Scotiabank Inaugural Global Technology Conference

Company Participants

- Will Grannis, Vice President and Chief Technology Officer, Google Cloud

Other Participants

- Steve Sparkes, Chief Information Security Officer and Head of Enterprise Infrastructure, Scotiabank

Presentation

Steve Sparkes {BIO 4981861 <GO>}

Good afternoon. Thank you for joining us for the first Scotiabank Tech Conference. Really glad to have everybody here. We're thrilled with the participation.

And I'm even more thrilled to be on the stage.

My name is Steve Sparkes, I'm the CISO and Head of Enterprise Infrastructure for Scotiabank.

And I'm joined today by Will Grannis.

Will is just an excellent, excellent, excellent individual all around.

But before we get into the Q&A that we've prepared, I'm just going to read the mandatory safe harbor statement for the benefit of everyone's safety.

So some of the statements that Mr. Grannis may make today could be considered forward-looking.

In fact, I actually hope that many of them will be. But some of them could be considered forward-looking.

These statements involve a number of risks and uncertainties.

Actual results could differ materially.

Please refer to Alphabet's Form 10-K, including the risk factors section, and the 10-Q.

Any forward-looking statements that Mr. Grannis makes are based on assumptions as of today.

And Alphabet undertakes no obligation to update them.

That is the end of the safe harbor statement.

So now I'd like to tell you a little bit about Will, for those of you that don't know him. I've been lucky enough to have a few interactions over the years.

And really, again, thrilled that he's here today. He's the Chief Technology Officer at Google Cloud, where he has a team of tech execs and engineers who are helping get the cloud to business.

Been there for.

Will Grannis {BIO 23031167 <GO>}

Almost nine years now.

Steve Sparkes {BIO 4981861 <GO>}

Almost nine years. From the early days of cloud.

And as all of us have seen, Google has invested substantially in cloud to make it a powerhouse platform. And Will is a big part of that.

Before joining Google, Will was an entrepreneur, a tech exec in a number of different sectors.

So it's not just finance, although that's obviously our particular sweet spot today, but been a developer and engineering leader, still very, very close to the tech -- a CEO, Director, has been in the past, graduated from West Point, jumped out of a bunch of planes, landed on a bunch of people, lived to tell the tale.

Questions And Answers

Q - Steve Sparkes {BIO 4981861 <GO>}

And so today, we're going to get to hear a little bit about what you do and how you see things developing. So we'll kick off just with a little bit of background of what does that role encompass for you as CTO for cloud?

A - Will Grannis {BIO 23031167 <GO>}

Great. Well, it is certainly an honor to be here with all of you. Steve, great to see you again. Great excuse to bump into you. And so my role in cloud, I lead a team.

There's two things. Number one, for our top customers and partners that are moving to the cloud. We help them navigate those first steps into the more complex patterns of adoption.

So -- and this is around the world, every geography, every part of the stacks, all the way from customers that are really interested to optimizing against infrastructure, all the way up to SaaS and everything in between the platform. An example of this, three or four years ago, we had customers asking us to create natural language interfaces to Documents. Does that sound familiar to any of you based on recent events?

And so over the period, of course, of years, we worked with customers to take the best of research that we're developing at Google and the products that we're developing at Google and helped them realize the potential of that technology. Now you can imagine working with all of our top customers across every industry and every geography. We also learned quite a few things while we're doing that.

And so the second part of what we do is we take these emerging patterns and we start testing them. And so over the years, we did kind of new product strategies, new entries into our road map based on the signals that we receive from the market, our customers and the work that we do.

An example of this is we did some work with Unilever around sustainable sourcing and palm oil and converted that into a bigger push around sustainability for all of Google Cloud, that now includes climate risk analytics. And there are now APIs that Google produces in the Data Commons, where you can get massive access to climate data through a simple API.

So those are the two branches of the work that we do. And one of the best parts of my job is getting to spend time with our customers and see the implementation in the real world because we're really an applied team. We care about making technology useful to all of you.

Q - Steve Sparkes {BIO 4981861 <GO>}

And so how -- that's a fantastic degree of customer engagement from a CTO. And so how has that scale of the team grown? I mean, you're running, what, over $30 billion run rate? That's a heck of a growth story. And how is your world grown to support that?

A - Will Grannis {BIO 23031167 <GO>}

Yes. Well, I think since Q3 of 2019, we've added roughly $24 billion in annualized run rate. So it's a different business than it was just a few years ago. And my history is, back to roughly nine years ago, when we were in this kind of first wave of cloud at Google, we were just starting to create the pathways for all of this technology out into the world.

And then there was a second wave, where it was creating kind of the platform that would allow any industry or any customer, any geo to solve their problems. And now we're really in the scale phase, where we've got an ecosystem that is rapidly growing, over 100,000 partners in our ecosystem today. So almost any implementation (inaudible) we have regional technology and we have a partner ecosystem that can help in the last-mile implementation, planning, training and the rest.

