Deloitte TECHTalks | EPISODE 17 | Sustainable AI

With Dr. Johannes Trüby, Partner, Deloitte France, Leader of Energy Markets and Modeling Team of Deloitte Economic Advisory and Geoff Tuff, Principal, Deloitte Consulting LLP, Sustainability and Climate Leader for Energy, Resources & Industrials (ERI), and US Hydrogen Practice Leader

Raquel Buscaino: Welcome to Deloitte TECHTalks. I'm your host, Raquel Buscaino, and I lead Deloitte's Novel and Exponential Technologies team where we sense and make sense of emerging tech.

On today's episode, we're going to be talking about the intersection of energy, consumption, sustainability, and AI. It's an extremely timely conversation, because, as the usage of AI increases and everything from training and running AI models to cooling data centers - the question of how we can do so sustainably is top of mind.

I have two incredible guests on the podcast today to help us get to the bottom of it. I'd like to welcome Johannes Trüby, a partner in Deloitte France, who is the Energy and Climate Economics leader, as well as Geoff Tuff, a principal at Deloitte Consulting LLP, who is a Sustainability Leader for Energy and Industrials.

Johannes, Geoff, welcome to the podcast. It is really so great to have you both here.

Geoff Tuff: Great to be here, thanks for having us.

Johannes Trüby: Great to be here. Thank you, Raquel.

Raquel Buscaino: Amazing. Well, Johannes, Geoff, both of you are clearly passionate about sustainability, and have been working in your respective fields for what I would say, is quite some time. Why is now a good time for us to be talking about the relationship between energy, consumption, sustainability, and AI? And Geoff we'll start with you.

Geoff Tuff: We're dealing with two of the most powerful forces behind or causes of transformation that we've seen in decades and decades in any sort of company out there. So, you've got sustainability. You've got digitalization. We're discovering that those two topics are increasingly intertwined and creating an even more powerful force for companies and all sorts of organizations to consider transforming themselves.

Digital tools like online conferencing messaging apps, all of those have significantly reduced the need for physical travel and it's kind of changed the way that many of us have worked, and that has led to lower carbon emissions. Similarly, Generative AI is completely refreshing how we, as a society work, and can offer opportunities to improve how we can operate and fight against climate change.

Those two factors working together are really going to be very powerful in determining what the future of sustainability looks for us. And as we'll talk about in this podcast there are some things that are really leading to good outcomes as rates of decarbonization, and there are things that bring challenges as well.

Johannes Trüby: Yeah, totally agree. We need to be cognizant of the fact that digital infrastructure, especially data centers, are based on an underlying physical infrastructure of be it networks, storage, cooling, equipment, and service, and all of those consume substantial amounts of electricity. And by

now, most of us have probably seen headlines warning us of the energy consumption of AI, and how that could threaten global electricity supply.

And what you read in the press is often either very alarming, or, on the contrary, might be toning down the risks. But in the end, no matter how you look at the impact, it's clear that first, the AI boom should be studied, and second, that it is crucial to commit to sustainable AI and data center practices. This could mean not only leveraging AI for green initiatives but also ensuring that the AI development in itself is really sustainable.

Raquel Buscaino: Both of you are really speaking to the fact that it is important that we have this conversation now. I'm hearing in the market two phrases that come up quite a bit. One is “green AI”, and the other is “AI for green”, and in this case green refers to sustainability. I know that those terms are different. So what do each of these concepts mean?

Geoff Tuff: I'll jump in on that one Raquel. I just want to say something that may be an assumption that maybe on people's minds. But we're past the phase where the world is wondering whether sustainability and decarbonization is going to happen, whether we will go through an energy transition. It's happening. We don't know everything about it. We don't know where we're going to land. We don't know the rate of change. We don't know which energies are going to end up being in the mix later on, but it is happening. The same is true for GenAI. We're not putting either one of those genies back in the box. And so what we now need to do is to challenge ourselves to think about “How can we deal with both of these forces responsibly.”

