Goldman Sachs Exchanges

Building AI Systems for Capital Markets

Chris Churchman, Head, Marquee, Co-Chair of the Global Banking & Markets AI Working Group

George Lee, Co-Head, Goldman Sachs Global Institute

Allison Nathan, Senior Strategist, Goldman Sachs Research

Date of recording: August 11, 2026

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

George, great to see you again.

George Lee: Good to see you.

Allison Nathan: So, today we're digging into another fascinating topic: what it takes to actually build products in the AI age, the ingredients, the challenges, and the approaches behind building solutions that first and foremost actually work, but also actually matter.

And I really just wanted to first get your thoughts, George, because as we've discussed many times, you're talking to companies, firms, clients about these topics every day.

So how much real progress do you see being made here?

George Lee: Well, I'm very excited for this discussion because in our prior sessions, we've spent a bunch of time talking about the foundational elements of AI, the infrastructural dimensions, and the topic of building applications in this environment I think is very timely and useful, and we have a great guest for that.

But look, I think there are a lot of questions in the ecosystem about what is the perimeter of the foundational dimensions of this relative to the application part of this stack. And so again, our guest will be able to untangle that for us in a really productive way.

Allison Nathan: Right. So let's bring in our guest. It is Chris Churchman. Chris is the head of Marquee, the firm's digital platform for institutional and corporate clients, and he's also co-chair of the firm's Global Banking & Markets AI working group. So you're on the front lines, Chris, on all of these matters in terms of bringing new AI products to the market.

So welcome to the show.

Chris Churchman: Thank you. Excited to be here.

George Lee: Great. Well, Chris, let's start first by just giving our listeners a little bit of background on what Marquee is, what your core mission is, and then let's talk a little bit about how you're starting to fold in some AI capabilities there.

Chris Churchman: Sure, yeah, so excited for this conversation. Thanks for having me. So Marquee is the firm's platform for institutional and corporate clients, and really our mission is to help our clients make decisions under great uncertainty. So we want to become an integral part of our clients' investment process. And so as you think about layering in gen AI capabilities into the platform, you can imagine what we've built and maybe it's actually a better way to describe the platform is how we've layered in gen AI capabilities, because you've got to start with the problem to be solved, not just, "Oh, let's use AI on that and that and that," which is people start with AI, but let's start with the investment process and how would a gen AI solution help you be the concierge that helps you down that investment process funnel?

So, you start on the top of the funnel, if you like, which is just millions of research articles on different companies, sectors, economic forecasts, et cetera. But then you've also got all of the

trading floor commentary about flows, trade ideas, et cetera. And so that's overwhelming for people to find what's relevant to them.

And then you've got all of the analytics that you'd need to look at, all of the data from dozens of vendors produced. We've got an ecosystem fortunately called Market View, which contains these-- You can think of it like Pinterest for capital markets, where each widget is produced by an expert in their domain.

They take the relevant vendor data, they then create an insight on that, and now we have like millions of those in the ecosystem. And then we've got pre-trade analytics. Like, okay, what trade am I going to do? I can do scenario analysis, I can do back testing, I can do all of these things. But from a user experience point of view, it's very difficult.

And I think the paradigm shift that we're seeing in software is that we've gone from a world where users learn software to where software learns users. And so now we're at the point where you can just express your intent to the system, and the system fully understands all of the capabilities of Marquee and can deliver them to the user.

Because that's always been the bottleneck - expressing intent in natural language is a very low bandwidth channel - but actually you mean way more than what you've actually said. And so, Marquee AI, which is only available internally right now, you ask Marquee a question which is very targeted to what you care about, then it will take what's happened recently to break it out into a number of relevant research topics.

For each of those research topics, it will then find all the relevant research, all the relevant trading floor commentary, all the relevant widgets in market view, so data analytics, and then assemble that into a cohesive thought. And then if it needs to do a calculation, it will pull the widget that contains that data, reliable data from an expert and then run a plan, do a calculation—like write and run some Python—and give back the answer, and then give you a full, well-grounded, every single sentence in the document is grounded to something someone at

the firm has said, or a calculation that you can audit. And that is the hardest bit to layer on at the end.

And then we can talk more about it, but the generative ability of it is key, but it's very, very, very difficult, and we can talk more about it to tame the hallucinations.

