



Persistent Podcast | Episode 03

Ronak Doshi

Everest Group

Gurvinder Sahni: Welcome to The Persistent Podcast and another exciting episode. This is a spot where we explore conversations that drive transformation and shape the future. I'm your host, Gurvinder [Sahni], I'm the Chief Marketing Pfficer at Persistent, and I'm joined by John Madden, who is our Global Head for Thought Leadership.

Gurvinder Sahni: We have a special show. And as you know that on this show, we uncover ideas, we explore insights, look at the latest innovations, things that are redefining industries and societies, from technology and business to culture and beyond. We will dive into the most pressing topics and trends of our time. John, who do we have today with us?

John Madden: We're very thankful and we'd like to welcome Ronak Doshi from Everest Group to The Persistent Podcast. Ronak is a partner in the technology practice covering digital transformation and applications. He covers horizontal technology areas, and he's also responsible for the global BFSI industry research. Ronak, thank you so much for being here, and welcome to The Persistent Podcast.

Gurvinder Sahni: So first of all, thank you for taking our time from your busy schedule, Ronak, and you know, there's a question in that comment as well. I know that you have the front seat at all these large tech conferences, exclusive access to what leaders are sharing, what big companies in technology are announcing. The in-person event thing is back, right?

Ronak Doshi: Yeah, I think in-person is back. It is good to be connecting with people, hearing their stories, and understanding where

they are in their journeys, right? And what I've seen is most of these in-person events has been about journeys, right? It could be about product journeys. It could be about customer journeys. It could be about transformation journeys. That for me has been great, right? So the storytelling has become the key crux of these events. And the story has been around the themes that we talk about. Change, as you said, transformation, use of digital technologies, AI, definitely a lot of Generative AI these days, and then the thought process around building the infrastructure for the future. I think that's become one of the more common themes, because if I have to do AI, and I don't have the right infrastructure, I'm not going to get there. So I'm seeing those to be the mega themes and with a very strong focus on the stories on what does it mean for business.

John Madden: Yeah, let me ask you, you talk about transformation. People are still talking about transformation, but transformation when times are good is one thing. Transformation in a time of pullback and a downturn is quite another, especially when it comes to AI and GenAI, because it does require a lot of investment as far as infrastructure and everything. So. I'm curious what you're seeing out there in terms of how are people limiting their transformation plans? I know they're getting a little sticker shock when it comes to, when it comes to realizing what they're going to have to invest to take advantage of AI and GenAI, what are you seeing out there?

Ronak Doshi: Yeah, so I call these as high velocity segments. So if you think about it, the focus is not about doing these big-bang five-year, seven-year, 10-year roadmaps of transformation. It's about: we want to get the change and we want to drive change at a very high speed in segments where it makes sense. So I have cost pressures, I have margin pressures and we see the macroeconomic environment. We see the overall uncertainty, be it geopolitical or otherwise. So we know for a fact we are in uncertain times and that is the reason why we want to be at the front of these high velocity things. And prioritizing investments in those is the transformation approach that we are seeing.

Ronak Doshi: So instead of saying, I'm going to do a three-year employee experience end-to-end transformation piece, I'm going to go after the high velocity items of where does it hurt the most. Where can I create the most value? And those are the areas where you go. So employee onboarding is a challenge or providing them a talent mobility path is a challenge or giving them a strong Learning and Development

experience is a challenge. I would rather focus on those versus trying to do that into an employee experience transformation. So transformations in high velocity segments, you can call them as micro transformations because the reason is there was always this waterfall way of thinking, I will create a plan and I'm going to stick to it, right?

Ronak Doshi: Now you're saying: I'm going to create a target-state vision, but it's going to be a moving target-state vision. Simply put, right? Eighteen months back, if I had put that target-state vision, I wouldn't have Generative AI in it. Today I do. So I can't be having a target-state vision, which is going to be static. It's a moving target-state vision with a set of high velocity areas or segments which are going to drive it. That's the areas where we're doing transformation. So that prioritization is becoming key, and that is where we're seeing most of the energy.

Gurvinder Sahni: So I totally echo the sentiment on what you shared. Ronak is the go to person in terms of if I ever struggle with a concept. So maybe during this discussion, I want to keep sprinkling some of my questions. We've all heard of large language models, right? What about small language models? Clarify that point for some of us that want to understand more about that.

