



## The Deloitte On Cloud Podcast

### Mike Kavis, Managing Director, Deloitte Consulting LLP

**Title:** Qualcomm's Atul Suri and Deloitte's Rahul Bajpai on how AI is bringing innovation to the edge

**Description:** In this episode, Deloitte's Mike Kavis talks with Qualcomm's Atul Suri and Deloitte's Rahul Bajpai about how AI is bringing innovation to edge computing. The trio discuss why companies are bringing AI capabilities to the edge, use cases for AI and edge, and how cloud enables this powerful combination. As to the why, it's simple: AI can help companies increase productivity and develop new business models that create new revenue streams—which will drive sustainable growth and innovation.

**Duration:** 00:24:11

**Mike Kavis:**

Hey, everyone. Welcome back to the podcast where we get real about cloud and AI technology. We discuss all the hot topics around cloud and AI with people in the field who've done the work. I'm your host and Chief Cloud Architect over at Deloitte, Mike Kavis. And today I am joined by two experts who specialize in AI at the edge, which is a topic I'm not really up on, so I'm looking forward to learning from these guys. So, first up is Atul Suri, vice president global system integrator and direct channel partnership at Qualcomm, and our very own Deloitte US growth offering leader for connected edge, Rahul Bajpai. And, so, we're just going to start off because a lot of my followers who listen are more cloud savvy than they are edge and AI savvy, so we're going to start a little generic, and we're going to talk a lot about AI at the edge, but we'll start with Atul. Why is this topic trending and getting so much action right now, and why is it such important discussion to have these days?

**Atul Suri:**

Well first off, thank you, Mike, for having me. I'm excited to be on your On Cloud podcast, and I hope I can provide some more granularity and some more clarity on this topic and some more familiarity about what Qualcomm is doing in our role at the edge. I'll start off with Qualcomm's vision. We are an intelligent computing everywhere company. What that means is we bring high-performance, low-power compute along with on device intelligence at every imaginable edge device. So, think about the phone in your pocket, the smartphone. To us, that's an edge device. Think about the PC that we use every day. That's an edge device. In an industrial and enterprise setting, think about the gateways, think about any of the infrastructure elements that connect to the hyperscaler cloud. Those are all edge devices.

Your question is very relevant. Why is AI at the edge important and why is it something that we're talking about and excited to talk about? In my mind, it unlocks greater operational efficiencies. In my mind, AI is bringing a new wave of use cases. It's going to increase productivity. It's going to enable new business models. It's going to create new revenue streams. And all of that is going to drive sustainable growth and innovation, even in the current macroeconomic climate. And, so, if you put all of that together, I think this is something that we're excited to talk about, we're excited to bring innovation to the edge, and excited to see where the future goes.

**Mike Kavis:**

Over to you, Rahul.

**Rahul Bajpai:**

Really well summarized and put, Atul, in terms of what the opportunity is ahead of us and how some of our clients in the enterprise segment are looking at edge AI to truly drive OT outcomes for their businesses. I think the one thing that comes to fore is how disparate the systems, applications, end points, devices, and the network infrastructure is today on the edge and the need for getting a common stack, a common architecture that is truly a cloud native

architecture which allows for different applications on the different systems to be able to communicate with each other so that we can actually drive insights on the edge.

To us and to several of our clients, edge infrastructure really is comprised of backbone network connectivity, a common cloudlike infrastructure, and the applications all residing on that architecture, so that they can drive not only ease of integration, but also ease of data extraction, correlation, and generating insights at the edge. The one thing also that comes to life with this technology is how we can drive real-time decision-making in domains or environments that require very low-latency, high-throughput, high-bandwidth type transaction management, which then allows the enterprise to start looking at some of the cutting-edge capabilities with computer vision, Generative AI, and so on.

**Mike Kavis:**

Yeah, and a lot of the space that we've talked about leading up to this was in the area of smart factory. I'm going to date myself here, but when I started in IT, it was on a mainframe in a steel plant and the manufacturing had sensors. And then we would get tape and we would run batch jobs, and then people would make decisions on that, but it was like a 24-to 48-hour turnaround. Now you have this intelligence on the edge and making these decisions real time, so explain some of the emerging use cases in the space that you see and how AI—because there's already been a level of intelligence in these sensors, but how is AI making it a game changer, whoever wants to field that one?

