FINISHED COPY

ITU AI FOR GOOD GLOBAL SUMMIT
POPOV ROOM
BREAKTHROUGH SESSION: SOCIAL GOOD DATA

GENEVA, SWITZERLAND
8 JUNE 2017
1530 CET

Services provided by:

Caption First, Inc.
P.O. Box 3066
Monument, CO 80132
1-877-825-5234
719-481-9835
www.captionfirst.com

***



***

>> MIGUEL LUENGO-OROZ: Bonjour. Hello. Okay. So we are going to start. Welcome, everybody, to the social good data session. My name is Miguel Luengo-Oroz, the chief scientist and this session is going to be framed by three questions. So the first one is which are the challenges and opportunities in data for social good, but to achieve this for everybody, everywhere.

The second is about the practicalities and the real challenges when we implement operationally all these innovations. The third one, what do we need to make sure there's diversity in the AI space and that we have global reach.

For that we have an incredible panel here, which is a panel with practitioners, with people that are actually working to solve this using artificial intelligence and machine learning but have a lot of expertise in the field. We have Moustapha Cisse, a researcher from the artificial intelligence group in Facebook. And we have libraries and archives Bromley, Chris

Earney, deputy Director in the innovation team in the UNHCR, High Commission of Refugees. We have Fei-Fei Li at the Stanford Lab and Leila Zig from Wikimedia Foundation, senior researcher there.

Lastly we have Dr. John Quinn. He is the scientist at the Pulse Lab Kampala. He has been working on artificial intelligence in Uganda over the past ten years, first in the university and later on at the Pulse Lab. He will start the session with a brief presentation on what is making rods for social good in a country like Uganda. Please.

>> JOHN QUINN: Great. Thank you, Miguel. Delighted to be here. Yes, I would like to briefly set the scene, very briefly. I would like to give two examples. Let me get my slides up.

So where is this coming from? From U.N. global pulse and this is an innovation initiative begun by the previous U.N. Secretary General, to harness social data for the public good and look at how the prolific resources can be used for social good. We have three labs, New York, Kampala and I'm the technical lead as Miguel said.

I would like to briefly highlight two categories of data, which I think are interesting in the context of social good and artificial intelligence. The first example given that we are here in ITU at telecom's data example seems appropriate. What you see here is a map of travel times within Uganda. This is several hundreds of millions of records and as people move around with their phones, making and receiving calls and using data, other types of things we can look at things like what is the flux of population. Where are people moving, how long does it take them to move around? We can see here the eke center is cam Pamela. Our labs are in many other parts of the U.N. system within Uganda and to kind of fill you in here in the north of the country there's a large scale refugee crisis, nearly a million refugees in the last year or so. This can be processed in realtime. So as we kind of roll forward the clock day-by-day we can see what is going on in response to weather events, things like what is happening deteriorating or improving road conditions and so on and so forth. You can see here there are logistical advantages and this helps to understand in realtime the issues to deal with disaster.

For instance, this is just one category of this type of data. I characterize this category as the result of what people do, people's behavior. This might manifest itself in terms of the byproducts of digital services. It can be various types of sensors, for example's people behaviors may result in satellite imagery and libraries and archives here is the expert on that.

The second type of category I want to highlight is to do with what people say when we are trying to understand the issues

around social good and promoting the global goals here. Then we can try to infer things from the sensors but also just see what people have to say about their own condition, right? This is another map of Uganda. This is the approximate locations of all of the radio stations within the country. There's a bit more than 200 of them.

And radio is an extremely interesting media. It is one of the main social media in Uganda because people phone in via mobile phones. That makes it a discussion and it is a Forum for people to discuss their needs, concerns, and so on. So it is an abundant data source. We estimate seven and a half million words are spoken we are today on Ugandan radio, ten times the complete works of Shakespeare. This is on a great range of topics. In areas which are prone to various types of disaster and experience various types of challenge. Extremely rich data source.

The issue with radio, an this is where the AI angle comes in. In its basic form it not very amenable to analysis or understanding. Let's imagine the Internet has no search functionality. In order to access the Internet, the pages of the entire web are paraded in front of you in an order and speed that you can't control. It might be quite difficult to really make the most of the resource there. That's kind of how radio is now, right? What we would like to do here is to add in the kind of searching and indices and categorizing functionality that makes the Internet useful and potentially unlock this information.

If we can do so, all the information coming out here can be a kind of massive sensor network. Then we can really look at things on this national or global scale which we would like to do.

So that's a challenge with radio. To do this processing. One of the things which is easiest, accessing it, the nice thing about this data source it permeates our beings literally all the time. We can pluck this out of the air and try to do something with it. We have these raspberry pis deployed around the fields in the country. The difficult bit is doing something with it. We need to do speech and language technology. We have to develop speech recognition in three Ugandan languages, very technically challenging particularly as there are few existing resources for these languages. And many of the examples we have been hearing of where AI is expected to have great impact depends on their being millions or hundreds of millions of training examples, and what if those don't exist? We have done that with language. We can start categorizing things and we find interesting stuff.

I'll conclude with a few examples. One is pertaining to the refugee crisis I referenced earlier, when we listen to what people say in your own words about the host communities where the refugee influx is coming, we get interesting issues that may not have been known otherwise. One, there is a great movement of people through wilderness areas into northern Uganda. It turns out that disturbs the wildlife. In particular hundreds of elephants seem to have come across and trampling people's crops and causing destruction there. We have a discussion like this, this is translated from, "elephants are spoiling crops like milo and sorghum. The people are threatened, so they will poison the elephants to prevent them destroying the crops."

You hear discussions about what type of poison is best to kill an elephant. These are things that no one had thought to survey, but this can raise the hypothesis that that might be something interesting to look at.

