

What McKinsey learned while creating its generative AI platform

To corral its vast amount of knowledge and drive new levels of productivity, the firm embarked on a journey to develop its generative AI platform, “Lilli.”



Nearly one hundred years of McKinsey's insights and knowledge serve as the source material for the firm's AI platform, "Lilli." In this episode of the *At the Edge* podcast, [Erik Roth](#), a McKinsey senior partner and global leader of growth and innovation, joins senior partner [Lareina Yee](#) to discuss how Lilli was developed and thoughtfully executed—and why immediate access to the firm's intellectual property has transformed the way McKinsey serves its clients.

This transcript has been edited for clarity and length. For more conversations on cutting-edge technology, follow the series on your preferred podcast platform.

The origins of Lilli

Lareina Yee: What is Lilli, and how did it become one of the most-used tools at McKinsey?

Erik Roth: Lilli was born from the question: "How do we help our colleagues at McKinsey access the deepest and broadest array of our best insights so they can activate them with our clients?" And from there, as we started that investigation, ChatGPT exploded.

And we had this notion of, "What if we train it on our own information?" Lilli started as a knowledge extraction and synthesis tool, but to your question, it has now become an orchestration layer that basically coordinates many different types of knowledge across and outside the firm.

Lareina Yee: How is Lilli different from other generative AI platforms?

Erik Roth: Lilli is different in a couple of ways. Number one, it is tuned to McKinsey and our client service, meaning that the way it recognizes the

intent of questions, and the way it fine-tunes answers to make them useful for our colleagues to bring to clients, is very "McKinsey-esque."

It does this through the combination of large and small models that we use in the platform. And that's probably the second thing that makes it different. It's not a RAG, or retrieval-augmented generation, instance. It's a combination of many technologies coming together in a software stack that enables us to do what I described earlier. So this notion of, "It's an orchestration layer," is real. It's very fit for purpose for an organization that needs to be thoughtful and mindful about regulation, security, access, and protecting its information.

Lareina Yee: As the parent and product leader of this feat, can you give us an inside view about how you approached the development of Lilli?

Erik Roth: When we started, it was an experiment. So it was very iterative, and the team was very small. There were only four of us to begin with. Now we're well over 150. But it was very much learning from our users. And in the early days, it was doing very classical, observational, ethnographic research on our colleagues.

It was very much, "Be focused on the problem, be user-centric." We had four domains, which were reasonably simple and have held up over time: "How do you have a high-performing team? How do you develop clients in a way that is gen AI or technology augmented? How do you deliver distinctive client service? And how do you maintain high-quality communications and connections once your projects are done?"

It was really those four domains. And within them, we of course have had lots of different frustrations and problems to solve, but we've stayed pretty true to that North Star all along.

The other way we keep it very user-centric is by ensuring that the user is always right in the middle of the development process itself. Our entire development pipeline, every single element of it, whether it's something in the stack that we're improving on the platform side or a new capability that is delivered through our users, is linked to a user-specific problem. There's nothing in the pipeline that doesn't have that on the back end. We keep that front and center, and I think that's served us well.

Lareina Yee: Can you tell me a little bit more about how you are personally supporting the adoption of Lilli at McKinsey?

Erik Roth: I ask everybody, "Have you used Lilli today?" The first question of every team meeting is, "Have you asked Lilli?" Oftentimes they have, and many times they haven't. So I think role modeling and actually using it ourselves as leaders of the firm is really important.

But behind that, there are 360-degree communications and adoption programs that go on, everything from integrating how to use Lilli into our earliest learning when you join the firm all the way to every quarter when we have our risk and legal assessments that the entire firm has to complete. And it is not an accident that Lilli is now a part of those. So every quarter you're reminded not just of the safe use of data but how to use Lilli within that construct.

And then in between we have Lilli user groups in offices. Ten offices have Lilli communities that send us ideas all the time. And we have training sessions. And then the platform itself has a how-to guide. You can click on the upper-left-hand corner, and every single thing about how to use the platform is right there for your own use.

Lareina Yee: What are some of the other, unexpected learnings about making this effective for users?

Erik Roth: We've just launched our agents framework. The McKinsey Tone of Voice agent is the number-one used agent in our beta group, which is about 2,500 folks right now. It takes any prose and translates it into what a McKinsey-quality writing sample would be. Whether it's going to a partner, a senior partner, or a client, it takes what is messy or confusing, particularly for those who are non-English-native speakers, and turns it into something that's immediately useful.

The consultant of the future

Lareina Yee: There is all this tremendous capability usage, these little but actually pretty big learnings. Can you paint a picture for us of how the McKinsey consultant, someone who people have interacted with in business for nearly a hundred years, will be different, if at all, over the next three years?

Erik Roth: We can make some guesses based on some early clues. First, the McKinsey consultant will be much more tech enabled. The amount of analytical work that they will do in a spreadsheet or through other means will be increasingly replaced by AI-related tools. Will that work be replaced 100 percent? I don't think we know that.

My hope is the future McKinsey consultant, and perhaps broader to the industry as well, will actually spend more time activating their insights as opposed to doing the analytics to create them. And that has implications for who we hire, what skill sets they have, and what experiences they might bring to a given project.