And so for our team, what it's really looked like is making sure that our customers are getting the most leverage that they can from the technology. And I guess, I could kind of break it into the stack, if you want, because this is -- in cloud, we kind of think of this as an end-to-end platform. So any company, any organization can achieve their goals in digital transformation or critical mission.

So that starts at the bottom, which is, kind of if you think of a typical stack, for those of you that are more technology-oriented, so infrastructure. We, today, have over 39 regions globally. We -- for example, we have a commitment by 2030 to operate entirely carbon-free energy, powering all of those data centers, more that we'll build between now and then. So pretty much making sure access to compute is ubiquitous and also sustainable as energy considerations weigh on people's minds, sustainability weighs more and more in company's minds.

If you move up the stack a little bit, so the second part of this kind of differentiated cloud that we put out is really around data and AI. And here is where you can see Google's legacy quite a bit really showing. We're making data accessible and useful in supercritical costs. And for those of you that -- many customers want -- all of our customers want -- they want their data and make it so that it's easily useful to them in building experiences for their customers or in refining their processes internally.

And they don't want to be constrained by the choice of a single vendor. And so for example, things that we've done over the years are a BigQuery. Has anybody actually heard of BigQuery here? It's one of our hallmark technologies. Okay. Great.

I saw some handful at the technology conference. It's great to see you.

BigQuery 2020, we released a capability called BigQuery Omni. And BigQuery Omni allows -- it sically co-locates computes and binaries inside of other clouds, like Amazon, Microsoft, and allows you to send a query across any cloud that you want, without having to egress your data, and bring that back and bring only the results back.

And you can imagine how important that is to companies who have data stores in multiple clouds. So now instead of having to deal with the cost of egress and these complex technical bits, we've solved this through this product we call BigQuery Omni. So that's the next layer of that the stack, this kind of platform layer.

Then if you move up, we've got what we call the security cloud. So this is end-to-end comprehensive security, not just infrastructure security, but also applications security. And we brought in a ton of new capabilities, one of which is Mandiant, to round out the threat mitigation consulting. And then we've also had virus Total Chronicle and a number of capabilities over the years to be able to characterize threats and reduce threat surfaces across the entire enterprise.

Then kind of moving towards the top, collaboration cloud, if you want to think about it like that. So infrastructure platform, infrastructure, data, AI and then up security now collaboration. This is where you get the work done. This is where you create. This is where your workflows exist in your company.

And so examples of this sort of like Workspace. And what's really interesting now is how AI and what we've been bringing to market this year, how AI is now cascading. It's kind of like a thread that runs through all of those from the infrastructure, the data, the platform, all the way up through SaaS.

Q - Steve Sparkes {BIO 4981861 <GO>}

And I think since we -- a few minutes in, and we haven't specifically hit generative AI, but we really must. Yes. There is. Yes, exactly. We have to put at least another couple of dollars in it. I think it is probably worth just starting with the foundational.

How do you think of it -- how do you think of Gen AI being different from Vanilla AI, if we can call it that? And then let's talk a bit about the foundational benefits that you -- that need to be in place for Gen AI to take off.

A - Will Grannis {BIO 23031167 <GO>}

Sure. So let's see. How do we take (inaudible) for generative AI (inaudible) Okay. Let's try it this way. Traditional AI -- we're going to do it this way.

Okay. All right. So in more traditional AI, what you're really doing is teaching machines to look for patterns and then you're using that to create like a classifier or a predictor. This includes most of the work around neural networks over the last 10 years and a lot of classical applications. So you can think of machine vision in this category.

So self-driving cars, you need to teach the machines to understand what traffic patterns and obstacles look like so that they don't come into contact.

Generative AI. Generative AI is really now these AI systems are creating content, and they're creating data that is similar to the data that's being trained on. An example of this is when you ask -- you can ask a chat application today to write you a poem. It doesn't really understand -- I guess some people would argue this, but doesn't really understand the artistic bits of poultry, but it can create content based on similarity of the words that you've used and words that are used in conjunction with tokens, words that are broken in tokens that would comprise the word poetry.

So in this way, generative AI is this kind of an extension and builds on a foundation of neural networks now creating content, not just thinking about predicting the next word, but actually creating the next word or creating an image.

Q - Steve Sparkes {BIO 4981861 <GO>}

And I think one of the challenges for all of us is at what point does it begin to hallucinate? And how can you control the sort of the error rate? And we talked a little bit in this morning's keynote about the legal example, where lawyers fell foul of having trusted ChatGPT's output despite saying, "Are you really telling the truth?" And of course, it lies again.

And -- so I think one of the things that we're thinking about is how do business leaders embrace the power and, at the same time, have some control and have the confidence that they're not about to make a misstep?

