RAW COPY

AI FOR GOOD GLOBAL SUMMIT
PROJECTS IN ACTIONS: TOWARDS AI AND DATA COMMONS
PART ONE

MAY 17, 2018

Services Provided By:
Caption First, Inc.
P.O. Box 3066
Monument, CO 80132
1-877-825-5234
+001-719-482-9835
www.captionfirst.com

***

***

thank you for the previous session, it was amazing, the projects we heard. In those tracks yesterday, there were two Rapporteurs that listened, participated, they captured the opportunities for data comments and today we'll hear from them about what they captured and how this was summarized. Before I would like to introduce as a first speaker and a welcome to the session, the Minister of state for artificial of intelligence of the United Arab Emirates. Please welcome.

>> Thank you for having this wonderful gathering. First and foremost, it is extremely imperative that we have such gatherings, they bring together two words that we're aware of today. Most technologies, they're being developed to create positive impact on the lives of humans everywhere. I think based on what we see today Artificial Intelligence, some people may show negative impact which we hope is never reached. I would like to start by first saying that humanity has gone through many faces of development, one was a race of Developing Countries, economies after the first, second industrial revolution. The race brought together the biggest talents, most resources and countries use energy to obtain economic and social prosperity. Many years down the line we discovered that this

was not a race between countries, but it was a race against time. The Climate Change came in full force and we realized that this development came at a cost.

Today we see that certain countries try to take steps on their own single handily but they were not successful. The success came when every single country came together to try to create a unified platform to combat the negative impacts of Climate Change and to make use of sustainable moves moving forward.

I think this is exactly about artificial intelligence as well. It is only when we come together to identify our moan men Tim to unify our knowledge and resources in the field that you would obtain the benefits, I truly believe and we truly believe that Artificial Intelligence has that lot of potential to do so much good and it is our vehicle, our tool to combat challenges like Climate Change, combat challenges like global inequality, the efficiency of aid across the world and we'll be able to do so. You have asked what has AI done so far, what have we as a country done to ensure that we're moving in that direction. I think I speak on behalf of the whole government saying that we can do a lot more and we will do a lot more. The first step that we did, we created a platform called the global Artificial Intelligence governance forum. Some of you have attended in February, Dubai, that platform brought together the greatest minds and discussed what can we do collectively to govern the technology positively moving forward, not looking at the long-term impacts, not talking about the issues that governments can't control, talking about governments can do today to ensure that technology progresses, but without effecting humans and citizens in a negative way. The outcomes of the platform, they're published and you have taken them apart, you left the them in the room yesterday and we're actively looking at the leveraging the technologies to see how to contribute to the Sustainable Development Goals to contribute towards combating Climate Change using Artificial Intelligence. We ask you, if you have an idea, there is some way to collaborate, don't hesitate to reach out to me or a member of my team. Thank you very much.

>> Thank you very much.

I would like to ask Urs to lead the reporting from the day 2 tracks from the -- he'll basely go through all of the tracks and come back for a panel session.

>> Excellent. Thank you so much. I hope you don't mind if I stand up, it is a bit more energizing as noted, some people may be tired this morning.

Here is a bit of an experiment. There are four Rapporteurs and there are several others in the audience to actually list

over the past three days and to take note of what's the role of data when it comes to AI for good. I think it is fair to say as we all know AI for good, it is only possible if we have data for good. What we have been doing, it is really to track the conversations, both in the plenary sessions, and also yesterday in the vertical tracks and cross projects to have the common themes and work towards maybe -- maybe it would be a roadmap 0 as we build towards what you may think of data comments for AI for good. What we're doing over it's next couple of minutes is to provide a snapshot and a bottom-up version 1 #. 0 of the roadmap as it is emerging. Most of the heavy lifting, it was done by the Rapporteurs. Thank you for that. My role, its to simply introduce the framework and then we'll hear from the different tracks some examples and how they map on to this framework emphasizing again this role of data. We also thought after hearing from the vertical tracks on the many projects that it may be helpful to add the horizontal view across the tracks. I'm sure many of you had the same experience and all four tracks are so interesting. You could only attend one, this is an opportunity to bring things a little bit back together. If I may have our slide deck, that would be great.

Excellent.