We see things to do with the delivery of healthcare services and this is a testimony from someone who says I went to the hospital and they made me sweep the yard before they would attend to me as a patient. This can be interesting in terms of monitoring service delivery. We hear things about, for example, again road infrastructure and travel. The problem for us is the road going from Palaro to Paibona, the problem is there is no connection there. That concludes the examples. I think that maybe sets the scene a little in terms of the data resources which are around. At this kind of global scale. There are global data resources that we can take advantage of if we can harness them in the right way. There is a few issues which I think Miguel alluded to earlier. One, getting from the prototypes to deployment is extremely difficult. And there are some programmes quite interesting reasons for that. The other is, to make the most of the global challenges, we need a global community of practice, right? I mean, these examples from Uganda are done with our team which is predominantly Ugandan. That's something that we are going to need to do in order to harness the data for the best.

I think that concludes my brief scene setting here. Thanks.

>> MIGUEL LUENGO-OROZ: One question before you go. How long takes like making this bridge and how complex is it?

>> Developing a prototype is a fairly quick thing. The whole project might be measures in years. I think what is interesting is to look at the different stages of this development and it is the rather unglamorous things that take the time. So in the course of a 12 or 24-month project, half of that time might be data access and data cleaning, basic data engineering might be the predominant part. Some of the machine learning programs, the actual machine learning, my favorite

part, but disappointingly you find a 20-year-old algorithm does the job well and after one week you're done. It might be possible to spend another year getting the extra 1 percent performance with state-of-the-art, that is a small part of the development time.

>> MIGUEL LUENGO-OROZ: Thank you, John.

So now that John has set the stage we are going to go through questions for all our panelists so they can share with us their experience. So Moustapha, which are the practical issues involving deploying an operational systems at scale? What have you learned from your experience in Facebook?

>> MOUSTAPHA CISSE: So I think deploying such systems at a scale of Facebook, there are challenges that maybe you do not encounter when you work at a different scale. So when John was mentioning, there definitely are technical challenges that may not exist in these particular cases.

What I would like to emphasize is for a company like Facebook, for example, and I'm sure it's true for many other companies, the data is very important indeed because it is what provides us the means to make services for the users. But even more important is the technology and the applications and everything that is output by the research centers and that the companies out-source. Put at the disposal of all the users so that there is this, I would say future circle that people can use all these applications and it is open source to develop applications for speech recognition, for example, that John would use to analyze the data coming from the radio. And improve his applications. And promote the use of these sort of data.

I think that is an aspect of the problem that I have not heard a lot in most of the discussions here, but it is important to have available data. I think most of the data that we need to solve the important problems is already available. Sometimes what is not available is the technology to, the right technology to process this data. And companies like Facebook or Google or others definitely have the expertise. It is interesting to see that more and more these companies are open sourcing the code and the technology to do that.

>> MIGUEL LUENGO-OROZ: Thank you. So talking about technology and the possibilities of open source software and AI capables, I would like to ask libraries and archives, he is principal analyst in UNICEF, he has been doing critical work for United Nations analyzing satellite images for many agencies. Tell us more about your experience. What is the reality today of those analysis?

>> LARS BROMLEY: Yes. So just to give some context, what my office does, we are basically an on demand satellite imagery

analysis service for the U.N. system. So all day, every day we are getting requests to get a satellite image of a location, analyze the satellite image and provide information based on the original request. So that might be something different and simple like analyzing a refugee settlement, analyzing destruction in inaccessible areas of South Sudan, looking for roadblocks along humanitarian convoy roads, stuff like that.

Really, the fundamental issue for me and my team, there's eight people who work with me, is simply time. So we get these requests constantly. We basically have to figure out the fastest way to answer the question. Within satellite imagery it's an interesting discussion to have here because I have heard people say that satellite imagery is the original big data. For decades we have been dealing with this issue of very large file sizes and analyzing large amounts of data. Now, a satellite image is not just a photograph taken from space. It is basically a relatively rigorous measurement of light bouncing off the surface of the earth and being captured by a sensor. You can think of a satellite image as a number set. Sometimes a large number set. If you had a satellite image, just imagine, of the coastal area somewhere. So you are looking at the ocean. You're looking at the land. Maybe there's a forest on the land. Theoretically you can train a computer to discriminate the ocean from the land from the forest. And in a perfect world any satellite image you got you would just push a button and basically the computer would go in and tell you what was refugee shelters, what was destruction, what was something else. However, the fundamental problem, first of all you have the atmosphere acting as a filter, a chaotic filter between the ground and sensor. Even if you got ten satellite images of the exact same location, as a number set each of them would be different, which is where the complexity of analysis comes in and the ability to automatically classify that image gets degraded.

I can tell you since I was a young man -- I have been approached by software companies who told me that they solved this sort of problem. We used to call it feature ex-extraction. We still call it feature extraction. But the marketing materials fall well short of what the actual capability of the software actually is. So again, coming back to my office and my daily routine of getting requests and answering questions, you are faced with this potential trap of do I spend eight hours doing the automated machine learning classification and then I come up with 80 percent of the answer, and then I've got to send an analyst back in there to clean everything up and come up with the actual precise data set. And that's simply what we can't afford to do. So nine times out of ten I'm basically giving the

problem to human intelligence and having human intelligence basically answer the question for me instead of going down that rabbit hole of automated analysis that then has to be cleaned up by a human. You can do a fascinating Ph.D. on that for several years. The problem is, I have about eight hours to answer the average question.

>> MIGUEL LUENGO-OROZ: Thank you, libraries and archives. It's very clear, real world challenge that though we might think it is solved it is not really solved.

Going more into the -- yes?

>> LARS BROMLEY: I will say we are collaborating with global pulse now on automated extraction methods. We are closer than anything I have ever seen before in getting good, reliable answers out of this. However, yeah, am I going to use this method in the office tomorrow? Probably not. Our colleagues are discussing it as we speak right now. So we are developing it. We are pushing it forward. But yeah, it is by no means what you are going to see as a finished product.