Ronak Doshi: So large language models are trained on huge data sets, so they are measured in terms of the number of parameters that they are trained on. And think of these as anywhere between 70 billion to, I don't know, they're going to get to 200, 300 billion parameters even more. But what happens is these are general purpose large language models. They can do a variety of tasks, but in an organizational context, in a business function context, I need the AI model to be trained on a specific set of activities.

Ronak Doshi: So think, think about it this way. So you are an insurer who has to do claims triaging, and you need to understand what is the state regulation that I need to look at as I'm doing the claims assessment. I need to be going through I think thousands of pages of documents which are created by each state regulator for each product, and I have to go through it. So instead of training, taking a very large language model, which is trained on everything in the world, I'm only going to train it on those sets of documents, so it's going to be a smaller large language model, which can understand natural language, but will give you answers specific to the data and the content it's trained on. So if I'm doing, say, for example, what should be the medical treatment

recommended? As part of the claims, it won't go to a regular WebMD or any other website and find data from it. It's going to find data from what the regulators have assigned [and determine] these are the prescribed treatments.

Ronak Doshi: So you're going to follow these rules. You're going to follow these compliances, and that's why you need it for the small language model. So you need them to be trained on that. The other aspect of it, and which I think you will find really interesting, is think about a lot of our devices, right? We've talked about these smaller, you know, form factor devices. Small language models could be stored on your phone. They could operate on your phone's infrastructure, and you don't need large GPUs and large infrastructure to run them, but they will do very specific tasks on your phone for your own data. And what if I want to create a future AI assistant, but that's only trained on my style of working. That's the data I wanted to be trained on.

Ronak Doshi: It is smaller in size, manageable in size, doesn't do everything, does a very select set of use cases of functions that you need. So it will be good for a compliance use case, for an experience use case, for claims use case, for a sales operations use case, but it won't do everything. It is cheaper. It is something which is very purpose-built, and it is something which you can manage better than the large language models, which you don't know what it is trained on. Currently the biggest issue is if it is trained on data, which I don't know, I don't know what output will create. But if I know that data set that it is trained on, I have more confidence on it. So the usage of it will increase.

John Madden: Does that mitigate or does that cut down on the need to clean the data? It's a less of a lift [for small models] than if you're doing some general purpose, large language model, it sounds like, but is the process by which you have to look at the data, find it, clean it, make sure it's integrated and secure. All that remains in place, it sounds like it's a much more concentrated and focused.

Ronak Doshi: Yes. So data quality is essential, John. So if I'm not getting the right quality of data on which I'm training the small language models, you're not going to get the right output. Do I have the data to take decisions on it? If I don't have that, and if it's not in good quality, I will not be able to progress well on the small language model.

Gurvinder Sahni: You know, on that theme, Ronak, you touched on the point of how do you enable AI? Because I think the whole point is that it cannot be externally driven, right? There's a lot of change in culture specifically around data discipline that needs to come in. And we want to touch a little bit upon that part as well. So how do you enable AI to deliver the results for you.

Ronak Doshi: Yeah. And that's a fantastic question, Gurvinder, because if you look at it, most of our conversations these days have been around AI in terms of using it or implementing it, but nobody's thinking, how do I get an organization ready to adopt it? So the big market, and again, for the likes of Persistent Systems out there, is enabling the organizations to use AI or to be able to implement AI. Everything from getting my processes in order to my people and change management in order to getting the right data – John, you and I spoke about this – and getting my systems in order, and infrastructure in order so that I could use AI is important. If I have a great AI tool, but I don't have the right process and change management in place, I don't have user adoption. If I don't have user adoption, I don't get the value out of it. And then it's going to be the same case with every other technology – we've invested so much in AI and the value realization is not there. So the investment you need to do into process transformation and modernization of data infrastructure and all applications to make AI useful is where the bulk of the investment has to be currently, because we are not ready to adopt AI as is across the board.

Ronak Doshi: And that is where I like the use cases of AI, which are helping drive adoption of existing technology. There are two things that AI does really well, right? One is AI is a new UI, right? User interface, AI allowing you to use your existing digital investments. [For example] I have a document management system that can be queried, but nobody is using it. Now I put a chatbot on top of it, which can read through all these documents, and I get more usage of it. So those are the ones where we see more of a success, but the ones where you don't have those initial digital investments done, or the modernization done, you need to do that step to be able to use AI better.