So let's start with Green AI. That refers to developing and using AI technologies and all the underlying infrastructure that Johannes just talked about in an energy-efficient manner. So, we got to optimize our resource use. We got to reduce emissions. But we have to acknowledge that as AI grows it is going to create more pressure on our power system.

Dealing with Green AI is mainly a technological operations procurement matter to making sure that we can access green electricity and commit, as we develop AI to strike a balance between the performance of the models, because we can always increase the performance. Or I'm assuming that we will continue to be able to improve performance exponentially. But as we do that that will challenge our energy systems. We need to balance the performance of those models with energy efficiency.

So, as a society, increasing computing power while remaining relatively flat in terms of energy consumption is historically something we've succeeded in. Whether or not we can continue to do it remains to be seen. But with the massive increase in computing power, this is likely to be a critical challenge that we continue to face.

So the other phrase you mentioned “AI for Green” is separate, sounds the same, but it involves applying AI, the technology, to help drive sustainability. And this can be everything from optimizing supply chains to reducing waste, helping deliver low carbon-energy, you know, there's a lot of human beings who have been thinking about these issues for some time. But as we use AI to solve somewhat have felt up until this point, intractable challenges to actually get us through the energy transition, we're going to make a

lot more progress. And so it's an interesting two sides of the coin that we're facing here because AI can be a great force for good in sustainability. But it's also going to create some challenges.

Raquel Buscaino: I appreciate the way you teed it up, whereas there are assumptions and givens we're operating in, the energy transition is happening. Generative AI is happening. So the question is, what are we going to do about it? And I also love the way you teed it up, too, with how "green AI" is about how we make usage of AI more sustainable, and "AI for green" is about the wide range of different ways we can use AI to further sustainable goals.

With that being said, I'd love to dive into Green AI first. So, Johannes, maybe I'll put this one to you. But what's top of mind for clients when it comes to thinking about how to adopt AI sustainably?

Johannes Trüby: I think in the very first place, it's important that companies understand their infrastructure and how it functions, how it uses energy. And for this, monitoring the energy consumption of a data center is really important. And for example, what they should do is they should ask themselves questions like, how much energy do I use? For which purposes? How much heat can I, for example, recycle? And with this data in mind, the operators can actually start assessing where to make savings, where they can reduce the energy consumptions, and they can start applying the best practices. For this kind of monitoring the most widely used metric is called Power Usage Effectiveness, very often abbreviated as PUE. And this metric measures the efficiency of an IT infrastructure; for example, how much the servers and the IT equipment consumes in terms of energy. And it compares it to the overall infrastructure consumption, which also includes cooling, lighting, and so on. Then once the facility is well understood, the clients may actually be focused more on green energy sourcing and on ideas about how to operate the data center more efficiently.

This first point, about the green energy sourcing tends to be well understood by many of the tech leaders which are already amongst the biggest contractors for green energy supply. And one typical way of contracting green energy is through Power Purchase Agreements and through Energy Attribute Certificates. These Energy Attribute Certificates are also called Guarantees of Origin or Renewable Energy Certificates depending on where you are in the world. Now, while it communicates on their clean energy procurement, we still see that we're far away from having data centers run 24/7 on low-carbon energy and achieving that goal of really low carbon data centers will require innovative solutions that come on top of what we're seeing with wind and solar electricity today and will most likely also result in an awful lot of batteries being used.

We're also seeing that some data centers are already running nuclear power while some others are actively looking at geothermal energy to provide stable power for data center operations. There is clearly still a lot of room for improvement, which not only involves being able to monitor how much clean energy is bought on a yearly basis, but also how you can use renewable energy on an hourly basis when the wind is blowing, or when the sun is shining, or when that's not the case .

Raquel Buscaino: The two things I heard there was: clients are usually focused on (1) green energy sourcing and (2) efficient operations. Which kind of to me means it's how you source the energy, and then it's how you use it.