The conversations about data, it is far ranging, and in the track presentations, the various project, data has come up several times in different context. I think it is fair to say that there is a notion, almost a kind of vision that you need data comments as we think about AI for the social good. As we went through our notes and had conversations yesterday with the various Rapporteurs we'll crystallize the layer to model with different functionalities as we have -- we're wondering, okay, the data common roadmap, what's it look like. At the core, we may call it a narrow version of account comments, the three layers you see here. At the very bottom, technical infrastructure layers, essentially where does the data live, across the tracks, of course, this is an important question, are we talking about data in the Cloud, talking about data repositories on government servers for instance or is it decentralized lecture blockchain table of infrastructure where the data lives that we may be using and deploying later on as we use data for social good.

Of course, data, in each track, there have been various types of data and we have been working a little bit on a taxonomy, there is talks about quantitative, qualitative data that can be used and there is a lot of conversation and different projects about structured and unstructured data and so forth. There are different sources of data. You think about the health track, the health information that can be gathered or

information that's relevant to health, it can be gathered from social media and the like.

The third layer then, it is really more about the metadata, the question of labeling, the label sets, again, recurring theme. For example, you may remember the slide from this morning where you did see red dots on the satellite image. How can you make sure that the red dots are indeed the water or whatever the snakebite, whatever the equivalent is. How do you ensure that the labels are correct and that you have also an interoperable tax onmy across projects that leverage some of these labels and formats of data.

Again, this is somehow the core functionality of what may be a concept of data comments as it emerges. These layers, they're, of course, interconnected. You will see that just in a second when you get to the examples, how they played together. In the conversations, it seems it is also clear for this narrow version of data comments the role of standards and standardization, it is really important. As we move up the layer stack as you will see, we'll add three layers in a few minutes, we may move away from standards towards other means, how to achieve the interoperability and define common languages and I wanted to flag that for this stack in particular, the notion of standards, it has played a role across stacks.

I wonder if Sean could give us an example from the satellite track that I will straight a little bit how the different layers play together and then we take it from there.

>> Thank you.

I think every U.N. agency present can appreciate the importance of on the ground data to being able to effectively run their programs. The same is actually true for satellites that aren't near any ground and when you want to be able to, for instance, label what crops are grown and what location, you need to have on the ground data that's labeled saying what crops are being intercropped in a particular location and on the basis of the labels, the satellite system can then produce an entire report for the country's regions, the globe, and what we need a lot more of in the course of building out satellites as a resource for achieving the Sustainable Development Goals it is on the ground data and one of the ways we can achieve that with infrastructures to have additional tools that allow for the crowdsourcing of the standardized format data that's Geo referenced and can be paired with the satellite data towards generating these global report cards. H.

One thing that Kim up for this, it the ability to use cellphones, that it is a very rich dataset that is derived directly on the ground and we have the ability to use the cellphones as a crowdsourcing platform that could be used not

only for AI approaches to problems but could be used as an analytic tool in general, there is no standardized way to field these surveys at this point that is a piece of infrastructure that could be built out for AI solutions and just general management of developmental programs in general.

>> Excellent. Thank you so much.

We were wondering if you could talk a little bit about the upper layer there when we were talking about the metadata and also labeling of datasets more generally. The trust in datasets which was mentioned also by the track leader earlier today, it was kind of a good illustration of what this layer could be about.

>> Exactly. Thank you. You know, as you all just heard there is many great ideas and projects that came up in the trust in AI conversation and one of the common threads that went through many of them related to how we label data in a particular operating, not as much as Sean was describing labeling specific data but labeling the datasets themselves, how do you bring greater transparency to the datasets so that when someone goes to use a dataset even if they weren't the ones who collected the data is that they have the relevant information that they need almost in the way that one looks at food nutrition label, the idea that one could create some kind of label or signal that could either be human readable, machine readable and then it would help people who were trying to apply the dataset better understand what some of the limitations are, how it was collected, when it was collected, the Providence of the data as a whole, and then that can act as a guardrail of sorts to help prevent using the data in ways that may not be appropriate, that may introduce unintentionally bias into the outcomes.

>> Great. Thank you so much. As we move forward, just again, the samples, some of these stories, there will be more detailed write-up of other examples and maybe they map on such a taxotomy, let me introduce the next three layers and they'll form the broader forms compared to the first layers were that were more narrow.

