality of our automatic event bulletin, we cannot stick to this timeliness requirement, so we need a longer time. And that's one of the reasons why we have been working for several years at improving this event-building module in our processing, and this is where we are applying machine learning.

So we are replacing the rule-based system that has been used in this event-building part of our processing system with the machine learning system, and we started several years ago a cooperation with Professor Stuart Russell at the University of California at Berkeley, who is actually leading the session on AI and satellite data. And the result of that cooperation was the so-called net Visa software, which is a software based on an extension of Bayesian networks as a methodology.

And just to very, at the very high level, explain what this is about. So essentially it's made up of a generative model that has features estimated through machine learning from past data, so from past reviewed event bulletins, and then it has an inference algorithm that determines the list of events that are most consistent with the model and that explain the detections at the stations. So this is what we are doing and why we are doing it.

And now I'll come to the topic of this panel, which is trust for the AI systems, and this is the definition of trust from the Oxford dictionary. So trust is firm belief in the reliability, truth, or ability of someone or something. And I will try to briefly touch on these three components of trust highlighted here from the point of view of our system.

So the first one is ability, and when we talk about ability, we generally mean -- we generally ask ourselves whether the system is able to do the task that it was intended for. And in our case, that task is to build a better event bulletin as a starting point for analysts than the rule-based system. So then we wanted to formalize what better means, and this has to do with defining ground truth data set against which we can measure performance of the algorithm. So our ground truth, we decided, is the reviewed event bulletin. We trust that what the analysts produce is correct and complete.

And from there on, we can then -- we have characterized the event set using measures that are well known from information theory, machine learning, statistics in terms of the percentage of ground truth that the algorithm finds. This is a number we want to maximize. And the percentage of false events that the algorithm builds. And this is what we want to minimize. So we call this overlap and inconsistency these two measures in our case.

So without wanting to describe this plot, just to say that there's a fairly sophisticated framework that we use to make complete characterizations of this event set, and we can then plot event sets in the space of this overlap and inconsistency, and we arrive at about 80% overlap with the ground truth set and about 40% inconsistency. So this is an improvement in overlap primarily of about 15% above -- over the rule-based algorithm.

So this is the primary measure of performance for us. We also have measures that are more specific to the domain, which is the quality of the events produced. So we want a good location, accuracy, a complete set of associations, and other things like this.

Then the second component of trust reliability is somehow inherently related to the notion of risk. So when we say we rely on something, we are also assuming that an undesired effect will not occur. And in our case, risk, the risk is that we will miss an event of interest. So man-made event of interest, a potential explosion. And from the beginning, one of the objections that were made against use of machine learning in this context was that the training data is strongly dominated by natural events. And so there's -- at least one can think that that might lead us to miss man-made events. This was addressed in two ways, already in the face of the design of the system through careful choice of the features to avoid this bias as much as possible, but also through targeted tests. And again, without attempting to explain this graph, what we did was take a set of events that were manmade in non-seismically active areas, and to then compare the performance of the rule-based algorithm with the machine learning algorithm using our measures that I talked about before. And so we found that also for this special event set, the machine learning algorithm does better.

Finally, belief. And belief -- I think we had some discussion this morning about trust versus confidence. Of course, doing the kind of systematic tests that I described goes a long way to increasing belief in the trustworthiness of an AI system, but there's still a component that has to do with stakeholder culture and how that influences perception. And in our case, the stakeholders are mainly two groups, the analysts who are going to directly use the outputs of the system, and experts from Member States or geophysicists, acousticians, who help us in developing the regime, and they have a regulatory and approval role in everything we do. And surprisingly it was the second group that was most skeptical against using machine learning, and to some extent, it was a matter of bridging the gap between these two scientific communities, geophysicists on one side and AI specialists on the other. And so we had these statements like the model is not physical and the system is a black box, which we have worked over the years to try to counter.

And so what were the most effective measures from what we've applied? First of all, transparency. So all our code is available to all the stakeholders, and it can be examined. That's more of a prerequisite, I would say, to trust. We have documentation as well as a tool that allows users to explore the model and to visualize the probability density functions that make up the model. And we also have some ability to explain individual events. So what the algorithm does, it assigns a score to the events it builds, and these can be decomposed into scores per feature, so it's possible to say to some extent what are the features that most contribute to explaining an event that was built.