Geoff Tuff: I did want to just reflect for the audience who may not be in the conversations with some of these tech players and some of the power developers, all the forces that Johannes just talked about are directly impacting really, really important capital allocation decisions. So that you know, the many of the big tech players are actually deciding where to put their data centers based on how they're going to get access to some of the power sources that Johannes discussed. These are millions and hundreds of millions of dollar decisions that are being made because we're assuming some constraints that we're going to have to work around. And so the more the more we can bring all of these solutions to the table, and give optionality to these companies the better off we are going to be.

Raquel Buscaino: Geoff, that's great because you mentioned the economics of this. Johannes, I'm wondering can you give us some frame of reference, as it relates to the financial costs, the environmental impacts and just contextualizing how big of an impact is this opportunity space we're talking about.

Johannes Trüby: Look, data centers, are actually very energy intensive beasts. So in fact, the electricity purchase represents between 10 and 20% of the total cost of ownership of a data center. And this data is actually taken from a white paper from Neil Rasmussen on the total cost of ownership for Data Center and Network room infrastructure. That in itself justifies all efforts to become as energy efficient as possible.

Still, energy consumption of data centers is currently on the rise, and has reached, according to the IEA, 380 terawatt hours of electricity consumption globally in 2023. And if you compare this with the energy statistics, for example, also of the IEA, then you will find that these 380 terawatt hours are about equivalent to the consumption of a country like Saudi Arabia. If you relate this also in terms of the IEA's global energy statistics with global electricity consumption, then you will find that last year's electricity consumption of data centers amounted to about 1.4% of global electricity consumption. Just to put this into perspective, what this means is 380 terawatt hours of electricity consumption corresponds to the output of 15 typical nuclear power stations. That's quite significant.

Geoff Tuff: That's a lot.

Johannes Trüby: The problem with emissions actually stems from the fact that most of these data centers still do not operate fully on low carbon energy based on my team's assessment in Deloitte France, we estimate that data center consumption of electricity could range between 1,700 and 3,500 terawatt hours by 2050, with almost 1,000 terawatt hours already reached by 2030 depending on how quickly the AI adoption takes up.

Raquel Buscaino: I mean, just recapping those numbers 2023: 380 terawatt hours. 2030: we expect 1,000. That's pretty wild. And then by 2050, I mean, what's 2050?

Johannes Trüby: So 2050 could range between 1,700 and 3,500 terawatt hours of electricity consumption. That really depends on how quickly AI takes up.

Raquel Buscaino: Wow!

Geoff Tuff: Yeah. And that is based on some assumptions around adoption and the nature of the models and what have you. And Johannes your experience may be different from mine in this, but I haven't yet seen at scale the rational conversation about what do we actually need to accomplish with AI? Do we need to assume that it has to grow, and data centers need to grow as exponentially as we're predicting right now because we're just assuming that AI will take off in all the ways that that everyone is talking about these days. And so hopefully, you know, if and as the world starts to understand the impact of that growth and what some of the constraints it may cause in other sectors. Hopefully, we can start to have that conversation at scale.

Raquel Buscaino: I think one of the ways that we can communicate the impact of this is...Terawatt hours don't mean a whole lot to me. But in terms of percentage of world's expected electricity demand. Can you take these numbers from Terawatt hours and convert them into, into total electricity demand?

Johannes Trüby: Absolutely. If we relate on my team's assessment for 2050, let's say 3,500 terawatt hours of electricity consumption to total consumption by 2050 then data centers would represent about 6% of the world's electricity consumption by 2050. Notwithstanding there are also loads of other electricity consuming applications that are going to increase. There is also steel making, the electrification of the transport sector, and so on.

Given the huge potential of electricity demand growth of AI and data centers, we need to make sure that the electricity that these data centers consume is low carbon, and as such, in order to stay on a trajectory consistent with the goals of the Paris Agreement, the IEA has shown, and we take this in our analysis that has been conducted by my team in Deloitte France, and showcase that the emissions from data center electricity consumption would need to peak by the 2030s and then go into decline.