You have the conversations across the tracks, looking at the data, data is far beyond the technical, it goes beyond data formats, the role of data for good actually includes things such as organizational practices emphasizing the need for collaboration, touches on the question, how do you incentivize different players, be it the government, be it private companies actually to share data, to create data in a meaningful way, to make it interoperable and the like. There is definitely an emphasis and I will hear an example in a second from a track on practices and communities of practices and even best practices

as we move forward.

A layer up, finally, towards the upper part of the stack, this notion of institutional arrangements and laws and policies are a key enabler or can be also a barrier to use data for the good. Think about Intellectual Property, the barrier to data sharing, sometimes, of course, it can be an enabler, speaking of comments, you think of the creative comments as kind of a way that enables the sharing and flow of information and data, it is an issue that came up in various tracks, but IP regime and how it interacts with the idea of data comments. Privacy, it has been mentioned several times in various conversations, data protection issues, data governance issues more broadly, these are some of the relevant forces of play that vary by context and needs to be taken into account if you want to unleash the full potential of data in the AI for good context. At the very top, somehow bringing back the -- as someone that works in the University, kind of excited, ultimately the role of education, it became another thread that resonated with all of the Rapporteurs and was brought up also this morning. The role of education, we have to educate the public, we have to educate the awareness, I need to work at times with local communities to concern around inclusion, diversity, it is very important values I think that empathize again the main points here as we think about the data comments conceptually, it is not only about the technical stuff, it is also really about the organizational and more so even the human issues. We thought we would feature two other examples maybe in the same spirit just to refer back or have link backs to the earlier examples. If you would be so kind as to maybe comment on the smart city conversation, the roam of accessibility, citizen empowerment, a great illustration of the organizational practices, but also on some of the the rights and Human Rights aspects.

>> Please.

>> Thank you. I'll try to illustrate the organizational level of the data comments thanks to the CTs. We have seen that there are a lot of problems here in the CT concerning the rights of the citizens and it was well explained for example by language, in cities -- you have different histories on language and sometimes old cities, they have always been multicultural, you have a lot of different languages, and sometimes it is something recent that happens with the globalization and you need to include everyone. That's why you have examples like the processer, they have to the projects in Latvia that helps people from Russian or English, other languages to public services, this is an example of how you can empower the citizens and then you have an overview of older problems, the citizens, right, you need somehow to have a place together, the experience of the

cities trying and experimenting and sharing the experiences, for example, when they fail, where they put the experience of failure, how they share with other cities. What they try to do, we trade what we said was a graveyard for experiments and we need an infrastructure for all of the cities to share the experience, if you expand this a little, it's the idea, how do you make it more easy for everybody to take part in this ethical design of problems, problem design, it is very important at this level and how can you do this so that everyone can take a part, even with Artificial Intelligence or something that's not really accessible for everyone. When you're at this everybody will, you create a safe space for everyone to part, that's when the idea of the sandbox came in. The sandbox is a place where you simulate with real cities and this could be a place where everyone can throw a tornado in a city, you see how it has happened. Thanks to all of the previous smart city initiatives, we have a lot of data already and if we pileup all the layers of data. Then we have an overview and we can create a place where people can try, people can fail, people can have success. If we create this infrastructure, what we call before the Internet of the cities now we can share all of this on the same platform and even create interdisciplinary teams so that even the students, researchers, they can it be looking at the practices, it is ethical even from design and then we can try more and more and allow ourselves to fail.

Thank you.

>> Thank you so much for this great illustration of some of these elements that we summarized in the Graph taxonomy and moving up, ultimately really to the had human layer, do you want to share -- we share one example of the rich discussion you had in health track that somehow again is symbolic for some of the issues of human knowledge and educational layer.

>> Be happy to do so. I was fortunate to be a part of the health conversation which was very much at the intersection of institutions, policy law and then what we call the human layer. Perhaps we have some context, some of that has to do with the fact and we're talking about the health space, we're talking about a rather complex ecosystem of stakeholders that not only encompasses government agencies, philanthropists, doctors, practitioners, patients, users, depending on how you look at that distinction, all of which have their own distinct models and that makes it somewhat complex as we look to the human layer. Especially because the data layer itself, it is also quite complex in a health context incorporating sources as varied as medical imaging to nurses' records, admission letters, a variety of other sources and social media was alluded to in this context which is helpful as well. As we look at pursuing

the data comments in the data context, that makes it more complicated and also means that especially looking to integrate in real languages in space, some of which are linguistic and others of which are perhaps disciplinary and so I think it was really encapsulated in the transparency of the concept of diagnostic tools which, you know, maybe used by clinicians in the field -- (no audio). (Audio has been muted).

