

This preliminary transcript is provided for the convenience of investors only, for a full recording please see the [Goldman Sachs Communacopia + Technology conference webcast](#).

Eric Sheridan: Our next speaker is Thomas Kurian, the CEO of Google Cloud. Thomas joined Google Cloud as CEO in November of 2018, bringing deep enterprise experience to the company. He's grown the business into the world's largest public cloud, with nearly 100 billion annual revenue run rate.

Some of the statements that Mr. Kurian may make today could be considered forward-looking. These statements involve a number of risks and uncertainties that could cause actual results to differ materially. Please refer to Alphabet's Forms 10-K and 10-Q, including the risk factors. Any forward-looking statements that Mr. Kurian makes are based on assumptions as of today, and Alphabet undertakes no obligation to update them.

I'll welcome Thomas to the stage to go through a handful of slides, and then we'll have a bit of a fireside chat. Thomas.

Thomas Kurian: Thank you, Eric.

Google Cloud is Alphabet's enterprise division. We take the products that Google builds and bring it to small enterprises, large enterprises, and governments around the world.

Over the last several quarters, we continue to see strong growth for our business. We're winning new customers roughly twice as fast as a year ago, deepening our relationship with existing customers, capturing market share, and as a result, growing both top line and operating income.

Now, the core of it comes down to the fact that we have highly differentiated products, and we remain the only provider that offers first-party solutions across the entire stack.

We offer our own accelerators, and as the demand for training, inference, and computation grows, we're seeing lots of customers choosing us for infrastructure.

On top of that, we build our own frontier models, and we've seen strong growth with Gemini in large companies, as well as in small companies, small and medium enterprises.

Organizations increasingly want to connect AI to their internal systems, and they want to do three things with them: Understand the data, secure their company, and automate processes or workflows within the organization.

And across our portfolio, we have strength in data management, and we connect our historical strength and data management of products like BigQuery and machine learning tools. We connect AI to that, in the same way we connect AI to our security portfolio to help people with defense, and we offer a suite of applications from our collaboration applications like Workspace, to domain-specific applications.

What's the value of having this diversified portfolio? First of all, it allows us to play in many large markets, each with individual large total addressable market segments.

Second, as monetization of AI evolves, it allows us to diversify and capture revenue and share across every way. Some people buy silicon from us. Some people use our models. Some people use our data platforms. And so it diversifies the revenue streams that we capture.

Third, because organizations want all of these things to work together, it differentiates our products from the point of view of quality, performance, and cost. Today we have over 17 product lines each over a billion dollars. They're growing quickly, and it shows you the diversification we have.

We've seen more than 2x growth year-on-year in new customer acquisitions, and also 2x growth quarter-on-quarter and year-over-year in deals over \$100 million, so large deals.

And when a customer gives us a commitment, say, for \$100, typically they spend more than 50% more than that. And so that's reflective of the adoption we're seeing and the growth we're seeing with our customer base.

Now, let's look at each piece briefly. At the infrastructure level, even this morning, there was an article showing the total cost of ownership benefits of our own silicon. We offer the best computational infrastructure for AI, and we offer three types of silicon: Nvidia GPUs, our own tensor processing units, and as increasingly as models generate code and need to run code, we also offer our own ARM processors to run that.

We offer 2.7 times better price performance for training, 80% better price performance for inference, 30% better price performance for CPUs. All of that allows us to

differentiate our portfolio from other providers. It allows us to offer solutions to financial markets and capital markets, so a lot of hedge funds. Deutsche Börse, for example, is using us.

And so beyond just the labs, capital market firms, high performance computing, people like Pfizer and a number of organizations in the government, the Genesis Mission in the United States government, for example, for energy, are using it for high performance computing, and of course, the number of labs as well.

Now, our infrastructure gives us three important advantages. First of all, these are highly differentiated products in large markets.

Second, they allow us to optimize the cost of our products on the silicon, so we co-design our products top-to-bottom. And so when we launch something like Gemini Flash 3.8 and people go: How come you guys are 2x better than anybody else in the market on inference per dollar? It's our tokens' real intelligence per dollar, it's because we can optimize that whole stack.

And third, it allows us to support a range of different business models, all of which have strong growth return on invested capital.