George Lee: Yeah. Let's just double-click on that now because in your role and for this service, as we ultimately extend it to clients, factuality, reliability, provenance is really important. I know that's been a really hard part of the work you've been doing. Maybe just double-click on that.

Chris Churchman: It's a really, really hard part of it. And frankly, it's also the most frustrating because what typically happens is you see someone with a flashy demo, right? And they'll take the frontier model, they will forget about all of the detailed stuff around grounding, and it will come up with incredibly confident, plausible answer, and it'll blow people's brains.

But the problem with that is that the model itself cannot distinguish between a fact and an interpolation. Like, it just it actually can't. Everything goes into the context window. It all goes through the same sausage machine, comes out the other end, and it really just doesn't know. It's not that it's misleading you. We just don't notice it because it knows so much. But when you try and apply it to institutional context, when the model doesn't know your actual data, then that's a big problem. But, the evolution of this has changed things dramatically because in the early days, for example, and this is why it's so frustrating to build in the age of AI.

So take ChatGPT, right? November 2022, right? We saw that and went, "Okay, well, model can't do certain things, right?" It makes stuff up if it doesn't know, and it can't use any tools, so it can't use a calculator or anything like that. And so you spend a lot of time thinking about how do we give it those capabilities.

And so you go, right, okay, it can only handle 4,000 tokens in the context window, so 3,000 words. So, you spend all of this time

breaking things up into chunks, embedding them in vector databases, which is basically like turning them into numbers that represent their meaning, filing them all away in like vector databases.

And then a year later, the context window goes to 128,000 tokens, and you're like, "Oh, well, that was pointless." And then it goes to a million tokens. And then you say, well, the early stages were interesting as well, because there was a very famous paper, I think we've talked about it, but "Lost in the Middle", which is that these models, basically, if you do the needle in the haystack test, which is that you hide some context in the middle somewhere. It can't retrieve it because it preferentially looks at the start and the end, and you think, oh, they're like humans. They pay attention to the start of a podcast and the end of the podcast, and they sort of drift in the middle, right? And that's why you break things up into chunks. So, but what's interesting about that is they've been trained on all human knowledge, right?

And so, if you look at all human documents, the start is very important, right? The start is very, very important, the end very, very important. So, it's learnt that, right? And so, Google, when they raised the context limits from one million tokens, which was amazing, they solved the needle in the haystack problem as well.

So, all of this finagling about put it at the start and the end, was all completely pointless because they could retrieve it. So I think as, as we've looked at dealing with hallucinations, there's been the problem is people like plausible, confident answers.

And if I told you that the system I built is right 90% of the time, right? 10% hallucinate rate. Which it is not, it doesn't hallucinate. But, if I built you a system like that, the burden on you to figure out which of the 10 facts is fabricated? So you have to check all the facts. All the facts

Allison Nathan: And is that what you do? I mean, how, how do you actually troubleshoot this in practice?

Chris Churchman: Well, it's very painful because what you do, and this is the thing, even off-the-shelf retail platforms, they give you sources. But actually, if you try and do what we've done in Marquee AI and say, attribute everything to a source, a specific line in each source, you can't do it with web grounding because it's not really using it. It's saying, "Here are my sources." And when we challenged it hard, at least it was honest. It was like, "Look, in the end, I'm better at sounding thorough than being thorough." Right? And so you really have to build the framework to force it to ground itself because it inherently can't do it itself, and it's not rewarded to either.

There was a famous paper like why LLMs hallucinate, and it's because they're doing multiple choice tests, and they're not rewarded for abstaining. They might as well guess because that is optimal, and so they're always going to hallucinate, and they can't distinguish between a fact and an extrapolation.

Allison Nathan: Do you think that we're ultimately going to get to a point, though, where we have complete confidence in these products built by these models?

Chris Churchman: No, I don't think we're going to get to the point where we're fully confident, but it touches on where I think the puck is going with this.

So, I think that we've gone through the prompt engineering, kind of let's tell it exactly what to do, right? We've gone through context engineering, which is okay, what if I can cobble together a bit of memory, a bit of data, a bit of prompts, a bit of playbooks. We're currently in the agentic engineering phase, which is okay, well, now we've got an agentic loop that can check its own work and iterate. And sure, you couldn't put one of those old models in this framework because the errors would compound, right? So, the fact that it can check its own work means that it is more reliable now than it ever was before. So those loops are important. But now I think what we're getting to is environment engineering.