>> MIGUEL LUENGO-OROZ: Thank you. So yes, going a little bit more into the human dimension, I would like to ask Chris, he has been working more than ten years with refugees in places like Darfur, Baghdad, and just back from Syria. In a context where privacy is so important, in a context where protection of refugees is such a complicated issues, where do you think that AI and data can help?

>> CHRIS EARNEY: That is a pretty big question. So I'll do my best to break it down into a couple of different areas.

For the UNHCR, we work for the refugees and other displaced people around the world. From the emergency situations such as the one we were talking about in Uganda, right away to three durable solutions. When people flee the country of origin, they cross a national boundary. Will they integrate immediately or be settled to a third country such as Australia or Canada, or will there be peace in the country they fled from and can they return?

Our spectrum through which we understand refugees and their needs with respect to protection is huge. Sometimes it can be a lifetime. For me, the question of the application of AI, how we understand, how we use and how we analyze information and therefore data can be broken down into probably three key areas. One of those areas is access. So how can we use AI to improve people's access to UNHCR, to our partners? How can we improve their access and therefore their agency in approaching us?

The secondary is how can we increase our reach? We are a small agency compared to the number of people who we are trying to provide protection and assistance to which numbers about 65 million people over 120 countries. Yes, we have a staff of

around 12,000 people, but that is not nearly enough to provide that protection and assistance in the most dignified possible manner.

For me it is a question of access, how we can improve access of refugees to UNHCR. Secondly, how we can increase our reach and, therefore, our impact.

If I look at AI today and I look at the data sets that UNHCR currently collects and interacts with and makes decisions on the basis of, we can already apply AI in many, many, many different scenarios. If I look at the really prosaic spaces of UNHCR, booking a trip to go to Lebanon to visit my colleagues to have a meeting about AI, for example, I can already think about 20 different ways that we could maybe use AI to replace people. To make some of our bureaucracy which is there to protect but also to empower us to make better decisions and to provide that protection and assistance more efficient.

If I look to the future, and I think about AI and our future as a U.N. agency we are not going to be dealing with fewer people. We are likely going to be dealing with more people. The rate of displacement isn't declining. The resources that we have at our fingertips to provide that protection and assistance, there is a question mark. Will those resources come to us, can we count on our large donors moving forward or not? When I think about artificial intelligence there, I think about more efficiency, increasing the dignity of refugees and other populations with fewer resources but with a better understanding of what their needs are.

So when I think about artificial intelligence in the future, I have to reflect upon our own experiences using artificial intelligence with our partners in global pulse and together with our partners in UNDESA more recently. The problems we have predicting population movements within the European refugee Christs understanding the intents of people moving through Europe in that crisis, understanding the sentiment of the host populations who are receiving millions of refugees on a daily basis and how they feel about these new people arriving in their territory. There are probably three key lessons that I take from that experience. Number one, partnership is key to UNHCR. We don't necessarily have the right skills, the right expertise, the right processes, it right tools, the right services to make the best that we can of artificial intelligence and big data. So partnerships for us is key. We need to remain very agile and we need to make sure that we are open to partnership and collaboration. That's one.

The second lesson as I think we learned over the past couple years interacting with AI is that privacy remains a critical area that we can not, cannot ignore. And we are a protection

agency. We deal with people's identities on a daily basis. We need to make sure when we are delivering protection that we keep people's privacy's as central to any investigations we make with AI.

And then thirdly, not engaging is not an option. Not engaging with AI is not an option for us as UNHCR. It is something that we have small successes in and it is something that we need to have larger successes in as we move forward into the next five to ten years. And then last thing is AI is really exciting. It can really, really improve people's life. It can have a positive impact on refugees, but you need to make sure you are trying to solve the right problem.

>> MIGUEL LUENGO-OROZ: Thank you, Chris. Partnerships, it seems like here we have already identified several problems with the matter. It is not target advertised, it's something that matters to a lot of people. I will ask Fei-Fei now about technology. Do we have the tools to attack these problems? Because we have to talk about a lot about deep learning. It has been the revolution, but you need a lot of data and you need a lot of training data. In this context, many times we don't have enough data. Many times we don't have enough training data. Sometimes we are talking about disasters. You don't have many disasters, to look back at what happened before. Maybe, this question for you next, what is the next frontier in AI? Maybe there is a need to create a new research agenda where you think about areas where you needless training data and can work with less.

>> FEI-FEI LI: Right. Thank you for asking that question. I think it put into context that while we have so much success and now excitement of AI, AI is a very young field. It is 60 years old. We are seeing one of the first waves of success.

But by and large there is a lot of open questions in terms of where this technology is going and in context of helping real world, solving real world problems. We are still very much in the middle of developing some of these solutions and techniques, not at all done.

As you said, in the past about seven years or so, the biggest success story in AI is really supervised learning and deep learning using big data. Big data meaning labeled data. During training time you have enough labeled data to train say a convolutional neural network or LSTM or whatever. That is a very good scenario that can already solve some of these problems, right. But if you look at how humans learn, especially many of you probably have children. I have two very, very young kids. One is five years old. One is one year old. If you look at the development of kids, learning comes in all different forms. There is one shot learning where you might be

offering all the capability of your kid looking at one tiger picture in the children's book and go to the zoo and recognize a completely different 3D tiger. That is the classic example of one-shot learning. There is transferred learning. The kid, you teach the kid to hold a cup. Next time you give a different cup, he or she can hold a totally different cup.

There is unsupervised learning. There is reinforcement learning, semi supervised learning. Listing all this to really give the sense of there is a lot more to be done in developing AI technology and there is -- while I think it is really wonderful to have the excitement and the attention on AI, I think there is also importance in recognizing we need to not only harvest what we have developed but also plant the seeds of developing the next wave of AI technology.

>> MIGUEL LUENGO-OROZ: Thank you, completely agree. Taking what you said in the previous session, we need to plant these seeds all over the world. We really need diversity and that no one is left behind, that all countries use these new technologies.