A - Will Grannis {BIO 23031167 <GO>}

Well, in my opinion, our experience is really approaching it from a full stack or a comprehensive approach. And what do I mean by that? The design principle behind Google Cloud's approach to AI is offering a complete AI stack. What does that mean? It means that the research that goes into these novel techniques, it means that the processors, both ours and others in the semiconductors and the chips.

The design and the optimization, we spend a lot of time there. But then we also spent a lot of time on the tooling.

Machine learning ops is a significant burden. Gen AI ops, a significant burden for any organization. So we also take it upon ourselves to create a platform that allows people to, without even knowing how to code, start their journey in generative AI. But then it also kind of goes up to the top of the stack, like I spoke about before. We want to give you an always-on collaborator.

So even if you're operating in SaaS, like you're in Workspace, and you're like, you know what, like you're a small business. Maybe you started a dog walking business. There's still time for you f that's your aspiration.

And the good news is Generative AI is here to help. And what it can do is you can express a desire like, "Hi, I really want to track my clients, but I don't know how to get started." And you can give what we call Duet AI, that's our always-on AI collaborator. You can give Duet AI that intent, and it will generate for you a tracking spreadsheet with the typical columns and the typical fields. If you're trying to run a dog-walking business, for example, without you having to think through the structural components or get started.

And so in many ways, at the top of the stack, it's about getting started with getting value without having to architect. We also have brought the same Duet AI into our platform. So for example, let's say, you're using BigQuery, which I hope you do. And

if you're not, there's still time. But with BigQuery, what really matters is and what matters in all analytics systems is structuring good queries.

You want to run over the right amount of data. You don't want to have a query that's not well structured. So it kind of overshoots the data, it's too costly or kind of runs on without bound.

And so Duet AI and GCP, for example, Google Cloud Platform, we have the capability within BigQuery for it to give you a recommended query for a question that you have. And so throughout the stack and for a company looking to get started, one, it's kind of marrying up your strategy and then being able to find your place in that stack.

So if you want to build a foundational model, you care a lot about like, for example, Google, you care a lot about our ability to tune infrastructure and give you optionality between like a TPU v5E, which is what Anthropic just published this week, what they're using tens of thousands of chips in these clusters.

You care a lot about efficiency when you're going for speed to market, but also you don't want to break the bank. But you could also be looking at GPUs, which we host from NVIDIA, and we have a range of them. We even published a -- we now made available an A3 supercomputer machine shape, where you get tens of thousands of NVIDA H100s available to run the most extensive workloads you can imagine.

And so if you're a model-building company, if you're an organization that really wants to innovate there, you need that infrastructure optionality, and you need a partner that really understands deep optimizations in the infrastructure layer. Maybe if you're a company that's like, "You know what, we're going to take foundational models that exist, and we're going to take our own data. We're going to combine them to create some competitive advantage." That's where you need that platform tier.

So what we call this is Vertex. And in Google Cloud, Vertex allows you, without even code, you don't need to know any code at all. You can go into the console. You can find the latest model, whether it's one from Google or you can find Llama two there. You can find models from Cohere, Antrophic, others, in our model garden, and you can get started combining those foundation models with your own data.

Well, then you're like, well, how do I get my own data going? Well, we have a technology within Vertex that allows you to point an indexer at structured and unstructured data. Again, no code, you're just pointing it to the data sources, and it will index it for you. Because, I imagine, in your work, regardless of what your background is in this room, you know that to produce content and produce insights, you also have to gather your knowledge, and you have to structure it in some way.

And that is actually one of the key barriers to using your own data to refine these foundational models is structuring it, whether you're going to bring it to like an

embedding, which is like the spatial relationship of your data. Or you're going to use a knowledge graph. Getting you to that point quickly is really, really important, so we provide that in Vertex.

And then I mentioned earlier, like if you just want to see what the stuff can do and you're not really sure where to kick the tires yet, that's where you can use Duet AI just to get immediate value from generative AI and just speeding up your workflow, getting to an image for a presentation. How many people in this room make presentations? And those of you that didn't raise your hand, you're lying. That's all right.

When you make a presentation, what is one of the key things in your workflow? Like let's say you're trying to convey a concept with an illustration or an image. Well, you have to go find the perfect one. And if it doesn't exist, it actually slows you down considerably in the workflow of creating that content.

It also is the same concept for an online marketer, right, who wants to go out and put out content but needs an image that is specific to them and that they have the appropriate copyright rights and other legal rights to. So in that way, it's really about speeding up that kind of -- that workflow. So that's how companies can get started is like picking your strategy and then picking a spot in the stack.

Q - Steve Sparkes {BIO 4981861 <GO>}

Yes. And I think you mentioned the access to vast amounts of compute, which, obviously, is a prerequisite for the model generation. And I think it's one of those weird things about this moment, where, in the past, you'd see startups emerging and challenging the incumbents. But in this case, the incumbents, actually, have access to the raw materials, whether it's data or compute.