My hope is the future McKinsey consultant, and perhaps broader to the industry as well, will actually spend more time activating their insights as opposed to doing the analytics to create them.

And all of that, I think, is up for discussion. But the future consultant will probably be a lot more empathetic in working with people. And I hope more diverse, because I think we can enable a much wider group of people to be able to do things that they haven't done before, and overall [have] higher quality in terms of the way they create impact.

Curiosity is key

Lareina Yee: One of the things that holds people back is data. And what we often hear is that companies say their data sets aren't good enough to make generative AI useful for them. What's your advice and perspective for other companies having this adventure?

Erik Roth: I was talking with a CEO, and her concern was that their data was not good enough. My advice to her was, "Yes, you need a data architecture. You need a data framework for thinking about how to tag and label things. You need a bit of a curation process." Data is one of the things that is required for this type of technology to be useful. And getting your data in order is

increasingly important if you want to take advantage of these technologies.

I think your question was a little bit about the human implications, and those are real. I am very optimistic. My best answer, for now, is that there will be parts of organizations that will be what I call the minimum viable organization: as few people as possible with as much technology enablement as relevant or appropriate.

That said, there are going to be other parts of the organization that are going to gain superpowers that will allow for upskilling of lower performers to average or top performers to do things they otherwise couldn't have done before. The technology won't do everything, but it'll do a lot of things, and I think one of the gaps is education and understanding what these technologies are all about and how they operate.

Lareina Yee: And the education and understanding are relatively new. It's not as if 24 months ago you could come out with a degree and expertise on this. You might have been one of the rare PhD students,

Don't be afraid of the technology. The more you learn about it, the more you're going to understand it and understand how it applies to your business.

perhaps. But you're an adult learner. You're a senior partner. What has been the education journey for you, and what advice do you have for leaders going down this path?

Erik Roth: First, I would say [dive in](#). Don't be afraid of the technology. The more you learn about it, the more you're going to understand it and understand how it applies to your business. The second is, be curious. Don't be afraid to ask questions. When I'm in meetings and I'm talking about gen AI, at some point I'll pause and say, "So everyone understands what GPT is, right?" It amazes me that every time, at least half the group doesn't.

Building deliberately

Lareina Yee: I would like to turn to a couple of more questions about the product itself. I think people would love to know, under the hood, how you've built this. Can you talk about how you prioritized testing versus development?

Erik Roth: There's always the challenge between building and learning. You can build faster than you can learn, because building is coding and assembling technology pieces. Learning is waiting

for your users to actually experience the product and provide feedback. So we've had to slow ourselves down quite a bit in terms of what we *could* build.

When we launched Lilli, we only launched it to about 2,500 colleagues. We intentionally said, "We're going to take a very measured pace. We're going to keep the group small. We're going to learn and turn them into evangelists. And we're going to add waves of users."

There are approaches or tactics embedded in the way we've built and developed this platform that have been emphasizing and prioritizing learning through a lot of iterative testing. We have alpha groups. We have beta groups. We have something called LilliX, which is doing experimental research. All of that is to really drive the highest adoption and the highest usage.

The innovation is just beginning

Lareina Yee: How are you thinking about Lilli and its capabilities around reasoning, precision, and predictive analytics?

Find more content like this on the
McKinsey Insights App



Scan • Download • Personalize



Erik Roth: These models may look like they reason, but they don't today. They emulate a lot of things. And when you start thinking about agents and chains of agents, they can actually replicate what might seem like a human line of logic and thinking. But we're not there yet. However, I mentioned LilliX. Inside LilliX, we're running a number of experiments with some of the leading thinkers on this topic in the world. So in this small group of folks, we actually are building models that can reason.

Lareina Yee: Two final questions, Erik. How did you come up with the name Lilli for your product?

Erik Roth: Lillian Dombrowski was the first woman to get her MBA while at the firm. She also helped create our archives way back in the 1940s—which is appropriate given what Lilli has built itself upon—and she was the instigator of at least two of our global practices.

And Lillian, who Lilli was named for, was an innovator, an entrepreneur, and someone who really was a pioneer at the firm way, way back during the 1940s. And our user community felt it was appropriate to choose her. And after they chose her, we shortened the name to Lilli.

Lareina Yee: Your history with McKinsey is around [innovation](#). So the last question is, as you think ahead, what brings you optimism about the age of AI?

Erik Roth: I think that AI will enable a new set of tech-enabled business models that have the possibility of not just improving human lives but also improving the way that businesses create value and contribute to society. I think we've been too caught up in legacy processes and legacy ways of working that, for a lot of good reasons, humans have had difficulty adopting.

The natural-language capabilities of LLMs [large language models], and the reality that you can interact with these technologies more like a human colleague or partner, is most likely going to change the way these technologies integrate themselves into how businesses and organizations function. So as I look ahead, I'm very optimistic that as long as we do this safely, and we don't forget that these models have biases, and we do it in a way that respects diversity, the possibility of AI and gen AI will create new possibilities that we may not fully imagine today.

Erik Roth is a senior partner in McKinsey's Connecticut office. **Lareina Yee** is a senior partner in McKinsey's Bay Area office and coleads McKinsey's alliances and ecosystems initiatives.

Copyright © 2024 McKinsey & Company. All rights reserved.