

Goldman Sachs AI Exchanges

Will falling costs drive new opportunities?

Kim Posnett, Global Co-head, Investment Banking,

Goldman Sachs

George Lee, Co-head, Goldman Sachs Global Institute,

Goldman Sachs

Allison Nathan, Senior Strategist, Goldman Sachs

Research

Date of recording: January 29, 2025

Allison Nathan: Welcome to Goldman Sachs Exchanges. I'm Allison Nathan. This year we've decided to look closer at the rise of AI and everything it could mean for companies, investors, and economies. So we're bringing you this series of special podcast episodes we're calling AI Exchanges, which I'll be hosting alongside my colleague George Lee.

George is the co-head of the Goldman Sachs Global Institute. He's the former CIO of Goldman Sachs. And before that, he was co-chairman of the Global Technology, Media, and Telecommunications Group in our Investment

Banking business. George, thanks for joining me for this series.

George Lee: Thank you, Allison. It's great to be doing this with you. Look forward to it.

Allison Nathan: I am super excited about this. The goal here is to help listeners understand the impacts that AI is having today and the ways it could change in the years ahead, the implications, as I just said, for businesses and economies. And George, there's really no better time to be having this conversation because of course we've recently seen some really fundamental questions about the AI build-out ripple throughout the markets quite powerfully. We all have observed the volatility we've seen in the markets this week. It's been well reported that a Chinabased AI company called DeepSeek has rolled out a lowcost AI tool which is raising questions about AI infrastructure and about the massive spending on AI by America's biggest tech companies.

George, this is right up your alley. You've been having a pretty active debate with Jim Covello, global head of Equity

Research at Goldman Sachs, about whether the benefits of AI will justify what was expected to be an enormous cost from these companies in terms of developing and supporting the technology. And that spending is well under way, as we all know. And you've been on the bullish side of that debate.

How does the emergence of this low-cost competitor shift that discussion? I'm just going to ask you, does it mean that Jim was right to be skeptical about the huge CapEx spend all along?

George Lee: First of all, Allison, it has been great to have this running discussion with Jim, and it's in such a momentous and dynamic time, as you've discussed. So it makes for a really fruitful and interesting dialogue. I would say, though, that to your point precisely, I would say, much to the contrary, I think this answers some of Jim's principal concerns. Not all of them but some of them.

Two of his understandable and well-founded concerns were the eye-watering capital costs of building this pre-training infrastructure, which you referenced. And then his belief

that the technology would never really get cheap enough to have broad utility inside the enterprise. And so this development, which may -- I would underscore "may" -- promise much more efficient ways of pre-training, may indeed deep into the future reduce the amount of capital we have to allocate to at least that part of this ecosystem.

And further, DeepSeek's pricing measures suggest that we will continue to see very steep declines in per-token costs that make the incremental cost of an intelligent token trend towards marginal zero, which is a very powerful concept.

Allison Nathan: But the question is lots of companies have been spending lots of money. Do you think that spending is essentially not going to be useful spending?

George Lee: Oh, no, again, to the contrary. First of all, this doesn't mark the end and perhaps it marks the beginning of even more pre-training activity by more people who can afford to embark on this at lower capital costs. And so, people talk about Jevons paradox as the price of something declines, abundance tends to increase.

Moreover, I think one important part of this debate is DeepSeek has performed some really interesting engineering hacks to address pre-training costs. These models have already shifted towards much more dense and abundant computing at inference time, at test time, and so that itself, if the price of intelligent tokens decline, it breeds abundant new use cases. And a lot of the computation is inference. I think all the infrastructure that we've sunk capital into now and in the next few years that are planned will be well used.

One could ask questions, as you look farther out the horizon years three, five, seven, 10, whether we'll need the same trajectory of capital costs, but that just comes down to a judgment on this idea of, "Hey, you've got a less expensive commodity. Will that breed more volumes that offset that price decline?" And so the history of technology would tell you that, indeed, that will be the dynamic. But again, I think this marks a really interesting advance in the technology, it addresses many of the concerns market observers like Jim have had, and promises the abundance of these tokens that for places like Goldman Sachs will open up new horizons of use cases.