And so it's really the model garden route is probably the way that we will see startups emerging. But I'm just wondering if you -- how do you think about making available that capacity to emergent companies? Because it feels like there's a very significant barrier to entry for them at the moment.

A - Will Grannis {BIO 23031167 <GO>}

Well, what's really interesting is that we're seeing both traditional enterprises and organizations and startups succeed in generative AI. And I'll give you kind of a view of their path, so somebody like a Deutsche Bank.

Q - Steve Sparkes {BIO 4981861 <GO>}

It's a small European competitor

A - Will Grannis {BIO 23031167 <GO>}

Yes. And you know, I just wanted to point out that --

Q - Steve Sparkes {BIO 4981861 <GO>}

For those of you who haven't heard.

A - Will Grannis {BIO 23031167 <GO>}

Yes. They are able to synthesize internal content. So you mentioned, what do they have? They have documents. They have content. They have analysts.

They have this know-how inside their organization. But they didn't have a really quick way to express a question or an idea or a natural language interface to this corpus of data that they have in these documents, and to get some reasonable results back quickly enough to spend time refining.

It was just -- in the past, it probably would have been faster just to have the humans go cascade a bunch of searches, like write all this stuff down and then put it out as content. But now, the pace and speed of a platform like Vertex allows them to index and to create these spatial relationships and this representation of their data fast enough, and in a workflow that now their analysts are reviewing the content internally that's being produced, and they're using it to synthesize a potentially new analysis for their customers.

Health care, another kind of a big traditional enterprise. Many of you've dealt with in the health care system very, very kind of classic enterprise issues around innovation. But a company like HCA Healthcare is using generative AI. They're doing live transcriptions so that doctors don't have to go back to their desks and type the notes. And they can get that live while they're with a patient and so they can spend more time on care.

So these are traditional organizations that are getting benefit from the speech to text and the natural language capabilities of generative AI today. From the startup perspective, and my experience both being a startup Founder, CEO and also kind of a big company person as well, startups, usually, they're seeded by people who have really, really specific knowledge about a very specific problem, and so their advantage is speed.

And so a lot of what we're seeing right now with startups and their success, I mentioned Anthropic earlier, Cohere, AI21 Labs, a lot of these organizations, they can get the scaled AI optimized compute from us so they can focus on solving a business problem or an idea. Like someone wants to put out a chatbot that they view as like a more responsible chatbot than some of the other alternatives that are out there so they can focus on what are the guardrails they want to put in.

What does the output look like that they want to generate, not where am I going to get these -- basically, these AI supercomputers. We take care of that for them. And so in a lot of ways, we're speeding up a startup's ability to get to their first article, their first output, and then they can tune it much, much faster.

Q - Steve Sparkes {BIO 4981861 <GO>}

And on that point about exposing generative AI content to the consumers, one of the things that we've been super thoughtful about at Scotia, one of my colleagues, Grace Lee, who leads the data analytics team, has spent a ton of time on the data ethics program. And we're acutely sensitive to the appropriate treatment of all forms of data, but in particular, consumer data. So we've got the human in the loop in all of our Gen AI programs and the use cases at the moment, and we've got a very bright line that we have not yet crossed for exposing generative AI content to our consumer base. And I'm just wondering, across the broader client base that you serve, where you're seeing that risk appetite and what checks and balances other clients may have been putting in place to make sure that their end users don't befall some of the worst characteristics of Gen AI?

A - Will Grannis {BIO 23031167 <GO>}

Well, responsible AI is something that, in my opinion, is associated with Google in a way. In 2017, we put out our first version of AI principles because we have been using AI in production at scale for years. And so in a way, we bumped into a lot of these issues very early on and have been developing guidance that we hope, and is constantly evolving, but we hope is helpful to organizations of all sizes. But we're product people, we're technology people, we're engineers. And being a technologist myself, I'm always thinking about how our customers can get leverage from this knowledge and this experience that we have.

So for example, I mentioned Vertex's AI platform. What it has -- one of the capabilities that Vertex has is it includes these responsible AI filters that get added to the runs of these generative models inside of the platform, and it will actually give these customers a preview of the potential brand, toxicity, other issues that might come out with this output before they send it into the next stage of their workflow. And that may sound like a little thing to some of you, but to create these filters and the ability to create this kind of sliding scale that an organization can decide, they can decide for themselves how much they want to push the boundaries of their own brand safety within their industry, their own competitiveness, what type of content they want to allow through and they don't. Just even getting them to a place where they can codify for themselves, what the slider bar, where it should be set is really, really important. And so we keep embedding responsible AI into the platform so our customers can benefit from it in kind of an automation in a tooling way because that's what scales versus having to consistently have humans overlooking every single thing that you're doing.

At some point, you need that leverage from the technology.

Q - Steve Sparkes {BIO 4981861 <GO>}

Yes. And one of the things that we've been finding is that it's a great augmentation for tasks. And you talked earlier about the opportunities for transcription and for rapid digesting of data. But we're not really seeing it as a displacement of actual roles in their entirety. And I think that there are -- do you see it as a side kick, as an

aid, as an intelligent partner in people's -- in knowledge workers' activities day-to-day.

