Deloitte TECHTalks | Season VIII | EPISODE 21 | The Evolving Role of the Chief Data Officer
With Ashish Verma, US Chief Data and Analytics Officer, Deloitte Consulting LLP

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.

In the past we've done several episodes on the topics of AI and Generative AI, but what we haven't done is taken a deep dive into what underlies these trends. And that is the world of data.

What does modern data look like? How are data strategies different or the same across industries? And how is the role of the CDO or the Chief Data Officer changing?

I'm delighted to be joined by Ashish Verma, Deloitte's US Chief Data and Analytics and Principal with Deloitte Consulting LLP, to discuss all that and more on today's episode. Ashish, welcome to the podcast, it's so great to have you here.

Ashish Verma: Yeah, thank you for having me. Sure we'll have an engaging dialogue.

Raquel Buscaino: Definitely. So, Ashish, you work with chief data officers all day long. Is that right?

Ashish Verma: Yes.That's correct.

Raquel Buscaino: In today's AI and Gen AI driven world, I think data has really become the critical asset for all sectors. How is the role of the Chief Data Officer changing in this changing tech climate we have?

Ashish Verma: Data is the essence of every algorithm, every machine learning model, whether it's Gen AI or any of the next generation of innovations, it's evident that data is driving this transformation. And amidst this whole boom of AI, it's pretty evident that the role of the CDO, which was historically, if you looked at sort of the way it evolved, it was about governance, making sure that the organization, as a whole, was in-line with right compliance, where the title was really not focused on the A of the CDAOs. So, if you start to observe what's transpiring in most organizations, the positioning of the office is evolved. There's an increased amount of focus on the "A", and you can think of the "A" as Analytics, and really, what has transpired is the CDAO is beginning to have a voice at least in two ways, right? Being part of the transformation journey and having a seat at the table as the Chief Data and Analytics Officer.

And what I think, in my opinion, is going to transpire as a trend- this is not limited to just AI or agents or LLMS [Large Language Models]- it's the suite of applications that are being built on top of this infrastructure, making sure that all the data and we'll get to what "all" means, all of the data that's needed to support these efforts is within the purview. So, think of them as becoming, you know, sort of having evolved from governance and compliance to architects of these transformations, as well as you know, having a seat at the table of these programs to pretty much make sure that any data that needs to drive this innovation, they understand the consequences, the complexities to make sure that when you're trying to extract value from this data set, that the data is there, and it's used in a responsible fashion.

Raquel Buscaino: That's such a helpful frame of mind, because there's a big difference between just protecting the business and transforming it and being the one responsible for that evolution as well. So, definitely a changing role for the Chief Data and Analytics Officer. But what about the data itself? So, you mentioned all the data, what does it mean to have AI-ready data? How does this translate in terms of a data strategy? Or what does a path to modernization even look like here?

Ashish Verma: So if you think of data as the fuel for AI, like most people understand that you need to source it, you need to cleanse it, you need to standardize it, and basically make sure that any algorithm that's written on the back of this data set is responsible because the outcome pretty much derived from data means that data needs to meet a certain level of standard or threshold. This is where a data strategy is very important. So, if you look at how organizations did data strategy, a good amount of data strategy of the past was: What is the internal data that I need to do in a certain domain, or a certain application and the fact that it originated within your four walls. The fundamental difference now is for the things that you're attempting to do with AI/Gen AI agents, it is pretty much making you look at, not just your own first party data, but data that shows up within your four-walls as a result of your interaction with your business partners, which we (within Deloitte) often call second party data.

But also think about the data that you buy, right? We, as Deloitte, buy a significant amount of 3rd party data and this is true for most organizations. There are data brokers that sell data sets for these purposes. Domain-centric data to fuel these experiments. But last, but not least, there's a category of data set that's called synthetic data that is usually used today to do PII [Personally Identifiable Information], PHI [Protected Health Information] masking, making sure that confidential information can be masked when you use it for the purposes of AI/Gen AI, so net-net, if you look at the evolution of “what data?”, you will also pretty soon realize that you don't have most of the data that you need, and in essence most people, or most CDOs in their roles are having to procure synthetic data to fill these gaps.

Raquel Buscaino: Really interesting. It's not just an increase in the amount of data, but an increase in the number of sources it comes from, and then how you ultimately decide to work with it. You know there's been a lot of talk about the role of agentic AI in this future. Does all of this type of data feed into the role of agents? How does that work in this future? How do the two play together?

Ashish Verma: I'll describe to me what an agent is, and most people think of it the same way. It's an autonomous task that's performed on behalf of a user or a system. So not all data will feed an agent, right? But if you are attempting to automate something that can be done, you know, as a result of giving some instructions on behalf of a user or a platform, it becomes interaction model to basically do something that, you know, you're so used to doing today -- as simple as calendaring your meeting invites, tying your expense reports to, you know, things that you intend to do from the standpoint of reconciliation, so, there's pretty good examples of how and where agents are deployed but think of it as a task being done on behalf of a user or a system.