But the most important, I think, was involving stakeholders in the testing and use of the algorithm, and we have something like five partner institutes in Member States with which we work, and they have done statistical analyses and compared the results of this software with their own regional and national bulletins.

And with this, I'd like to just put up these concepts that have to do with trust in a more general context that I discussed in this presentation. Thank you.

(Applause)

>> Thank you very much, Elena. Fascinating talk and a real-life case study of the kinds of things that we have been talking about here today.

Our third speaker in this session is Joe Westby, who is a researcher on technology and human rights with Amnesty International, and Joe is going to be speaking to us on a human rights framework for trustworthy and accountable AI.

Welcome, Joe.

>> JOE WESTBY: Thank you. Thank you for the opportunity to talk to you today. It's been a fascinating, important Summit, and there's been many interesting discussions around AI for Good and AI ethics.

But I want to be a little bit provocative today and say there are two concepts which I feel haven't been explored enough at the Summit and which I think are particularly critical when we are talking about trust in AI. The first one you probably won't be surprised to hear, as I work for Amnesty International, is human rights. When we are talking about AI and ethics, human rights is the only ethical framework that is universal based on binding laws that virtually every country has signed up to. As such, I think if we are talking about trustworthy AIs, we should be striving for human rights-compliant AI. Human rights have significant moral force and legitimacy in this space and should be the basis for a lot of discussions. I also think they have a lot to offer around some of these hard questions which we've already been discussing around what we mean by fairness, what we mean by holding algorithms accountable.

AI presents huge opportunities for human rights, and we've heard many interesting projects which I think talk to that. And I should say that Amnesty International is also using machine learning technology in our human rights research, so we are very optimistic about this technology, but we are also very aware of the inherent risks to human rights, particularly around privacy, discrimination, the right to work, and the use of AI technology in policing and warfare. And I should also want to stress that marginalized communities are the most at risk in this space. It's not just a concern for the future; it's already happening now. We have predictive policing systems currently in operation which are fueling a crackdown against ethnic minorities in China or entrenching discrimination against black communities in the USA.

Using autonomous weapons systems raises fundamental questions around accountability in warfare and the laws of war.

So this brings me to the second concept, which I feel has been a little bit missing, which is power. AI is going to further concentrate power into the hands of a few countries and companies.

Price Waterhouse Coopers estimates that 70% of the economic benefits of AI will flow to China and the U.S., and a handful of major companies are already leading investment in AI innovation and already have a monopoly on much of the data which is the fuel for AI technology.

Now, if we want to ensure trustworthy AI, how do we ensure that those in power actually build and develop systems that are deserving of trust and that are not harmful and violating human rights?

So human rights are fundamentally a way to empower individuals whose rights have been undermined and to hold the powerful to account. And we've learned through years of experience that the way to do that is through binding laws and regulations. Even with the goodwill of the major tech companies, we can't rely on corporate self-regulation alone in this space.

To give an example from another sector, something from my previous work, the (?) plaza garment factory collapse in Bangladesh killed more than a thousand people and exposed what happens in a highly unregulated industry. In response, governments and countries rushed to introduce new safeguards. But with AI, we can't wait for the first big disaster. The technology is moving too quickly, as we've heard from other speakers, and the potential impacts are too great.

So regulation is often seen as a bad word, and I know that many are skeptical about regulation in the tech sector. So I think we need to be really careful around government control of the Internet. Current attempts by Russia to censor and block parts of the Internet at the moment are drastically undermining freedom of expression. But this shouldn't be confused with appropriate legal safeguards that are critical in order to protect people's rights.

During the first industrial revolution, the enormous impacts of that, of the revolution, meant that it was necessary to introduce a broad set of social protections, factory regulations, environmental protections, welfare, living conditions protections. These didn't stifle the industrial revolution, and we need a similar kind of set of regulations and laws today to manage the fourth industrial revolution. But on the positive side, I don't think we should reinvent the wheel. This is where I come back to human rights law, which is binding and already established, and there's a lot of jurisprudence and standards which already apply this this space and should be the starting point. But we will need to ensure that we further interpret these to make sure they can be applied sensibly in the context of AI and Big Data and to bridge some of these policy technical gaps that were referred to earlier in the previous session.