I would like to ask Leila about her research in Wikimedia about diversity and knowledge gaps.

>> LEILA ZIG: Yes. First I want to say a few words about why we are talking about which can peed I can't. You have seen the content on Wikipedia but you may not be familiar with the machinery of artificial intelligence that is going behind the project. So volunteers on which can peed, close to 200,000 of them come every month and do at least one edit. But hundreds of them come every month and do develop on projects. They work on vandalism recognition technology that can help them, detect vandalism, copyright violation detection, these are the things that editors do on Wikipedia. At Wikimedia Foundation we are interested in machine learning specifically and artificial intelligence broadly for basically helping us identify, for example, harassments on Wikipedia talk pages or identifying what articles to be created next and helping editors to determine what content to create let and em developing labels, quality labels to help people identify what are the higher and lower quality projects. AI is more hidden for our projects but something that is deeply in the heart of what we do in Wikipedia and Wikimedia. We run into a lot of challenges p and there are lessons to be learned. I share a few of them here. One of them is the issue of free open licenses. So if you don't want to leave people behind, you need to make sure that the research that you do in artificial intelligence, the result that you share with other people is one, not behind pay wall because not many of us are willing or able to pay the 40 Euros to read your Article. But also not only is it not behind pay wall, but also

freely licensed. This is an important component of empowering people not to become the consumer of your research, but become the kind of people that build on what you do. I give you a very simple example. In terms of like freely licensed. If I publish a paper and this paper is not under free license and basically putting a significant barrier for a teacher in Kenya who wants to use a plot in my paper in his other her classroom, they need to get permission of me and the rest of the authors to use that content in their class because fair use doesn't apply everywhere, right?

So number one, open free license. This doesn't apply only on the research publications that you have but also applies on the code that is behind your research, the data that is behind your research, the labels that is behind your research. You need to truly engage people to understand the thought process and the research and all the underpinnings of what you have done for people to be able to use them.

The second thing I want to talk about are legal challenges to digitalization. This is something that Wikipedia and Wikimedia projects deal with very frequently. We talked yesterday about the importance and issues around digitalization in the countries that we are going to. I want to share with you one example of challenges that we have. So on Wikipedia, there are Articles about business cultural heritage, monuments across the world that are documented by Wikipedia editors. As you can imagine, the editors put the text, they basically gather knowledge about these monuments in text, but they also want to share the knowledge visually. What they do, they want to be able to take a photo of a monument and share it on Wikipedia. Now, the machinery is there. We have tens of thousands of volunteers who are willing to go out and take a photo of a monument and upload it on Wikimedia comments. For us to be able to use it on Wikipedia. Kind of the citizens are involved. They want to do the work. They want to do it for you for free, for free knowledge.

What stops them is that they live in countries that don't have freedom of panorama, freedom of panorama is a provision of copyright law that would allow any of us to go out and take a photo of a public in a public place and legally share it under a free, open license. Now, we don't have to go to the most remote places in the world to run into this problem. Freedom of panorama doesn't exist in many countries. The familiar examples for us is Italy. If you go to Italy and take a photo of a public monument you cannot load it on Wikipedia comments. Greece is the same. When you talk about not leaving people behind, this is not necessarily Africa or places that don't have many of the basic things we are used to. We are talking about

very familiar countries and these legal challenges have to be removed. That is one of my hopes from United Nations being behind this, union UNESCO being behind this, to put your energy and effort for removing some of the legal challenges for digitalization of content and training sets and everything.

The third thing that has worked well in the context of Wikipedia, we have learned that centralization works but you have to provide the infrastructure, the basic infrastructure and AIPPIs for people who use Wikipedia content, to use basically the different AIPPIs, the data that is associated with Wikipedia. When we leave it up to people to decide what to do with that data and how to you it, there are hundreds, thousands of developers and third-party developers that use Wikipedia content, which project survives, which project thrives ends up being a natural selection process. You allow many people to use it. Whichever project gets a lot of people behind it is going to be the will project that is going to survive.

And the last thing I want to say is kind of the message of hope. In the sense that we talk about ethics. We talk about privacy. We talk about security and all of these are topics that are very dear to our heart and topics that we have to deal with on a daily basis when we do project research on Wikipedia project. When you empower people to do the right thing, they actually do the right thing. When you give them the resources to be the consumers but to engage with your project and when the goal of the project is something that has social impact, the majority of people do the right thing. Wikipedia is one of the best examples of that.

>> MIGUEL LUENGO-OROZ: Thank you. Talking about empowering citizens, do you think AI can be a tool for leap frogging technologies? Can a farmer in Uganda with a smartphone have a personalized suggestion for which crops have to plant when or where and can be as believable as the guys with the track at MIT? We haven't talked about this.

>> MOUSTAPHA CISSE: This is an important question. To iterate on what Leila was saying, sharing the data, sharing the code and whatever technology is very important, indeed. But I think there is more important threat that the community is facing as a whole. All the AI community currently is biased towards solving white male problems basically. And if you have attended the sessions since yesterday, just to give an example, we have been talking a lot about the main threat of AI being job displacement problem and the very example that people cite when they talk about these problems is self driving cars or self driving cars replacing all the professional truck drivers.

So if you look at a problem deeply, the number of professional truck drivers in the U.S. is 3.5 million. And 96

percent of these 3.5 million truck drivers are male. Which means like it is mainly a white male problem.

In contrast if you consider a country like Uganda, 37 million people and 73 percent of that population is agricultural workforce of the country, basically 30 million. What they do mostly is, one of the most important plants cultivated there is Casava which is very important for food security, et cetera. So there are threats. I know John here has been working on that, dealing with the problems related to white flies, et cetera, which are very important problem. If we manage to solve these kind of problems we would have impacted 30 million people and their families. So maybe even more than that. 30 million people out of 37 million people. We are not doing that. We are pouring huge amount of resources to solve a problem that at best concern 3 million people out of 300 million people in the U.S.