Allison Nathan: Right. And more cost-effective use cases. And essentially could we see the applications and the adoption of them speed up?

George Lee: Absolutely. And, you know, Jim cites in our dialogue a few applications where the technology is useful for us, and yet it remains prohibitively expensive relative to human capital. And this actually I think changes that equation. And again, this will embrace new use cases we can't even really imagine today, which will be fun and interesting. And again, this is part of a repeated history of the way that technology happens. This feels like a very discontinuous moment because of the sharpness of decline and potential capital costs and token costs. But if you scope back out and you view it in the history of Moore's Law over the last 120 years, even spanning outside of the Silicon Age, or this phenomenon in general, I think it's a measurable but small blip in a steeply declining overall curve.

Allison Nathan: Well, it's really interesting. Let's bring Kim Posnett into this conversation. Kim is the global cohead of Investment Banking in our Global Banking & Markets business and the former global head of the Technology, Media, and Telecommunications Group.

George Lee: Kim, welcome to the discussion. Can't think of a better guest.

Kim Posnett: Thank you for having me, guys.

George Lee: Do you think about this phenomenon the same way? Allison and I just had a fun discussion about the trade-off between price and volume and Jevons paradox. Do you see it the same way? And are your clients thinking about it the same way?

Kim Posnett: I do, not surprisingly, George. If you put aside the global race for AI supremacy -- that's a separate conversation -- I think there's -- and you and I talked about this over the past few days and weeks -- this unambiguously good news. The cost of compute is coming down dramatically. The price per token is coming down dramatically. That means these models are becoming more

cost efficient. It is great for the world that this will be cheaper for all of us, point one.

And Jevons paradox, you mentioned, I think that is absolutely in play where you see increased efficiency that will lead to increased adoption and consumption. There are so many examples across the business landscape where you can see expanded use cases. We've all been talking about automating repetitive tasks, but imagine automating complex processes. So legal assistance, financial services, scientific research.

You know, I was just with an AI researcher last week talking about modeling the immune system and modeling the brain. Think about the implications for the health care industry if we're able to achieve that. As an example, my favorite example, I think this -- I want to know what yours is, too -- is ubiquitous conversational AI. So sort of like personal assistants for everyone in every context.

Personally, professionally. I do think you'll see that. There are so many more use cases we could go into, but those are some examples.

George Lee: Agreed. I think the conversational interface is very powerful and requires a shift in some ways in how you use the technology. And I find myself with my AirPods walking down the street, talking to voice assistant and looking a little peculiar in the process, but it's a very powerful modality to get access to that intelligence for sure.

Also, a parallel phenomenon that people are talking about - - this got drowned out for just a minute by this whole DeepSeek episode -- is the rise of agents and some of the new approaches there. What do you think about that? Are we early in that? Is that a whole new vector of improvement for these models?

Kim Posnett: I think we are early days in AI agents. I believe also that they will be ubiquitous over time. Who knows what the time frame is? You tell me. I think you agree.

George Lee: I do. And I've spent the past weekend playing with Claude Computer Use and the new operator product from OpenAI. And it's very early. It's sort of a proto experience, but it hints at something that's very powerful.

Allison Nathan: Sorry. Can I just ask, when we say "AI agents," for people who are not that close to AI --

Kim Posnett: Go ahead, George.

Allison Nathan: -- what are we talking about?

George Lee: First of all, I'm so glad you asked that question because there is a very broad set of definitions around agents. I'll give you two. One broader, which is a system of models, computation, and resources that complete linked tasks and allow you to complete more multistep complex tasks in business or personal life. So the canonical example is I want to take a trip to Phoenix. Help me book the flight. Help me book a hotel. Help me book a rent-a-car. And it's autonomously executed.

