DR: [00:00] I feel like people aren't realizing just how huge of an impact this idea of AI-RAN can really have for the telco.

Chris: [00:06] Yeah, so working with SoftBank, we're seeing that for every capital dollar that is being invested in this new infrastructure, they can generate $5 of revenue over a five-year period.

DR: [00:18] That's amazing.

Chris: [00:19] From a net profit perspective, SoftBank actually calculated out an eye-popping number of 219% profit margin considering all the CapEx and OpEx costs. So this is a huge statement, dropping additional dollars to the bottom line.

Announcer: [00:38] This is "Telco in 20," a podcast that helps telco execs achieve a competitive advantage with AI and the public cloud. It is hosted by Danielle Rios, also known as DR. Today we're talking to Chris Penrose, the Global Vice President of Business Development for Telco at NVIDIA.

DR: [00:58] Hi, guys. I'm DR. According to research firm Rewheel, most mobile network operators use less than 20% of their available network capacity on an average day, even though most utilized networks will use a little more than 50%. This happens because these networks are considered essential services. They can never go down. They must handle emergency communications all while providing an excellent experience for their subscribers.

[01:24] Most of the time that extra capacity goes unused, yet the CapEx has been spent and the network is up and running. What if you can monetize that unused capacity? Well, NVIDIA has an idea for you that they call AI-RAN. The idea is to use your extra network capacity for AI workloads selling it to enterprises. You'd be in total control of these workloads, being able to schedule them to scale up and scale down as necessary while reserving capacity for the network when you need it.

[01:53] The results have been eye-popping. NVIDIA tested the idea out on SoftBank's network and demonstrated they can run the network at 90% capacity while consuming 40% less power. They estimate that for every dollar invested in AI-RAN

infrastructure, telcos can earn $5 in AI revenue. That's a 5X return on investment. Amazing! Today I'm talking with Chris Penrose, NVIDIA's Global Vice President of Business Development for Telco.

[02:19] We're going to break down exactly how this new AI-RAN concept works, what it means for telcos looking to monetize their unused network capacity, and how this could completely change the game for telcos on their biggest CapEx expenditure. So let's take 20. Chris Penrose is Global VP of Business Development for Telco at NVIDIA. Hi, Chris. Welcome to "Telco in 20."

Chris: [02:43] Hey, DR. It's great to be here. Thanks so much for having me today.

DR: [02:46] Yeah, I am so excited to have NVIDIA on the podcast. You guys have been making mega news it feels like every single day and making some big moves in telco, and so super great to have you on the podcast. But let's first start with why is NVIDIA focused on the telco industry?

Chris: [03:03] That's a great question, and let me start off by saying NVIDIA is really a full-stack accelerated computing and AI platform company, but we go to market in a vertical way. Our job is to really look at every single industry and how can we accelerate the workloads in that industry in a dramatic fashion with our computing and software capabilities. And so my role at NVIDIA is to look through the telco lens.

[03:28] I spent 30 years at AT&T, so I'm a long telco vet, and having the opportunity to then come over and work in the telco vertical to figure out how we can apply our technologies to help telcos really transform their operations, build new monetization models or next generation networks. That's what we do. And candidly, the final piece is that telco is a huge market.

[03:48] Telco spends $300 billion annually on the capital that they're putting into the networks and over a trillion dollars in operations. And so we really think that this is an awesome market for NVIDIA to be able to apply our technology to help telcos.

DR: [04:01] Yeah, and I think AI is a really huge opportunity. The adoption of the public cloud was my first foray into telco, and I thought that was a really big idea. But now you couple cloud technologies and AI. And for telco, the latest developments with AI represent this gift from heaven for them to revitalize their business and really restore the glory of the telco industry.

[04:24] So at MWC earlier this year, NVIDIA joined forces with some key telco players like Nokia, Ericsson, and Samsung, the hyperscalers like AWS and Microsoft, and then telcos like T-Mobile and SoftBank to form something that you guys are calling the AI-RAN Alliance. And so what are the key goals of this alliance? What are you guys trying to accomplish?

Chris: [04:45] So there's really three key things, DR, that we're trying to do. The first is what we call AI for the RAN. And so if you think about how can we apply AI technologies to dramatically improve the performance of the network, improve energy efficiency of the network, even spectral efficiency, these are all ways where we can make the network run better with AI. The second place we're focused is on what we call AI and the RAN.