The size of our accelerator business, our TPU business is more than twice that of the next-largest hyperscaler.

Second, our payback period on AI servers in aggregate is less than two years, and on our own silicon, it's half of GPUs¹. We have strong payback periods, and the majority of our infrastructure contracts, the total contract value of the infrastructure, is long-term committed five-year contracts.

Now, on this platform, people have started using models beyond just answering questions or chat for task execution. You give the model an objective and you call that an agent, and the agent is given an objective. It decomposes that objective, uses a set of tools to connect into a company, and execute the objective for you and give you back a final answer.

So, our platform is called Gemini Enterprise. It is used by over 90% of the Fortune 100, and thousands of small businesses.

¹ Revised to correct an inaccuracy

It's used in a very specific way. People want to use it as a reasoning agent. So, break the plan, understand the steps that are needed, reason on it, then execute the steps.

So it uses a reasoning agent to understand all the information in the company, to then automate that workflow process. And when it does it, you want strong controls. What kinds of controls? Companies are worried about security. They're worried about auditing, what these agents are doing. They want to manage costs and set budget caps. We have all those controls. And we allow people to use the right model for the right task, so you don't have to always use the most expensive model, saving people a lot of money in doing so.

We have a range of companies from insurance, Signal Iduna there is the largest insurance company in Germany. They use it for claims and underwriting analysis. We've got PepsiCo using it for supply/demand planning. Macy's is using it for retail commerce, and there's a whole range of companies doing it.

Now, on this infrastructure, we're also building domain-specific solutions that's built on this foundational platform. What do these domain-specific solutions do? They do the three important things that people want to use agents for. Help me understand my data, protect me from cybersecurity threats, and transform my applications and processes.

For each of these, we bring together a stack of capabilities. Think of it as you're at PepsiCo and running a demand plan, and you want to model the forecast. The first thing you need to know is where's all my data. So we offer a platform that stores and manages a lot of data and can connect to other clouds to get the data.

Second, tell me which of my different products are seeing growth, and in which countries and what metrics. Is it revenue? Is it inventory? What is growing? So you need something called a catalog that gives you the definitions of those things. You then need to execute a machine learning model to run the calculation for your forecast and surface up the answer.

We do that super well. Example is our cost to run these models, machine learning models, 2.5 times better than the number one in the industry. And second, we do that with much greater accuracy, because the definitions of your financial metrics that Pepsi is looking at is in the catalog.

Same thing from a security point of view. When cyber risk grows because of threats from AI models, there are two things people care about: How fast can you find

vulnerabilities and how quickly can you repair them.

We do that in the same way. By connecting AI to our Wiz platform. Wiz gives you the same capabilities. It helps you define what are your applications that need to be protected, which ones are high risk, have they been compromised, help me remediate them, and then test that I've actually fixed the issue. And we do the same in a number of application domains.

So we are not just building a base platform. We're also integrating it into the core workflows of a company to help people understand their information, to help them modernize the processes, and to help them protect themselves.

The benefits of these solutions. First of all, it allows us to bring the investment we're making in Gemini into many more markets. Number two, it helps us enhance the frontier. If you look at the frontier, the big challenges are can you give the model new skills, like forecasting, or long horizon planning? Can you help the model understand private data? Can you help the model behave in compliance with guardrails for safety? So it allows us to build that frontier.

And lastly, it allows us to leverage the strength that we've had in Google Cloud's platform for many years, the strength we have with our analytics business and our database business, the strength we have with our cybersecurity portfolio, and the strength we have with our applications to bring AI along with that to a broad set of customers.

90% of the Fortune 100 use our Gemini Enterprise portfolio. Separately, 90% of the Fortune 100 also use us for cyber defense. 80% of Google Cloud customers now use our AI products, and those that do use our AI products use 1.8 times as many products as those that do not, showing us being able to strongly upsell and cross-sell into this base.

The lifetime value of a customer who uses our Gemini portfolio in the cloud, over a five-year period we estimate to be 1.5 times. And so all these are examples of things that we see where we're delivering strong return on invested capital.