In that vain, I want to use this opportunity to plug an initiative which we are actually just announcing today in Toronto, which is called the Toronto Declaration. It was developed by a group of human rights machine learning experts, and it begins to outline how states and companies' existing human rights responsibilities to prevent discrimination in particular -- we are focusing on discrimination -- how these should be met in practice when designing and implementing machine learning systems. We are aiming to use that as a starting point to develop more detailed principles and guidance which can guide companies when designing and developing these systems and which should inform policymaking.

We want to use this initiative to complement and build on the other important work and efforts developed standards and guidance for AI and ethics, like the IEEE, which was mentioned earlier by AI Now and others. These point a way for how we can address the human rights impacts of AI in this space and what specific measures companies and governments should take. For example, carrying out algorithmic impact assessments or ensuring the public bodies do not use black box algorithms. The critical thing is that these need to be enforceable and human rights have the teeth to do that is what I would argue.

So I really think in conclusion that human rights needs to be at the heart of discussions around AI and ethics, and I would really welcome the opportunity to discuss the Toronto Declaration and the steps that we are taking moving forward to kind of develop what this looks like in practice. Thanks.

(Applause)

>> Thank you very much, Joe. We now have about 15 minutes or so for Q&A. I am not sure if this thing is going to be feeding me questions from the app or not, but first of all, are there -- does somebody have a question from the audience? Yes.

>> (Off microphone)

>> Could you use your microphone, please? Just wait till the red

light comes on.

>> Okay. The question is for Joe. Hi. I am curious about the Toronto Declaration. Who -- are there various organizations behind it, and I understand the goals. It sounds very complementary to the one that's going on in Montreal. So was it developed by consultation with the citizens? Is it with the vector -- who and what is behind this? It's a very interesting initiative. Thank you.

>> JOE WESTBY: Is this mic -- yeah. Yeah, thanks for your interest. So it's really a starting point. It was developed by Amnesty International and Access Now, working with a group of machine learning and human rights experts, and we had a kind of event at Day 0 of the Rights Con conference yesterday, and then we are presenting it in the conference today. It's really as it sets out the existing obligations of states and companies in this space and starts to build more detailed guidance for what that actually looks like in the context of discrimination and machine learning. But we are certainly aware of the Montreal Declaration and have been in contact, I think, with some of the people involved, and you know, would see it as very much complementary to that and other efforts that I mentioned.

>> Thank you.

>> Yes.

Is the light on?

>> Yeah, it's on. Okay.

Question for you as well. Thanks so much for the information about the Toronto Declaration. In selecting the focus on and around discrimination, you know, I sense in a way one of the dangers of the rise of AI and its impact on human rights and redefining human rights is that we are redefining the ways in which human rights are not so much universal anymore but culturally adaptable. So for example, access to credit in a country like India where the imperative of development is way higher than what we see here in Switzerland or in Europe creates a condition whereby you would have more social acceptance for a form of discrimination. There a tension between access to credit or access to microinsurance and what we see as tolerance for discrimination.

How do you see that conversation moving forward, and how do you -- how would you recommend we deal with that tension? Because that tension is very much there. And I am quoting and talking about India because it has the critical mass as a market to move, you know, the tension away from or understanding of how universal certain rights are and certain definitions are, such as human dignity, for example.

>> JOE WESTBY: Sure. Thanks. That's an interesting question. I mean, I think, obviously, our position is that human rights are universal and universally binding, and not -- but there is a certain leeway for cultural sensitivities within that. To use your example of discrimination, it is not an absolute right. Human rights law allows for discrimination in certain very prescribed circumstances, and there are human rights test and legal tests behind

when that would be permitted. And the same goes for privacy and freedom of expression. These are -- human rights is already very used to dealing with niece kind of trade-offs and has the I would say capacity and jurisprudence already to deal with some of these questions. But I do think, as you say, we are in an emerging new space where there are new challenges presented by this technology, so we'll have to address what that means for existing human rights standards and human rights frameworks.