**Atul Suri:**

Yeah, I'll go first, and I'm sure Rahul's going to add more to it. So, like you said, Mike, the edge is becoming more intelligent. What does that mean? In the example that you gave, the edge always had digital technology. You had sensors. Those sensors were creating an output. Intelligence at the edge really means that we're bringing real-time data and analysis. And the purpose of doing all that is to make time-critical decisions, in real time, that will bring a further improvement in productivity, will bring a further enhancement of workflows, it will make processes that, like you said in your example, that take two days happen instantaneously. So, this idea of immediacy, this idea of computing and real-time data analytics driving additional performance and productivity gains, to me that's what this whole thing is about.

**Rahul Bajpai:**

Yeah, and a couple of things just from a use case perspective that come to mind given that we have trialed some of those in our smart factory. One is real-time quality detection and sensing, which really allows to get as narrow and as deep as we need to in terms of monitoring an assembly line, looking at errors in machine parts, in the form factor, or even looking at the thickness of a paint coat on the door of a car in a car factory. Being able to track this in near real time as the production assembly line progresses and creating that automated feedback loop or that feedback control that allows the anomaly to be detected, analyzed, and fixed without having a lot of manual or human intervention, that is the end goal of any predictive quality sensing and detection.

Another one that also comes to mind is outdoor plant in terms of efficiency of AGVs or shipments that come in, and the ability to track how much time the inventory sits outside, how much time does it need to actually be sorted and distributed to the right floor and the right aisle, and the ability to correlate that with RMAs, replenishment of parts, all in near real time speeds up the inventory management and the supply chain ecosystem for an enterprise.

**Atul Suri:**

I love all those use cases, by the way. I think as Rahul started off, and I think you mentioned this is really driving OT transformation. So, every business today is looking at how to create more economic value given the current fixed processes and workflows that we have today, and this whole operational transformation at the edge is really driven by bringing new AI capability, so great examples.

**Mike Kavis:**

Yeah, one thing I was going to note is, being a cloud guy, a lot of the AI discussions in my space have been more around Ops, but what you guys are talking about is more around emerging use cases and the new value props and stuff. But I'm sure there's a component of Ops too. Like as you mentioned at the beginning, all these connected devices are so hard to manage, but now you can use intelligence to manage it. So, the question here is, was it born from the need to manage or was it born from the need to change the game and provide additional value, or is it kind of both things?

**Atul Suri:**

To me, digital transformation is really centered around economic value creation for an enterprise. The choice of technology is a catalyst to deliver the performance. It's not a hurdle to overcome. So, the first place where we start is we look at enterprises that have a need for creating top-line or bottom-line economic value, and we look at what are the ways in which they can be automated using existing devices and what level of connectivity, what level of AI, what level of immediacy, resiliency that we can introduce along with real-time data and analytics that we talk about that can really transform these businesses.

As Rahul brought up, again, quality sensing and defect detection at the edge is a fantastic use case. We've been having computer-vision-based AI that was happening in the cloud for quite a while in smart factories, but the ability of the technology, the evolution of the technology to bring AI workloads, so you talk about Ops, I'm going to talk about TOPS, which is the ability to run AI at the edge so that you can take some of those AI workloads that were happening in the cloud, for example, quality inspection and defect detection where a camera is consistently taking an image of the object going on a conveyor belt and it's analyzing in real time at the edge or a KPI, whether it's defect in the size or shape of the object that is going down the conveyor belt.

So, this idea that we can bring high-performance, low-power compute says that you can run AI at certain TOPS at the edge and you can create that real-time analytics and real-time economic value creation. You can imagine the question to ask yourself is, "What does it do for the enterprise?" I love this idea, this paradigm of asking a so-what question. So, what if we bring AI at the edge? And really what it means is you can now become much more efficient in detecting those defects, you can become much faster in the number of products that you can detect in the same amount of time. So, you talk about increased productivity, which again is a driver for economic value creation. You talk about a better response rate, so you talk about a better defect detect

ratio, which again means a higher-quality product. So, all these things, in my mind, translate to creating a better economic value, in which case they will get accelerated, they'll get adopted, there's no hurdle to overcome.

**Rahul Bajpai:**

Yeah, and Atul, from our clients' lens, all the dimensions you mentioned in terms of achieving those benefits or outcomes, the way we help them realize the need for the investment up front, because, remember, all this technology, the entire ecosystem with computer vision, camera analytics, defining, designing, and architecting a new edge infrastructure does come with a cost component, but insofar that cost is attributed to the gains and the ROI and the benefits across better inventory management, better quality control, less time to market, it almost becomes a self-funded sort of life cycle in which the upfront investment needed is justified by these savings, and the benefits that you gain from the top two, top three use cases in that example you were giving from a factory setting.