The applications that we're talking about are more, for now, more consumer oriented and basically give the instructions to these applications to resupply something that you need for your house from Amazon. And it brings up a webpage. It takes a picture of the webpage. It's able to discern those

pixels, identify text entry boxes, buttons. It takes a hold of your cursor and begins to execute on your behalf. And it's really extraordinary to watch.

It is early. And by the admission of the people who are developing these capabilities and they want to enroll people in refining it, one market observer, though, had a very funny characterization of it, which is you ask it for a task, it brings up a browser, it starts executing on your behalf in websites, and it's so slow and deliberate and herky-jerky while it's doing it. Reminds you of teaching your grandparents how to use the Web in 1997. But nonetheless, an inspiring direction of travel for the technology.

Kim Posnett: But I think it's an important question because you've got this sort of near vertical advancement of these AI models -- this is an example of that -- which is driving this increased demand for -- and you mentioned this earlier, Allison -- scalability, efficiency, sustainability. And the increased scale and complexity of these AI models today require huge amounts of capital. Equally, they require huge amounts of energy, powered

land, and data centers. And that leads to your question around CapEx spend. It's why I agree with George's answer on the CapEx spent today. Was that for naught? I don't think so.

I'll give you a data point, which is fascinating. If you look at the CapEx spend of -- I'll just pick four companies -- Amazon, Alphabet, Meta, and Microsoft. Across the four of them, they spent over \$116 billion of CapEx in 2022. That was the year that ChatGPT was launched to the public. Roll forward two years later, 2024, they spent just under \$200 billion. That's almost double in a two-year time frame. And you've seen the recent announcements. Meta has announced that they'll spend \$60-65 billion on AI-related CapEx spend this year. Microsoft, \$80 billion. You saw the announcement, the Stargate AI infrastructure JV announcement just last week.

So there is I think a huge amount of CapEx spend that is appropriate right now, given these near vertical advancements. And I agree with George, the question is in three, four, five years, as these models become more

efficient, it's unclear what the CapEx requirements will be in the medium term.

Allison Nathan: But we are hearing very low CapEx spend for this low-cost Chinese-based competitor, so are companies actually rethinking the amount of dollars that will need to go towards this technology?

Kim Posnett: Over the medium and long term, perhaps. And I think it relates directly to what we see on the efficiency curve of these models, and I don't think we know the answer yet.

George Lee: I agree with that. So Kim, one of the other predicates, one of the ingredients you pour into the top of these models to create this intelligence, in addition to power and data center capacity, you cited is data itself. And in your career as a banker, you've done a lot of very datacentric transactions. You know that ecosystem well. Observers say that we're running out of human-generated broadly available data very quickly, if not already. So that sets the mind I think of model makers towards synthetic

data generation or unlocking data that's behind firewalls or are protected and proprietary.

Is there any chance that sort of economy of data emerges around that?

Kim Posnett: Yes, I think so. And my view on this has evolved over the past few years. What is the single greatest bottleneck to AI? Is it data or is it power? I think maybe a year ago I would have said data. I think today I would say power. So anyway, we can debate that.

But I do think that the landscape of data economies is emerging and evolving. So there's new data marketplaces, as George alludes to, but there's also the reshaping of existing markets. And so last year, you started to see new creative partnerships form and data licensing deals. So that was publisher partnerships. That was social media partnerships. That was stock photography partnerships. And I think you'll continue to see that because data has become so valuable.

And then on whether we've run out of data, I think you'll start to see things like synthetic data marketplaces. So that's where AI-generated data mimics real-time data. Think like your medical records and you use that synthetic data to train models. Or personal data marketplaces where you can opt in and sell your own personal data to a business to train a model.

George Lee: It's a great segue to deal-making and what you're seeing in Silicon Valley and around the world in capital formation and new companies and potential for IPOs and sales. There's a whole set of fellow travelers alongside these model providers and infrastructure providers that give infrastructure and tooling and security. Are you seeing the emergence of a lot of those companies? And are they growing faster than prior generations that you and I might have worked with over the years?

