Goldman Sach Exchanges

AI Exchanges: The Role of Data

Neema Raphael, Chief Data Officer, Head, Data Engineering, Goldman Sachs

George Lee, Co-Head, Goldman Sachs Global Institute

Allison Nathan, Senior Strategist, Goldman Sachs Research

Date of recording: September 25, 2025

Allison Nathan: Welcome to Goldman Sachs Exchanges. I'm Allison Nathan, and I'm here with George Lee, the co-head of the Goldman Sachs Global Institute. Together, we're co-hosting a series of episodes exploring the rise of AI and everything it could mean for companies, investors, and economies.

George, good to see you again.

George Lee: Great to see you, Allison. Good to be here.

Allison Nathan: So George, we've had several conversations about how AI is shaping the economic and business landscape, but today we want to actually get a little bit more under the hood and talk about the

technology itself. And in particular, the role that data will play in enabling or possibly stalling its progress.

We have a terrific guest to dive into these issues with us, Neema Raphael, the chief data officer and head of Data Engineering here at Goldman Sachs. Neema, welcome to AI Exchanges.

Neema Raphael: Thanks for having me. Excited to be here.

Allison Nathan: So before we dig into the topic at hand, first just tell us a little bit about how you got here, your career journey, and what your current role here at Goldman Sachs entails.

Neema Raphael: Yeah. This is 20-plus years for me at Goldman. I started right out of college. Studied Computer Science. And during that, I sort of realized that I wanted to apply technology to a domain that I had not known before. And so really finance was sort of like this black box to me. I came to Goldman, met amazing people, started as an analyst here, software engineering, you know, typing at the keyboard, writing code.

And then really the data thing came I'd say five years in. Global Financial Crisis, 2008. Lehman Brothers collapses. And a group of technologists called core strats at the time was going around the firm saying, "Hey, you know, we have to figure out what our exposure to Lehman is. We have to figure out our liquidity profile, see what's going on at Goldman."

And the way they had structured it was to try to get all of the data from the front office, middle office, and back office together in one place to, sort of, figure out the end-to-end exposure to Lehman in a very technology- and data-heavy way. And that was sort of, like, the genesis of I'd say my data journey here.

That project actually was super interesting because we had heard other banks and other financial institutions actually have to go into their filing cabinets to dig out their ISDAs that were signed with Lehman to figure out what their contracts were. We luckily had a lot of our data sort of corralled in one place, and actually that database we built, it was called Copter, ended up becoming this place not only that people realized, like, the power of data not just being

sort of like an exhaust but actually an enabler for the business.

And then not only did people recognize, okay, this database sort of saved the firm in some interesting way, but then when we gave that same data to traders, sales people, strats on the desks, quants on the desks, people started coming up with new innovative ways to use that data for helping our clients and just running the firm a lot more efficiently. And so it became this sort of launching pad for people outside of technology to say, "Hmm, data maybe can be a powerful concept here."

George Lee: That's great. And so in your role as chief data officer now, you oversee all of that. Other things you've done in your career have been involved with what we used to call in the dim, dark past machine learning. The point is AI has been around. We've used it at the firm. It's been broadly proliferated, but the rise of generative AI has garnered so much attention. Is it fundamentally different than the journey we've been on? Or is it just an extension of the continuum of good old-fashioned AI?

Neema Raphael: Yeah, I think I'd say a little bit of

both because it feels like some sort of step-change function from the historical. You know, I always talk about the first 50-60 years of computer science, sitting down and humans have to code rules to tell the computer what to do. And we talk about determinism. Like the rules were deterministic. If you push this button, please do this. Or if you type these keys, please do that. And so it was this really fundamental shift I guess in machine learning in general which is, like, learn by example instead of learn by rules.

And so in some ways, the generative AI stuff is just continuation of learn by example, but I don't think people naturally saw it go from, "Hey, I could learn maybe how to predict some patterns," to now the computer could create anything. And so there's a little bit of that continuum. Like, "Hey, if we just feed the machine more and more examples, more and more data, it could start learning things," is probably the path of continuum. But the sort of step-change was, like, "Oh, well, can we feed it and create images? Create audio? Create language?" And so I think that's sort of the novel step-change in the generative part.

George Lee: And you illustrated something I think is very fundamental in terms of company culture in this shift,

which is we're used to deterministic computing. For a given input, the outputs are correct, repeatable, and traceable. We're no longer in that sphere. As you pointed out, these are probabilistic machines. Something emerges from it that you can't trace and is often right but not always.

Talk about the mindset difference inside an organization of getting business users in particular to be comfortable with that.