[05:13] And so think about this as traditionally telco networks were purpose-built and that means they're also designed for the peak hour, but it also means that they are highly underutilized in the evenings and weekends. And this is not the most efficient way to be able to get the most out of your investment.

DR: [05:32] Yeah, absolutely.

Chris: [05:33] And so the idea is can we treat the radio access network just like any other workload and be able to put on a shared infrastructure the ability to run the telco network and at the same time be able to run AI and generative AI workloads to get maximum utilization. So AI and RAN is that focal point.

[05:52] And then the final one that we're focused on is what we call AI on RAN. And this is how can we really support all the amazing new applications of AI and generative AI that are going to be running on top of these networks and do that in the most performant and efficient way to deliver the best experiences out of those solutions.

DR: [06:11] And I think that leads really nicely into this announcement that you did in mid-November where it felt like your CEO Jensen Huang was everything, everywhere, all at once, dropping all these amazing announcements in Asia. And so you guys recently had the NVIDIA Japan Summit and you dropped this news about some work that you're doing with SoftBank.

[06:34] And then he went over to Indonesia to talk about the sovereign AI factory that NVIDIA is building with IOH for Indonesia's industries and over 277 million Bahasa speakers. And so let's start with that huge SoftBank announcement. What were the highlights coming out of that?

Chris: [06:50] Yeah, so this has been really exciting because SoftBank was one of the first telcos in the world to take a look at the vision of how can we bring AI and RAN together and begin to work closely with us to turn this into reality. And what we are so excited to announce is that we now have the first live field trial where we were able to successfully demonstrate the ability to run the radio access network and AI and generative AI applications all off of a common set of infrastructure.

[07:18] We were able to get to the same levels of carrier-grade performance with regard to the radio access network, as well as be able to demonstrate the effective ability to move between the RAN and AI workloads, orchestrate that based upon the demand that was being needed, and also came up with some really powerful new energy efficiency metrics on how this is actually being applied into the network. So we're super excited to have that opportunity to make that big announcement with SoftBank.

DR: [07:49] And I think what's so great about that announcement, and we'll link to it in the show notes, is that you guys shared some really eye-popping business value metrics that will benefit the telco. And so can you walk us through what were those metrics? Because I feel like people aren't seeing that and realizing just how huge of an impact this idea of AI-RAN can really have for the telco.

Chris: [08:10] Yeah, I'm glad you asked. This is such an important thing to get across. And notoriously, you're putting these big investments into the network to be able to deliver these basic

services, but now that same infrastructure can help drive monetization and additional revenues on top of providing the current services that it was traditionally built to do. So what we announced are three big numbers that came out of this.

[08:34] The first was that working with SoftBank looking at the opportunity to deliver generative AI services in the market, we're seeing that for every capital dollar that is being invested in this new infrastructure, they can generate $5 of revenue over a five-year period.

DR: [08:51] Yeah, that's amazing.

Chris: [08:51] And so this is new incremental revenue for the telcos. The second piece is that from a net profit perspective, SoftBank actually calculated out an eye-popping number of 219% profit margin after putting in the investment considering all the CapEx and OpEx costs. So this is a huge statement with regard to the bottom line. And then equally exciting is what does it do with regard to sustainability?

[09:17] And they were able to show that they are expecting a 40% less power consumption than the best-in-class radio access network offerings in the market today by using this new latest and greatest infrastructure. And power obviously is a huge cost for the telcos as well. So doing good from a sustainability perspective and also dropping additional dollars to the bottom line.

DR: [09:40] And it feels like, and again just layman terms, but just like with cloud computing and what the hyperscalers did, it's this concept of elasticity a little bit in the network. They're over-provisioned because they have to be, they have to be carrier-grade and handle all this capacity, but most of the time it's underutilized.

[09:57] And so taking that extra capacity and then leveraging it for AI workloads for enterprises that are looking for AI infrastructure. Is that basically the idea? And so you're getting a lot more mileage for the dollars that you're investing and really utilizing the full capacity of that built network.

Chris: [10:14] That's exactly right. When you look at how the cloud companies operate their infrastructure, they're achieving in the high 90s of percent utilization of that. And what we've seen oftentimes with the telco networks, it might be on average between 35, 40, 45% utilization of that capital investment because it's only being utilized fully during those certain times of the day.

