# Transcript Day 1 video 2

# Session 2: Foundations/Motivations

### Discuss Reports

Table 1 Adam Martin

all right okay we were free associating on all the various problems that we were hearing both based on Yolanda's presentation and then just our personal experiences here so our list of kind of problems that need to be focused on or thought about include communicating to and educating users on the **knowns and unknowns** within our the models that we're using and kind of spoke to **Yolanda's ontology of models** and how communicating knowns and unknowns should be somehow part of that communicating **uncertainty and margins of error** that can be easily propagated throughout workflows you know is a key issue it comes up in the work of folks here help you know as these models and and **methods become more democratized how do we ensure a suitability for particular datasets or situations** and does that require domain expertise or methodology methodological expertise and so at least from a methods perspective maybe you can help identify suitability of models based on the data that you're bringing to the table and just avoiding common problems like overfitting some of the another challenge around that came up when we're thinking about autonomous vehicles is you know knowing when you're working with interpolated data versus extrapolated data from models I won't expand on that because I don't understand as much and then just you know concerns about knowing when your samples are representative and making sure that there are ways to check that you know we when we you know just I think validating that concern that was voiced in the panel

Table 2 Andre 3:10

so I've made the argument that **AI is a social construct** not because I think so but because our analysis showed so we one of our projects recently we annulled we did a domain language analysis of more than a hundred thousand documents in data science and analytics and it clearly showed that: Machine learning is real - it is a set of technologies; Deep Learning is real CN NS are rea; AI is a social construct is this societal construct it's not a set of technologies. Now there are other indications of that like one recent study of European companies showed that at least forty percent of those that say they used a I don't and even that is probably optimistic then I think the observation was also that it was a very nice setting here for communication about like what is machine learning what is deep learning and I think there's maybe a **lack of venues** to have exactly the conversations that we had to build that up from how does it work to what are the use cases and the third theme was generally **knowledge representation** I think where we had a lot of commonalities the the need to basically analyze large so and there was one question I had about knowledge-based whatever the end of that title was I would have called it model-based for Yolanda gills presentation that's right because knowledge comes in many forms and was a very limited set of knowledge types that were being used to drive it interesting physics based model as opposed to the many other internal and externalized artifacts of human knowledge that should be used

as we're moving up to the next table let me do a shout out for Andre though he's doing some work for OGC that will be revealed real quickly a **base map for geospatial technology** and so we're looking forward to that as a will put it on the public website and it'll be really quite a resource for the community to understand these concepts so I look forward to that

Table 3 Ajay Gupte 6:00

we had a good conversation on the table and had a number of things that we thought were interesting in the in session too one was **the quality and source of training data** really is a key issue also identifying what the correct the **best or the good sources of data both data sets** and data sources really are and we thought it was interesting that the level of confidence in the data and the outcome is related to the application **some applications need more and some less level of confidenc**e. **Model integration and scalability** was also of interest we had a long conversation about that about how some models may simply not be scalable and that communication ultimately was the key broadly communicating the data to others to decision makers or end-users that confidence in the data of the models and the ultimate impact in in our analysis thank you

Table 4 Lauren – 7:20

we talked a lot about hype. We at the top of this hype cycle and will we find our way to a little bit more reality and we talked a lot about that and mixed kind of ideas but that there's a lot of value I mean and part of the part of that led us to discussion about the fact that we haven't talked much about theory and that in that theory plays a really important role in analysis and perhaps less of a role in data science and not necessarily making a value judgment about that but that **perhaps there's should be more theory at play** particularly in the problems that we're trying to solve in the geospatial context and we also talked kind of in a related vein about **education and what does it mean to build up the next generation of data scientists** and what does the data science program look like and kind of as an interdisciplinary approach and what what are the foundations of that what should a someone coming out of school the data science to degree know how do you how do you **fill in all those circles in the Venn diagram** so that someone gets out with enough knowledge to be a productive data scientist at an organization and and the importance of subject matter expertise and data scientists working with **subject matter experts** in whatever the field insurance healthcare particularly we were talking at our table but that the role that that plays in validating what what we're learning because it's really easy for a data scientist without that subject matter expertise to sit here but but but you know **the r-squared is really high but garbage**

Table 5 9:32

we had a couple different topics of discussion I think one of the big ones because we had some folks representing and space agencies in Europe talking about through **international standards and data sharing across borders** and just you know what what it takes to **create the community will to set standards and make data available** and what some of the political pitfalls might be in making that happen. Then we got into a discussion about you know **graph query language** and kind of probably deep into the weeds of the metadata that are required to make that happen

Individual 10:19

I work for the **California Public Utilities Commission** it's a state agency in California for people who aren't local. We regulate electric utilities, gas, all of rail, telecommunications and transportation, including uber lyft and limousine services. I challenge everybody here in this room of masterminds to do what they can and to get out with what this the PhD from ESRI said at the **local level is where this data is most important** and most useful and it's unfortunate that we don't have the expertise but that's in this room and a lot of the sounds like the private companies are taking advantage of all of this technology but there really **needs to be a connection between private and government** and there are some government agencies that will have for example you know one every user one geospatial expert I would say for that agency but still may not even come from a background of any specific science or engineering may have just been an analyst who learned how to use the tools. I challenge you all to think about how you can participate in your local governments and whether it's your company giving back some way to you're gonna see a we donated so much money to the food bank will **donate geospatial expertise** because it's really important. I don't want to alarm anybody but you would be surprised that even at the catastrophic for catastrophic events in the state of California that California callaloo yes they are overwhelmed with the job that they have and they have they're trying to get data out for their for the governor for the politicians and for their leaders when there is a catastrophic event and they don't have time to focus on the actual data or the science they're **just trying to get pictures out** as much as best they can and now they're working thankfully with the National Guard and Bill Air Force Base who are who are using some of their intelligence officers and personnel for situational awareness and they have some of the expertise but even they're still learning. I was just talking to this gentleman here who spoke earlier about working with Fish and Wildlife and tracking the elk that was a research project you weren't actually being paid for that although he has he does it did say he has worked for the Air Force in the Navy but that information or that work isn't transmitting down to the local States for example emergency management or it could be very useful. so anyways I just want to put it out there to you all to challenge you to think about the local governments where you may or may not be able to fit in and offer your expertise even if it's on the volunteer basis even if you call up your local city councilman every single County in the state of California has to have an emergency response plan it's a public document anybody can comment and participate in the process so I would just encourage you all to be thinking about that as well for your local communities and it we're a long ways away from trying to gain the expertise at the local level I'm just going to be honest. I was shocked that my director actually let me come to this for example we have a big data collaborative group at the PUC of employees that are trying to work on data warehousing storage and a little analytics and mapping within the PUC but it's just a handful of us and we're pretty much is volunteering doing that on our own to try to bring technology into the CPUC. So anyways thank you for this good comments good trust right and there's a company called bright I've run out of Chicago so anyway it's just a concept I thought I'd throw out right now and we can talk about it I was going to point to your session tomorrow afternoon to get after kind of what should be the next steps focusing on local can be certainly part of the recommendations exactly all their data trust

Table 6 15:42

We discussed briefly about the very desirable situation of having **standardized data that is that a consolidated location** or agreed-upon location that would be easy to access which really hits home we work for **HERE technologies** so the open location platform that we're trying to build out and also talked a little bit about the challenge of **maintaining a training set of data for content that has a temporal characteristic to it that changes over time** they need to continuously re-annotate and continuously we look at make sure that you got a representation of ground truth. I'm very important to our transportation discussion tomorrow and from d-o-t and have a good conversation.

Table 7 16:30

we started with the discussion of how **there are so many domains coming together there's GIS, there's machine learning, there's data analytics, as the social concept of AI and how do we put those things together** and how do we help our local governments and how did speak tech who can we can throw down money and just get whatever data we want essentially vs. academics who struggle to get data and struggle and it's there's a dichotomy there and like how do we similarly how do we have a data platform where maybe it's easier to share data with select few groups that we want to share data with and is there a **standard licensing** that can be used maybe this is where OGC can help us and is there a place that we can put out data and know that it's safe and but with that we talked about uncertainty and like how can we measure that and how can we push that through our data and with the metadata. The specific example was given and sorry Apple maps but like everyone's experienced either on Apple because we're at Google that weren't say Google hey you're on this rally it's gonna take you five minutes less time if you take this all their route in that moment of what do I do and it's never any faster you know what if there was the suggestion from Ed was like what if there was an uncertainty associated with that just making sure that you know if there's an 80% chance miss it's a 20% chance Kosta you know what you're gonna do so the overarching I guess theme of our of our discussion was **sharing more whether it's data or whether it's modeling but then also making sure that we have an idea of what the quality is** and you know how stale is our data and how good is your model and being able to communicate that as well did you guys

Individual Mark Reichardt 18:32

I'm intrigued in about a couple of areas first again the issue of **uncertainty how do you explain that in context to the decision-maker** on the ground particularly at the local level that the model is predicted here's the uncertainty or here's the reliability of this information that's huge and I'm not sure we've even scratched the surface on how to communicate that and I think that was brought up earlier. The other is bias **New York City has gone through a tremendous activity of understanding bias and machine learning** how do we how do we detect it how do we manage it how do we explain it in terms of the decisions that are made from third-party models by local government without really good knowledge of what's behind the machine learning I think these are really integral contextual issues for the community. Then one of the things I observed with the Houston storms not too long ago key is data inundation model inundation it's understanding how we organize the response and appropriate swim lanes **so the first responder gets only the information that he or she needs from someone who is processing upstream just in time** and I don't think we've sorted all of that out as well as a community is how do we work a straight the use of AI models and other activities down to different swim lanes and a response mode so I just offer up those ideas.