Neema Raphael: I'd say a little bit in finance people maybe have understood that because of our pricing models and derivative pricing. I mean, it was also stochastic in that way anyway. So there was always a little bit of, okay, like, the world is nondeterministic and so prices are nondeterministic, the markets are nondeterministic, economies are nondeterministic. So I think there was maybe a willingness to sort of understand that here in the finance world.

But I agree. I think when non-engineers sit at a computer, they sort of want a thing to be a repeatable pattern. That's how we build workflows here. That's how we build client

insights or anything we do here to help our clients.

So I think it's really about teaching people this isn't just some magic crystal ball, right? What it's really doing is taking a lot of examples and giving you an extrapolation from those examples.

Allison Nathan: Let me just ask a follow-up to that, though, because we've had a lot of conversations on this podcast about the ultimate potential of the technology. There's so much hype around it. We're having another, I think, leg up in the hype in the last month or two here. Given what you know about the technology, do you think it's over hyped or maybe even under hyped?

Neema Raphael: Yeah, so as George knows, I'm always a little bit of a skeptic of new technology. Historically, we've talked a lot about blockchain and things like that. And that was supposed to revolutionize, and it was the next thing to revolutionize. And look, I think from an AI perspective, it's obvious that it's real. It's here to stay. There is absolutely a hype to it.

But also, when you go on your phone and you ask Claude,

Gemini, GPT, take a picture and you ask, like, "What is this?" Or you ask, "Give me some research on a topic I'm curious about," and you get great answers and you research more, it's definitely, definitely real in the sort of consumer world I think.

I think where the hype -- I don't know. I would say it slightly differently than "hype." I'd say the potential I think in the enterprise is still to be seen. I think there's some really slam-dunk use cases we've seen, right? Like, agent coding, for example, is a thing that sort of flipped my brain from "this might be vaporware" to, like, "wow, this is, like, really real." When I sat down at the computer and I was, like, coding with an agent and it was helping me with problems that I've never been able to solve before, I was, like, wow, this is incredibly powerful as a superhuman ability, you know, like, amplifying my abilities. So I think there's definitely real there.

I think from an enterprise perspective, the thing to be seen is where can people harness their data and their enterprise data and the proprietary data they have to make some differentiation in the enterprise space? That's the "to be seen" part.

Allison Nathan: But we're only a couple of years into this newer generation of these models. Do you foresee a future where we actually do, though, run out of data? I mean, we're early here, but is that ahead?

Neema Raphael: I would frame it a different way. We've already run out of data. We've already run out of data. When you read about the new models, the undertone of what people say -- and you've seen this in, like, the DeepSeek moment and things like that -- is, like, everyone wonders how did they do that with less money? One of the big hypotheses is they trained against another model, right? So it already incorporated the previous thing.

I think the real interesting thing is going to be how previous models then shape what the next iteration of the world is going to look like in this way.

Allison Nathan: So let me reframe my question, which is more that do you think this is going to restrain the potential of the technology?

Neema Raphael: No, I don't think so. The explosive

nature of the synthetic data and the fact that now the computer can generate infinite amount of more data, again, I think there'll be a sort of a cursor of what people call, like, “AI slop” versus maybe more insightful data. But I don't think it's going to be a massive constraint only because a lot of trapped enterprise data still has not been harnessed. And I think you see that in the work that we're doing at Goldman.

For example, like, we want to help our sales people, our traders, our quants, our PMs to, sort of, again, get that superhuman capability, that information synthesis, being able to help with their hypothesis. And there's still a lot of data here at Goldman that could be used for that. So I think from a consumer world model, I think it's interesting we've definitely in the synthetic sort of explosion of data. But from an enterprise perspective, I think there's still a lot of juice I'd say to be squeezed in that.

George Lee: Yeah, I would echo that. I think these machines have come an enormous distance in their quality, and they've done it largely on the back of publicly available and synthetically generated data. The amount of data that lives behind firewalls, trapped inside corporate repositories

that's highly salient to garnering business value, that has yet to be unlocked. It's the work that Neema is doing here.

Then there are also other horizons. Think about all the video data in the world. Think about spinning up virtual environments where you're creating a platform for virtual robots to generate their own data about understanding the world. I think while we've exhausted one pool of data, there are many others to go attack.

Neema Raphael: I think, to get a little philosophical out of my realm, but I think what might be interesting is people might think there might be a creative plateau. I mean, if all of the data is synthetically generated, right? Then how much human data could then be incorporated? New human data, new human intellect, new human creativity. I think that'll be an interesting thing to watch from a philosophical perspective.