Raquel Buscaino: It's a really helpful definition because if you think about the next evolution for AI, I think Generative AI was definitely a step in our future and now everyone's talking about Agentic AI and seeing what value it can unlock for them. I'd love to dive into a bit of the challenges and risks. So, you had mentioned earlier some of the complexities of unlocking value from data. For our listeners, would you mind breaking down what some of the challenges, the organizations, and even the CDOs in particular

may need to navigate to achieve their ambitions, given the number of data sources and how much data they need now?

Ashish Verma: Yeah, I think, I think you've correctly phrased it. The volume of data is significant. So, most people don't realize this until they start this experimentation for the volume of data sets that you need. And then you realize that it's not just volume it needs to be correctly labeled, and of high quality, right? But the right set of attributions it needs to be fairly diverse, representing the domain of the problem you're trying to solve. Or else you end up with over-indexing on the outcome of a particular narrative. And this is when people talk about hallucination. So, if you're going to attempt to get to a diverse set of outcomes that represents the problem domain you need to look at, not just volume, but even attribution and the right labeling.

Now, if you look at what organizations have been doing for the better part of the data strategies in the past, you never dealt with sort of the issues of storage because it was limited volume, you never dealt with the issues of, things that happen beyond an ERP package or a CRM suite, because pretty much most data that originated in any domain - - finance, supply chain, talent-- all happened as a result of process centric software instantiating that data set.

Given sort of what is transpiring in this world now, and if you extrapolate sort of what that means, it means that most of the data that you need today for the things that you'll define do not actually happen just within your walls. It happens outside of your walls. The question really becomes, okay, so if this is coming as a firehose from all the usage-generated content that we're so familiar with, it'll become very evident that what we are creating, that multi-variate data set which is what AI and Gen AI is being built for, is actually very diverse from your technologies that you run, the platforms that you run, and the data strategies that you had.

So, two things have transpired. The kind of data has changed and where it's coming from has changed. So really, the data strategy now needs to reflect how do you deal with this aspect and what does that mean for the choices that you need to make.

Raquel Buscaino: I love the point you mentioned, where you don't know how much data you need until you actually start experimenting, because it's almost like you don't know what you don't know yet. And so yeah, you might need a little bit, but in most cases, you really need a lot, and it's not until you start digging a little bit that you'll uncover what you really need and what's really valuable to you, and what sources you can, and need to pull from.

Ashish Verma: That's right.

Raquel Buscaino: So, given that-- challenges and risks-- what are some of the steps that organizations may want to take today in order to help them prepare for an AI and Gen AI future? I know that we've done a lot of work at Deloitte, so we could even start with Deloitte's own journey, but would just love to dive a bit more into some of the use cases and first steps we could take or first steps organizations could take?

Ashish Verma: So, I'll tell you, at Deloitte the biggest thing that we had to start with this -- first and foremost -- know where the data is, what the controls are? Who has access to it? Who's contributing to

it? How have we cleansed it? And have we sort of made sure that the use cases are commensurate with the data that we need? And in reality, when we started to articulate that data strategy on paper, it became fairly evident that the ROI needed to be commensurate with the broader investment strategy, or else nobody had the appetite to sort of do this holistically at the enterprise level. We needed to pretty much make sure that the annotation of the data set for usage, needed to be explicitly clear.

And when we started to sequence the use cases, we realized that we needed to cast and that fairly wide. Our first party data, the 3rd party data brokerage purchases that we do and synthetic data. And in reality, we've never turned off the procurement strategy on that data set, because the minute somebody shows up with a use case, whether it's our own business, or some of my peers that run these experiments, they all show up with sort of a request that forces us to go back to the procurement strategy for another kind of data set that we didn't have, or to enrich something that we need as a result of this experimentation.

So, I would say these are the things that you're going to probably get used to. If you want to unlock value from this data set, you're going to have to build the strategy to procure this data set on a fairly frequent basis. And then you're going to have to make sure that every use case that's tiered up is building sort of the return on the value, not the cost value but the economic value of what you're intending to do in your portfolio with this data set.

Raquel Buscaino: I love the way you said it, because your data procurement strategy isn't a goal. It's a process. It's never going to end.

Ashish Verma: That is correct.

Raquel Buscaino: Let's dive into a couple of industry examples, because I think AI has really taken the world by storm in the last couple years especially, and now people are really wanting to see the true value from it, and specifically for their industry, how it moves the needle. So, can you share any use cases, any examples of data modernization efforts that have really moved the needle for our clients or the industry in general?