# Session 3: Ripe Trends

* Moderator: Jay Theodore, CTO, Enterprise Technologies, Esri
* Rapporteur: K. Kim, AIST and GeoAI DWG chair

## Presentation on Ripe Trends – AI at The Edge – Philippe Cases

21:00

Philip is the CEO of Topio networks and spoke software he focuses on building market intelligence platform dedicated to tracking innovation he spent 16 years our venture cap in Europe and then in the US focused on information technology specializes in AI mobility and smart cities authoring or co-authoring landscape describing these industries frequent speaker at industry conferences topics are usually smart city and mobility but he's going to talk about AI at the edge you know

Thank you so much for having me I mean it feels like having spent the day with you it feels like you're actually in the trenches building those models building you know gathering all that data and making it real I'm on the other side I'm trying to understand where is the innovation happening what I'm going to do today is we have about to **release a set of surveys** we cease to love our customers Twilio and influx data and so I'm going to give you a preview of that so please I'm going to be trouble with my customers and so what we're going to do is we're going to we are going to see you know where is weary actually IOT and an edge apathy and and therefore where is AI gonna happen in the next the next future so very quickly you know we studied in 2015 so we are four year old we will focus on the business of connected things we studied with wearables right so Club edge computing is - is going to move for the future and for the path right so see us think about the yet this vide labs and that's us and so that's what we are tracking right now we have we have about 1.2 million subscribers on our platform

what we do is we a little bit like we write so a wave in the ocean but now when it comes to a beach is gonna sumbitch is going to be a big leap big wave and so what we do is we we we understand the same same thing **we identify the emerging trends** we map timing we met the critical touch cons and opportunities we then create and you see landscapes so **we'd love to have your landscape on the geospatial** and we develop a new trilogy by impacted communities that's where we have 1.2 million subscribers and then we enable companies and expert voice is to contribute content the story. With that we enable entrepreneurial starter but also be companies like to really understand how they can catch the next technology with you know what is the use case what is the verticals and how do you how do you go after those articles and we do that we have about now 15 customers and 1.2 million 1.2 million subscribers

This is the agenda all right so I'm going to give you a quick review of available market then we are going to look at a AT at the Edge at a glance and then we are going to talk about a moving forward at the edge a few observation and then we'll do Q&A; and hopefully we can I think it's going to be a nice segue to to the panel and we're going to have a more interesting conversation on the subject

So very quickly I think I think this this slide said oh all right that's the level of excitement people have artificial intelligence and you know maybe maybe deep learning right the term that we are using is AI so it's not the level of excitement that AI has in the industry. The reason is encapsulated in that slide right so if you think **the CEOs of the big companies 40% of the top performers think that AI is going to be a game-changer for the industry.** That's full time that's just to give you a sense of how important that is for the industry as a whole right so people love seeing that AI is in the game changer. so yes so the issue is not about whether there is demand **the issue is can we fulfill that demand** is there is there an infrastructure is there a set of tools is there a set of technology that enables you to really capture all that map and so if you look at and you know I understand it may be a little bit if you look at the way radar is actually by now so one intern enterprise now 10 or more AI applications governor studies you know the bigger the be AI the game right so the beat the big AI application was processed optimizations or a lot of industry and so that's that's where we are today right and

so now if we look at the other side where is AI on the Edge? So this is we are so you're still having so in order for the wave to I know there is a lot of even more data and so and that's where you see right so very small right and so if you look now the good news is that if you look now at the rhythm of adoption what you can [Music] so so so still still some some time to go but there is nothing in our study in our studies that show that it's not going to happen it's just a matter of the maturity of the project so now if you look at if you look at kind of the the data right so what are the data that are being collected today so what's interesting is that it's a mix of it's a mix of things that people are productive so of course the data itself is about 66 person but video people are collecting so this we percent of the projects that have been studied includes video 28 personality context and so so you're seeing more than data being collected and 75% of the data is actually time series and then of course you know I mean we've been talking about we've been talking about quality of data we'll be talking about accuracy the first word is descriptive analytics right so if you're looking at just being able to describe when the environment or if you if you're looking at politics and the best the best work and you're looking at this cave analytics you you may be able to cut corner there so again right so when you're looking at implementing solution start with the script even going back and build those tanks but instead of starting with specific event and then going back up going back down because like Steve is probably the hardest of the and and then the

[gap in transcript] slide 16 - 32:50

industries and and so I mean I realize I may be going fast is that okay I mean can you can you hear me is it is it okay and so so now if you do get

[gap in transcript]

in terms of implementing those those projects I know there is a Mesa I mean does that make sense in terms of what you're seeing as well in the industry is there any people that are now okay so now use case the the best **use case** for implementing 45 percent of the projects abilities actually because if you

[gap in transcript]

Moving Forward slide 19 – 35:00

so now switching so if so if you're looking at **it is very early** will pick up in 2022. Use case and verticals are very you have a lot of activity and then the use case are very you know right so smart and

so moving forward what's going to happen right so what's going to happen is massive amount of data being today all right so we are looking at we're talking about 10 terabytes of data four terabytes of data Airlines 2.5 terabytes of data so so it creates it creates it creates new require and the new requirements are emerging the new requirements we've talked about right so buyers so they then so when your massive **when you're sending massive amount of data through the network connectivity becomes an issue** and the major points so you know do you get do you have access to the data that you need and so if you look at it **so that requires a new infrastructure** and so if you look at the current infrastructure that's in space

slide 22 - what happen is you send you send data to the networks and networks and data to the to the cloud the cloud aggregate the data and the data is being mined on the smart data and then insight are being sent to the to the cars and so if you then insights are being sent to the smart data and so you get to a point where you have data that are basically going from one place so that's kind of the cloud model.

Slide 24 - **The edge model is a different** what happened here is talk about that what happens is the the the edge process he process the data create the inside the inside is being sent to the consumer labs and then the edge send also anomaly to the edge infrastructure the **anomaly sent to the to the cloud and** then the cloud creates that that that sends the information to the smart data and small data process the anomaly and create update to the software which is then sent to the to the device right so it's a it's a completely different model **where the data is being processed at the edge** and only located in the cloud and so so it looked like this and so now if you look at

Slide 26 - **so now you're going to have this new infrastructure** right so if you look at what are the key

Slide 28 - the key points for that infrastructure the first one is where you store all the data where do you process the data right so data storage is a big point the second one is the models but then so including this concept of a **digital twin** right so you're gonna create this digital presentation of entire cities, entire manufacturing capabilities, and that the information is going to be stored everywhere. you're gonna have to know where is information needed in order for me to add info you know in order for me to process information about that specific site and then of course the cost right so we've talked about the cost and the other thing is is we need cheaper smaller and more efficient hardware in order to to build that infrastructure

Slide 29 - so now whether without kind of the some of the key **blockers** and I know we're going to talk about that a little bit more during the panel so it's a way to introduce the pan also one of the key blocker is access to data as we discuss right the data is still there is data X you know that that is available today but it's not the comply the second thing that people are talking about is things about looking exam is the lack of talent there is not enough talent so you know University please you know train as many data scientist as you can the infrastructure is not is not ready as well and then the last thing is of course ethics and principle and biases associated with that.

I think we need to come up with further cross this specific one I think we need to come up with some type of a two-ring principle right wherever we would compare the barriers that human have compared to the values that a I have because right now we are comparing AI to an absolute standard and that's not I don't think that's appropriate right so we need to come up with way too to think about to think about things

So that's kind of my last slide so you know I was able to do this in 15 minutes pretty smart pretty fast. thank you

41:22 - Q&A

Question (Andre): some people have started to refer to as **IOT as standing for the internet of threats** yep. We hosted an annual meeting of the **managed security services** industry and we had a room about this size of professionals in cybersecurity and we asked knowing what the answer would be how many of them have connected devices at home – nest, Alexa, whatever - and it was zero percent in fact they were laughing. That we would even suggest that they might so in the blockers and the slowdown of adoption **where does cybersecurity play?**

Philippe**:** The answer is yes absolutely. That's one of the other questions [gap]

Question: [gap] I mean you're we're sitting at Google right now and Google cloud is pushing things and stuff like that so where does that number coming friendly where do you think that's going to switch from on-premise to potentially cloud and woods that time so I think [GAP]

Question: [gap] if you you know just think about it there's 5000 games in a day and then you multiply that by two and half hours and you have that at 4k just the cost of getting that in let alone what you do with it once it's in the it's not a guaranteed network **to get it into the cloud and so there's no scenario by which we can even just get the data to where it would need to go** from a distributed global you know SATs so it means **the cloud generally is off limits until all of our analytics and all of our kind of aggregate** I think I saw that chart before we're you know this kind of you save chart our aggregate analytics and statistics and that's all in a cloud so what we can do we do but there's a tremendous amount of stuff that I think the cloud is just not there yet.