George Lee: For sure. You know, one or two of our prior guests have made the observation germane to this discussion that the quality of outputs from these models, particularly in enterprise settings, is highly dependent on the quality of the data that you're sourcing and referencing

inside the business. Do you agree with that? It kind of goes to this what's the value of these behind-the-firewall data stores. Maybe just illustrate a little bit of that, how we can make models smarter with our own proprietary data stores.

Neema Raphael: Yeah, I think first again you got to remember what this thing is doing, what this machine is doing, right? Whatever data patterns you are feeding this machine is what it's going to learn and what it's going to extrapolate from. And so I think from an enterprise value perspective, cleaning your data, normalizing it, having the semantics of that data well understood, how it links to other pieces of data, all of this stuff is what's going to allow enterprises to level up from what we think the consumers get to what enterprise value could be created.

Allison Nathan: So what are some things that Goldman is doing to unlock that value?

Neema Raphael: Yeah, look, I don't think people had always thought of data as sort of like this thing that could give more insight to the world. I mean, it's always historically been thought of as, like, business exhaust in

some way, right? Like, a trader executes a trade, they're sort of like, "Okay, I'm done. Now I'm just managing the risk." But there's a whole machine behind that about what happens after that, all the workflows that happen after that, and before that.

And so the real challenges are getting that disparate data into some place where you could organize it in a sane way and then normalize it in ways where the data is correct when you ask it a question, it's linked to the other facts of the world when you want to navigate from that fact to another fact. And so all of these challenges are really, you know, that's why the role of data engineering was even created. People were like we need a practice of engineering that's like software for data. And so just like people write code in a specific way and there's specific architectures and engineering practices to that is the same in data. You have to sort of understand what the data actually means. You have to understand are these two concepts the same? Are they linked differently?

And so really the challenge is understanding the data, understanding the business context of the data, and then being able to normalize it in a way that makes sense for the

business to consume it.

Allison Nathan: Can you actually use the models to help you organize the data?

Neema Raphael: Absolutely.

Allison Nathan: Is there some synergy happening there?

Neema Raphael: Definitely. Like, people have built software agents, people have built engineering agents to do, like, this cleansing, this normalization, this linking. So absolutely in the same way where we're seeing software being created by these agents, there's also a feedback loop of data cleansing and normalization and wrangling, so that's a good insight.

Allison Nathan: Neema, we often close these interviews by asking our guests how they might actually use AI themselves. Either we talked a lot about you are using it in the office, but even personally what do you find is the most interesting and helpful usage?

Neema Raphael: Yeah, you're asking a tech nerd, so

obviously the tech nerd to the coding answer is, like, the base case. But I also have a three-and-a-half-year-old son and he's in his "why" phase, which is awesome. But I'm just, like, I run out of, like, the turtles of the "why." And so, like, actually a lot of times he's like, "What is that? Why is that?" And actually bouncing ideas off of him with the agents I think is really cool. And I think it's been powerful for me.

So now he asks me questions. I ask the AI questions. We learn together about questions he's curious about. So I love that.

Allison Nathan: And as he gets older, he's going to be able to ask himself. So when my teenagers ask me questions, I say, "Look it up. Use AI. Figure it out."

Neema Raphael: Let me Google it up for you.

Allison Nathan: Exactly.

George Lee: Though part of Neema's genius is he gets to freeride on the knowledge acquisition of his son.

Neema Raphael: That's right! Exactly!

George Lee: And be involved. So I love that.

Neema Raphael: Exactly.

Allison Nathan: Well, thanks very much, Neema. That was a fascinating conversation.

Neema Raphael: Thanks. Thanks for having me.

Allison Nathan: I mean, George, Neema had so much insight. What really stood out to you the most about the conversation?

George Lee: Well, as I predicted, you know, grounded, objective, thoughtful. I agree it was a great discussion. You know, people sometimes peer pass the data problem. But as Neema I think illustrated well, it really lies at the heart of bringing value from these systems in business. So onward and upward.

Allison Nathan: As always, thanks for the conversation, George. Always great talking to you.

George Lee: Thank you. Thank you, Neema.

Neema Raphael: Thank you both. It was awesome to be here.

Allison Nathan: This episode of Exchanges was recorded on September 25th, 2025. I'm Allison Nathan.

any purpose. Each name of a third-party organization mentioned 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.

A transcript is provided for convenience and may differ from the original video or audio content. Goldman Sachs is not responsible for any errors in the transcript. This material 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.

Disclosures applicable to research with respect to issuers, if any, mentioned herein are available through your Goldman Sachs representative or at www.GS.com/research/hedge.html.