How do we make sure that the environments that agents are operating are secure? Recent headlines notwithstanding. But also, that they have access to all the relevant tools and institutional knowledge in a fully entitled way. And then I think the next phase, which shasn't been coined yet, but I think we're going there, which is like mandate engineering, which is like how is what are AI allowed to do under whose authority, in whose name?

And I think that's the cutting edge. I think once there's an insurance market or a liability market around what these agents do, then I think you'll see the autonomy of them increase. But they'll still make mistakes, and it's like trusting any human. It's the same mechanism, but we don't have the recourse that we do with humans.

George Lee: It's fascinating to reflect actually all that you describe, I think very eloquently, that progression. It's been less than four years. Those are eras of AI development that have impacted this four-year period. And as you say, the confidence interval in the reliability of these systems has expanded a lot. They may never close to perfection based on the probabilistic nature of the technology. But wow, amazing progress in different regimes of research in that short period of time.

Allison Nathan: Absolutely. The other question you brought up was how do you keep up? We are going to continue to see evolution of these models, so what are you doing to get ahead of that. Or can you do anything to get ahead of it?

Chris Churchman: I don’t think you can get ahead of it. I guess I'd make two, two. One is about the demo to start with, which is that, you can take a frontier model and connect it to data, and it will blow your mind. And that's the proof of concept. But like a demo is judged on its best day, whereas a product is judged on its worst day, right? And all of the work is the gap between that. And that is boring, hard, actually pretty thankless work. And so, if you try and build around the deficiencies of the model, you will get burnt eventually, and you will waste your time.

So, for example, take the context window. All of that work to build RAG pipelines, chunking, vector databases, all of that, just to curate. Now you just throw out the whole book at it, right? Or multiple books. The whole library at it instead of throwing chunks at it. Like, so build for the model that you will have at launch, not for the model that you have when you're starting your development. So that's the first lesson I've learned. So assume that those... So all of those reasoning scaffolds we built, once reasoning models came out, totally pointless. All of those ways that we'd connected the LLM to our tools and our data, totally pointless once you had function calling and MCP.

And so, you waste all of your effort. Don't bet against the model, basically. That's the key. And then the other thing is build what the model can't learn, like can't know, like by the millions of users they have. And so that's very institutional specific. What doesn't it know? Doesn't know our entitlements, doesn't know our mandates, doesn't know the way that we connect our data together, et cetera.

Cause everything else is going to get consumed by the model, and you'll have wasted your time.

George Lee: It's fascinating. As you say, building these systems in an environment where the foundation layers is shifting a lot, the regimes, the architectures are shifting. I think you've done a very good job of rolling with those changes, but also keeping your mind on some first principles around what you want to build, and in particular, as you said, user intent and experience.

Maybe work backwards a little bit from what are those first principles of user experience that you're trying to incorporate in the system, all while the playing field shifts around you?

Chris Churchman: Yeah. I think it's about, it's partly about how do you ensure trust and transparency? Because trust, we all know intuitively, is something that takes years to build and is lost very quickly.

And so when you think about these gen AI capabilities, you can generate all sorts of answers.

So the question is how do you generate an answer that someone will stand behind? You know, you have to think very deeply about what a firm like Goldman Sachs does, right? We don't just generate advice or analytics or possible scenarios.

We stand behind one of those and take a view in highly uncertain times. And so then the question becomes: how do you build in such a way that our human capital can fully understand and stand behind the analysis that was generated by AI? And that has to be done in the design phase. As in it has to be done thoughtfully.

Because I think there's a huge danger here that in the era of AI, we outsource our reasoning to these models, and we have cognitive atrophy that stops us being able to reason from first principles ourselves. I think there's a real danger of that. So if you think about the history of technology, we have always built tools that are external to ourselves that do a capability that we had better than us, and then those skills atrophy, and then eventually we forget we even had them, right?

So think about like wayfinding, running hundreds of miles on the savannah and finding your way back, right? Totally lost that capability. Remembering things, like bards used to be able to recite entire books. Now you don't need to remember anymore. You just read it in a book. That was the printing press, right? Then you said, "Okay, well, now," even when I was at university, I had to find textbooks for information. Now, you don't even need to go anywhere or information's not scarce anymore, so you don't need to remember that, but reasoning is still important. You still need to reason about and structure it into an argument.