[gap]

46:58

## Jay Theodore, CTO, Enterprise Technologies, Esri

before we start the panel discussion that oh it will be good to see some how we were at ORNL a few weeks ago where we talked about geo AI and what would mean like if you have to handle a hundred trillion pixels flowing every few seconds what kind of architectures emerge it was sort of like a moonshot kind of discussion. I thought it'll be good for us as we **talk about right trends** you know just talk about what it means

So I'm Jay for about 26 years obviously you know like many of you we've seen the trends come and go and very often it's the adoption of the trends at different cycles that matter in terms of how you use it and it's not often that the trend stands all by itself very often it is it's like a **tsunami discussion** right the title was tsunami and I was looking at it's caused by natural things like earthquakes which cannot be predicted sometimes it's by weather which can be somewhat predicted sometimes it's like detonation which is totally artificial you know it just comes up by itself right because you implanted something and you detonate it right so I think similarly we can say trends are also that way **trends don't stand alone by itself when you talk about AI and ML it's not all by itself that the trend is going to show up**

So let's just talk about the **hype cycle** itself. I won't describe what this is because you all know what it is but the key thing is there's a lot of disillusionment disappointments because it's tough to get it right ok that's what we're going through right now. There's a lot of marketing hype with AI and ML of course what is the AI which is what is ml what is DL that itself is like up for discussion because a lot of folks don't understand the differences the way I try to categorize is **it's okay to be a little loose in some of these thoughts** it's okay that the term AI exists because AI brings these technical things like ml and deal closer to human understanding it's trying to say that this **is augmented intelligence in a form that humans can understand that's all it's trying to say so don't over read the AI aspect of it and try to like map it with something or other that we are more definite of about.** Where this data science fit in? I think it's trying to get us out of this trough into being productive that's where data science comes in and the tough part is that we have to solve important problems if we need to make this trend useful and for that what we need is domain X peace without **domain expertise** we cannot bypass and make a I come to life in a meaningful way so if this is the case how do we go about doing it

so I want to start off with **defining spatial data science** itself as organizations we focus only on one or a few but not often on pretty much everything else so if you take spatial data science itself it starts like we said with beta engineering that's the tough part we talked about spending 70% of the time in trying to massage the data or do daily janitorial work as opposed to useful you know analysis work right so data engineering it starts off with there and then of course we visualize and explore the data sometimes you visualize explore to find out if it's valid data you know is it useful data is it data that's got a lot of noise that you have to filter out and then of course you go in and do your spatial analysis you do the modeling and scripting you bring it into production you work with all kinds of data real time big data and so on but the last part of collaboration which is communication is critical none of these things of any use if you can't collaborate and communicate tell the stories that you found in the process of doing spatial data science

so if we talk about the experience of data science itself a common way I'm not saying this is the only way is notebooks we talked about it we had presenters talk about it but **notebooks** are a way to experience data science to do your experiments to make them repeatable shareable and so on but what do you work with you work with various spatial analysis tools you might work with a lot of open libraries that are out there scientific libraries and of course you bring AI and ml into this so it's sort of like a experiment that you conduct within a **notebook and then when it becomes meaningful you bring it to production**

now I want to switch gears a little bit after making that definition on if you need to scale this bring it to production make it useful one have a discussion about **why the edge matters** so today if you see you have all the satellite imagery the drone imagery and IOT and everything else going into some cloud a cloud of your choice so very often you're sort of **like tied to a cloud a public cloud a kind of cloud** and then when you talk about **cloud native** it's like have you made your code native for AWS native for GCP native for as you know fill in the blanks right that's how it works today. **But it's all going to one place** it's one data center in one region in one location and so how much of it can go to one place as opposed to the surface of the earth not to scale of course here but if you take a city that wants to get all of these coming to life that's just a tiny place on the earth and you have a massive surface area to cover so doesn't it make sense that as you have to cover the surface area that everything is going local to that place? now this might just be like a hypothetical scenario right but when it comes down to **building AI models** I think this is the way it will come to production we can't build global models overnight or within a limited time frame because **what matters is the training** if you need to remove a bias from a model whether the bias is can you detect rooftops of buildings or can you reduce let's say crime in a set specific area based on demographics movement data and so on you have a lot of training data within a certain small region I'll call it the edge the edge that you define as a precinct or as a city or as a state you train the models there **you deploy the models locally at the edge** don't attempt to create global models it'll take forever so those are the reasons why we are driving towards the edge as opposed to cloud of course the other reason voice you know latency and the volume of the data and so on but as we talk about the **trend the trend is often defined not just by their physical aspects of things but also the practical aspects of social acceptance** so that's why I'm bringing this edge case

similarly when you do that not all things are equal at the edge we have **edge devices** everything from our phones and a lot of these sensors are very smart I'll just classify them as edge devices as opposed to there are machine to machine things services I'll say at the edge where humans are not interacting at the edge that is really responsible for the state of the road in a certain place so that as fires go by the sensors on these are getting detected whether it doesn't matter in your overall calculation of going from point A to point B across the state whether that is happening there what matters is the other cars that are coming around in that area what's really going on and then you give it to them in real time that's what matters so the edge in that case it is sort of like a machine to machine communication at the edge then the **third case** which I referred to as **edge portals** is actually for human interaction so if you for example want the complete experience a human experience you don't want to go to the cloud but you use the models at the edge you have portals that have built-in models applications they tell the ends stories at the edge you don't need to run them in a centralized manner

that doesn't mean that the cloud goes away it just means that the cloud now extends to the edge and there are many techniques that are coming in from a software architecture perspective which tries to abstract that we sort of like moving from your lock down to a public cloud scenario to going multi cloud and then going to the edge the same way I can mention one technology for example kubernetes kubernetes tries to create an abstract layer where if you write software to that you cannot only run in any cloud of your choice because all of them have accepted that as a kind of a standard platform but you can also get it to the edge through cue beds so there are new techniques coming up where you can think of the cloud as infrastructure and then there's going to be more infrastructure at the edge where the cloud providers themselves will move to the edge like we're seeing happening there is a Geo data box edge there similarly from AWS and so on all of them moving to the edge because you want to have the receivers also at the edge the first phase of edge I would say we're just responsible for receiving and disseminating the second phase of the edge which is really what we refer to as **edge computing** is actually executing getting the job done at the edge itself and then **if it's meaningful it'll send some aggregated information up to the cloud** and then the cloud is probably getting a global view of the system as opposed to a local view so if you want to predict what will happen in the next few years you might go to the cloud and get the **answer if you want to predict something at the edge in five seconds** it's what I'm going to go to the cloud and then come back so here's a few examples and just because let's say training models you want to **batch inferencing you probably go to the cloud if you want to do interactive inferencing you ask at the edges** because that model at the edge is probably governed better to run that the edge is probably more compliant to the local laws of what you can use and what you can't use that's also they will you address privacy and bias and so on because if you're trying to solve a global problem with obvious that it's going to take much longer than localizing the problem right

similarly we want workflows to be end-to-end okay so when we talk about deep learning and it's using imagery we have to talk about the full workflow where you collect the data you prep the data you use it for training and then you build the model you validate them and you deploy them and then once you deploy them you probably want to run the inferencing at scale this is the full **end-to-end workflow** this typically might happen in the cloud because you want to have the full system in the cloud this is what probably happens in terms of batch inferencing now how does that come to the edge itself you pick and choose what you want to do at the edge and then you deploy it at the edge you don't do the full length to end workflow like you were doing in the cloud you just bring the ones that are relevant so for example you might train the model at the edge because that kind of citizen data may not be able to go to the cloud in a different region or a different country so **you just train the model at the edge and you from inferencing so there is no privacy law that will broken** and you probably didn't even push that citizen data up to the cloud in the first place

so really what we're saying is today we have cloud intelligence and we have edge device intelligence so you can pull out your phones you can search for your daughter or sounds or your spouse's name and you can get the list of all the photographs or search by place searched by face you know similar faces **edge intelligence on the device** then you have the cloud intelligence where everything goes up to the cloud you run the models and you have it there but there's a huge gap and this huge gap is because a lot of IOT data that we talk about whether it's sensor data or data that's collected in so many other ways it's just wasted today because we don't know what to do with it because **at the edge a single piece of data is very important, at the cloud it's the aggregator that's usually** **important** or you can't eat maintaining that tiny piece of data so what we need to do is fill this gap right I'd refer to that as edge intelligence so when you do add intelligence it's not again a black or white it's a combination you can do a combination of like you see here a device and the edge together doing some training or inferencing everything could be done at the edge or it could be a combination between the edge and the cloud so it's sort of like this full spectrum so whenever we have these conversations about you know this might be good for this or this might be good for that what it's really all right what is coming out of it is really **a continuum of architectures** that support these cases so it's again not just the cloud or not just the on-prem it's not in your data center or in the public cloud it's all of the above it's just that as humans or as a software architects we tend to focus on wherever the deficiency is we highlight that and keep paying attention to that until it becomes the norm right so today I would say yes it's accepted that the **cloud is serving the needs that it claim to serve of course like you said it's expensive and so on so it's probably bent for some elastic scenario is some burst scenarios and so on but if you know exactly what your job size is it's probably done better elsewher**e so there are combinations of these things that'll arise

again when we talk about **geo AI which is using ML techniques with data science** there are many fields that are going to be affected and when they affected think about the impact as things like these converge: 5g, edge computing, and many other things that are probably not listed here so think about all of these things because sometimes it's not just one factor that is going to make a difference it's going to be many it's like cost versus value right sometimes the cost is significant the value is very less but if many things come together then at larger scale the value can stay high but the cost comes down quite a bit so