This is the state of where we are. We are basically solving white male problems. I think this is the most important threat that we have to tackle and make sure that the whole community is aware of this problem because we are only solving problems that we are exposed to, basically. And some people are to be blamed. Fei-Fei has spent a lot of money, for example, collecting Image Net. Image Net has been instrumental in the progress that has been, that we have witnessed in deep learning. No question. But then are the same resources spent to build data sets or to build technologies to solve other kinds of problems that are more important to other type of populations in the world?

I have serious doubt about that. And the concern is that we are here this week because we think AI is in a unique position to have a particular impact in people's lives. But if we think we are in a unique position, then we have unique responsibilities. And I have serious doubts that we are at best matching these responsibilities. We are not. In fact, I really believe we are not trying to solve these problems.

This is a concern that I wanted to share. I hope it will be, people will talk about it in different areas and probably can hopefully have some impact somewhere. But it is a very serious concern.

>> MIGUEL LUENGO-OROZ: Thank you, Moustapha. I think the panel cannot agree more with your words. Now we are going to have some questions from the audience. Let's take four, for instance. One, two, three, four over there. I don't know where the mics are.

>> AUDIENCE: Hello. I'm from McGill University, the Alfred project. We are applying deep learning principles to psychiatry. I am from Canada. One of the things I wanted to talk about, Moustapha talked about it in your closing remarks. We are talking about access to data sets and technologies. So

make it truly, to use the term that was used a little earlier, probably by Dr. Fei-Fei Li was Democratization. So the Montreal Delegation, we have been giving a lot of thought to it. We have certain principles that we've sort of arrived at for a Democratization. I wanted to share some of those and get your opinions on that.

So we received, in terms of accessibility, lowering the costs, making it representative, having standardisation of the data and interoperability. Giving the people voice in this for the data sets and data technologies and most importantly having some educational measures that are there that empower people to actually work with these data sets and technologies because right now I think Peter Lee from Microsoft yesterday said talent is very specialized and artisanal and hard to be accessible to everyone.

My question for the panel, what do you see your role or the roles that your organisations can play in making these principles happen? If I've missed any principle, like please feel free to add to that.

>> MIGUEL LUENGO-OROZ: Thanks. We have another question? Just behind?

>> AUDIENCE: Hi, my name I'm from C step, based in Bangalore. My question is with reference to the use of AI for SDGs. And so if you look at Developing Countries context, where you are looking at AI implementation for sectors like agriculture, health, much of the data which is collected today is manually by the people or workers in the field. There is a level where they are collecting data, collecting manually. The distribution of tablets or smartphones has not percolated to that level.

In that sense the quality of data pretty much depends on the individual who is at that spot. And so the efficacy of the government information scheme depends upon quality of data which is flowing in through these channels.

My question is, have you guys encountered similar situations in collecting of data which potentially may not be that credible? What techniques could be used to address them?

>> MIGUEL LUENGO-OROZ: Thank you. Over there?

>> AUDIENCE: Hi. I'm from the university. I study neuroscience and I work in Ottawa. My question is to John Quinn regarding your work in Uganda. I wonder if Ugandan computer science students are trained in grad school or are they involved in the work that you do in Uganda?

>> MIGUEL LUENGO-OROZ: We are going to have the last one for this batch.

One of you.

>> AUDIENCE: My name is just continue from the internal displacement monitoring center here in Geneva, just across the street.

First, to thank all of the panelists for their candid and thoughtful remarks. I think this is my favorite panel I've attended so far. Congratulations on that.

I want to pick up on Moustapha's last point which I think is really the elephant in the room these last couple days and ties into something that Chris mentioned about the need for efficiency gains.

I think that the key question that keeps coming back to me is essentially, to return to Moustapha's question, what is the solution. I hear about all the potential use cases. How do we decide where the investments are made, what problems ought to be solved first and what order do resources need to be bear on those. One of the things I haven't yet heard and wonder what your reflexes would be on this, what about just using AI to develop our understanding of the problems in the first place rather than leaping into solutions. I feel like if we don't have a clear and granular understanding of the systemic and root causes of these problems we will be throwing solutions at symptoms and not really addressing the issues at the heart of the matter. And if you guys can speak to that, I would be grateful. Thanks.

>> MIGUEL LUENGO-OROZ: Thank you. We are going to try to answer these questions. And we might start with John, as you have a particular question for you.

>> JOHN QUINN: Thank you. Yes. I appreciate that question about the Ugandan grad students. Pulse Lab Kampala is comprised of former graduate students who are Ugandan Ph.D. and Master's holders themselves. One thing that is suggest full, Ugandans who had computer science training on sandwich university, with the main university in Uganda spending maybe 50 percent of their time in a European or U.S. university, for example, picking up the skills and coming back and applying it to their local context.

And I attribute most of the success of BLK such as we have had to the fact that the majority of it is people who are from that culture an also understand the technical tools that have those tools in their tool box and absolutely it all comes back to what Moustapha is saying. Unless there's a global community of practitioners, it's difficult to have global impact.

I think one of the reasons is that, as Moustapha was saying, the problems just don't occur to people who aren't from a particular place where the problems occur for someone who is embedded in a culture and also knows about technological possibilities. They can think up stuff that might not occur to

others. And there's a second effect where they may understand the problem much better and often much as I have been the technical lead in Kampala, the people in the team there often are able to pinpoint what the real problem is and the best way to go about it. And I'm often surprised by that.

When working on problems addressed in a culture which is not the one that we are originally from, it is easy to have these cognitive biases where our inexperience leads us to overestimate our own ability to solve the problem. And in the nine or ten years I spent in Uganda, the more I'm there, the more I realize how little I know and how important it is to have things like these Ugandan graduate students grappling with the problem. Thank you for the question.