The one other digital transformation imperative for most of our clients is looking at both solution stacks from the hyperscalers, so the cloud becoming an extension to the edge as well as non-hyperscalers where on prem infrastructure with custom, I would say, orchestration, aggregation allows our clients to be a bit more nimble and agile in the way they deploy the edge capabilities. And of course, tying it back with the device ecosystem where the sensors, the cameras, the endpoints, tablets for example, then become true intelligent endpoints where, as another example, if a foreman or a workman is walking around with a tablet in the shop floor for assessing what to do with a defective part, in context, again using Generative AI on the edge, they can be made aware of a similar part that may be in stock, the actual guide and troubleshooting mechanism to go fix that part or replace it, and automate the replenishment of that part all within one, I would say, event or one scan, read, analyze, and act upon sort of life cycle.

**Atul Suri:**

Yeah, and the best part about AI, which you just captured, is because it's continuing to evolve so rapidly, the use case with which you started, you can continue to add use cases on the same infrastructure, so thereby further creating a better ROI for the investment in the technology to go and deploy that in the first place. So, in the defect detection example that we've been discussing, the camera had a singular purpose to detect the number of defects, but now that you have AI running at the edge, you can now change those workflows, you can change those models, you can bring new innovation through all of the enhancements that are happening, and you can use that same camera infrastructure to solve other business problems related to where that installation has happened. So, again, the ability of AI evolving so rapidly creates a further economic tailwind to go and deliver on these use cases.

**Mike Kavis:**

Yeah, I want to touch on that point there. I'm going to date myself again, but adopting data warehousing back in the day, big data, MDM, all those things were a huge upfront investment that you kind of went dark on the business for a while to build all that. And it sounds like with these use cases, you can more simply and quickly apply a use case to solve a specific problem, and as you just noted, then expand for multiple use cases. Is that the case here?

**Atul Suri:**

Absolutely. Absolutely. I think we're a firm believer that as you bring more intelligence, as you bring real-time data at the edge, you can solve so many different problems. For example, let's continue talking about a smart distribution center or a plant or a warehouse and there the first premise to get better productivity was to bring connectivity and to bring the devices online so that they could be better managed, whether it was management for route planning or for better productivity gains, but now what we're seeing—because now you have continuous data. So, remember, AI feeds on a large data set. So, the AI algorithms need a very large continuous data set to continue to create those economic gains. So, in the example that we're discussing on smart warehouses or distribution centers or smart plants, the first paradigm was let's connect the devices. So, once you connected the devices and applied AI at the edge, you were able to get things like visibility, things like tracking, things like monitoring. But now with AI evolving—and remember, AI feeds on continuous data. AI is trained on large data sets. So, now these devices are continuously connected, they're continuously sending data. You can now use AI models to create the next level of optimization, which is utilization trends. So, now you can predict how likely it is for that device to be utilized in performing a certain activity and thereby creating and unlocking even further productivity gains because now you can do shared resources, you can do shared device and productivity gains across the whole workflow because you've got so much data from the device on which you can further implement and advance those AI models.

**Mike Kavis:**

Yeah, I think what's amazing here is what we're trying to solve aren't new problems, just there's new technology to do it. Like I mentioned at the beginning, we would have to run VSAM files and tape mounts and all that stuff. But now that you can do this stuff in real time, that's what's real exciting to me from someone who had to go through this batch processing for all these years. And like you say, once you open up those floodgates of value creation, you have all the data. It's just like what problem do you want to solve next. So, Rahul, give me some examples of some client success stories applying some of these technologies that you're seeing, some of the common ones.

**Rahul Bajpai:**

Yeah, absolutely. So, the one I would start off with retail and customer experience, specifically in retail stores as we have helped our clients think through what the architecture is, and more importantly, what are the set of stackable use cases leveraging edge AI to be able to analyze customer behavior in real time, to be able to manage inventory by detecting out-of-stock items, and for them to offer personalized shopping experiences through smart mirrors, smart screens, or even recommendation systems. That has shown early signs of success in terms of increasing operational efficiency and also being able to introduce new products and services.

As I've mentioned before, in smart manufacturing, we have been through a couple of trials with our clients on how edge AI can truly optimize through very, I would say, near-real-time monitoring the specific manufacturing processes, not only from a detection of anomalies and defects, but also be able to predict some of the equipment failures before they happen, which ultimately reduces your overall downtime and maintenance costs. So, those are some examples that come to mind more recently as we've been engaging with our clients.