>> Approaches to what policy works best in what context as we try to use the data for the good.

In that sense we hope it is kind of a helpful contribution to structure a really rich, wonderfully diverse conversation with special thanks to all the Rapporteurs.

Thank you.

>> Thank you very much. A wonderful job. They framed it in an interesting way to understand the stacks. A question I think that comes to mind, it is a notion of public, private data, what part of it should be available, Open Sourced for everyone to use and what should be controlled. I would like to turn to His Excellency on public data opinion and what public data should be available and what is your reading and thoughts on private data.

>> Thank you for the question.

I would like to first thank my fellow panelists for the wonderful illustration I think that we'll tap into that in and try to build programs based on that. Thank you for sharing that. With regard to public, private data, I think it goes back to the citizen, when it comes to what public data could be shared or could be used, it is extremely crucial for us to engage what they consider private and public. The issue today we're seeing with standards that are happening, it is because citizens are not engaged and they feel cheated with the data that they could create and with what they consider their privacy. What we're trying to do in the U.A.E., he would like to have a massive survey that will take more than 4 to 6 months to get from the citizens what they consider as extremely private data compared to what they consider is private data but they're willing to share. With regards to the private sector, I really think there should be more transparency based on everything that we have seen on the different platforms. It is imperative for the private sector to at least inform the citizen what type of data they're collecting, what they have, we're seeing more and more of that. It is about doing what's morally, ethically right, not doing things that create the right amount of data, not engaging the assistance.

>> You mentioned moral and ethical aspects of it, are you doing research in U.A.E. on ethical guidelines or governance of data or have you any plans for that?

>> We are actually planning on doing something. We actually came to this conference to explore ways to do so. What we want to do, we want to see Artificial Intelligence from three separate buckets, the first, it is the optimal way to collect, store, share data. The second, it is optimal way of using Artificial Intelligence in a had way that does not compete with moral, ethical guidelines and the third is using Artificial Intelligence to combat the biggest challenges as I said in my speech.

Based on that, we want to try to bring together the talent, we want to bring together the resources that people are creating to use all of the collective intelligence to achieve what you want to achieve as humans. We're we're looking to create things for SDGs where all the talent that you have, all the knowledge of the data, it can be can be put together to find a solution that does not benefit the U.A.E. on its own but benefits every country, every citizen that wants to be a part of this.

>> I would like to ask you a last question, based on the framework presented here, which is a nice stack of the data comments and I think you want to have more data comments, trust, repositories, would you see an opportunity to use this from a stack to enable more data comments in different sectors such as health, education, transportation, finance.

>> I do completely believe that this framework works. I think that the U.A.E. are looking at us from top to bottom. The first thing we're trying to do, creating human talent. As you know, some countries have a lot of experts and a lot more AI talent than others and the U.A.E., we're creating talent there. With regards to the comments, yes, we're looking at creating datasets that are available in our healthcare and education industry and other industries as well. Our energy industry, it generates a lot of data that could be leveraged to create solutions for the U.A.E. and globally. We're definitely looking into that and you'll be hearing a lot more of that coming out soon.

>> Wonderful.

As we talk about data repositories and comments, we cannot think about capacity building. ITU is a strong advocate of capacity building. I would like to ask the director of ITU about ways to standardize either ways of accessing data in standardized format or trusted data in terms of governance, and in your experience, what is the roadmap to create more standardized ways to access data.