[10:37] And so if we can now take that additional capacity and reallocate that to be able to use it to do other things, particularly generate additional dollars and/or potentially driving additional value for their own use cases, that really unlocks the full potential of going down this common infrastructure.

DR: [10:56] And how long would it take to convert a network into an AI-RAN network?

Chris: [11:00] Now that's a great question. So at NVIDIA, as I was mentioning, we really build these platforms and tools, but we don't bring the end final solution out to the market. We enable our partners to be able to create amazing products with our foundational technologies. And so as you know, many of the telcos around the world are using obviously your Ericsson's and Nokia's and Samsung's and Fujitsu's to be able to ultimately offer the radio access network to their end customers.

[11:29] And so that's the work that we are deeply in right now is how can we enable those partners to be able to leverage this new capability of AI-RAN and be able to take their software ported on top of this infrastructure. And so Fujitsu is the first one to effectively do this and actually deliver it into the market. And that's what we saw with Softbank.

[11:50] But you probably saw several weeks ago, we made a big announcement with T-Mobile and Nokia and Ericsson going down the same path to begin to port their software on top of this new NVIDIA AI Aerial architecture that will ultimately allow them to bring these services into the telcos at scale.

DR: [12:10] Yeah, awesome. And so let's talk about the other announcement Jensen made. Indonesia tech leaders announced the Sahabat-AI model family built with NVIDIA AI to advance

sovereign AI in the country. What can you tell us about this vision?

Chris: [12:23] Yeah, so we've been working very closely with IOH and Lintasarta, which is a subsidiary to be able to stand up an AI factory in the country of Indonesia to support the agenda for Indonesia overall. And the first thing is how can we make Generative AI available to all the citizens in their language? And so the work that was done, Tech Mahindra actually worked closely with IOH to build out this large language model utilizing NVIDIA's software tools and hardware capabilities to be able to make this as an open source large language model for Indonesia.

[13:06] And what's exciting then is that this foundational model is going to be able to be used by all other businesses and system integrators and others that want to be able to build different generative AI solutions for Indonesia. We absolutely believe that every country should be looking at what they need to do to create the right foundations and the right infrastructure to be able to participate in the generative AI wave.

[13:35] That every country needs to have sovereign AI, and every country should be producing their own intelligence. And so it's exciting to see that IOH has leaned in to become that national champion to be able to stand up the infrastructure for the country and as well as begin to bring additional services on top of that. And I would say that they're a great example of what we're seeing actually happen, not just there, but around the world.

[14:03] Telcos are oftentimes the biggest trusted partner to the government for a lot of the services they are traditionally providing. They build out infrastructure. They've got relationships with businesses. And all of these things lend themselves to being a great partner to the country's governments to be able to stand up the right infrastructure to advance the AI agenda.

DR: [14:25] And so this is actually my next question. A lot of telcos missed the cloud computing wave despite having invested in building significant massive data centers to support their networks. They're still behind on cloud, but now we're like, okay,

they're going to be able to do with AI. What makes NVIDIA think that telcos are positioned to succeed with AI infrastructure this time around?

Chris: [14:47] Yeah, so I think there's a couple of things that we believe put the telcos in a unique position to be successful. First and foremost, this is a multi-use investment, and I say that because of this. Every telco is looking at how they can apply generative AI to their own operations to drive operational efficiency. And that means that they obviously need the infrastructure to be able to advance what they want to do internally to make better customer experience, to drive improvements in network operation, to drive better employee productivity, and the list goes on.

[15:19] So there's already going to be some internal demand for compute and generative AI capabilities for themselves. The second thing is obviously they do have relationships with the governments, with businesses, and the importance of every business and every government wanting to apply generative AI with a trusted partner that give them the sovereignty, the security, and the performance that they need to count on.

[15:47] And telcos, again, I think are in a good place to be able to offer the ability to bring this capability into the market. And then the final piece that also makes a ton of sense is that if that same infrastructure that they're investing in today can be used in the future to run the radio access network on top of, now you're eliminating the need to buy purpose-built infrastructure just to run the radio access network.

[16:14] You're going to take advantage of one capital investment that can be used to run the network and also serve your own needs and your end customer's needs. And so I think this whole idea that you've got multiple ways in which the telcos can use this makes it a very different situation. And then I think the last piece is that this also does require telcos to invest in making sure they've got the right skills and the right structure to be able to go and offer those services effectively in the market.