A - Will Grannis {BIO 23031167 <GO>}

Absolutely. We're in an era now -- and I've been building data products, data-centered products, AI-centered products. My undergrad is in linear algebra, which was never cool until now. So I feel like I'm back. There was like -- there was a winter there for me.

But now it's cool again. And what's really interesting is the accessibility to advanced capabilities that is being created in all layers of the stack. I mean just the computation. One of the reasons why neural networks took off so prominently roughly 8, 9, 10 years ago is because the computational efficiency was at a place where you could reasonably now start to run your own networks at scale. So that created kind of this first wave.

Now we're in the second wave, where generative AI has kind of made AI more accessible. It doesn't matter if you're in IT, if you're in the business, if you're in finance, if you're in any part of a company now, you can get the benefit of AI because it converts your kind of natural language way of interfacing, rationalizing the world into computer language and machine language. And this is pretty profound. I think about it as Duet AI. All you have to do is tell it what you're trying to accomplish in the surface of spreadsheets or creating a slide presentation or even within the operations of an application on a cloud, and it will help guide you to more efficient queries.

Hi, maybe you're using the wrong size disc. Based on the IO properties of this application, maybe you've overspent on this and maybe you ought to look at a different shape or you're not really accessing this data all that often. Maybe you should look at archival storage versus running it on this kind of hot, high-performance storage all the time. And those little things start to really add up.

I was working with one customer in the gaming industry, and they were really interested in this concept of a digital concierge. And in this meeting with the leadership team, it was really fascinating because now it used to be a presentation where like the CEO and the CTO and the CIO would go up and they give a presentation about like here's how you do all the stuff and everybody would be like, wow, that seems like really weird. There's a lot of math and a lot of computers and a lot of network and a lot of stuff, and that's really not relevant to a lot of the kind of like day-to-day experience building that a lot of the business is trying to do. And one of the things generative AI has done is kind of broken down the silos between these functions allowed more people to participate in the creation of a digital concierge or in the creation of experiences for customers. And I guess one of the experiences for me in AI that really was an eye opener was how we were working with NASA.

And they're like, "Hi, we have some interns, and we'd like these interns to be more productive during the summer. " And they're like cool, okay. How do we do that? And they literally like they sleep on the floor because they run these jobs and they're

looking for a planet, they're looking for explants, this particular group. And they're like they have to sleep on the floor because these machine jobs, they run for a while, they may stall out, they don't work, and so they have to be there all the time.

It's really sad being an intern there. And just by providing a hosted notebook with some small GPU attached and orchestrated by Kubernetes in a cloud that could kind of like scale up and scale down, they discovered a whole bunch of exo planets without having to sleep on the floor, in the office, happier interns, scientific discovery, groundbreaking new discoveries in science, and all just by providing kind of the simple platform for AI. I mean that's the era that we're in right now.

Q - Steve Sparkes {BIO 4981861 <GO>}

All right. Well, I'm going to bring you back down to something that's of immediate concern to me was just how can we use AI to look for security holes? And do you see it as an arms race between the threat actors using AI to find those holes before the defenders can plug them?

A - Will Grannis {BIO 23031167 <GO>}

Okay. So let's see. We're in your domain area now.

Q - Steve Sparkes {BIO 4981861 <GO>}

Just a little bit.

A - Will Grannis {BIO 23031167 <GO>}

All right, big CECL. I'm going to try to do my best for you.

Q - Steve Sparkes {BIO 4981861 <GO>}

We need help.

A - Will Grannis {BIO 23031167 <GO>}

To your statement, it was never made. Okay. I think it's -- what is it, like $10.5 trillion in kind of cyber crime by 2035-ish on an annualized basis? So yes, this is a really big deal. And at least, what we're doing, try to be helpful in this moment.

And what we've been talking about is, one, making sure that there's comprehensive security for applications and programs that customers entrust to Google Cloud. So for example, we have a security command center that now integrates a lot of telemetry coming off of it. It used to be you kind of have -- and you would know this extremely well, Steve. You have kind of applications that deal with security for different parts of the stack. So like endpoint security, application security, network security.

And one of the things that we've done in Google Cloud is we deployed this thing called Security Command Center, and it's kind of this wrapper that gives you telemetry into every layer of the stack, from an application all the way down to the

infrastructure. And that's really important because these threat actors are very capable, and they're constantly changing tactics. And so the telemetry and the data and the log is actually super important because the tools will evolve. But having that foundation of data is really important, so we're creating a lot of telemetry there. But even more importantly, in my opinion, and gaining us some leverage is we've created domain-specific models like Sec-PaLM 2, where we take the expertise of Google threat groups, where we take telemetry and logs and events from.