So, to close, our product differentiation stems from the fact that we are still the only provider that offers first-party solutions across the entire portfolio. Each layer of that gives us the ability to play in a large addressable market. We have strong positions in each of them, and that's what's driving our growth in new customer acquisitions,

deepening relationships with existing customers, growing our partner ecosystem, and driving in aggregate top line and operating income growth. Thank you.

Eric Sheridan: Thank you, Thomas.

So, Thomas, you touched upon a lot of key themes in those slides that I want to come back to, the first being a vertically integrated technology stack. It's so critical to what you talked about today. It comes up a fair bit with respect to the company's approach to market.

Talk a little bit about that integrated stack and how it has a financial benefit for your business as well as it stands to benefit your customers.

Thomas Kurian: So, I'll give you a real example to illustrate. Citigroup is building their wealth advisor on our Gemini platform, Gemini Enterprise platform, and it is designed to offer people financial advice.

When you look at that, we're combining four important capabilities. The first one is the person who is getting financial advice sees an avatar, which is remarkably lifelike, offering them financial advice. They can ask it questions. They can assign it an objective, like my kids are going to college; can you tell me how to moderate my cash flow. That agent is built using Gemini Enterprise, and it uses the strength of our TPU infrastructure to provide extremely efficient, realtime streaming, because you want the agent and the avatar to behave properly.

Second, they're reasoning on an objective. It is looking up financial information about the person who's asking the question, and it needs to be 100% accurate. It cannot give you wrong financial advice, because you'll lose trust in it. That uses our data platform for reasoning on financial information.

Third, you clearly don't want such a system hacked, for obvious reasons. So, it uses our security tools to protect all this. And then because it's going to scale, you need to be extremely cost-efficient in ensuring that these models run and the tokens are served in a very cost-efficient way.

So, that's just one example. There are thousands of these companies who are using this platform top to bottom, and our -- the financial benefit for us, we get to play by offering customers the whole stack, if they want to. We can also play in every individual market segment as it grows. As the way that AI is monetized continues to evolve, is it a

-- do you monetize the model to an API, do you monetize the model through agents, do you monetize the model by driving more cycles on your data and security platforms, do you monetize by driving more silicon? We get to play in all those markets. And that diversification gives us a lot more capability to reach more customers and more markets, as well as grow.

Eric Sheridan: Okay. I want to turn to the topic of AI accelerators and how you're positioning your strategy for AI accelerators relative to traditional computing systems. What are the sources of demand you're seeing, and how do you think about meeting that demand with a combination of your own TPUs and external GPUs?

Thomas Kurian: So we -- you know, we've been in this space for a long time. This is I think our 12th year of building accelerators. There are a number of factors that we're focused on. First of all, there are a variety of different types of accelerators people need for different types of models. You can have sparse models, dense models, mixture of experts. You need sparse cores for certain things. So one type of accelerator definitely does not solve everything that customers need. That's number one.

Number two, we offer specific silicon for training and specific silicon for inference, because the needs of these systems are quite different. For example, training typically requires a larger scale cluster with different kinds of connectivity across the different nodes of the cluster. Inference typically requires a lot more access to disks, because it's loading data off-disk when it reads and writes. So there's different kinds of silicon we offer for training and inference.

From a demand point of view, we see different things, from historically it's not just now the AI labs, but many capital markets firms, because they see the efficiency with which their algorithms for numerical trading and research can run on these chips. We see them migrating -- historically high performance computing, which is solid state computational fluid dynamics, et cetera, are also shifting on these platforms, and so we see a range of these.

And for us the opportunity is to offer this obviously to our Gemini team, and that's our most important priority, continued leadership at the frontier for our models, but we also offer these to our other customers. We offer it in a range of configurations, and some of it runs in our cloud, but we also offer it in other people's data centers, because people want it close to their data, and across every business model we have strong returns in this business.

Eric Sheridan: Okay, understood. Probably one of the most often asked questions I get from investors is about TPU strategy broadly. How do you as a company think about offering TPU systems to third parties and neoclouds, and how do you think through the return on the invested capital from that activity relative to meeting your own demand needs?

Thomas Kurian: As we get into more markets, one of the things we realized was there are many places where customers are putting these machines where it needs to sit, near their own systems. So, for example, if you look at high performance computing, people have collected data for years and years, and there's a ton of data. And telling them please copy all that to the cloud in order to reason on it or build a high performance competing application is not viable. So we put our systems in their data centers.