>> Thank you for the question. It is a difficult question. We have the essential part of the AI based environment, but looking at the association of data space, it is huge and I don't know how many data types we have now, one sector of this has

environment, sectors, we recognize more than 600 different data types now. Taking into account this data, the complexity, the difficulties, so we are in ITU-T, we organized a focus group on data processing and management. They tried to collect all of this data and they identify all of the frameworks, what is the data-related frameworks and I'm happy to see our discussion, it is quite difficult with the study results and we tried to find out how we can facilitate the interoperability of the the data performance. We find solutions to exchange me Da Silva data levels and if this is good, we can have a more comfortable situation to exchange data, which means the exchange of data. Also it appears, one of the concerns, it is -- it is the exchange of the data, it isn't important, data should be shared, but there is still -- we don't have any classification of data concerning that because I don't know how much data I shared with others, is there anyone knowing about how much data you share with others, whenever you use the mobile device, how many data from your mobile device would go to where, maybe go to mobile device operator or a service operator, providers, how many datas, which data, it is coming from a mobile device, even the laptops, the moment gateway, do we have any idea. No idea for me. Even when I use on my device, my home gateways, I don't know any -- about any updates of the software if I can climb over parts, it is not allowed to update. I have to agree to everything, and then it agrees to update, which means as a user we don't have any way to manage my data. Is it right. Give me a challenge here. This kind of concern, it is raised, it is a good subject for the future if you really want data as common, data, you know the data, it is essential to operate the AI algorithms, we need all of that. We need this ITU as a standard organization, we do all of our efforts to address this data and to be more comfortably way to trust the data.

>> Wonderful. Thank you very much.

>> A question for Urs from David, do you mind asking the question? I think it was covered. It was a question of ownership of data. It was in the stack and it was more question of the balance between commercial ownership of data or public ownership of data and how do you incentivize companies to put as much commercial data in the public domain as possible and I think it was covered in the discussions.

>> Wonderful.

>> A question about data coming from mark jeffrey, do you want to ask the question directly?

>> Good morning, everyone. We talked about the quality of data and issues of bias and things like that. My question is that data is not free. It costs money to collect, colate, sort, clean, debias, protect, validate, store, make available so how

can it be made available for free. Who will pay for all of that?

>> Any suggestions? Anyone who want to --

>> I think the data is increasing as we look at it for more and more platforms. The same thing with creating APs, there will be be incentives to create datasets, you as an individual may be able to rent out or seabed your data to platforms that will pay you to use your data. I do believe certain platforms need to be created by government to ensure that the collection and consultation of data is protected and done in a way that the humans -- also the citizens are engaged in, with regard to measuring the private and public sector, the public sector, they have some sort of impact where you can change the views of the government and we can have an impact on whether the government is doing, when it comes to the private sector, it is very difficult for us to do so. There has to be a platform where all data can be consolidated and you as a citizen get revenue for creating more data and making more data. The private sector, it is the one that creates the economic return when this comes to the platform and with regard to the government, we use that to regulate the ownership, the using, the impact of the data on both citizens and the country and maybe all combined.

>> My view, it is slightly different. Even we allow to share the data, the use of it, that doesn't mean I gave all of this data free of charge. Without sharing of data, it is smart operation, you have to forget it. Data should be shared, data should be protected, data should be shared, without sharing data, make the smart operation as possible. It means in the other directions, after we have shared this data, you have smarter services. The issue is how we can make a balance, this is a non-issue, it is a really good balance, we wondered if someone used my data, they make a big business so it looks like they're stealing the assets. This is not true. Something we have to benefit from those sharing of data so how do we make the balance of sharing value, it is now an issue, I don't believe that this is really effective and free of charge.

>> Any other comment on that?

>> Briefly, I would just acknowledge how hard of a problem this is, I think the question, the notion of data quality, I think it is an ecosystem level challenge facing, and one way to look at it, it is really -- you already started to map out some of the tools available from other areas so one thing that comes to mind, it is what can we learn for instance from environmental movements where we have concerns of comments and the qualities of commons and what could be transferred in terms of policies or approaches to incentivize some sort of an ecology of in this AI context.It think that requires a lot of creative thinking and

also the siloing of previous experiences whether it is Open Source software or environmental movements or whatever is in the toolbox. I think we'll not get around also a fair amount of experimentation and see what works and where and in what context.

>> Wonderful.

>> Maybe we can have two more questions. One first question and a closing question. Please.

>> Thank you very much for the great panel.

This is not a question, but a proposal to the floor, you know, I'm from WHO, here as as senior adviser. I think we're faced capacity building and health sector quite well and the innovative solutions. I think similar approaches can be adopted here. I'll tell you a case in point. I think if you think of it.