And now we're delegating reasoning. And so I think that my North Star here is, I feel a great responsibility building products for the firm such that in 10 years time, a new joiner to the firm, their friends say to them, "Oh my goodness you can handle that much reasoning?

It's like getting a job as a fighter pilot versus a bus driver. It's like, wow, there's a lot coming at you, but you must be really

smart and really good at reasoning under uncertainty to be able to handle that. And so, I think that has to be a proactive choice to think about how do we empower human reasoning, not delegate it.

Allison Nathan: Fascinating. So you're saying that if technology evolves as it has historically, we would lose it. So, we actually have to really emphasize that.

Chris Churchman: Like working in the fields, right? I go to Barry's in the morning. You know, like we have to build, build gyms because I sit in an office for 12 hours a day, right?

So you need to actively... And so I think that's comes back to like the apprenticeship culture that we have to proactively push. I don't think that's going to happen by accident because a lot of the way that juniors learn is this apprenticeship culture. All of these things like, take a junior trader who was seeing that price request on a chat, working out what does that even mean?

How do I put it into the pricer? Well, what price am I going to put back? Et cetera. So we can absolutely automate that, but then do we get the senior traders that fully understand? There was a lot of conversations that spun off the back of how to price that. So you learn by doing, and a lot of knowledge is tacit.

It was never written down. And I think that's where we really need to double-click on is to make sure we don't lose that tacit and intuitive knowledge that some of our best people have today, but to ensure the next generation have it too.

George Lee: So Chris, what's on your wish list going forward or your roadmap for this platform? And we've talked about the rapid evolution and improvement in fundamental technology underneath this. As that plays out, what additional features and qualities do you want to add to the platform?

Chris Churchman: I'd say there's sort of two mindsets that I think are important to distinguish when you think about innovation and AI, and most people use the first one. So, there's an automation mindset, and then there's a re-conception

mindset, and most of the work at the moment is in the automation mindset. So what's an automation mindset? It's you take an existing process, you look at the cognitive bottlenecks in that process, whether it's reading, reviewing, et cetera, that are done by human. You replace that with AI and then you get the same answer, just faster, right?

The risk with that is you lock in legacy. The benefits of that is you get some efficiency gains. So most people are doing that right now.

Then there's the re-conception mindset, and this is where you get orders of magnitude improvement. Where you forget about the existing processes that were fossils of a previous age, and you say, "Okay, what's the real job to be done here?" Right? What's the real...

You have to have deep domain expertise to do that. And then you say "Okay, let's assume intelligence is incredibly scalable and elastic and abundant." Okay. What questions could we now answer that we would never be able to answer before because of limited human bandwidth or coordination, et cetera? And let's reconceive what's possible.

Allison Nathan: So what are some examples of things you're working on right now?

Chris Churchman: So let's take so there's the capability lens. So what capabilities do we have that can be made much more powerful in the age of AI? For example, we have something called visual structuring, right? Which is a intuitive visual interface that can price any derivative, for example, but it doesn't price them.

It looks at the scenarios, how it's going to evolve over time, when did it pay out, how does it back test. And so, think about wrapping that capability in an AI agent that thoroughly understands how that tool works and how to interpret the output.

Well, now you're away to the races because the bottleneck is not teaching a user how to use the tool and how to interpret its outputs.

You teach the agent that, and then the user just has to express intent. So how do we bring to bear the real capabilities that we have as a firm, uniquely as a firm, in an area of AI, to help our clients make better decisions under uncertainty? That has to be the North Star.

George Lee: Fascinating. Maybe we just close. It's been a fantastic discussion. We've talked a lot about the limits of the technology, how they've evolved, how you're navigating them. Maybe just a very broad question: are you a believer in the continued scalability and progress of this technology paradigm, or do you see some sort of unanticipated glass ceiling ahead of us?

Chris Churchman: Yeah, great question. I am a believer, yeah and I think for two separate reasons. I think firstly, on the next token prediction, there was a period of time where people were dismissing this technology: "Oh, all it does is predict the next word." Well, turns out that is one of the most deep training functions you could give something. So for example, for me to correctly predict the next word of "Allison turned to George and said," I need to know who Allison is, who George is, what situation we're in right now, like what typically is said on podcasts, et cetera. So I need to build a world model to be able to do that job really well, and I think that's been the surprising thing, is that the more we train these things, they are actually building synthetic world models, and a lot of the doubters said that that wasn't going to be case.