Ashish Verma: Sure, I'll give it two lenses, right? So, I'll give it the lens of the reg (regulatory) framework, because if you look at financial services right at the heart of what regulation or reg framework is attempting to do is make sure that anything that they're doing regulation on, you're going to have to bring data to support the fact that whether you meet the mandate or not, whether that's financial liquidity, or, you know, an example of a consumer data being used. So, for any sector that you're dealing with, you'll realize that an aspect of this is regulation which is about you know how we're using the data, is there a human in the loop? Those kinds of things, right?

But let's look at, investment on this topic and let's look at what happens if you look beyond regulation. So, if you look beyond regulations, and think healthcare-life sciences specifically, life sciences. And you look at disease pathology, right? Disease pathology, or western medicine is based on pharmacology. This is the reason when you go to a doctor, they order a bunch of tests for you. What they're attempting to figure out is your biomarkers, to figure out can they diagnose what's gone wrong with your biomarkers for this disease pathology. The model of that is fundamentally changing. If you've heard about AI for protein folding, and if you've heard about the 200 million proteins that have been mapped, you will be

able to sort of decipher what I'm about to go into. They pretty much are now able to identify in the 3D structure of proteins what disease does to it. That means fundamentally disease pathology is not going to be biometrics. It's going to be, the observation of what's turned incorrect in your protein synthesis that your body does to be able to provide disease pathology. That means R&D changes in pharma companies. How I do drug research? How do I do drug development? How do I do clinical trials? How do I submit for clinical trials? How do I manufacturing / supply chain?

So, you can see sort of the impact of something so fundamental that is happening in industries or sectors, which is sort of a consistent story. I can give the same examples in healthcare and consumer, and so on, and so forth. But the narrative is fairly straightforward from the standpoint of what this means.

Raquel Buscaino: I really love the AI protein folding example, because I think one of the interesting things that AI can do in my opinion, is almost upend the scientific process a little bit, not in terms of a revolution, but in terms of an evolution, because in the past we might have just simply had to one hypothesis and test it. But with AI we're able to say, "Hey, we're going to send you 20 hypotheses, tell me which ones you think have a higher probability of success. We're going to test those ones specifically." And so, by doing a better front-end analysis, if you will, of what the possible opportunities could be, we're more likely to end up with a hypothesis that can be validated at the end.

Ashish Verma: That's correct. That's exactly right.

Raquel Buscaino: It's truly profound how much innovation takes step at every point in the process, not how AI can be used. So, I do love the Life sciences example here. As we look into the future, let's say 5 years forward, which seems crazy, because if you would have asked me 5 years ago what these 5 years would look like I couldn't have told you. But as we look 5 years forward, what are some of the most exciting data-related advancements that you are expecting organizations to adopt? What do you think will be table stakes in 5 years' time?

Ashish Verma: You said it, 5 years, very long-time horizon given.

Raquel Buscaino: Good luck!

Ashish Verma: Exactly right. So, I will tell you right, like the pace at which tech disruption is happening. I don't think timeframe, I think horizon. Let's assume today is horizon one, right? Really, what I need to worry about is horizon 2, horizon 3. And I think, looking forward, some of the few things that are going to become obvious to us is, when you hear about sort of this disruption, and you hear about what happens to the computing infrastructure as a result of this disruption, it's going to be like how it is when you build a house. You don't think about water, electricity or sewer, just like you're not going to think about compute and storage. And this whole thing about GPU, CPU is going to be a moot point, right? It's commodity at that point in time.

What that means is like all the things that, you know, we as society, are expecting to happen at scale, whether it's the 2-nanometer chip design that's going to change electricity consumption and heating cooling technology in a data center. These are things that are teething problems for us to get to the journey that we're talking about.

So, I don't think we'll be talking about the business of AI, you know, we're going to be talking about what does AI for business mean at scale? And what are the things that are happening in the space that materially impact what you did just like in my example of a life sciences company in this time horizon. You're not going to see drug development look like what it used to look like. Autonomous cars that people talk about that, that consumes all of this infrastructure to be able to self-drive with cameras and all the other sensing technologies, and those are things that will be mainstream.

So, and the reason why this is important is, if you think about what's happening underneath the hood or in the back office of these things is, data is being collected at scale and volume and being run through this infrastructure to implement the strategy. So, I think the CDO role, which we started with, which is sort of going to be at the heart of these transformations will become very, very evident. That's my 5-year, prediction or horizon prediction depending upon how you consume it.

Raquel Buscaino: Amazing. Yeah, I often say that the future is coming faster than ever before. And I think that's true nowhere than in the world of AI and data. So, thank you for coming on sharing your perspective, and for such a great discussion.

Ashish Verma: Thank you for your time.

Raquel Buscaino: Well, to all our Tech Savvy listeners out there, if you enjoyed this episode, please share and subscribe, and if you'd like to learn more about data modernization, AI-driven data strategy, or the evolving role of the Chief Data Officer, you can follow myself and Ashish 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.