[16:42] And so that's a piece that we also work closely with telcos that want to go down this path is how do we train up the skills and then bring the ecosystem partners in that have

expertise across industry to help them be successful on that demand side of things.

DR: [16:56] Well, absolutely. They're taking on a couple of big initiatives simultaneously, but certainly, this concept shows tons of promise, and I hope it's wildly successful. They're also tackling the network API idea, which is opening up their networks and extending them to developers. And so we're pursuing these different strategies. Is one going to pay off more easily than the others?

[17:18] I can't wait to see how it all turns out. But exactly what you said, how they're transforming their operations internally and then how they're extending it to their customers. I mean, if they can leverage the investment in the RAN for this new business idea and they're building it anyways and they have anywhere between 50 and 70% extra capacity that they can now leverage to run AI workloads, it will be amazing.

Chris: [17:38] Yeah.

DR: [17:39] Speaking of successful business ideas, we both live in Texas. I live in Austin, you live in Dallas, and we both love food. And recently the Michelin Guide that ranks the best restaurants came to Texas for the first time. Texas actually has some amazing restaurants.

[17:55] One restaurant in Dallas called Tatsu received a Michelin star and seven restaurants in Austin got a star. No twos or threes, just ones. And so if you could give the Michelin Guide advice, which restaurant would you recommend to put on the list in Dallas?

Chris: [18:10] So they did get a few of these right. We didn't get too many stars. We did get several recommended. And on that list are some of my favorites, like Georgie, Monarch, and Quarter Acre, but I'll call two shout-outs. Gemma under the Bib Gourmand section and that's right around the corner from my house, and I highly recommend that as one of my favorite places.

[18:30] And then one that wasn't on the list that is one of my all-time favorites is called Town Hearth. It's one of the most

exciting steak houses in the arts district of Dallas, but that one would be one I would expect to see in the future.

DR: [18:41] Yeah, no, for me, it's Uchi.

Chris: [18:42] Uchi, yeah.

DR: [18:44] I think Uchi is maybe the best sushi restaurant in the United States and didn't get a star, and I think everyone in Austin was pretty much shocked about it. They're kind of like the nexus of producing excellent chefs that have gone on to build other amazing restaurants. So kind of bummed about that. I will go on my own little personal protest and eat there every week like I do.

Chris: [19:04] There we go.

DR: [19:05] And continue to love that restaurant. Well, Chris, this was such an amazing enlightening conversation about what NVIDIA is thinking about, and so thank you so much for coming onto the podcast. It was a great conversation.

Chris: [19:15] It was awesome being here, DR, and we're so excited to have the opportunity. So you have a wonderful day.

DR: [19:20] Yeah, awesome. Thanks. Stick around. We end each podcast with a "Telco in 20" takeaway. I've got two minutes to tell you something you need to know. At the moment, there's a lot of big new monetization ideas floating around telco. Chris and I just talked about selling underutilized network capacity to enterprises to run their AI workloads. The other big idea in telco is around network APIs where McKinsey & Company projects that it can unlock up to $300 billion in new revenue over the next five to seven years.

[19:56] These are both great ideas, but remember what former CEO and co-founder of AOL Steve Case said about implementing strategy, a compelling vision without the ability to execute is just a hallucination. For both of these initiatives to succeed, they hinge on one critical factor, your enterprise sales team's ability to execute. Telcos have amazing relationships with governments and enterprises.

[20:21] You've worked hard to become their trusted partners, but selling network APIs and AI workloads is not the same as selling traditional telco services. Most telco enterprise sales teams will struggle to find the right buying persona and will have a tough time communicating the value of these new technologies. So here's my advice. Invest in your enterprise sales team now.

[20:41] The operators that transform their sales capabilities will be the ones that actually capture the billions in revenue that's on the horizon. Need telco-specific CPQ software to help your team create quotes for these new products? I may have just acquired the telco software company, CloudSense, that has a solution that can help your team. DM me on LinkedIn or X @TelcoDR to continue this conversation.

[21:02] Until then, tune into more "Telco in 20" episodes. Like and follow and leave us a five-star review. Don't forget to sign up for my must-read email newsletter on TelcoDR.com and check out our killer YouTube channel. Later, nerds.