## Panel on Ripe Trends 1:04

### \* Anand Kannan, Pitney Bowes - Perspective: Data science, an interdisciplinary approach

I want everyone to know that I wrote these slides before Todd spoke and Laurence summarized the tables observations I actually already deleted some content so that should save some time

Just to complete the introduction my day job is managing a bunch of geospatial software engineers we produce a product so and I happen to as part of my duties as a fellow in the declarer eight I work with data modeler data engineers and other parts of the organization as well so some of what I'm going to present here or come from that perspective right so so I'm not going to go through why the motivation for skill set diversity we gon be been through that but I'm going to focus on the challenge in the opportunity right so many universities my son goes to Waterloo I have a friends in Ufa we hire interns from U of T and Ryerson University of Toronto that is so we have these universities do fantastic job like the quality that they produce is very very good but in a very narrow sense **many students graduate knowing one of the several discipline we need to succeed in our jobs** if your job is to build a model that a bank uses for let's say customer segments in revenue prediction and things like that you need a little bit of domain expertise Mike Lauren said but **the technology space you need is quite wide deep learning geomatics sequel statistics operations research graphically be big data this is subset right there are several things you need to know** so the interesting thing the opportunity for the organization is when you build skill teams to solve problems whether it's a software component or a model or a data engineering problem it's a nice opportunity to deliberately build a team that's diverse talent wise right so you that's actually a little bit hard for all of us to accept because we are we are all too well-trained in one thing and it's kind of not easy not to see the value of something else I can't count the number of times I've said myself hey I've been doing this for twenty years no not the same way right it's actually quite eye-opening for you to get a new 22 year old University student come and teach you things that actually the last time I hired an intern the first thing he said is interview went really well both of us seem to agree that it's a good relationship but one of the things he asked us **you're not going to make me program in Java are you I** said it **I fell off my chair** ik what would you ask that it was like well my comfortable is Python I can build you up it it happened to be what I had in mind I was rude but just kinds to kind of change that he said that yeah so it's important for us to remember that our skills whatever however proud we are of them it's like only subsets of the tools you have **the good tools men understands all the tools what's available and when you know you don't know you're missing some tools go acquire them** I mean that's how you 300 shooting I also said it involves learning and unlearning I mean the geospatial geeks that I work with I've been telling all of them saying hey you're I know you're great Java programmer but they asked you to write a code till if a picture is a cat or not how long will it take there was a university student who can do it in like half an hour you're using source that already exists elsewhere so that that's justification for us to go out and learn the things that are that we are missing and the other thing is unlearning I again not being very quick to say it took I wanted it done this I did this 20 years ago I say that again and again I stop myself it's so

**the mandatory Canadian map** so so I suddenly realize it's no map so I had to put something put it in there very quickly I'm going to talk about **some projects that we thought we did well in machine learning and I'm going to talk about what's common between** them I'll go bottom-up we actually train your networks automate making some map entering decisions typically when we sill mapping data products some cartographer sits and like laborious lis figures out everything about every layer there is it's all the thing it means to do and just that's doesn't scale I mean if you go to another country there's several reasons why all that has to be read in the data suppliers are different local cultural setting is different even if map look and feel is identical the decisions are entirely different could we make 80% of the map workspace redundant so that actually turned out to be we didn't do it intentionally but one of the things we did right was **we got one of this like interns were very smart knew all the tools in the CNN day a trade and tied him up with the cartography software engineer specializing in cartographic decisions that worked beautifully right** so we didn't set out to do it but we did it just kills it at that time but later it we turned out how that uneven in cheating work so **likewise I mean things like free format address parsing using neural network somebody did the deep learning course in the geocoding team** so you know what like all these if-then-else conditions I do it at work for four countries and we have like hundred and how many countries right so those things don't work very well like I mean its case statements if-then-else statements get very ugly very quickly so would you let me retrain train this models now those models ship is the thing so it's it's actually a happy coincidence that somebody actually had the motivation to go something outside the day job right likewise I mean I'm gonna rush through this because I'm taking too much time so the other is a lot of the decisions we made when we there are lunchtime we talk about the points of interest when a new point of interest comes from a new data source is something like that you to say he's this record the same as what I had in other records we had a Glee cold like you wouldn't believe right all that is now like going through neural networks **it's it's a kind thing of like somebody who's completely new to the team looking at the problem in a different way** so it's actually there's a little bit of lesson so when you do make this change **it's not going to be comfortable either for the managers or the business people** right managers will need to sign up for deliberately for some setting up some storming in the team for people who are Mis come terminology no **storming norming performing** right so so it's throwing his team back but that's for a good cause you should sign up for it likewise the business people usually pay a premium for predictability I mean whether you're training model or whether you're actually starting machine learning thing you're signing up for an iterative thing so software engineers want to like design the whole thing from end to end and anything you don't anticipate everything completely you're a failure right it's a completely different mindset for engineers likewise for business people as well don't project the end try to get it feminine be ready to cut it off of it doesn't work but try not to predict the whole thing so that so it's a little bit of training learning unlearning involved in everyone's part that's my point of view. thank you thanks

1:12

### \* Milind Naphade, CTO, Metropolis - NVIDIA - Perspective: AI-IOT and Location

Alan next is Milind he's the CTO of metropolis Nvidia he leads the technology and innovation strategy over there and has deep expertise in applying AI to solve various business and scientific challenges go ahead Mellon ok let's try this thank you

good afternoon I think there was a bunch of things I was going to show you about how cool you know **sensor data when it comes together with system of record data** you know we can really get insights for multiple domains but after looking at the slides J had I think that motivation is not needed anymore

I'll start with a question I know I'm going to get because I am representing Nvidia we are talking about compute we are talking about compute at the edge compute in the cloud and so on and so forth and this is fresh I just read this early this morning that opening I came up with this study which says that till 2012 that's the era that they are now calling where it was okay you could go with Moore's law and the doubling the two year doubling was okay but they are basically now showing with examples that in just the last seven years we are needing more compute unique up you to **double every three months three and a half months so this is just going to get more demanding** so we know for a fact that a lot of what we also saw earlier about the various use cases this will translate so we are going to need more and more air compute the hype cycle or no hype cycle you know the compute need is definitely increasing some of this will come out as products you know functioning products when people create these technologies **some of those will die as data science experiments but the need for compute is definitely going up**

I'm gonna skip this because I was going to talk about how my my approach to geospatial is more as a user I know we have seen AI talked about deep learning technologies to really annotate the environment itself the underlying map and you know in the digital layer enrich the environment context but we are looking at it far more from it from a perspective of **marrying the sensor data that we see with the and data that is coming from a Building Information model or you know layers of GIS information** of ultra environments and when you look at sensor data you know it is just I will just show you with an example the amount of data the both the volume and the rate at which these data is coming at us is so humongous that as was discussed earlier going to the cloud is actually not an option you have to actually push the processing closer to the data then push the data to the cloud and and this is something that **will impact so many other verticals in terms of getting situational awareness** so what we are working on at Nvidia metropolis is basically a stack goes all the way from hardware and software **to help developers AI end-to-end solution developers build solutions for situational awareness** that's really what the platform is meant for so