So, I think as we scale it. And then the second thing is, I think I've been blown away by if you go back to first principles to see through the noise, it's really the transformer paper in 2017, "Attention Is All You Need," which basically said that actually you can have every word pay attention every other word in a context window, which is amazing.

But then the Kaplan Scaling Laws plus the Chinchilla paper basically told you that these things, if you do a small experiment, then to see whether it works, you can then scale it up and it will work at a larger scale, up to 10 orders of magnitude bigger.

And so what that means is two things. Firstly, it justifies spending trillions of dollars on infrastructure because you can say look, that's not a wild bet. There's a line here that's held a long time. So, if we do 10 times, we build a trillion-dollar data center, we would expect this. So you can pitch that to investors.

But importantly the labs, their number one priority above all else is to automate themselves, which is automate AI R&D, right? And so, what that means is, what's the bottleneck in AI R&D? It's not reading the papers. It's not interpreting the results. Absolutely not. It's the running of experiments.

And so if an AI can run actual experiments at small scale many times and the scaling laws hold, then it means that there's no actual bottleneck. And so they can actually expedite the number of experiments they do and know that those will scale. And so I think that is the next paradigm here.

I think the big missing piece as I look at this is, firstly, self-learning, it hasn't come yet. Catastrophic forgetting. And so all of these memory architectures that are being built, it's possible at some point they get enveloped in the model. The second thing is I don't think we found the form factor. Like I think we're in phase of AI where television is invented and we start showing radio shows on TV. We haven't got the native form factor, we're in the command line versus Windows and mouse kind of era. So, where's the native form factor of AI? And then we also haven't figured out like how to, going back to this apprenticeship point, like how do we ensure the next generation that grows up in the systems we're building today, we build a future that we're proud, a legacy that we're proud of.

Allison Nathan: Interesting. Thanks again, Chris, for joining us and sharing all these insights.

Chris Churchman: You're welcome.

George Lee: Always fun to talk to you about this stuff, Chris. Thank you for being here.

Allison Nathan: George, anything that was particularly striking to you? Any main takeaways

George Lee: It's a learned discussion on the history and trajectory of generative AI and great to hear it from someone who is actually building fundamental product with, with the technology.

I loved his framing of automation versus reconsideration, first principles, new experiences. I often refer to that as a lot of the effort here is breeding faster horses versus building the automobile. And then I think the final segment on just a deep belief from a practitioner that the bitter lesson, the scaling laws will hold, that we'll continue to travel up a curve of improvement. I think it's an optimistic message, and let's hope he's right.

Allison Nathan: Right. And I think I have a deep appreciation for everything that Chris does and the effort and challenges that it comes with building these new products, which are very exciting, but take a lot of effort, a lot of work, and a lot of trial and error.

George Lee: In an uncertain environment, as he described.

Allison Nathan: And George, great talking to you as always.

George Lee: As always.

Allison Nathan: This episode of Goldman Sachs Exchanges was recorded on Tuesday, August 11th, 2026. I'm Allison Nathan. Thanks for listening.

The opinions and views expressed herein are as of the date of publication, subject to change without notice, and may not necessarily reflect the institutional views of Goldman Sachs or its affiliates. The material provided is intended for informational purposes only, and does not constitute investment, legal, or tax

advice, a recommendation from any Goldman Sachs entity to take any particular action or be used as a basis for any other investment decision, or an offer or solicitation to purchase or sell any securities or financial products. Any forward-looking statements, case studies, computations or examples set forth herein are for illustrative purposes only. Past performance is not indicative of future results. Neither Goldman Sachs nor any of its affiliates make any representations or warranties, express or implied, as to the accuracy or completeness of the statements or information contained herein and disclaim any liability whatsoever for reliance on such information for 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 sponsorship, affiliation, endorsement, ownership or license rights between any such company and Goldman Sachs. 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 information relating to Goldman Sachs Global Banking & Markets, if any, mentioned herein, are available here.

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

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

Date of Recording August 11, 2026.

© 2026 Goldman Sachs.

All rights reserved.