>> I am going to give myself a question at this point, and it's a question for Elena about this protocol she was explaining to us for using AI for better detection of possible nuclear tests. Elena, I remember at the time of the remarkable detection of gravitational waves a couple years ago, reading about the protocols that they used, you can see where I was reminded of that, this attempt to discuss -- to detect very minor kind of perturbations in the physical environment. Apparently in their protocol, built into it is the use of deliberate false signals that are fed into the system so almost everybody involved in the system doesn't know whether a particular signal they are looking at is one of these false ones which has been deliberately fed in or not. It's an interesting case in the context we are talking about because it's a deliberate injection of mistrust in order to Mr. Reliability, and I understand in the case of the -- to build reliability. I understand in the case of the actual one, which was the first sensational detection of a black hole merger, most people on the team assumed that it was one of these false ones

that they were looking at. And it was only after all this analysis had been done that they were able to find out that that wasn't actually the case.

Does your organization do anything of that kind?

>> ELENA TOMUTA: So are you asking whether we exercise with getting false data and whether we can --

>> But false data so that the people analyzing the data don't know that it's false data at the stage at which they are doing the analysis.

>> ELENA TOMUTA: Right. So okay. Maybe just to take a step back and give a bit more background here. We have -- one of the things we try to do to ensure confidence in the data is we are required to implement a public infrastructure so that all of the data is signed at the station, and the reason the possibility that one of the countries that host these stations deliberately introduced a false signal or removed something from the data that would cause the -- so the -- that would modify the data, effectively. And second, we are dealing with a lot of noise. So for instance, we have anthropogenic noise and a lot of noise sources that the analysts have to be able to discriminate with. So while we don't have the type of exercise that you've described, we constantly have to weed out a lot of signals that look normal initially for the automatic processing but that actually are noise simply.

>> That's very interesting because clearly the LIGO people have to deal with noise too. That must be a major problem. It may well are that the additional security dimension for which you use the key

signatures and so on would prevent the use of the technique that they are using of deliberate false signals in this case.

Okay. Let's move back to the audience. Rafael.

Could you use the microphone, please, Rafael.

>> This was a question for Hagit. I am also an engineer educator, and I wonder if you could give your ideas, your experience on how to motivate engineering students to engage with ethical issues. What I find is that often engineering students are very keen to the mathematical, physical engineering problems and see the other as a destruction of part of the curriculum.

>> HAGIT MESSER-YARON: I can only answer from my personal experience, which is different. I find out that the students are very excited about something different. It's only a small part of their curriculum, but it opens their mind. And sometimes even after five or ten years, they come back to me, and they say this was the best course that we had because really in many cases they are not aware of the issues that they tackle. And if you just get them interested, then they are brilliant, they are young, they like it, and you just need to give them the opportunity. It is much harder to convince the heads of the faculty to waste time on such courses. But once you convince them, then for the students themselves, it's very important and they are very open to it. This is my personal experience. It may be different in different places.

>> Did you have a question?

>> Yes. I am waiting. Okay. Thank you. My question actually

goes to Hagit as well. I am very -- you introduced the methodology of implementing EAD, this program. But I am just wondering whether engineers themselves are really capable of doing that kind of thing. For instance, you talk about privacy. But this is a quite controversial issue what privacy actually means. For instance, like genetic data, would you say that is a kind of private data or somehow public? So in these very controversial cases, I am just wondering whether engineers, they are really capable of debeat and also making a decision of these kind of things and then to -- what you really expect, infusing these human values into their design of AI systems or robotic systems. Thank you.

>> HAGIT MESSER-YARON: This is a very good question. But focusing on misuse, I think the basic idea is that you only use -- you only log and save the data which is needed for your algorithms. And then you don't care about the rest. It's very important. Sometimes it's very easy to say okay, I will save the photo as it is, you know, for future documentation, et cetera. But if you don't need all of it for the algorithms, for the product you really want to come up with, you should be aware of potential misuse of extra data, extra information, et cetera, et cetera. So the engineer is not the one to make the decision. What is the risk with the different kind of data. But the engineer is definitely capable of making decision about what he or she needs to implement their task. And they need to do the minimum needed, not more than it. Then the rest comes with legislation, with regulation. Most of the questions that you are

referring to are in different level. But the role of the engineer is really to minimize the use of data only to the thing that he or she really needs for the specific task. That's the way I see it with this very specific example. There are many others.