just an example right we are this is something we actually are doing we are already deploying using kubernetes what we call the HDX stack this was just announced by Jensen at the Mobile World Congress in LA a few days ago and what this stack really does is it basically **allows for your workloads to run at the edge** and these workloads can be addressing hefty data like video data so if you look at a the Levi's stadium I don't know how many of you know they have a thousand cameras if you look at a small city like the one that we are working in Iowa five mile by five mile city it has more than a thousand cameras just imagine right that's terabytes of data per day that is being generated you just don't have the option of sending that data to the cloud even if you did I mean various issues privacy security but just cost inject that ones some simple question makes the determination for that CIO of the city saying I can send this to the cloud so you have to take the processing to the edge right so what we are doing here is we are basically deploying an end-to-end application for traffic monitoring and analytics and what that does is it actually detects objects again just looking at if you detect 50 objects per frame that's 1.5 million messages per second so all this data is now I would **say this is the sensor plus geospatial because every object that I detect currently in that frame by calibrating my camera I can actually tell you what is the GPS coordinate** so we are actually getting more data from cameras about location information of objects then what you get from you know I started I started large projects at IBM Research you know using this data right tower log data and GPS data from apps etc but the rate at which even the messages forget about the pixels that we are getting by **analyzing cameras** is just your mungus there are other data sources to write like lighter 3d point cloud so on and so forth and we really cannot make sense of all of this if we are just limiting ourselves to the pixel domain to the image plane **we have to convert this information into the real world map based understanding of the context to say what's going on so** for example I don't know how many of you can see this you know I wish I had the video with me this is actually a real anomaly this nobody was hurt in the making of this video but this is a real anomaly where a **vehicle kept going in the wrong lane** right and it kept driving in the wrong lane and we actually ended up detecting it in multiple cameras this was an experimental system the city did not have access to these insights you know life but if they had they could have actually done things in real time like you know switching the traffic lights to red escorting police vehicle you know escorting this vehicle off of the road so on and so forth so lives could be saved right so in the various things that we talk about in terms of spatio-temporal awareness situational awareness insights some of these as was discussed are more global in nature so we can actually figure out what is the you know how does the traffic necessity look like although we of course have Google to tell us that anyway but these kinds of things such as accidents anomalies or optimizing the traffic based on how what kind of vehicles are actually experiencing more time etc you I have to go to other forms of sensing and so my my main takeaway from this is there is that is **definitely compute coming as close to the data as possible now some people may call it edge some people may call it fog** you know different terms but there is a continuum and we'll have to actually process it in different places the other part is the networks you know **a really good vehicle detection model in the Midwest may look very different than a really good vehicle detection and tracking model in Shanghai or Beijing or Mumbai so the localization and in fact I would say creation of bias that makes the accuracy improvements in those particular environments will actually force us to actually have an ensemble of models where we switch between different things for example in this city** itself it's already started snowing we have had three snowstorms so we are anticipating we'll get more data to do detection of vehicles while the snowstorm is on but there are other use cases right where you know do we still see the road networks right or do we see them covered you know how has the snow been plowed throughout the city **all these kinds of additional insights the city can continue to get if they have you know this geospatial situational awareness that marries sensors to location**

a second thing that we have done in Metropolis is we have created a small little library called **cuSpatial** and the purpose of this library is nothing more than GPU acceleration of very common space spatial operations right so we are taking **the most commonly used operations for example point in polygons such a simple thing right I'm learning from my geography experts and my GIS experts** that you know that's the bread and butter of a lot of the underlying operations that happen we are able to accelerate it by ten thousand a factor of ten thousand right and so we talked to our colleagues at Esri and they say that if you actually have that a level of acceleration because of the architecture because of GPUs and CUDA things that were not interactive before things that needed a lot of indexing can actually become interactive and so that has a lot of promise in terms of how data scientists - there was a question should GIS engineers, data experts also think of themselves as data scientist I think **this is one way in which that can accelerate because they can get instantaneous results of whatever hypothesis they want to test**. this is just clustering again it's you know several orders of magnitude better using this it's a free library anybody can download it here are some of the links now everything we create in terms of the software stack is free for people to use and download

I heard people talking about data sets and the lack of temporal data sets so I wanted to just add one more point which is we have for the last four years now including this been running a challenge core a **ICT challenge and it's all about spatial temporal data the primary source is video** but we have other metadata also that we couple with that this will be the fourth year we typically run it at cvpr because our first you know 300 participating things all were computer region teams but we are getting strong interest from civil engineering and transportation departments to you know for those who are actually interested in just looking at this as a spatial temporal data set you know I would welcome for you to take a look and reach out to me thank you thanks Milind

AI City Workshop at CVPR 2020 will specifically focus on ITS problems

1:23

### \* Devaki Raj, CrowdAI

next up to present would be Devaki Raj she's a CEO and founder of CrowdAI, Forbes 30 under 30 and 2019 Inkster Deon under 30 2017 I mean wow good okay so she worked on AI for social good projects former data scientist and part-time product manager. Google used to be your home

My name is Devaki. I run CrowdAI and what we do out crowd AI is builds computer vision and deep learning technology for any type of imagery and video to help our customers solve their largest operational challenges so we work with customers across utilities telecoms insurance as well as oil and gas but **what I'm here to talk to you about is a little bit about how we're trying to solve some of the largest operational challenges and that often occurs after major natural disasters** so I want to talk to you about the work that we do taking different types of third-party imagery and as you guys know in the last in the last five years alone there's been an explosion of different imagery types right now you know we're talking satellite but we are also talking about drone we're talking about aerial providers and it's not just imagery that it's exploding there's a lot of satellites that are being launched for things such as AIS so right now I want to show you a couple of examples of the work that we do in terms of the trajectory we took it as a company to **exploit all the different imagery types and then solve large-scale challenges** in this case I'll focus primarily on natural disasters so

we started about three and a half years ago after graduated one from Y Combinator and that was a time in which Planet just came out with a lot of their satellites a huge selling point what for Planet was primarily even though the resolution is low they've got global daily coverage and so when you think about natural disasters that's often what you need you need daily coverage so when **hurricane Harvey** hit we wanted to do something so what we did right here is we took planet imagery over Houston and this is pre disaster and then we took the three meter resolution data after her Harvie hit so you could see that there's extensive amounts of flooding the reason this is particularly interesting is because often floods as you guys know once floods hit there's two problems when you think about satellite imagery the first is that there's a lot of clouds so subtly immature really doesn't take a lot of the imagery or can't satellites can't capture a lot of the images but to floodings floods tend to dissipate obviously unfortunately for Houston the floods did anticipate but there was imagery captured almost kind of within the after a day after hurricane Harvey get hit and so what we did at CrowdAI was use our **Road detector** so we had worked on building a global roads detector on Planet for the last two years before hurricane Harvey hit so we **trained our model in over 126 different countries on Planet** and primarily why we create so much training data sets beforehand is because if a disaster occurs no matter where it is in the world we want our models to scale so what we did was in green you could see give me one second what we did was **map all the roads on the previous imagery and then we mapped all the roads on the post flooded imagery and then we did a matrix of the two** so the red indicates all the roads that are flooded and the green indicates all the roads that are still not flooded what we then did was convert this you know TIFF file format into GeoJSON vectors and provided this data to first responders

I want to talk to you a little bit more of a different example so this is on NOAA imagery and so this is a different resolution this is about 20 centimeter resolution and this particular example is the work that we did **for a telecom company** so this telecoms company is a six largest telecoms company in the US they service a la a lot of different states in the Midwest they knew that a **hurricane Michael** was going to be passing over one of the areas that they serviced I mean that key in this case is **Panama City Florida so NOAA imagery** is taken obviously by NOAA and it's aerial **aerial footage** so it's underneath clouds generally and NOAA footage is generally taken over most continental us kind of us disasters so what we did was given that we knew that hurricane Michael was going to hit we took past NOAA imagery created training data so we could have a so we could have a model that was ready when hurricane Michael hit so what we did was what once hurricane Michael hit NOAA imagery was actually available a couple days after Panama City Florida was was devastated and so what we took was all that imagery we mapped all of the buildings **on pre imagery and then transferred that pre imagery buildings on to post imagery and then classified each building based on definitions that this telecoms company wanted** so what we did was we **mapped almost 18,000 buildings** in a couple of minutes first getting the building footprint and then providing our customer with locations of buildings as well as **addresses of buildings from anywhere in terms of disaster from roof tile missing all the way through to entire building missing and what they then did was prioritize where they send their field service folks on the grou**nd so we see this scaling across obviously not only just post disasters but had the potentiality to have be helpful for disaster management and risk purposes pre-disaster the thing about hurricane Michael is actually pretty interesting is if you notice that a lot of the houses along the beach weren't devastated it was mostly a lot of the inland inland buildings which is opposite from what one normally thinks after a hurricane

got a couple more examples this example obviously hits more close to home obviously as you guys know a lot of **California still under a lot of wildfire warning and wild fires still happening** so we obviously being a California company wanted to be able to do something about it so this is on a different resolution image **this is on Digital Globe** and so we'd worked with digital globe for last five years **and trained our models in again 125 different countries** again why we trained our models extensively globally frequently is because if a disaster occurs we don't want to wait to train a new model so to create train data train a new model and see how it works so this is the work that we did post **Campfire** and so we mapped about **25,000 buildings off of the imagery** that digital globe had on their open data platform so this is at 30 centimeter resolution we mapped all the buildings on pre imagery over this sorry that it takes of polygons a while to live there's 25,000 polygons that are currently loading right now and then we mapped all the buildings post fire and you can see that our models are robust over nadir angles because we train our models on different types of major angles and this imagery happened to have a pretty severe nadir angle and then we identified all the addresses of the houses of the polygons that were no longer visible to the model and what we're able to do is kind of scale this in a crisis an entire county and provide this to first responders we then take in County information so these are all parcels and then kind of **classified it by level of damage** so this is all our post disaster work and wildfires are particularly interesting because **you can't necessarily just aggregate and say this entire area has been destroyed because if you see individual houses in the same neighborhood some are standing and then some are completely destroyed** and

one final example of the work that we do is kind of **risk mitigation** so as you all know **Santa Rosa fires** were we're fairly devastating in late I think 2016 2017 but this is imagery that's taken on near map a year after all the fires and you can see that **people are starting to rebuild** so a car today what we're trying to do is understand we know that there's a lot of we know that fires are still continuously going to happen so how do you how do we speak with the homeowners how do we speak with the insurance agencies how do you speak with all the you know people that do the construction to say **given these risk factors mean that you have a higher risk for wildfire** potentially you're gonna build your your your houses in a different way so what we did was took the same area that we knew that was devastated by Santa Rosa mapped out all the building footprints and then took it one step further right so we mapped out burnt trees healthy trees utility poles and also use the **Cal Fire's definitions for zone edge** around what is a kind of zone 0 through zone 3 wildfire zone specifically to understand whether people are essentially rebuilding their houses that were lost in a fire in potential and risky manners and so that's why we're we want to take this one step further **not just to mitigate risk after a disaster but mitigate risk potentially before a disaster** so a crowd a we don't just work on natural disaster use cases obviously that's something that's really close to home and and really important for us but the work that we do can extend not only just to kind of nonprofit or Emergency Response but across a lot of commercial customers as well.