Raquel Buscaino: So often when we think about the problems. And just as we have right now, we've talked about them at that global scale. But I think most of the decisions around sustainability happen at that national or even local business level. So maybe we can switch gears and talk a little bit more about that. But, Geoff, what are some of the challenges that happen at the local level that we should be thinking about?

Geoff Tuff: Yeah, well, so you're exactly right. I mean, let's think about how our power sector runs today. You know, we do not have global utilities, we have for the most part highly diverse and distributed utilities. Sometimes, you know, here in the U.S. we've got them covering, sometimes very small regions, and certain parts of certain states. And decisions around what to build happen at the local level. There are national considerations, especially for some countries that are, that are faced with, you know, questions around energy security and what they're using their entire power sector for. And so let me start actually at the national level, and I'll go to the local level. But Ireland is a good example of a country that does need to think about it at the national level. Today, as I understand it, about 20% of the country's power consumption is going into data centers. And that's a number that could rise to upwards of 30% by 2026 in just a couple of years. So this is my understanding from an article from the DCD or the Data Center Dynamics on Irish data centers consumption of electricity.

That, as you might imagine has caused important energy security concerns and considerations and conversations which has prompted the Irish Government to impose a, essentially, a moratorium on data

centers around Dublin until 2028. That's a, that's a big deal. But you know, 30% of your power going into a single sector is also a big deal. So that the heavy consumption of the data centers also contributes to things like emissions and in a country like Ireland that still relies largely on gas fired power plants. That's going to be the case until they are able to decarbonize their power sector, which is exactly the point that Johannes made before.

If we think, then at the local level, I'll pick an example of a region here in the U.S.--Northern Virginia. So that's one of, if not the largest data center market in the U.S. Current power consumption actually exceeds what was planned for just five years ago, sometimes by as much as 10 to 20% so, and this data comes from the IEA, if we think about from a planning perspective, and what the bill timelines, and reality about the capital needed to build these things, if you're blowing through your targets every 3, 4, 5 years, then there's no way you can plan for that, and there's no way that that you're ever going to be able to build to the capacity that you assume you're going to need. So from an investment perspective, in terms of that just great capacity and power generation, and how much we need to be building that's really a very big deal. This has created tension in lots of other places, lots of both regions and national power authorities are either essentially saying, we need to stop building data centers, or we need to at least pause for some period of time.

And it happens at the business level as well. So if we think about what the increasing power consumption has looked like. For some businesses they hadn't anticipated how much they were going to need in terms of power, and how much of that was going to have to come from non- decarbonized sources, and that's led them to blow through some of their climate plans and their commitments, all because they just have not done a great job of predicting. And so the last thing I'll say on this, is just about some of the tech leaders, some of them have actually seen their emissions jump primarily because of AI taking off. That's a huge, huge deal. And nothing that they had anticipated back when they were setting their targets around Net Zero.

Raquel Buscaino: Given the stakes here, what do the legislative frameworks to help address these challenges look like? Because in a market that is constantly changing, Geoff, to all the points that you just mentioned. It's hard to predict where this is heading. How do you even go about legislating or providing guidelines to help guide businesses, local governments, and national governments in the right direction? Johannes, I'd love to hear your thoughts.

Johannes Trüby: As much as technology is evolving very quickly. The policies are also evolving quickly. We already see some important policies implemented in legislative frameworks. For example, we find policies about mandating increased transparency about data center consumption. That's well underway, for example, in the European Union with the Energy Efficiency Directives. And we also find that in China, with the 3 Year Action Plan on new data centers. I also believe that this is likely to be followed soon in the U.S. and elsewhere in Asia, where data center electricity consumption is really growing rapidly. And only when a certain level of transparency is reached we can have an appropriate understanding of how we can tackle the issue of rising electricity consumption.

Some countries are already mandating concrete efficiency figures for new and existing data centers, and while these measures are far from perfect, my team estimates that these measures alone could save more than 400 terawatt hours per year by 2050.