Similarly, capital markets firms, particularly those that are doing quantitative research, want it close to the venue where they're executing, because from the research they want to execute. So rather than saying please bring all that to our cloud, we say we can put it in your data center.

Now, we offer two or three types of business models. One is a subscription in our cloud. Second is a capital purchase. When you're buying the system to run in your data center, you buy it as a capital purchase from us, similar to buying a hardware system. And third, we also offer it through Neocloud. In fact, we set one up with Blackstone to offer that as a neocloud.

The benefit of these things. One is we have diversification of the way that we monetize it. For example, when you buy it as hardware, we don't need to spend capital expense for data centers, because you're putting it in your data center, and so we don't have the responsibility for that power and space, so we don't need to manage that capital expense. We also see strong free cash flow when you buy that to put in your premises. So we see a range of these business models, and we're diversifying it specifically to manage that.

Just as importantly, we do long-term contracts. So, for example, we're very smart on how we think about pre-positioning inventory. In a number of cases, we may do long-term capital purchases so that the unit cost of components are very efficient for us.

And third, we are also prepositioning, for example, the way that customers buy and use these systems by making sure they're signing long-term contracts with us.

So, all of these different elements on the infrastructure side gives us the ability to diversify our portfolio, and because of the strength we have with the product line, we're seeing a lot of demand for it.

Eric Sheridan: You talked in your presentation around Gemini Enterprise, and at Cloud Next, you talked about it being widely deployed as an agent platform, and you also talked about it being a stimulant for paid user growth quarter-over-quarter. Can you talk a little bit about how Gemini Enterprise is being positioned commercially, how customers are adopting it, and what you see as durable differentiation compared to model-first companies that maybe you're competing with, with something like Gemini Enterprise.

Thomas Kurian: We have today over a billion consumer users. We have millions of small businesses using Gemini through our collaboration tools and Google Cloud platforms, and we also have very large customers using Gemini to run their core functions in their companies. So, as I said, 90% of the Fortune 100.

What is it that we do uniquely?

If you think about a person wanting to use an AI system for building an agent, typically there are four tasks they're trying to do with it: Write code; analyze data; change of process, for example, change how I do my back office reconciliation for cash accounting; or protect me. And "protect me" meaning please protect me from external threats, and please protect me from an agent running amok within my company.

What Gemini Enterprise does is it applies -- it has three important differentiations. One, it gives you a single agent that provides all of these capabilities. It can analyze data, it can generate content, it can protect you from a security point of view, et cetera.

Second, in order to do it well, you actually need to hook up the AI system into the internal processes and systems of your company. So, as I said, we do give you tools to understand your metric definitions, so that when we give you a financial answer, it's 100% right. When we give you security defense, rather than saying I can scan all the applications in your company, which you are free to do, but it will take you a lot of time and a lot of resources, we can be very precise on which applications do you prioritize and how do you fix.

And third, we also understand process definitions in lots of enterprise systems.

Now, the benefit also is that we give you a choice of model. Many people ask: Why do you offer a choice in models? Very simple. Take cyber. No single model finds all the issues from a cyber point of view. So if you scan with just a single model, the risk you run is -- think of it as you will be protected from only those vulnerabilities that that model finds.

More importantly, no model finds a superset of all the other models. If you really want to defend yourself, you have to use multiple passes. One pass with our model, one pass with another commercial model, a third pass with an open source model. And particularly the reason we also say these open sources, many times the attackers are going to use that, so the best way to defend yourself is to figure out what they're going to attack with.

And so that choice gives you actually better capability than if you just had a monoculture with one model.

Eric Sheridan: Understood. Maybe just building on that and sticking with cybersecurity for one quick follow-up. How would you see your strategy with respect to cybersecurity evolving specifically to address AI-driven threats?

Thomas Kurian: Most people asked us when we acquired Wiz, why did we acquire Wiz? We saw in 2023 that as models learned to code, because the most initial skill that models were doing was learning to write code, that they could understand code and configuration of systems, and once they understand code and configuration of systems, they could become extremely proficient in finding vulnerabilities.

So, we acquired Wiz as a platform because it gives us four things. First, it connects to a lot of clouds, not just ours, but enterprise-wide, and it captures the configuration of your systems and what software is running on those systems.