John Madden: A follow-up question specifically around BFSI, Ronak. We certainly we have a lot of conversations with clients and Persistent BFSI clients in terms of how they want to enable the AI and the use cases that they want. I'm curious, especially for say like the big banks, which have been in chatbots for a while. How are they evolving sort of

their existing customer service? The UI that you mentioned – how are they really turbocharging what they already had?

Ronak Doshi: So if you think about from a banking standpoint, there are a few things that is happening. One is they're expanding the number of channels through which they're engaging customers and the number of integrations they're doing with the broader ecosystem to embed financial products in the customer's lifestyle. So this theme around banking as a lifestyle or embedded finance is something that is picking up. So any life experiences or any business experiences that I want to run. I have a small business: I need your treasury solutions; I need forex solutions. I need some advanced credit or other pieces.

Ronak Doshi: All of these products can be made available to me as I'm running my business. So that is the piece that everybody is doing now. What banks and financial services institutions have done is, as you think about embedded finance and these multiple channels through which I'm consuming banking products, they're putting these AI agents or chatbots to do a lot of self service. So that is the first line of activity that they're doing. You need certain services that you can now start consuming without reaching out to a call center or going to the bank and other pieces through your app, through your ERP system. So if I'm a small business, I'm using an ERP system. The bank is providing a chatbot API integration where you can directly connect to it and get services done. You can ask these queries on the ERP system, and it is directly integrating with the bank. So that for us has become the use of AI. On the experience side, making [customers'] access to their own data, their own services and products as much as easy as possible is the area where they're doing it.

So I call that as a SUPER experience. And if you think about each of the [letters] of SUPER, the S part is important. It is Secured, right? So that's the first thing that they're talking about. The second thing which is the Ubiquitous. It is across all channels, across areas where I need it? The P is the Personalization, with AI now allowing me to understand these were my past inquiries. These are the most common activities that I do; it will remember, and it will guide me, and it will basically make sure that I can now do a lot of this even without prompts. And then the E part is Easy. It has to be simple. So the UI of AI is that piece where it is allowing it to be easy. And then lastly, Responsive. It cannot be something which if I call a call center and then I go to my chatbot and

then the chatbot doesn't remember what I spoke about with the call center, it cannot happen.

So it is also integrating all of these data sets. It remembers the context of what you had over the past few touchpoints with the bank. And that for us is what AI is now enabling. So banks are investing heavily into this across talent, in terms of technology process and getting that right.

John Madden: That's really interesting. And I love a good acronym. That's really helpful too, thanks for that.

Gurvinder Sahni: So Ronak, I want to discuss something called a BYOM, we were talking offline the other day about the whole concept of Bring Your Own Model, you know, and to be very clear, this is an AI model, maybe you can you talk a little bit about again. So what is BYOM?

Ronak Doshi: So one of the things we're seeing is there are a lot of applications being created that uses multiple large language models based on your preference. So if you want to use an open source model, you want to use a proprietary model, you want to use your homegrown model, they're allowing you to do that. So as we are doing this, what we realize is: I like the application, I like the functionality, I like what it is doing, but I want it to be using an LLM model, which is mine, or I have trained it, or it is a custom model that I've built.

And that flexibility in the architecture is what Bring Your Own Model is doing. So think about it and we'll take the banking example again. So the bank has created their own large language model. It is trained on all the marketing and product collaterals. Now I'm using several different technologies in my marketing stack, which each offers me great feature functionality. I don't want to throw all of those away. I want to keep using that entire marketing workflow that I've built. With all the marketing tech stack, now the large language model should be mine, and it has to be integrated into all of these. So I want you to follow my content branding guidelines and generate images, any other form of content videos and other things using that large language model.

It is compliant. It is a bank-created one. I have more trust on it, so I get the functionality and the marketing stack, and my people know how to use it. The experience doesn't change for them, but what they get as an AI output is one that is trained on my model. The same can apply in

compliance. The same can apply in healthcare, in manufacturing operations, everywhere else.

So that for me is where Bring Your Own Model is exciting because there will be a few organizations which believe there is a value in building their own model versus using a third party. Not all of them, very few of them, but the ones who do would want to bring their own model and the application providers have to enable that. If they don't, they miss out on this opportunity.