Another important step is for policymakers and industry players to collaborate, to help encourage increased sustainability in the development of AI. This could mean developing new projects in regions where it really makes sense, and operating them in a manner that is consistent with the power system needs.

Raquel Buscaino: What could that look like in practice Johannes?

Johannes Trüby: So this could include, for example, dynamic electricity, retail prices, higher flexibility in the operation of data centers, or other innovative measures. Let me just give you an example, a cool example. Consider the possibility of achieving spatial temporal load shifting. Meaning, starting to shift computing jobs not only in time but also in space, in places where they can use direct wind and solar power. Just imagine the wind is fading away, say in the Netherlands. What you would do is to move your computing jobs from your data centers in Amsterdam to data centers, let's say in Barcelona, where the sun is baking at that very moment. Such proposed measures would not only be beneficial from a climate perspective, but could also contribute to lower the operators grid electricity bills by more than 5% based on our model.

Raquel Buscaino: Well, that's awesome. The themes that I'm hearing with what you just said. The first is transparency. One of the quotes that I like says “you can't manage what you don't measure.” So how do we make sure that we're measuring it and sharing that across the bounds? Second thing I heard was agility. You know your example with spatial temporal load shifting, right? How do you stay agile? And the 3rd is innovation. We need innovation in the sector. I know it's been a conversation throughout this podcast but that's critical here, too.

As we think about wrapping up here one of the things that we mentioned early on that we haven't touched on yet is AI for Green right. I would love to just do the quickest of hits with both of you. Geoff, will actually start with you, what's most exciting for you when it comes to AI for Green?

Geoff Tuff: I mean honestly every everything that it can accelerate anywhere where we know that we need to lower our carbon footprint or change the practices that the world uses today that contribute to higher than necessary emissions. So I mean, AI can be applied to reducing forest fires because it can be used in vegetation management. That's a great application for it. It can improve agricultural practices. It can help make sure that we're wasting less in our food value chain. I mean just about anything that that we, as people are struggling with today to try to get to our Net Zero goals, AI is going to be helpful for. It really is an interesting dilemma. It is a great force for good if we apply it in the right way.

But if we apply it unchecked, and if we just assume that we need to exponentially grow computing power everywhere all at once, then it will end up creating more problems in it than it helps solve. And so you know I leave it to people much smarter like me, like Johannes, to come up with the practical solutions of what exactly do we do to shift where we're doing our computing or to match the right computing power to the right applications. We're on a path right now where I'm not seeing those conversations happening at scale. But they're happening in pockets, and they're certainly happening at the local levels like we talked about before. Hopefully, in forums like this, we can get more people talking about it, more people understanding both the pros and the cons of unchecked AI growth.

Raquel Buscaino: Well, said, Johannes?

Johannes Trüby: Maybe I can just add a bit of an example to make this a little bit more tangible on how AI can contribute to making operations greener. You can, for example, imagine that you got a set of cameras that looks at a highway and uses visual recognition of cars and license plates to then learn and understand when a specific electric vehicle would charge. And that depends, for example, on, you know, like where the car is coming from, where the car is going, what type of car it is, what the range of the car is. And if you do that continuously on a highway, on many highways, your electricity system will get a really good understanding which charge points will be used where, where you need to expand them, but also when you need to dispatch certain power stations in order to optimally serve the charging of those electric vehicles.

Raquel Buscaino: So interesting because there's so many things that were prohibitive before that now with AI, we can start to actually do and make progress on. So I love the way that you've tee that example up because it's really flipping the script on what is possible with this.

Johannes, Geoff, and thank you both for such insightful discussion. I mean, we've covered a ton of ground on today's podcast from the energy demands of AI to sustainable practices, regulatory frameworks. So thank you both so much.

And to all our Tech Savvy listeners, if you enjoyed this episode, please share and subscribe. And if you'd like to learn more about sustainable AI. You can follow myself, Johannes and Geoff, to stay up to date. Our socials are listed in the episode description. Thanks for tuning in, and I'll see you on our next episode. Until then, stay savvy.