We're pretty -- it's no surprise, we're a pretty big target as well for cyber events. And we take all this know-how, we take this data and we train these foundational models, in this case, PaLM 2, our pathways language model 2, and we kind of bring the security angle to it, combine this data. And now Sec-PaLM two is a model that's available to customers of Google Cloud.

So you can bring -- it's already kind of tuned for the domain. And then you can use your own observations, telemetry and data to further refine it and create even new capabilities within your own organization, which may even be differentiated at the firm level, which is an entirely new form of leverage. So definitely focused on cybersecurity. I mentioned that's one of the foundational layers of the stack in Google Cloud.

Mandiant and the addition of, really, that threat action gave us insight into. Anybody deal with -- you're not allowed to raise your hand, Steve. But anybody deal with malware detection threats in your job? So yes, so you may know that there's this Yara kind of framework language that is all around malware identification and threat detection. It is like a new programming line.

There's like a different language completely. And so one of the things that we did inside of our security console is we've used Sec-PaLM two to create auto-generated summaries of malware events and threat mitigation and remediation in natural language. So you don't have to be a cybersecurity professional and understand the semantics and all the nuances of Yara to actually get started on threat understanding and threat remediation. So what does this mean? It means we're leveling up the security capabilities of an organization, both in the tooling and the people.

And in my opinion, the biggest risk is the lack of cybersecurity skills that are available, advanced cybersecurity skills that are available. So you're going to need that leverage from your tools and your platform. And you're going to need your entry-level folks to have more capabilities without the 10 or 15 years of threat-hunting experience that you would typically need to operate at that level.

So even just these auto-generated summaries can quickly orient people on remediation. And that could save, in real workflow terms, I mean, this could save hours, days, weeks, of threat curation and what to do next.

Q - Steve Sparkes {BIO 4981861 <GO>}

Right. And those hours, days and weeks really count because the rate of exploits is accelerating constantly. So any time you can shave off the detection and remediation

cycle is incredibly valuable. And in the past, we might try and use simulations and follow some of the other more conventional learning methodologies. But being able to address the lower level of the stack is absolutely a value add, for sure.

One of the threats that we're -- one of the opportunities and the threats that we're interested in and welcome your opinion on is the use of Gen AI as a co-development platform. Because we've seen, and I was talking to Mike earlier, and we -- their acceleration of commit time was impressive. And I don't want to quote the number in case it's too sensitive, but it's certainly in the double digits percentage acceleration of the time to release code. And conversely, the risk of inappropriate code being injected through that process is something else that we're concerned about. So I wonder, firstly, to what extent you -- within your own teams are you using augmented development techniques?

And then secondly, how are you protecting against that risk?

A - Will Grannis {BIO 23031167 <GO>}

So code completion, code generation, understanding the providence of certain code bits, and this is all foundational to what we've been working on for years at Google. And it shows up in a couple of different ways. One is through a partner like Replan, one of the most successful IDEs, integrated development environments for developers, tens of millions of developers on it right now. And they're using our models underneath the hood to help their developers with cogeneration code completion and really, really speeding things up. But even within -- let's say, you're in the GCP environment and using a development environment there, this is where Duet comes back into play.

And I mentioned some earlier examples around workspace and like slides and sheets, but we also have -- Duet can help in code understanding, in addition to completion and generation, which, if you've ever been a developer and showed up to an organization and been told, "Hi. Go check out this code base and tell me what this means," you have felt a significant amount of panic and pain immediately. And then you're trying to find out what the intent or the design principles or what this code was intended to do can actually take a really long time. And so one of the capabilities that we're really excited about in Duet, and it kind of underpins us, is just understanding code and summarizing what this code is trying to do. In the Nexstar annual conference here, we showed examples of Duet being able to create summaries of this is what this code is intended to do and also be able to look for cases of like risky dependencies, outdated frameworks, other things that can also kind of help you in refactoring.

Q - Steve Sparkes {BIO 4981861 <GO>}

Yes. And as a former programmer, now recovered, I can say that I amplify Will's thoughts and I pity the poor people that had to read my code to try and figure out what it was doing, but that's a long time ago. Let's go back a little bit earlier. You talked about the capabilities of Omni -- BigQuery Omni across multi-cloud. I think multi-cloud is obviously -- it's a thing that we are all evaluating and embracing.

So in common with a lot of other organizations, we've got a hybrid environment, where we've got a bunch of compute still on and data is still on-prem. We've got footprints of both data and compute in multiple other clouds. So maybe just spend a moment on thinking about is that BigQuery Omni an example of how there's some, I wouldn't say collaboration co-opetition, where there's an opportunity to have some horizontal capabilities extend across the multiple clouds? And how should we be thinking about taking best advantage beyond Omni be?