It also builds a risk score on which of these applications are most likely to be compromised. The first thing it gives you is an incredibly efficient way to prioritize what you need to look for.

Second, it allows us to then scan the stuff that you are most worried about, and identify vulnerabilities, and our Gemini platform hooked up to Wiz is being used by many, many companies now to scan and find vulnerabilities.

Finding vulnerabilities is only half the problem. In fact, it's less than half the problem, because repairing it is the bigger part of the problem. So we built with Wiz a product called CodeMender, which is repairing. The name pretty much says "code mender." It repairs your code.

And then Wiz also gives you the ability to also test it.

But that cycle is what has driven over 90% of the Fortune 100 to now adopt our cyber tools to defend themselves from threats due to AI models. You can only defend a threat from an AI model by using a combination of a security platform and an AI system. An AI system by itself cannot protect you. A security platform by itself cannot protect you. You need that combination, and we're uniquely capable of offering that.

Eric Sheridan: Okay. We have a few minutes left, but I do want to sneak in a couple more topics.

You recently implemented a strategy that forward-deploys engineers to help with implementation. Talk to me a little bit, Thomas, about the challenges of implementation and the opportunity that sits if you forward-deploy engineering talent to try to help drive a wider array of enterprise adoption.

Thomas Kurian: We've never been a Services company. In fact, our Services business as a percentage of our total business is the smallest of any hyperscaler. We are a products and solutions company.

So, where do we use Services and our forward-deployed engineers? Our forward-deployed engineers are focused on three important things. The first one, working with the top-most customers in each industry to identify where to push the frontier of our model. So, for example, in capital markets, it's forecasting. In Legal, it's correctness of the actual results that are being presented by a model, because in Legal, you need to be 100% factual. Otherwise, you will have a problem with the lawyers.

So, first thing they do is find the frontier.

Second, they build the tools to allow us to capture the frontier, and what we mean by that is, for example, when we go into a financial market, understanding the actual private data for that company and being able to connect the model with it, our forward-deployed engineers build the tools to do that part.

We make those tools available to all companies, but they are the ones, because of their expertise, they understand here is a specific thing and here's how we're going to go implement it.

Third, they build the training and certification program to scale the business through partners. Even this morning, we announced an agreement with Accenture, where Accenture is building a large Gemini Enterprise business group focused on this, and they're working with us, because our goal has never been -- it's always to enable the partner ecosystem, and our forward-deployed engineers learn what customers are doing, build the tools and certifications, and then we scale it through our partners.

Eric Sheridan: We've only got about two or three minutes left here. Maybe I'll jump to one last final one, which dovetails on the back of what you just said there about Accenture. Google Cloud has announced sort of these new and expanded partnerships in the way in which you're bringing all of these products and Services to market. Talk a little bit about some of the initiatives or the verticals that you're most excited about that contribute to growth as you think about Google Cloud over the next couple years.

Thomas Kurian: So, we're focused on partnerships in three different domains, okay? First one is in specific industries. We've seen that the needs of specific industries are different. For example, the way that a life sciences company wants to use our models, like what Merck is doing using Gemini Enterprise for both science and for corporate operations. So there are eight industries that we are very focused on pursuing, because we see them as having unique needs that help us advance the frontier. That's number one.

And in those, we will work not just with leading companies, but also with the independent software vendor and SaaS ecosystems in that market to build solutions.

Second is the ability to scale the Services community. This is both traditional system integrators, but also new breed AI specialists, and we have that down by market, by country, so that we have it mapped and we have a lot of work going on with that. That's number two.

Number three, we're also working with specific data providers and people building datasets for specific industries. So, for example, in financial markets, Bloomberg, FactSet -- there's a whole range of them -- MSCI, et cetera, they've put their information in our platform so that people can reason much more quickly on that information, and depending on the domain in life sciences, for example, in healthcare, there are different

providers that offer that information, and the goal is to make that information available so that agents and models can reason on them much more quickly and easily, and that's a very focused strategy we have by industry, by country, and by customer segment.

Eric Sheridan: Alright. Well, I think we're going to leave it there. Please join me in thanking Thomas for being part of the conference this year.