Gurvinder Sahni: What about security, Ronak, then? I mean, when I hear about bringing your own model, the next thing that I think of is security. Can you talk a little bit about that as well?

Ronak Doshi: Yeah, and one of the reasons for bringing your own model may be security and security from very different lenses. So I want a model which is trained on data, which is compliant per my organization policy. I want it to be, making sure that I'm not using any personally identifiable information. I'm not going against anybody's copyrighted IP, and I'm making sure what it generates is something which is organizationally approved through our frameworks and guidelines.

So the security of the Bring Your Own Model increases because what it allows me to do is: I have a model that I trust. I have been working with it. So I have more trust in an organizational framework, and I want to do that. Now, the other piece of the security is the application stack misusing the model that you've built. And that is where the application security comes into picture, which is the applications and the tech stack that I'm working with. They have to build those controls that they cannot abuse models that are built by other clients of theirs because if they start doing that, they lose [customers'] trust and they will not get business after that.

So one of the pieces of Bring Your Own Model is the ones who are providing this capability of Bring Your Own Model have to have dedicated security to protect somebody else's IP that is getting embedded as part of their solutions, and they should be okay with it. And for me, that is the piece of security which is important. But otherwise, I think Bring Your Own Model from a risk perspective, from a security perspective, is going to be much better because you have more control on it.

Gurvinder Sahni: What's your view on the whole open source brigade, with the launch of Meta's AI. And the 2nd part of the question is what about beyond text? Because as we are seeing, the ability to create those images and, and videos, etc.

Ronak Doshi: So one of the things we've seen with open sources, they have been a fertile ground for innovation because they bring the best from the industry to kind of solve problems that they enjoy and that they love. And it creates an environment where everybody benefits out of it. And that's been the thought process and all these open-source large language models, which is we don't want these large language models to be restricted to a few corporations and enterprises who make disproportionate advantage of it. So it's basically commoditizing or democratizing access to AI models in a way. So that is the place where I believe the open source brigade, as you said, is great.

Again, we've seen examples of great innovation, right? Be it Red Hat and others that we've already seen that innovation comes from those quarters. So from a large language model, we will expect a lot more activity around that because it's a fertile ground for innovation. It is a great training ground. It is a great experimentation ground. You can do a lot of that with the guidelines of the community, so you add to it. But you also take a lot out of it because you get to collaborate with a lot of similar or diverse sets of people. You're working with on these projects, and it influences enterprise tech strategy.

To your other question, which gets me really excited is we've always talked about text and code as the two big use cases so far in terms of large language models. Yes, it can generate text, it can generate code, and it can generate images. I think we've seen images, with Adobe Firefly and others, right? So those have been the three big use cases. I think now it's opening up a lot more. I can generate video, I can generate animated graphics, I can generate 3D models or 3D CAD designs. So the future instrumentation designs on spacecrafts or in space stations are going to be designed using these large language models because they can be trained on existing IP that's out there.

And then they can help you build these faster chips and chip design. If we are able to train a large language model on historical chip designs and certain principles, can it then design the future chips, right? And the same from 3D visualizations of genes and then therapeutic medicines, can I basically do molecular generations right out of large language

models, which we are already seeing in drug discovery. So I think the 3D video and other content forms that will get created out of these large language models is also something exciting. Now there are challenges on the way because you don't have all of this IP in public domain. Nobody's going to keep their chip designs and semiconductor designs in a public domain. So you can't train it out there. So you have to train it in organizations who are building it. So that is where it will get exciting in terms of the usage of those.

John Madden: You know, I'm curious, Ronak, you talk about whether it's large language models or small language models or bring your own model or the IP that's out there from open source. It's not just about those components, it's about talent too, right? And, and there's always a talent war going on, but it seems particularly acute right now when it comes to AI. Enterprises don't have the internal capability that they need. That's why they come to us and look for certain skills and services.

John Madden: I'm curious what you're seeing in terms of the talent war and how this is different from, say, past talent wars, whether it was around cloud or analytics or something else. How are enterprises coping? How are they trying to augment their AI skills overall to take advantage of everything we've been talking about.

Ronak Doshi: Absolutely. And again, talent is one of my favorite topics to talk about when it comes to AI. If you think about it, if you take a step back, AI does, I think three things when it comes to talent.