A - Will Grannis {BIO 23031167 <GO>}

So I think about this because I spend a lot of time with customers in my job, my roles before coming to Google were all in the enterprise large organization, like startups. So I find a problem, I go try to solve it, and then I'd realize for scale, I have to do it from inside a large organization. I always start with what's the business problem you're trying to solve because that really dictates how you view multi-cloud as an organization. So for example, disaster recovery, continuity of operations. That's one form of potential multi-cloud.

You may decide that it's in your best interest that you want to run -- you want to have data stores across multiple clouds to meet regulatory requirements and for Coop and DR purposes.

You may also decide that you have a bunch of data in different clouds or even on your laptop or in your own data centers, and you may decide that you want to have optionality to not have to maybe converge all of them into one big data lake or data ocean or data universe. I don't know what's next. But something like Alloy DB Omni, which we just released earlier this year, in postgresql database, and it can run on your laptop, it can run in your own data centers, it can run in our cloud, it can run across other people's clouds. And it's the continuation of this thread of, if you want to have consistency in implementation architecture, but you don't want to -- now is not the right time to make a choice to constrain the number of clouds or hybrid or even your own laptop.

There are tools that we're making available and databases and analytical tools to run across all of those different services without you having to commit to a convergence event. And so the way I think about it is, if you want to have -- the business problem you're trying to solve should drive how you think about multi-cloud and then seeking out a solution that matches that business imperative rather than saying multi-cloud is a tenant onto itself and try to project it into a solution.

Q - Steve Sparkes {BIO 4981861 <GO>}

Yes. And I think it's really interesting when you look at the -- you shoot for the lowest common denominator, do you find -- there are some really terrific companies that are multi that support the -- in particular, security operation line, thinking of Concourse Labs who operate across multiple clouds.

And you can see that there's a clear distinction between those companies that have deliberately aimed to be multi-cloud from the outset to support that capability versus those that were sort of incubated on one and then grudgingly accepted that

people are going to have multiple clouds, and they were obliged to support multiple different platforms. But --

A - Will Grannis {BIO 23031167 <GO>}

Well, choice, again, as a technologist, I want to solve my problems as quickly as possible at the least friction. And this is a thread that actually runs through. I mean, if you rewind the clock eight or nine years through Google Cloud's history, you can see the earliest beginnings of this in the surfacing of Tensor Flows and open source framework for machine learning. You can see this in Kubernetes for container orchestration. You can see it through these omni versions of data analytics and data stores.

You can see it across Apogee, which is a multi-cloud API management capability. I mean, this is a consistent drumbeat and demonstration of a principle of being open and giving choice that now even is now extended into the generative AI era where you pick your model, you pick the workflow, right? You pick what data stores you want to bring to these models and do so safely in your own environment without having to expose them. And so this continuous drumbeat of choice is core DNA for Google and Google Cloud.

Q - Steve Sparkes {BIO 4981861 <GO>}

Yes. And that partnership engagement model is one that I think I was keen for you, just to double-click, even just slightly more on that. Do you have a group that's dedicated to figuring out how to drive those partnerships? And who to bring in? And how -- what are the selection criteria you use?

Just a little bit more on that.

A - Will Grannis {BIO 23031167 <GO>}

It was probably the biggest -- so I mentioned there's kind of been three waves of this business of kind of making sure the engineering primitives were there, making sure that we could serve every geo and every customer the breadth of the platform. And now we're in the scale phase of making sure that like we can scale up and our customers can scale up without any roadblocks. And our ecosystem has grown considerably. I mean we are at now over 100 -- and I said this early. I can't say it again because it actually -- it's a pretty big change.

Over 100,000 partners are in our ecosystem for Google Cloud. And this can range from the independent software vendors who use our capabilities and they kind of -- we ride the channel with them through their software deployments like a Workday or SAP or DocuSign.

But it's also in consulting companies and integrators who help solve problems. Most companies already have relationships with other organizations. They have this trust built up over the years, and they want to continue to leverage their preferred partners. And so we've now made it available. So Deloitte and Accenture, Wipro, they all have thousands, if not tens of thousands of trained individuals on Google

Cloud now, ready to help solve problems that are deep implementation problems, maybe you want to -- you're looking at maybe mainframe modernization or you're looking at these problems that typically are very complex.

They're multiyear. They involve integrators. They involve potentially cloud providers. They involve on-prem. Those are the types of situations that work now we're capable of moving very quickly with you.

And kind of it and to continue kind of up-leveling the entire industry, we've also spent an enormous amount of time in training. Over 100,000 machine learning courses are now in the world sponsored or brought to the world by Google and Google Cloud and hundreds of thousands of practitioners now kind of are now in the world that understand both AI, machine learning and Google Cloud. And we think that that's a real business benefit to organizations at scale, right? Because these implementations are complex, and they need to happen over a period -- a's longer period of time. So that really deep understanding of your business, your industry, your geography, nuances and regulatory concerns and the rest.

We launched like our sovereign clouds in Europe. We launched in partnerships with two large partners over there who help us kind of manage on behalf of the countries in which they operate.