Kim Posnett: Yeah, so if you talk to CEOs -- and I'll just focus on the US -- across the US corporate landscape, I think many would say over the past few years they felt headwinds to growth from a monetary policy standpoint, from a regulatory standpoint. And if you ask them today what their perspectives are, and I think there's

a general tone of optimism and a belief that the monetary policy and regulatory environment will ease, which will allow them to be more forward leaning on growth, on investment, on M&A, on IPOs, etc. So I think that the backdrop to dealmaking, especially in tech, is quite constructive today. In the early days of this year -- it's only been three or four weeks -- you've seen a lot of strategic activity. I expect that strategic activity to continue and accelerate throughout the year.

As it relates to AI specifically, I think of AI dealmaking through sort of two lenses. One is capital markets and financing, and the other one is M&A. And on capital markets and financing, we've already touched on a lot of the thematics that investors are focused on, which is CapEx spend, ROI, global supremacy, who will win, who will lose. I do think generally investors are bullish still on AI. There's emerging questions as last week proved.

And then on M&A, I actually think you've seen already AI-driven strategic M&A. There's a bunch of examples from last year. So I think you'll see more strategic M&A specifically related to AI this coming year.

Allison Nathan: And, Kim, you briefly mentioned power as a constraint as well, so talk to us about what you've been learning about that and why you're more concerned.

Kim Posnett: As I said, I debated in my mind what's the bigger constraining, data or power? I now think it's -- do you think -- do you agree with me that it's power?

George Lee: My answer to whether it's data or power would be yes. And I'm not trying to balance them.

Kim Posnett: That's quite clarifying, thank you. But, you know, historically, we've seen decades, literally decades, of sub-3% annual growth in base load power demand in the US, as an example, okay? And now you're seeing this unprecedented tectonic shift in demand for power related to AI. And just to dimensionalize that a little bit, AI servers require -- I don't know -- 10x the amount of power as a traditional server order of magnitude. And you're seeing these companies, the hyperscalers, build these AI data center campuses that are multi-gigawatt centers, okay?

Just to put it in context, that's what it takes to power entire cities. And so I just think we can't underestimate the amount of power needed to run these highly complicated and complex AI systems today.

George Lee: And will likely be a source -- one of the interesting parts about that is it and of itself will be part of a demand function for innovation and power delivery. And so scaling green sources, battery that allows you to store and use that in a less intermittent way, small module nuclear fusion. Again, it's innovation in an industry that's been relatively static spawned by this demand.

Kim, maybe let's end with something that you and I talk a lot about, which is we focused on the steep improvements in this technology, the potential for it to be more broadly useful at lower cost in the world, and yet we see a little bit of a lag of enterprise adoption. I am hopeful this a year where we're going to see that inflect upwards. What are you hearing from clients? What are you observing? And then maybe, last, do you have any interesting personal use cases of it that you can share?

Kim Posnett: Oh my gosh, I use it all the time. Yeah, I think that last year people were still testing and learning and trying to understand applications to their own businesses. And I think this year is the year of true enterprise adoption and scaling. And so I think this will be an important year to see how much enterprise adoption there is across AI and love to hear your views.

And yeah, I just think that there's so many examples I could use, both personally and professionally, about how I integrate AI into my life. I don't even know where to begin. But I, like you, am walking down the street, talking to my AI like my imaginary friend.

George Lee: It's funny because we laugh about it and yet you have to recognize it was only two years ago --

Kim Posnett: I know.

George Lee: -- that this capability was loosed upon the earth in the form of the initial launch of ChatGPT. And we worry about the lag in enterprise adoption. We wrestle

with the amount of capital and the costs associated with this. And yet you look around and there's a generational dimension of this, too.

Kim Posnett: I totally agree.

George Lee: You look around and people joining the workforce or in schools, the way that they fluidly use this technology and are just perhaps in small quanta to begin with but a wedge opening up of more productive. A little bit smarter. A little bit more responsive. And so I think it's again it's just a glimmer of where this will take us hopefully in the enterprise and in our personal lives.

Kim Posnett: And can you telegraph anything to come around Goldman and enterprise adoption around AI?