 so if you have any questions I'm happy to have to answer them

### \* Jim Stokes, MAXAR

next up is Jim Stokes Jim is the VP for analytics for AI and m/l products to at MAxar. Great thanks appreciate it how's everybody doing one of these will work good okay yeah

some observations from the edge so I'll leave the the analytics AI and mental product group max are it's a new group we are basically a company that's evolved out of a bunch of companies most prominently probably know great imagery they're being modeled by CrowdAI. I but we actually do a lot of work with a lot of other data types to most most everybody thinks of us as imagery guys but we do a lot of other stuff and IOT data placed wrong and a lot of different things that we that we work we've worked with machine data in agribusiness so when you look at a combine it has potentially today **a big John Deere combine and has gazillions of points coming off of observing observations in the field** **we've adapted to an IOT model legacy sensors like stream gauges that are now connected to the Interne**t just go ahead and call it a part of IOT we can measure a stream volume weather data so we've got a **weather analytics team** that provides products and it uses field observations that come from devices in the field that make those detection and so when we look at this data we realize you know the standards are just not really there we've got we've got a lot of a lot of challenges in that space so you know my perspective that that's a place especially that **could use some thought leadership and guidance from a standards community** you know the data a lot of times little different from imagery data these observations are small but the volume is huge so it prevents a different set of challenges and dealing with that data and in many cases it's also very high velocity and with that that high volume and it and it represents an extremely great opportunity to be able to **change the paradigm of analytics to be more focused on that edge** but at the same time that volume creates data engineering challenges and many of my colleagues have addressed a lot of these different observations and different perspectives but from my standpoint that that data transaction and I really liked what you had presented about that **gap between the edge and then the cloud we can't we have to have both** **and how we bridge that gap is really important** and we we've had to build a lot of different things to make that gap filled in because they just don't exist right so we're living in a world where all of this stuff is very **much like Philippe said we're so early this is baby our AI is is of baby mind capability** right **AI right now is just tools** right when you really think about machine learning and the other components they're just tools really to kind of help us so these these technologies you know really standards are important one of the big things we also observe in in the data and the normalization issue is the time factor so I wasn't around maybe for some early conversation earlier in the day around temporal but **we see time is one of the critical elements that comes through to be able to normalize** a lot of these and then make key observations that are extremely important from that data so we've actually been kicking around some concepts and written up couple **white papers on the aspect of a time information system** much like a GIS geographic information system an actual platform and a methodology for dealing with those temporal datasets by presenting standards and methods and I think that would be something from our perspective that could be great in an open source world too to have the world kind of pile in on ways normalizing across time **pedigree** is critical we see **tremendous amount of spoofing** especially in like mobile device data that's used LBS a lot of people get paid by the unit so they come up with ways to add more units to their pie and if you're not able to detect where that observation came from then you're gonna get paid for that so spoofing is a big problem **a lot of fake data** generated that we see and then you can't even trust the device necessarily right the device depending on its own pedigree may be throwing off some sort of errant data or data that's just not right based on its trying to make the performance bet I mean we all saw what happened with VW right I mean **reputable companies still do things that they probably shouldn't do** and you can't just trust the chip you can't just trust it right so you have to be sure that you think through where all those different pedigree aspects can come in terms of the data **privacy** you know big deal right when we're looking at a lot of things that we can do with device data that comes from these it's really important that it's used for good we were able to discover a lot of pattern in an anonymized way and use that location data to help us understand other things or better ways to use our satellite constellation or new models but the real key is really to deal with that privacy in such a way and I think that's one of our roadblocks along with policy and governance more so than even the tech right the **tech can get us there it's it's those aspects of that privacy governance and otherwise that are gonna be an issue** and this is this last point here is for us you know we work with satellite imagery and one of our big paradigms is there's **not enough eyeballs** to look at the images to understand what's in the images so we didn't need the machine to make life and data easy for us that's the value proposition of potentially using machine learning to pull features or automatically detect objects and imagery but really there's other novel and unique ways to use the tool right and back to the point of it being a tool you know **using machine learning methods as a filter to help feed the analyst** having been an analyst most of my career having learned from a guy who was a senior in US government who was an **all source analyst** and had a slogan on the back of his card that said **what part of all don't you understand** and we talked about all data earlier you know I argue that we need to yearn for the best data not all the data right so let's use the machine learning as that tool in the process put it at the edge figure out ways in the near term to deploy that tool I think that's one of the big things we've observed in our you know artificial intelligence strategy from a product team is that what can we do now right and maybe I come to the table with that little disruptive thought of we've had a lot of discussions about big things and different things but how do we use the tool now to help get the most out of these different volumous content that just present these daily challenges so that's my perspective and thanks for having me

### Group Discussion

1:43

the way we're going to do this is I have a few questions that I'm going to put up on the screen and we'll have the panelists make their comments about it and then after we finish a few questions we'll open it up for everyone to ask so just think of these as just to get the conversation going so

Question: the first one is **what is the future of AI as we head into 202**0 what key problems will be addressed in 2020 **which ones will linger around and prevent adoption by the data science community and organizations at large** accepting it as a mainstream aspect will it be ethics garnon's explain ability or just the fear of the unknowns which is sort of like cropping up to the top quite often in terms of this adoption that will factor in as the thing that slows down the pace of AI infusion into everything that we do in data science

Anand: yeah so short answer is all of those if you ask pimmy for us me for a specific answer what I will say is the biggest challenge is lack of data I mean there are companies today built for data they have a natural advantage there are other companies that have deep enough pockets they can acquire the data that they need to make progress in AI but and there are small companies who I mean they were cases when she talked about you're actually very nice applications niche applications razor-sharp focused they identify a problem that they it's helpful to solve and they know how to do it right but there is a big ecosystem in the middle right so there are there are companies that are small enough their data set is not big enough to make decisions with confidence there are cities municipalities I was in a discussion in the Schulich School of Business in Canada and they happened to set things to two or three city administrators and they were talking about it what fell for it fell to me is like their respective towns they were all pretty close to each other there's similar socio-economic structures and so on but their respective towns don't have in a critical mass and data but together they might so what might happen is we are going to see a little bit of **ecosystems evolving by compulsion and may not be strategic** that may not have we are doing it chosen but that's the only way they can get enough data to make decisions right now another example is **collaboration between companies** yesterday there was a news bit about Toyota collaborating with the Japanese with the Department to provide wiper data alright so this is not something that happened by this thing suddenly somebody identified a mutually beneficial opportunity they went with it so we should expect conglomerates forming across companies in other there is insurance and health industry right so there may be collaboration that we don't anticipate today that is going to be the next step to watch so look out for unusual collaboration exactly anyone else want to comment on this so what I would yeah so

Milind: what I would say is there are some real good successes that it's I mean when we talk about air this is really something that's going to transform for decades it's not at the risk of saying this question is you know very short-sighted it's very short-sighted we are going through a totally different paradigm of what computation means you know what I mean in terms of workflows changing and so forth so we are not I would say we are in the first innings right and I don't know if it's you know we are just seeing the first base right so there are already notable successes for example conversational yeah he's doing really well you see more and more most of my iPhone answers whatsapp messages SMS is by itself I don't even seem to be driving the response to an SMS that I get you know I already suggested so if you get something from you take it to the pinch of salt it's not the real millon so so conversationally is doing well I think vision part of it it has found very good niche applications self-driving will happen it will happen but I think that the main thrust that at least I personally will push for in 2020 so that that's what I can answer there are **a lot of what I would say hello world experiments** and we need real world solutions right and so **pushing a lot of these experiments into real production into real you know operational deployments** is where I think more and more of us will see our trust being and that will bring with it its own share of pain points which have nothing to do with the rhythm or the actual science it's just **the core engineering of building end-to-end systems** and so far we have nicely segregated ourselves into I am a data scientist I am a design engineer and whatever right but at the end you know **a city or a retailer or will want an outcome** so I think we have to start gearing towards the and to that end **benchmarks data sets with labels** you know these things actually really keep us honest you know each of us you know in whatever way we can should be contributing towards creating this benchmark data source so the **end users can actually start having confidence that this really works** and I'll give you an example you know there are DOTs I'm sure there is somebody here from transportation they have a big fight with us you know saying no radar is it you know we don't need we don't trust cameras to do as good in terms of vehicle counting it or is it and **so it took us years to actually prove to them that vehicle camera based vehicle counting is as accurate as they need to be** for making this that was something that had to be proven the assumption was it doesn't