El ehealth and medicine, where health, you know, specialists, they were not in the country, accessed across the country through various mechanisms, I think for AI, the similar approach can be taken, so if you think about tele medicine for -- if you want to call it teli, with the ai, the experts, they're spread all over the world to be facilitated by processes for countries that are implementing solutions. This is a viable option, how do we broker that process, and I think not every country can produce a better highly talented AI engineer to set in the country waiting for a job to come, that's probably not going to happen. I think we have to really think of a way to advance this field, otherwise it is a distribution of countries benefiting more in the richer part of the world and countries that don't have resources, they won't be able to advance in the area. The teli approach to accessing resources, human resources and consultations that could be facilitated for countries that do not have -- that could perhaps not be an option for jump starting this supply of engineers.

Thank you.

>> A great suggestion.

We'll report it back.

In closing, I would like to ask each of you if you wanted to disseminate and create more security and privacy and governance, what's in one sentence an incentive that you can think about to promote more existence of this.

>> Thank you for the question. I think it is a very wise approach if it can be done, it should be done. With regards to global data commons, I think what we should do, we should share the steps taken to create the commons first, independent with all countries and we are looking at launching platforms that any data is built on top of and any AI could block in that. I think globally blockchain platforms that talk to each other when

everyone is integrated where the platforms it tap into the data and determine what datasets they tapped into, what decisions were made and give access to certain platforms and AI tools. In the regards to incentivization, it is empirical -- imperative that we use platforms like the U.N. to push countries to more sharing of data. I don't think one country can do this on its own, I don't think one countriy show on the importance, but if a non-government, multilateral organization, it shows the importance for collective humanity, all of us, what benefits can we generate, what solutions can we create and what opportunities can we leverage and what challenges will be overcome. I think that's the big question.

Thank you very much. A good answer.

>> To make it short, what I want to say, the unrerefrigerated agencies, this definitely should be -- this is our roles. We have to do, how we can make sure that the use of data is for good.

>> Briefly, I think a lot has to do with powerful narratives that demonstrate why it is so important to unlock data silos and really use it for the social good. I think this event was an amazing source of such narrative. Thank you to the organizers for having us.

Thank you.

>> I would like to say one sentence as well if you have an additional comment.

Incentives that could make it possible.

>> You took mine, there were great stories we heard, what struck me, part of the incentive for creating a data commons, it is not just focusing on the data in the quantitative sentence and all of the stories and narratives we heard that are about the value of qualitative data and the role that plays in advancing and helping AI for good.

>> Thank you.

>> In the health context, coming out of the conversations I was a part of, I think it is a fantastic opportunity to be able to connect an ecosystem by way of the infrastructure that the data commons would provide, not just with regard to connecting different types of data, some visual, some more language based and different tools that could be used for that, but also to connect with the community that comprises people who were working on health issues from government agencies to entrepreneurs that want to get more involved in fixing the problems and obviously patients themselves looking for answers. Community focused.

>> Thank you.

>> My point is close to what Urs said and based on the narrative side and point of view. I would think that maybe

people should have a chance to participate through the solving of global issues through the data comments because it would be a good place to centralize all of the forces and I think that maybe the first step is to demystify the stories around AI. That would be a good way to help people take part.

Thank you.

>> As AI researcher, I would say that the greatest incentive or ad for data commons is the ability to work on applied problems rather than abstract ones and to take the complexity out of modeling the data and applied problems is a great asset for moving into applied space and actually producing good in the world. Thank you. Thank you. Wonderful. Thank you very much. A round of applause for the panel. Thank you for your hard work.

>> Before you leave, the next Agenda item, 11:30, exactly at 11:30, it is that we're going to celebrate the the ITU world telecommunication and society day. We do this every year. This is on the 17th of May and the anniversary of ITU which was founded and it is pure coincidence this coincides with our Summit. You're all invited to an exciting programme with women that have been to space, it should be really cool. We're expecting additional guests. I'm trying to limit the amount of movement that's necessary so I have a favor to ask you. If colleagues in the first three rows -- well, first I would like to ask my colleagues from ITU, that they go into the overflow room, if you take your belongings now with you and you move to the overflow room, and if I could ask the colleagues in the first three rows that they take their belongings and we see where we place them best. There may still be seats available in the first three rows afterwards. I think that may limit the amount of action that's necessary. My colleagues, ITU, all move to the overflow rooms and for the ones in the first three rows, we see where we receive them. You can sake the seats that's going to be made available.

Thank you.