>> MIGUEL LUENGO-OROZ: To complement John's response, John has been part of one interesting initiative called data science for Africa, which he is starting to gather AI experts and machine learning experts all around the continent. I encourage you to look for it in the Internet, digital science Africa.

Now, I lose my papers but the question around data access. Maybe Fei-Fei and Leila can speak about?

>> FEI-FEI LI: Yes, I cannot agree more that we need to democratize data more and I did didn't memorize the list of requirements, but access, standardisation, fair representation, all these are really important points. I think one of the most important roles that some of the current institutions can play, I think, is really the effort to open source data set. And in this particular case I see academia leading nonprofits as well as government and United Nations can play a huge role.

We can put a lot of pressure on private companies, but it is very difficult for private company to play the leading role in Democratization of data. They do have legal constraints and other issues, especially if they answer to their users. That's a fair thing. So I particularly want to encourage more efforts from the above institutions to come together and open source data. This would tie to the last question and Moustapha's point of are we only solving the small slice of the problem -- white male problems.

(Laughter.)

>> FEI-FEI LI: I think in a way, if we get more efforts in Democratizing data, that has a diverse origin and the diverse context from satellite imagery of Africa to deforestation in Amazon to human traffic data. Once these data are available we are solving the global problems beyond the white male problems. And we can actually call to action from everybody around the world, from data scientists. There is a lot of data scientists and machine learning developers who are eager to solve nonwhite

male problems, but they don't have access to data. That is a huge bottleneck.

I think that would be one stone, many birds. That is my comment.

>> MOUSTAPHA CISSE: I agree with most of what you said. I know for a fact that there are important challenges to be tackled for which the data is example. John, for example, has done amazing work on collecting various data and solving, for example, this white flies detection in Casava crops. And also he has done a lot of work on automatic diagnosis, collecting blood samples for automatic diagnosis of malaria. The data is there. Maybe not of a great quality. I totally agree because people have not invested as much time and effort to make data of the same quality as the one you did, for example. And this is very important to advance technology.

But the main problem in my opinion is that we are just not exposed to these problems. Somehow these problems are not part of what is generating the dollars. Let's face it. And I think that is the key to the problem.

It is very related to the last question that we had, which I would like to say a few words about, which is -- what are the solutions now? In my opinion no nonprofit organisation or the U.N. or other nonprofit organisation will solve this problem. I think the real solution will be in entrepreneurship. Because when there are entrepreneurs who tackle social -- social entrepreneurs mainly who are interested in having a positive impact in society but who also are interested in making profit, then they will tackle these problems and improve the availability of the data and the use of the technology to solve these problems.

So the main hurdle for entrepreneurship in these areas, most of these areas is the availability of funding. That's the main issue. So where institutions like the U.N. or any organisation that is part of the U.N. can help is definitely there. How to improve the availability of funding for entrepreneurs in this specific field, in these particular areas to increase the diversity of problems that are solved. I think that's where the main challenges are.

>> CHRIS EARNEY: I would like to build upon that and build on your question, Justin, especially from displaced people's perspectives. First of all when we talk about problems and challenges and things that need to be solved, there's a power dynamic there that I think we are talking around and about a little bit. Through organisations such as UNHCR, if we are going to be -- we need to be radically open about the problems that we are trying to solve. We need to be a little bit more agnostic, in my opinion, about the problems that we are raising

and bringing to the fore. Therefore, the solutions that we are trying to bring to displaced populations around the world. Part of that has to do with access to data. UNHCR like many other organisations gathers so much data on a daily basis. We actually don't know what to do with all of it. We gather financial data from our implementing partners. We gather population data from the tens of millions of displaced people. Biodata. We have financial data. We have archives. We are sitting on a data gold mine that we don't necessarily know what to do with.

The World Bank has open data protocols and they share data quite openly. And they release that data to the world. People can do very creative things with that data. Find out trends that the World Bank wouldn't otherwise be able to identify. With their own resources. From the UNHCR perspective we need to be a little bit more open with the data that we have. Not personal data of refugees and displaced people, but there are trends and there are data sets that we should be releasing on a routine basis to the public. And, thirdly, this brings me to the last point. If we are agnostic about the problems we are identifying and agnostic about the solutions being developed, we should have a role and a much stronger role in terms of how we interact with entrepreneurs, how we interact with the private sector. Because we have a responsibility once we identify those problems and the that affect 65 million people, we have the responsibility and accountability to those 65 million people to get the best solution to match their needs. We need to have some skin in the game and need to be interacting at the least with entrepreneurs in the private sector so we can identify the needs and we have the best minds that the world has to offer to try to meet them in addition to our own staff.

>> MIGUEL LUENGO-OROZ: One question for libraries and archives is, what do you think, what should be done in the satellite industry in the remote sensing industry with the data? So more people can have access to that data in the cases where the access should be given?

I guess it's a very particular space where many times it's sensitive data. How would you move that space forward?

>> LARS BROMLEY: There's a lot of interesting parallels between the history of remote sensing and satellite imagery and technology transfer and building capacity around the world for the use of these technologies. The exact same conversations were probably had on these technologies a long time ago as we are having about the AI issue and how to transfer them.

Honestly, I mean for satellite imagery there's different types. There's a lot of it which is freely available. In terms of capacity building and learning, don't think that access is an

issue really. There's that as a satellite -- NASA, European space agency imagery, which anyone can go out and get it. In terms of learning the basics of using the technologies, then that is really not an issue.

The issue comes in when we get into the commercial providers. So the high resolution satellites, the ones where I can actually see individual refugee shelters, individual vehicles, or something like that, these generally come from commercial sources. And that's where things really do get a lot more complicated because just like any company, the companies are not looking to give their products away for free. Now, each company does have a venue to get free imagery, but it is a donation vehicle that is designed to make you into a client eventually. There's no getting away from that from what I have seen.

There are big data buys by parts of the U.S. government where they will make imagery available for free and for everyone to use. That tends to be the exception rather than the rule.