Question: how far will **Hardware play a role in propagating data science in AI** to be an integral part of everyday life what are some limitations and challenges today which we feel that should be addressed or being addressed is the next generation of applications poised to benefit from these advancements so milind I mean you come up with new hardware right a lot of software doesn't run on it

Milind: you know so so I would say that which which says that you know we are really expecting a doubling in computing need every three and a half months so no forget Moore's Law right so we do need to go beyond that architecture we do need architectural advances which we are seeing but it's not just hardware because if you notice you know for example I only speak on behalf of Nvidia on top of that GPU computing architecture is **a stack called CUDA and that's what makes any code developed by an application developer work against any of our hardware** so as as much as hardware getting better and the software above it getting better is concerned which is going to happen the fact that for an end user or in this case a data scientist or data engineer being able to write one you know one piece of code and having it work everywhere is actually vitally important and and that's what will actually help in pushing the next phase of benefits from the application. Is media working with the universities to increase the you know academic environments you know **turning out good Cuda programmers** absolutely absolutely so we have called **DLI deep Learning Institute** we have a lot of initiatives where some of these courses are offered completely free of cost and online and actually we have a we have an extremely strong partnership program with universities so we a colleague of mine keeps the statistics on you know how many how many more developers are now DLI certified and all that so we actually do believe that this this part of curriculum will actually now become pervasive and I'm seeing this happening you know it's not even you know PhD students anymore you know **even my kids now want to go and buy you know a Jetson nano** and and use it to program and write code to figure out if you know there was a cat thank you figure out if you know if a cat was you know our backyard in the night and did something right so I think the overall I think we are pushing this knowledge and **pushing this learning to earlier and earlier stages of the education cycle** which to me you know you are saying about learning from interns I'm pretty sure I'm going to **learn from my kids** by the time you know they are in college more than what I can teach them I thought they already knew yeah they know it all all right and

Stokes: you know raspberry PI or you know a simpler version of that I think on this it's like it depends it's like okay there's some there's some hardware challenges and issues that are we still need to get better at certain things but there's actually we found that these **things on the edge can be simpler** right and the fact that we've been able to actually maybe take a step back it becomes more of **an architecture and an application set of requirements that really drive how that distribution of that hardware** and in some cases thing you know it's back to the future right we've gone through these different phases where it's like back to client-server all these different aspects of how we get there so right now I think we're in this mode of having to even **simplify on the hardware side with these devices to just get the simplest form of the data and the content you've got a you've got a comms and a network challenge that's maybe the thing that you don't need to simplify but mesh networks are coming mesh networks with SATCOM** are coming that changes all this paradigm of you know the aspect of the edge right and then there have been things around called **FPGA is associated with CPU no offense to my GPU buddies over here but they've been around for a long time that can do a lot of things we've just not taken advantage** necessarily right so I think that there's parts of hardware that actually were way ahead and still haven't even been exploited to their fullest years and years later but there's still some **gaps on communications** really the big data piece performance that's yeah and

Devaki: to that part I think a lot of the production ization right now for people to really productionize the data we're productionize the output of they are **fairly savvy in terms of what hardware** plays a role in that so some of our customers specifically in the humanitarian and disaster response side for example we do wild fire detection and real time off of drones they said if it doesn't run on a p 100 at 15 frames per second they can't really productionize it so right now a lot of it is kind of also it kind of drove our research **the hardware constraints as well as obviously the customer specs drove our constraints in terms of how we build AI models** and that content potentially will continue to happen as Hardware plays a large role in kind of defining productions Asian requirements especially on part of the AI

Question: so the next one is what's the **most meaningful and satisfying project** you've been involved in applying data science you know both the principles and the practices and why did you find it fulfilling

Devaki: well I'll start given most of my talk was on natural disaster response so for us being a California company obviously wildfires as top of mind I mean if you're if you're in San Francisco or Oakland you know just last week you can you could see the smoke coming from the fires um so for us it's being able **to use AI to solve some of the most fundamental challenges that humanity will face** going forward it's not I mean obviously we want to both as a company do good but also do well for my new financial aspectus so it feels really good to have the ability to apply AI to solve these problems also **attracts great talent right people want to know that the work that they're doing actually solves important problems** and so for us it's it's a win-win not only we're doing good we attract and retain incredible talent but also it's kind of where the future is going

Response (Miland?): there's been a number of them but the one that is most gratifying is we used we used machine learning to understand how people move in a city and what modalities they use and then we actually married that to the existing public transit infrastructure that they had and we found out that you know there was this one city where you know people just kept dragging the bus route the point we're in to end you know into end time to go from the one end of the route to another was 75 minutes frequency was once every two hours and the city was losing money the state was losing money and the federal money they were putting in you know it was just a really terrible situation and because service was low a utilization was even lower and it was just a death spiral and just based on data we actually completely reoptimize their transit and now we see that it's very successful the occupancy the the ridership has gone up 37 percent in two and journey time has gone down by 40 percent and fuel utilization has also reduced by 38 percent so that's I mean I worked on many more complex projects but this is **the most gratifying because when you see what data science can do and change an outcome or a city which is now not having to cut the subsidized bus service** right which which actually affects their neediest people the most

Response (Anand?): I'm gonna cheat a bit in answer a slightly different question so I have what I call the envy list of projects I wish I was involved in and I haven't been last week there is a story from **NPR about African elephants** 7 elephants did you know one read it okay I'm gonna tell you this right it's very quickly it's storytelling yeah right it's they were trying to do a census of African anyhow their perception as it is going down but they wanted to be sure statistics wise I mean they started the most obvious way even I mean all the **they used aerial photography and try to count the elephant's didn't work because elephants like to hide** they live in the forest and it is hard to be sure that they got all the elephants in so what they did was to plant **microphones** all over the thing and try to get a network and use statistics to try and extrapolate for the rest of the jungle and **they heard something that they didn't expect to hear gunshots** right so they retrain the model to detect and shots and trigger a proactive activation system and they were able to prove that that actually cut the decline and they turn the curve around so I hope to say before I retire that I was involved in on such project

2:00

Question: next one is what's your opinion what in your opinion is the **biggest challenge we face** in the world today which can be answered but if you had to pick one data science ai and some trend and just force of hints it could be quantum computing blockchain I had to put it up there it's computing 5g so you know my opinion about it only pick one of the technology you can pick although so I mean if you can talk to all of them great yeah

Stokes: I like big challenges so apparently it's the biggest challenge in the world that's a pretty big challenge I you know to me I think where my heart on this is the **convergence of environmental impact of population expansion** right and I think our you know our world with the combination of all of these technologies in a lot of what we've been talking about here today being able to have model integration data that is high velocity sharing and collaborating across those models such that AI is more learning on the results of the first phase models to the point that we make discoveries so that we can actually mitigate the problem right we can get predictive and prescriptive on that value chain Phillippe used a classic on that value chain so to me I think there definitely is a finite set of resources I think we'd all agree on that we'd also agree that the population is booming out of our control right and being able to help solve those challenges I think I think the pieces are there to be able to do that the observable engines and tools and sensors are there to do that **we're just not at the point to put it all together** and again I think it's it's all these things that eventually come together I think we live in a day in a time that's amazing though you think about the fact that **we can do almost lab coat level research and use frameworks and modern computing techniques and access the data that comes from the remote places in the world and even the universe and turn that into a usable product for many people in very short period of time that's just transformational** so I think when these things start to come together **in model integration** and other ways to kind of have the community be able to share that across standards we'll will be able to start to tackle that that challenge that's my take

Response: there are so many so many right so one one would be hard to pick but I if there is one that actually has pervasive impact on many fields I would say **with the help of these technologies and their convergence if we can scale if we can scale expertise** right we have very few really good doctors in the world but we have 7 billion people needing that we have very few really good teachers in the world but we have you know the entire student community needing that we have a lot of food that gets wasted and then a lot of people that don't get food so that's more of a supply demand and so I think a ways of meeting demand you know to the population that we are going to with the resources that we have without you know without destroying destroying our ecosystem that's where I think just like a sustainability of human existence itself fundamentals that's a great answer

Response: I just wanted to say concur it's a great answer

Questions: any suggestions for introducing the field of geospatial data scientists we know data scientist exist we know geospatial analyst exist what about geospatial data scientists explicitly firstly is it necessary and then you know it's both to recognize the trend as well as you know it's to increase the number of practitioners any comments on that

Stokes: to be honest with you we started saying that years ago we actually when data science came on it we said well we're geospatial data the site when the data science trend started coming **geo is hard by the way it adds a level of complexity that's way harder than people realize** and when you go through the generic course of data science you're not even close yet right and so I think it is a **field that needs specific components of education and training** there's nuance and beyond the nuance then there's big rocks behind that as an iceberg behind geospatial data to actually get the true value out of it and so to me that this is this is a big I'm passionate about this one it's **really important that there is an education and field of study around data science** that's focused on geospatial what we'd all benefit from it