The first one is it changes the way we do work currently. So it impacts all talent, right? So think about it, the existing talent and workforce needs to get upskilled to be using AI, even if they don't know how to build it, but to use AI in their activity. Be it I'm in HR, I'm in marketing, I'm in sales, I'm in supply chain. Any function I am in, I now need to start using AI. So that's the first impact on talent. How do I get a workforce that is upskilled in talent in AI. If it is not, I miss out, right? So that's one piece of the puzzle.

The second piece of the puzzle is there is a massive talent supply shortage when it comes to AI skills, specifically in these areas of Generative AI and AI, because these have been around for the last five or six years but have gotten to prominence over the last two and a half years or so. So that is where the talent is still building. We don't have an

education system that is geared toward creating the mass talent that we have currently on Java developers or testing and other areas in IT. We don't have that engine of talent generation. So that is where you'll have to manufacture talent of people who can work on these AI skills.

The third is, which is very important that's going to be the bigger challenge of there are new rules that need to be created because of AI. So, if you think about it, I need somebody who is now an AI compliance leader or a Responsible AI leader, an AI architect leader for Responsible AI. I need a data ethnographer that can understand data in the context of AI. We are seeing a bunch of new roles getting created which traditionally did not have a job description. There was nobody with those roles ever, but now to be able to adopt AI to drive that change management, that is where we are seeing investments into that. So the combination of these three is going to be a multifold impact on talent. And I think there is going to be an AI talent war; I think it's already started. So we are seeing a lot more of that.

John Madden: I was going to mention that in terms of upskilling ourselves, we're doing a lot of that internally at Persistent, too, across marketing and software development. And by doing that, we're learning best practices and learning that kind of enterprise experience, as far as what our clients are dealing with all the time in terms of trying to upskill their people.

Gurvinder Sahni: So there are about 16, 000 people that have been trained on AI, and plus we're also looking at certifying folks on external certifications. And we pride ourselves in terms of the talent that we bring to our customers. And I think it's very interesting because we're hearing exactly similar set of things that as we go and talk to our customers, they're finding a great partner [in Persistent], and I know that thanks to Everest Group, we've been rated well by you and your extended team as well in terms of how we are showcasing when it comes to helping our customers take AI from playful to impactful, right?

Some fun questions [now]; Ronak, you and I share a passion. We follow each other on Instagram. I know that you were talking about AI and cooking. I got a little confused. Can you help decipher that?

Ronak Doshi: There are these tools which are being developed which can help you to do these daily chores better, right? So from grocery shopping to figuring out, I have this in my refrigerator and pantry, what

all I can make out of it, giving you recipes... how do you optimize the steps to manage? So it's not just about giving you the recipe of it, but also giving you the steps involved and smart tricks involved on how to do it. And that for me is important. And think about it: We already have smart devices in our home that are already listening to us, and they might be future in watching us. They can basically say, oh, you're cutting it this way. This is not the right way, right? You could do it differently. And they could actually guide you with live training on how to cook. And then anybody who says, I don't know how to cook, oh, there's an AI which can train you on how to cook. I don't know what to cook for today. Oh, there's an AI which can tell you what you can cook based on what you have in your grocery and pantry today. That for me was the thought process that I was thinking through in terms of a lot of our daily activities can get very intelligent and interesting at times.

And I know there'll be a lot of folks who are saying, I don't want AI to be doing that. It's a specialized skills and other things but think about it. There are, I don't know, thousands of millions of people who don't enjoy that piece of cooking, but with AI and the intelligence and the activity around it might just become fun... [AI has] gotten people to do more physical activity. It's gotten people to track their health better, sleep better. Why not what they eat better? So you cook the food and then the AI will tell you, oh, this is the number of calories you have now. So again, possibilities are there.

Gurvinder Sahni: I know our next episode with you will be from the kitchen. You know, this was so much fun. So, Ronak, thank you again for joining us, for another amazing show. We had so much fun and absolutely thank you for sharing your insights and the learnings from your experience. We hope to have you back and like I said, maybe the next episode will be from the kitchen.

And a big thank you to our listeners and viewers for joining us today. Remember we want to hear from you, to learn more email podcasts@persistent.com. I am Gurvinder Sahni, joined by John Madden my cohost, thank you, and we look forward to seeing you next time.