Another issue that kind of der lies all this which again is a discussion that probably has been going on for decades is simply Internet speeds around the world. One of the saddest things I have to do every week or so is when someone from rural Uganda writes me and says I want to, can you give me a satellite image? I'm with an NGO and I'm doing this and I have to write back and say well, this is an 800 megabyte file. Can you download it? They write me and say, million computer says it will take 67 weeks to download. I can't imagine anything with AI is low bandwidth. We are going to keep on having these same sorts of discussions.

With all this stuff, the devil is really in the details. Is satellite imagery available? Yes. Anyone out there in the world in any university can start making use of it right now. Is all satellite imagery available? No, that's where you have to figure out the ins and outs. Figure out what you might need funding for, what there's free sources for. None of this work is ever easy.

It really is one consequence of being on the frontier of this stuff is the frontier is not always a happy, comfortable place. It can be a pretty difficult area to operate in. But that's up to the user, to the counterpart, to not just assume that it is going to be easy. And then it's up to organisations like ours to definitely make it as easy as possible.

>> MIGUEL LUENGO-OROZ: Thank you. So we are reaching the end of the session. I think that one of the guidelines or topics that have emerged from this panel and from the questions from the audience is the necessity to map the problems that

matter. Okay, the problems that matter globally, not just the white male problems.

Secondary guideline after that mapping could be creating the open data sets that could allow to develop the technologies to start resolving these problems.

We still have time for a couple more questions.

>> AUDIENCE: I have been trying to write some notes to make it as cogent as possible. I think it is another quite big question. In a way so that we don't lose the main messages from this panel I'm going to suggest that I think there's an important need to put it into an economic context.

Because an economic context I think relates to market shaping, not the fact that you have to pay for things, but actually markets are created and markets drive demand, and demand creates opportunities for producers and manufacturers, innovators and entrepreneurs. I think the essential question in the AI context has to be something along the lines what is the market shaping pathway to enable AI powered social good data. Somebody needs to grapple with that and try and unpack it.

From my perspective, in global health, we know that global health is inextricably linked to the health of the market. But we know that markets fall short and in poor countries particularly. You know, developers may not see enough demand to develop new products. Manufacturers may not know what is the best way to produce and distributors may not think there's enough profit to be made.

The question about the richness of all of this, and this significant power and scientific value and global impact this data can have in Developing Countries, in countries with fragile resources, is what is the way to actually shape a market approach that is relevant and fitted to those economic circumstances? I think it is a very specific call to say another layer of expertise needs to be brought in. It can't be marshaled from all of us who are very passionate in this context. It needs the critical lens of market shaping economics.

>> MIGUEL LUENGO-OROZ: Thanks. Another question?

>> AUDIENCE: Hello, testing, great. Yeah, I had a question that was really building upon a point Fei-Fei made. I'm Matt Lessig. I'm from Data Dot World, a collaboration platform. It is about the value that collaborations hold that would have a few benefits. We have seen, sharing agricultural efficiency data, sharing across countries, climates and crops that are put into agricultural yield. More food, that's good. There's UPS around transportation damage that package humanitarian food delivery and U.S., Uber sharing a lot of data that city planners are using for investment and planning.

Given the potential impact of this, how do we encourage and even incentivize corporations to open up data? A lot of data that they consider kind of data exhaust, to the broader community so you can tap into all of these data scientists and machine learning experts who have a desire to lean in and start solving a broader set of problems?

>> AUDIENCE: My name is Tanya I work in the office of the high commission for human rights. My question regards the role of AI in prevention efforts. When we talk about problems with refugees, with migrants, we are already into the damage control zone.

I'm interested in the project in Uganda and the radio. The monitoring of the radio station because if I remember correctly, the genocide in Rwanda started exactly through the radio station in 1994. Do you monitor incitement to hatred? What are the benchmarks if you collect this data? It's important.

The same applies to Facebook. If I send you an email and ask you for data on invitement to hatred, would you be able to provide this data? It's interesting, you say there's plenty of data. We struggle with the lack of data when it comes to racial discrimination according to the international standards.

>> MIGUEL LUENGO-OROZ: Thank you. We don't have time for more questions, unfortunately. So I want to propose to the panelists that they make their last statement, the sum of the intake, one of the questions that the audience has posed.

>> LARS BROMLEY: Starting with the most.

>> JOHN QUINN: Starting with the most recent one on radio, these are interesting opportunities and perhaps in the context of refugees, certainly those kind of issues come up.

With radio we have begun exploring more practical issues like exposure to natural disaster and that kind of thing. One thing we are very conscious of is radio is these issues to do with privacy and data protection and so on. And we should try to carry out what we call a privacy risk assessment project that we do. What we concluded was that we would start off this project with the most kind of practical and politically neutral topics, as it were, to get a sense of how this whole thing works. There are various kind of issues to do with data protection in that kind of analysis.

But interesting opportunities nonetheless.

There was a question about the kind of world data and encouraging these resources to become more abundant and encourage participation. One thing we see which is interesting is this concept of data philanthropy and businesses providing their data for social good. And there have been a few instances of that where that is really got some traction. A lot of the most significant data sets in the world, especially this kind of

global scale that we are considering in this meeting, are the property of industry and where, to the extent they are willing to make that available, that can be a quite powerful resource.

>> MIGUEL LUENGO-OROZ: Chris, your latest statement? Or any comments on the questions?

>> CHRIS EARNEY: Yes. It still comes, I think, the gentleman I think from WHO, am I right? Yeah.

>> AUDIENCE: (Speaker away from microphone.)

>> CHRIS EARNEY: Sure, health, great.