George Lee: Sure, yeah. Well, I mean, it's in the news. We launched our GS AI Assistant which allows more people across the firm to get access to leading-edge models, to be able to use them in a safer, more reliable, and more compliant way, which befitting our role as a regulated financial institution, it's important.

And again, early days. AI is, in many ways, prolific throughout the firm, but this is the broadest and most general purpose offering we've made. And I think it will be really interesting to see in the coming months what use cases emerge, what innovations, what inventions, what creativity is brought to bear, particularly by our junior people.

Kim Posnett: Yep.

Allison Nathan: George, Kim, this has been a fascinating conversation. Thanks so much for joining us.

Kim Posnett: Thank you for having me, guys.

George Lee: Kim, great to kick this series off with you. Couldn't imagine a better guest.

Kim Posnett: Super fun. Thank you.

George Lee: Thank you.

Allison Nathan: And George, if I take away anything from this conversation it's that there is a lot of good news in these recent developments even if the market's been very volatile around them. And we've said this for a long time now, we are in the early stages of this, so there will be many more evolutions to come.

George Lee: Agree with you. Obviously, as you noted in the beginning, I do have a bullish take on this. But lest I be accused of being a perpetual bull, I think this is in some ways a source of optimism for the future trajectory of the technology. It also poses some questions about the fundamental economics for various participants. As Kim said, it raises questions about longer-term capital spend and how companies and infrastructure builders think about it.

But I think for the near term, this is pretty much good news, and it's part and parcel of fast scaling of an innovative technology and will be incredibly fun to watch.

Allison Nathan: Well, George, this has been fun. I'm looking forward to continuing the conversation in future episodes.

George Lee: Great. Thank you.

Allison Nathan: This episode of Goldman Sachs Exchanges was recorded on Wednesday, January 29th. Thanks for listening.

The opinions and views expressed in this program may not necessarily reflect the institutional views of Goldman Sachs or its affiliates. This program should not be copied, distributed, published, or reproduced in whole or in part or disclosed by any recipient to any other person without the express written consent of Goldman Sachs. Each name of a third-party organization mentioned in this program is the property of the company to which it relates, is used here strictly for informational and identification purposes only, and is not used to imply any ownership or license rights between any such company and Goldman Sachs. The content of this program does not constitute a recommendation from any Goldman Sachs entity to the

recipient, and is provided for informational purposes only. Goldman Sachs is not providing any financial, economic, legal, investment, accounting, or tax advice through this program or to its recipient. Certain information contained in this program constitutes “forward-looking” statements, and there is no guarantee that these results will be achieved. Goldman Sachs has no obligation to provide updates or changes to the information in this program. Past performance does not guarantee future results, which may vary. Neither Goldman Sachs nor any of its affiliates makes any representation or warranty, express or implied, as to the accuracy or completeness of the statements or any information contained in this program and any liability therefore; including in respect of direct, indirect, or consequential loss or damage is expressly disclaimed.

This transcript should not be copied, distributed, published, or reproduced, in whole or in part, or disclosed by any recipient to any other person. The information contained in this transcript does not constitute a recommendation from any Goldman Sachs entity to the recipient. Neither Goldman Sachs nor any of its affiliates

makes any representation or warranty, express or implied, as to the accuracy or completeness of the statements or any information contained in this transcript and any liability therefor (including in respect of direct, indirect, or consequential loss or damage) are expressly disclaimed. The views expressed in this transcript are not necessarily those of Goldman Sachs, and Goldman Sachs is not providing any financial, economic, legal, accounting, or tax advice or recommendations in this transcript. In addition, the receipt of this transcript by any recipient is not to be taken as constituting the giving of investment advice by Goldman Sachs to that recipient, nor to constitute such person a client of any Goldman Sachs entity. This transcript is provided in conjunction with the associated video/audio content for convenience. The content of this transcript may differ from the associated video/audio, please consult the original content as the definitive source. Goldman Sachs is not responsible for any errors in the transcript.