Devaki: so I think **geospatial analysts are domain experts and I think that data scientists tend to sit more horizontal** across it I think the way we can marry the two betters because you know you don't see data scientists necessarily learning geospatial technology but **you expect you spatial technologists to learn data science** so how do you make sure that they could do both while still remaining domain experts so I think there needs to be kind of a meeting in the middle of the two which is **you have geospatial scientists that are armed tools that data scientists builds that are geospatial specific so they can they could meet them halfway** right I want to make sure that domain expertise remains still domain expertise without having to kind of water that down because you wanted them to do so many things I don't know exactly how that works or how that would be implemented in real world but I think the value of having geospatial scientists is because you know they're really good at what they do and so you can't really water that down by asking them to do all these other things so you're **saying it's a team of two and not one person who expands into some of the newer techniques of data science** yeah I think that or potentially **data science tools become more flexible for people that have domain expertise** so even bringing down kind of the research element sub data science and something that's more tangible for geospatial analysts so who's feeding into whom I'm not being prescriptive about how it should be done and more okay OK opining on it

Response: yeah two parts to the answer oneness yes like to your point you don't want to lose one expertise at the expense of another and wait till the geospatial guy who also learns tensorflow find out like tensorflow pays like three times more money so he's going to switch to optimizing website clicks and you know where does it leave us right the second part of the answer is that is slightly flippant a second but other answers I I'm not saying this is widespread but some universities already do that I was talking to some placement people in Waterloo they did run a co-op program the co-op coordinate it would happen to be from the Department of Environmental Sciences so they have a program in what they call **geomatics that's a cross-disciplinary thing so they're allowed to take courses from three different departments engineering mathematics which is the computer science as well and environmental sciences** so I think the the they have very little mandatory content in their course it's pretty much what you what makes you happy to get them and just get the right number of credits use whatever combination I haven't really explored that much in the sense like I don't know what people do after they graduate and stuff that's something for me I want to find out and I think that would be good model if you I don't know how many universities do something comparable but that's for me it is a relatively recent discovery okay yep

we'll open up for questions for the panel anyone with the question

Response: may I answer the last question? I am an academic there are at least 50 programs like that here in the United States and from one of them and I could be I would say **the most successful ones are over there in Europe because they have less they are less afraid of reaching out into information science** so there are lots of geo programs that are not affected yes science geo-informatics programs over there yeah thank you

Response: the only thing I there's this extra piece kind of on in addition to I told I totally agree that no one we were talking about at our table no one could know everything right so if we that expectation is ridiculous and ever there people have to have their area of expertise but there are and it's one thing that it's kind of interesting we haven't really talked about much but you know there's machine learning and AI and for the most part we're talking about taking kind of traditional ml and AI and applying it to spatial data but there are inherently genetically spatial methods that are exclusively a that take geography into account I mean it's started by saying that **things that are close together more related than things that are farther apart it like invalidates lots of Statistics** right off the bat and we're not thinking about spatial in the way we develop machine learning algorithms in the way because those data science programs themselves aren't spatial so it's not so much about GIS you know spatial analysts becoming data scientists but it's like **we need some spatial injected into some of those computer science statistics data science programs** **also** alright it's not a question I'm also quite passionate about this

Response: yeah I agree with that I think you make a good point who do **we want the geospatial analyst to stay domain-specific and there are geospatial analysts** that then are sub domain specific within that we're talking about more of a much more technical developer type role that takes on a data science type career it's a different persona in the grand scheme than a then a been than the geospatial analyst in my opinion so it's just my observation on that one yeah

Response: I just add I to **that person who really helped us develop the cuda spatial library actually started out as a GIS expert which is PhD in GIS is a professor at City College New York** but then over a period of times these things were not as dynamically interesting in that domain he trained himself on CUDA eight years ago ten years ago and he actually had been working on GPU acceleration of some of the special operations and when we decided we really needed this in the metropolis platform I reached out to him you know he just came in and he helped us so we actually cheated we benefited from somebody who was a GIS expert also over his career became and now the benefit of that is not everybody has to go through that and experts can now just use but that kind of tool tooling will need to happen and that can only happen when the first phase of you know let's all do click through gets over and people start looking at other problems to be solved

Question: yeah the reason why I put up this quest and was during user conference that we had earlier this year there was a user who came and asked something along the lines of GIS analyst was created or the past world were 10 15 20 years and there's a pay scale that's associated Python developers are there at the pay scale data scientists are coming at a different pay scale I want to see my career grow I went and spoke to them I said I learned Python I'm a GIS analyst they said you can either choose GIS analyst pay scale or the Python developer space scale what they were asking for is **I provide much more value than either of these in isolation** I provide the value of a data scientist who can give you the answers the solution the outcomes but just because I predated this this trend would I have to be left behind so that's actually the reason why I'm bringing this up that we not only have an obligation to make everything correct and perfect and so in terms of the boundaries that we build but also in terms of **making careers out of people who can deliver the outcomes** so that that's the origin I mean this I'm just stating this as this little beyond the conversation that we have so just consider that yeah question so

Response: I was just gonna add a comment that you know **it's a hopeful sign in any field when we get enough depths that we actually have to start to specialize** so the same thing happened to medicine you can't be a specialist in everything it's a good thing that we're at this crossroads but I wonder if this is the right question really I would want I mean so I'm also a professor at UCLA and **I teach my data scientists how to work on interdisciplinary teams** so I want them to be respectful and understanding of the scope that storytellers bring and what geospatial experts bring and what policymakers bring rather than make anyone of them feel that they have to be all things to all people **as long as they're conversant with each other** and able to come up with the whole thing it's very rare that we need that one superstar you know there's there was a Leonardo da Vinci but there are a lot of other inventors after

Audience question: I think there was a paper published recently or not recently but um talking about how **autonomous vehicle algorithms were manipulated by stickers pasted on stop signs** right so I'm just wondering for geospatial data is there a certain part of the industry that has a better position to **actively monitor and prevent bad actors or bad data** into seeping into the models and changing them

Response: it's unfortunate that we all have to bear the burden of the cost associated with combating all these different cyber challenges you know it drives the cost up in the grand scheme for all of us and with your requisite model or data type in your industry your sector you have to focus on what those vulnerabilities are we suffer it with satellite imagery in a lot of different ways it's not only the data but we've got gazillion-dollar machines in space right and that's a whole other thing related to ensuring the integrity from a picture in space to the ground to then all the next things that happen right we we have to take care of that all ourselves so unfortunately **it is on the onus of the operator of whatever that model content element and then you know the the autonomous car systems and the providers of those are gonna have to deal with that themselves or wrangle it out between source content being having a pedigree and that company has to certify it** and insure it so it's at each one of the source levels if you will yeah. That's a mic drop

Audience Question: one I have any of you worked on any **projects with interconnected infrastructure** waste water drinking water electricity gas natural gas petroleum any project any projects like that

Jim Stokes: yeah I have so networks you know let's say electric power infrastructure networks so what would be your question

Question: then so interconnected yes of this resource so for example if you have a wastewater treatment plant that's using SCADA and the internet goes down or there's no cell service or if you're looking at power generation people think that if the electricity goes off that's a bad thing and it is but they don't realize that if the natural gas is empowering the electric generation station there's no there's no electricity so now if your gas is very important if there's no water there's no electrical generation either so any projects where you've worked with any local agencies where you're trying to **understand the interconnectivity of the infrastructure** and especially for emergency response and recovery is where it comes into play most

Stokes: so that's that's a big concern across a lot of different angles the first thing that comes to my mind or my colleagues from ESRI and fundamental to GIS right and especially you know putting together multiple networks and being able to understand where a portion of that network fail what's the downstream effects and they fundamentally have **technology that can cascade model effects** and all those things built into that just it's it's a lot of work to build the network but once it's there it can be smart and intelligent and you can do interdependencies but I'll defer to ESRI to respond you answered I mean

Jay: I think you're familiar that we have the utility networks for we're also going through a major overhaul and redesign and we've shipped some new parts of it which is everything a service based as opposed to you have to work directly with the database in a client-server model and the idea there is you can go disconnect it with one of these and still make it effective that's where I think the first generation of going disconnected happened between let's say the servers and the actual devices the field devices so it was a person going disconnected who didn't have the you know data availability because of the network right I meant the wireless network so big they been disconnected I think we have seen **the next-generation of that happening which is between machines going disconnected which is what the edge we were discussing earlier that you still have a system of truth that is in the center but then you're able to disseminate that to all the edges** so let's say in this case where the internet goes down or power goes down in different places everyone has the last known truth or **the last known system of record to still sort of like have emergency operations run** so it's not just a device that's going to have that but even a utility network and then the interconnectivity is like what you said you know the gas lines and the electric lines and everything else want layered one on top of each other I mean right now we are going to a phase where yes it's still individual they're not heavily interconnected but you can see that happening more and more so

Thank You panel great discussions we can do a check next year okay

So we've worked very hard today and so the reward for that is a reception so please come over to Stein's beer garden it's right off of Castro in downtown Mountain View and join folks there and then we'll be back here 8:30 again check in tomorrow morning 9 o'clock and we'll start talking about applications and next steps so thank you all for today and we'll see you tomorrow