No, I think the point raised is something that is certainly resonating with us, how you influence the market moving forward. The secondly, we have been talking about access to data. Recent example that we are kind of grappling with, we are trying to put together that predictive analytics model, looking at movements within and outside of Somalia, as it happens. One of the pieces of data that we definitely would help us out an awful lot would be the remittances data from a company called Dahopcha, based on the remittances system that Somalis have been using for probably millennia. If we could add that to our meet lodge cam data sets that we enjoy, that WMO shared with us from various societies in the will region, if we could add that to our own registration data in hosting companies in places, if we can have access to that data where people are currently sending their money or people are anticipating sending money moving forward, that would help us out enormously. This is one example. We can use this to add to the algorithms, to feed into our algorithms, to learn from them, to start putting out predictions for movements moving forward. Then adapt.

If we don't have that with the private sector, if we don't have that dialogue and if the private sector don't feel like sharing this responsibility with organisations such as the UNHCR and our partners, then we are kind of dead in the water. We'll do what we can with the data that we have and we'll try to apply great minds such as these to the issue, but ultimately we have to talk about the shared responsibility agnostic of organisation. I that is the reality that we should be living in. Over to you, Moustapha.

>> MOUSTAPHA CISSE: Yes, just to answer quickly to the question. I think forcing companies to private sector, as Fei-Fei said, to share data is very problematic because most of the time the data that you think belongs to the company does not belong to the company. One example is Facebook. The data that you share on Facebook does not belong to Facebook. It belongs to yourself. Has Facebook got the right to share it? I don't think so. There are serious privacy issues here. Even there are motivations arguments. This could be for good purposes. There are a lot of potential benefit on having all this data

shared publicly in the world but there are serious questions regarding privacy and fairness for sharing all these data. And I think those who want this data to be shared are all at the same time those who care a lot about the privacy. So can we reconcile this? This is a question I don't have any answer to.

>> FEI-FEI LI: First of all, I'm aware of the work from Data Dot World, good job. And just to answer your specific question about how do we incentivize private sector and also in my opinion governments, these are the places to share more data as Moustapha was saying. First of all, I think it is important for governments and private companies to be highly protective of user data. It is not fair for them to just want to share that.

But having said that, I think the world of data, especially in the age of AI is getting very interesting because there is an ecosystem that would naturally incentivize certain type of sharing data. Companies, some companies are realizing data is not only just data loan. Data is embedded in the ecosystem of machine learning developers and technical solutions. If you look at the different sectors of the world, some of the companies do not have enough machine learning talents yet. Yet they are in the world of intense competition of machine learning solutions and data science solutions. The way to jump start themselves is to actually under proper security and privacy protection, share data so they can engage in the greater developer community to solve the problem for then. And in the meantime, they obtain solutions that actually matter for their business. Cackle.gov started hosting competitions and they expanded their data sets and invite companies and academia to share data. And then in the meantime they have developed this tremendous machine learning science developer community that constantly put forward to these data sets. I think that's extremely value for companies. If they can make their data clean and free of privacy issues, I think there is actually incentive for companies, governments to recognize the benefit of sharing data.

>> LEILA ZIG: I'll say one thing about Facebook, as far as I understood that privacy statements and terms of use by signing up to Facebook, you give away a lot of rights about your data to Facebook. I'm surprised if Facebook cannot share much data. I understand if you don't want to share user level data, but I think there is a lot that can be done, that can be shared.

To the question about companies, I think this is an important question which Wikimedia deals with because we are looking for open data. Generally I have seen two approaches working. One, if the company is small there are incentives for branding basically by opening up your data. You know, if the next series of academic papers are going to be based on your

data, you will get visibility to the work that you do in the company. It can be one incentive.

The other thing I've seen, at least being discussed and I have never seen this being implemented yet is that if you are a very large company, let's say top five to ten companies that have access to large amounts of data, one incentive can be to basically make the market more level in the sense of competition. So if you are really good, if you have a data set that is very valuable, you know that few of your competitors also have that. You know that you also have a lot of other services that you can offer. So you are not really, your whole business model does not rely on one data set you have, but releasing and opening that publicly you kill competition. You open it to the world and say everybody can use it. This is a topic I have seen discussed when companies consider big data donations. That is something that is reasonable and can be explored.

>> LARS BROMLEY: I'll just end up with some broad thoughts about what it means to try and do this stuff.

My office is an operational office. We pride ourselves on basically being available all day, every day, to answer questions using satellite imagery. That's basically what we are funded to do. However, at the same time we have this whole research an development component that we do also have to get into because we have to get better at what we are doing. We have to use the new tools. That is a really, really, really difficult thing to achieve in an office like ours, to have both that operational aspect and that R&D component where you are trying to learn about and do new stuff. That's definitely a big challenge for any U.N. organisation or NGO that is doing something similar.

Some of the ways that we try to deal with it, we A, try to get more efficient at our day-to-day operations so that I can take 10 percent of someone's time, 15 percent of someone's time and throw them into a new topic to learn about something new.

Another way it commonly happens is basically with partnerships. Global pulse and UNHCR here on the stage, we worked with both of them. And it has been a really interesting synergy about how what we do every day is extremely helpful to them. Then what they are exploring for the future is helpful to us.

However, those sorts of relationships are very, very difficult to build. You first have toe find kind of cough conspirators to engage with you on this stuff. Don't tell the boss, all that sort of stuff. Yeah, it is not easy. It doesn't just happen. When it does happen, it can work out very well.

>> MIGUEL LUENGO-OROZ: Thank you. So we have seen many real world problems. I think we, now we have some ideas of guidelines what we need to do. I think the summary of the session, we need to map the real problems. Everybody needs to be aware of which are the problems that we have more impact at the global level. Once we have mapped, we need to mobilize the pertinent research communities, open the pertinent data sets. Really new technologies can be developed for something more than the typical massive white male problems.

From there we need to shape the markets. That could be the next step. So I would like to take this sentence from one of the first Facebook employees that says it cannot happen that the best minds of my generations working on making people click on ADD.s. The challenges are clear and we just need to put things to work. So thank you to the panel and thank you to the audience. We are done.

(Applause.)

***



***