Q - Steve Sparkes {BIO 4981861 <GO>}

Yes. One thing I've been fascinated about is to see how Phil Venables has built out the CISO practice within Google Cloud. And I've known Phil for a long time, weighing back in the Goldman days when I was at Morgan Stanley, he was at Goldman, and we're always the healthy competitive tension, but in the cyber arena, it was always a cooperation. And then as he built out the risk practice and then ultimately became the CISO for Google Cloud. I'm curious how much time do you spend hearing from him about threats that he'd like you to build some capabilities to defend against and vice versa?

When do you go knock on his door and say, "Hi, here's something that you ought to be taking advantage of."

A - Will Grannis {BIO 23031167 <GO>}

Well, like I imagine all of your organizations, the relationship between the CTO and the CISO has grown very close in Google and Google Cloud over the years. And I go back to something like a Sec-PaLM 2. Creating a domain-specific model is the union of domain expertise and knowledge and telemetry that someone like Phil would be dealing with on a day-to-day basis and these complex policies that have to get embedded into the response and the guidance that these models might give or the content they might generate. But then also like the technical underpinnings of the platform and the infrastructure and the computation that makes those feasible to operate, creating a duet AI capability inside of a security console that can summarize threats as they come in, in near real time is actually a very computationally complex and intensive process. And so Phil doesn't want to sweat the technical infrastructure,

and I would do a very poor job of describing in detail the nuances of the security concerns and roles.

Q - Steve Sparkes {BIO 4981861 <GO>}

Yes. Awesome. So I'm curious, based on everything that you've seen and everything you know about what's going on, what are the things that you're most looking forward to in the next sort of -- bearing in mind the safe harbor statement, not necessarily Google products, but the -- over the next six months to a year, what's getting you most excited?

A - Will Grannis {BIO 23031167 <GO>}

There are three things that are happening right now that I am extremely interested in and excited about and, in many ways, I think reflects how the cloud market itself has reset itself in the AI era. And that is, number one, the list of companies today that we just briefly talked about, they're in almost every industry, every size, almost every use case you can imagine. And so the access that's being created to advanced computation is really, really staggering. I have two daughters, 21 and 17, and they're in school, various schools and college and high school right now. And they're able to utilize generative AI, sometimes with authorization, sometimes without authorization because the next generation does find ways around.

But they're leveraging these tools and asking questions and getting answers. And they have access to like the world's knowledge in a way that's been distilled and synthesized so that they don't have to like try to ask 5,000 questions because you -- if any of you are raising kids, you know that they give up after not getting the answer to the first or the second question. And so now there's this like democratization and access to the most advanced computation that I've seen in my career, and it's just happening in a way that feels almost invisible and really easy to use. And that's happening at kind of the individual level.

Then there is this massive wave of discovery happening. For those of you that are following AlphaFold and the second generation of AlphaFold, that has already had a dramatic impact on how we think about the building blocks of life and how treatments could be envisioned and might be envisioned in the future and precision medicine, for example, which is something I'm very excited about. I mentioned the scientific discovery outside of our planet. And the thing with exoplanets. Well, if any of you have heard of this organization called the B612 Foundation, they went and took these computational tools and AI and they went and found asteroids that -- just from old data, not anything new, not new telemetry.

But they found asteroids using some -- the first versions, they had to like literally look at images and use experts to try to figure out whether that constitute an asteroid. Now they're finding them. They've found hundreds in a very quick period of time. And you can think about the ramifications of being able to characterize potential existential threats in that way in the future just is so profound. And then, also, the kind of the third one that I think a lot about is you have medicine and you have space.

But then we also have just in the last week, DeepMind. Google has released a way of characterizing crystal structures, millions of new simulated crystalline structures, which might someday power the next wave of development of artificial crystals for semiconductors, solar power and the rest. And so the pace and the cadence of these discoveries is so quick, is so rapid. And I met -- I'm in such a wonderful place to be that I get to see the convergence of the research, the computational realism and these organizations who might only be dozens of people, but now they have the output of an organization 10x to 20x their size. It's really just an incredible time to be a technologist.

Q - Steve Sparkes {BIO 4981861 <GO>}

It really is. And I think it's also a time for us to consider what are the unique characteristics of any company, what is it that they sell? What is it that they produce? How do they do that? And I think we're going to be facing the question of if Gen AI can deal with a lot of the basic toil in our knowledge worker tasks, is it going to be like a calculator that lets people focus on value add?

Or does it just make us stupid over time? And do we lose the ability to innovate? And is it the friction and the toil that actually creates that struggle? Does it create new ideas and innovation? I don't think we're going to know the answer to that one for a little while.

But it's going to be an amazing journey.

And Will, I just want to thank you so much for the time today.

A - Will Grannis {BIO 23031167 <GO>}

Real pleasure, Steve.

Thanks to all of you for being here as well.

Q - Steve Sparkes {BIO 4981861 <GO>}

Thanks, everybody.