**Mike Kavis:**

Those are pretty cool. So, I'm going to ask you both now—we'll start with Atul—get your crystal balls out. Last question here. We've got about five minutes left. Whip out your crystal ball. We're early. What is the space going to look like in the next few years? Right now, we're tackling some pretty cool use cases, but I'm sure it's just the tip of the iceberg, so where do you think this is going over the next few years?

**Atul Suri:**

Yeah, excellent question. I wish I had a crystal ball on many things, but I'll give it a try. Like you said, we're in the early phases of driving accelerated adoption of AI at the edge and digital transformation, however, if you look at the percentage of the 2022 global GDP that relied on digital technologies, more than 60 percent of the global GDP relied on digital technologies. So, what that means is enterprises are becoming digital, and that trend—the crystal ball will tell you that that trend is going to continue to accelerate.

And as enterprises become more digital, what that means is they're going to become more connected, those devices are going to become more connected devices, and once you have the infrastructure and connectivity and digital infrastructure at an enterprise, what that does is it drives really the opportunity that we see in front of us, which is driving AI at the edge. In fact, we like to talk about Generative AI, of course. We didn't talk much about it today, but we think of Qualcomm has evolved and innovated through multiple "Gs," the 1G, 2G, 3G, 4G, 5G in wireless communication. We like to think of Generative AI as another "G" that's going to further enhance and evolve enterprises' workflows along with connected devices.

**Rahul Bajpai:**

Yeah, and from our perspective, again, no crystal ball I promise, but when we start looking at some of the advancements in hardware, so, just the ability to have more powerful, energy-efficient edge computing hardware with specialized AI chips, that allows us to kind of handle complex AI tasks like deep learning with minimal sort of power consumption and latency. I think that will definitely start driving a lot more adoption and increased integration with IOT because edge AI definitely will play a very, very important role in processing data from billions of IOT devices in near real time, but also the backbone technology. The rollout of 5G has shown a lot of promise in improving the speed and reliability of data transmission, but it also allows us to enable edge AI applications to perform these complex tasks in real time.

I'll say our software engineering, data engineering, sort of domain that continues to evolve in terms of the AI algorithms that we have, how to make them more efficient, how do you actually broaden the scope to include a lot more applications with the right data privacy and security packed in. That hopefully allows our enterprises to start looking at autonomous operations from vehicles to drones to robotics and just start reliably operating them in environments with even limited or no connectivity as distributed edge. So, a lot more to come in this space. I'm a firm believer that ultimately technology is there to improve human lives, and so the ability for edge AI to drive innovation in healthcare and in the safety of our cities and bringing in sustainability and energy efficiency to the way we operate our infrastructure and retail or manufacturing or healthcare locations. I think all that comes to life as the technology evolves.

**Mike Kavis:**

So, my closing thought here, I've been doing these podcasts for a while. About three, four years ago when cloud was really, really hot, we talked about everyone's becoming a software company. And my takeaway today is everyone's becoming a connected AI company is kind of the next step. So, any closing thoughts on that idea before we head out?

**Atul Suri:**

Absolutely. I think you've hit the nail on the head. We believe in the transformation potential that's ahead of us by bringing AI to the edge, so I couldn't agree with you more. Thank you so much for having me.

**Mike Kavis:**

If you enjoyed this podcast, make sure to like us, leave a review, and subscribe. You can also check out our past episodes wherever you listen to your favorite podcasts. You can always find me on Twitter @MadGreek65 or reach out to me directly at MKavis@Deloitte.com. Feel free to write to me with questions we can address on future episodes. Thanks for listening, and we'll see you next time!

**Operator:**

This podcast is produced by Deloitte. The views and opinions expressed by podcast speakers and guests are solely their own and do not reflect the opinions of Deloitte. This podcast provides general information only and is not intended to constitute advice or services of any kind. For additional information about Deloitte, go to [Deloitte.com/about](https://www.deloitte.com/about).

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

**Visit the On Cloud library**

[www.deloitte.com/us/cloud-podcast](https://www.deloitte.com/us/cloud-podcast)

**About Deloitte**

As used in this podcast, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see [www.deloitte.com/us/about](http://www.deloitte.com/us/about) for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see [www.deloitte.com/about](http://www.deloitte.com/about) to learn more about our global network of member firms. Copyright © 2024 Deloitte Development LLC. All rights reserved.