>> (Off microphone)

>> That's why we have different levels starting with the tough legislation. Then you have standardization, which is the kind of soft legislation. At the end of the day, there are -- and that's what I wanted to emphasize -- the everyday work of the personnel that can actually make decisions with big implication that the legislature and the others, by the time they see the potential use of it, it will be five years and damage can be done already. But some of it can be avoided on the technical level, on the practical work of the engineer -- the individual engineer. And this is, for me, a key issue because by the time the policymakers are aware of the risk and the problems, it's already there with the quick evolution of technology, it may be too late if you leave it only for others to take responsibility. Engineers should take responsibility of what they output they are using. This is my view.

>> Let's move on. We have time for one or two more questions. I think Yousef, then the person after you.

>> I think what you say is very, very important. I don't think engineers are in a position to decide what's the right way to go forward, but they have to spot when they are implementing an ethical decision. So I find this is a very subtle difference that a lot of

people don't understand. I assume that's what you meant.

And the other question is more towards you. I mean, human rights do not apply only to AI. So I feel like what you were talking is really technological progress. So my question to you would be is where is AI special in that it needs a special treatment? Because for example, all the bio guys doing genetic engineering and stuff, it's just not as sexy as AI anymore, but they still continue doing it, and same implications.

>> JOE WESTBY: Yeah, that's a really good point. I wouldn't say that AI was in some way unique. I would say that all of those different areas of technology would be subject to human rights as well, and we should also consider the human rights implications of those. And indeed, at Amnesty, we look at broadly different ways in which new technologies impact on human rights.

I think that what we have looked at with the AI is the -- and particularly with machine learning -- is what are the inherent risks to this technology? So for example, some of them have been raised already, like what about if you are using a data set which is already biased or discriminatory, and then that further entrenches discrimination if the system is not set up properly? Or with autonomous weapons systems, there is an inherent risk to human rights if the systems are making critical decisions about life and death on the battlefield, where is the line of accountability and who do you hold to account when those systems are in place? That's why in that particular context, we have called for a ban on fully autonomous weapons systems because we feel that they cannot be used in a way that's in line with human rights.

>> I am sorry, sir, we have two minutes left, and I promised this person over here that she can have a question, so I will make this the last question.

>> Also for Joe. Regarding the trade-off between privacy and say using data for good, like to spot food shortages, et cetera, yesterday at Robert Kirkpatrick said privacy is a right, but so is food and water. Can you elaborate on how and who makes these trade-offs when necessary between human rights and opportunities to use AI for good?

>> JOE WESTBY: Sure. Yeah, I think that's a really good question, obviously very pertinent to these discussions. I mean, I think without wanting to get too technical, the test with privacy is there can be legitimate interferences with the rights of privacy under human rights laws, so long as they are necessary for legitimate aim, such as for enhancing access to food. And if they are set out properly in law and if they are done proportionally. So that's the kind of framework in which we would assess those kind of trade-offs, and that would have to be the government that was making the decision when implementing one of those kinds of systems, but fundamentally, they would be subject to challenge in the courts if it didn't meet those tests.

>> Okay. Thank you very much. In a moment I am going to hand over to Stephen Cave to tell you about the breakout sessions happening downstairs in the next hour, but first please join me in thanking

our three panelists for fascinating presentations.

(Applause)

>> Thank you very much, Huw. Thank you very much, ladies and gentlemen.

Now, today you have heard nine fantastic, innovative, interesting, important projects be presented over the course of the day, and you've heard many other stimulating thoughts about building and earning trust in AI, both from the floor and from our panelists. But now is the most important hour of the day because it's the hour when you get to feed in what you think about these projects and the directions they should be taking and your ideas for how we can copy them and develop them and build on them.

So I know you are all tired, but it's time to join in, and it will be energizing, and it will be the perfect preparation for the reception later on this evening. So the breakout sessions are going to be held downstairs directly below here, so if you go to the staircase or the lift just over there and go down one floor, that's where you came in, where you registered, where you have to click your badge. The breakout sessions are just there.

If you make your way down there calmly and efficiently, pretend it's a fire drill, then when we get down there, then we will divide up and I will give you more instructions, and we'll get going. Please don't get lost on the way. Hope to see you all downstairs in a moment. Thank you.