McKinsey & Company | QuantumBlack AI by McKinsey

Technology, Media & Telecommunications Practice

The promise and the reality of gen AI agents in the enterprise

Generative AI technology is improving so quickly that a range of new capabilities are rapidly coming online, but only for those who can understand how to use them.

May 2024

The evolution of generative AI (gen AI) has opened the door to great opportunities across organizations, particularly regarding gen AI agents—AI-powered software entities that plan and perform tasks or aid humans by delivering specific services on their behalf. So far, adoption at scale across businesses has faced difficulties because of data quality, employee distrust, and cost of implementation. In addition, capabilities have raced ahead of leaders' capacity to imagine how these agents could be used to transform work.

However, as gen AI technologies progress and the next-generation agents emerge, we expect more use cases to be unlocked, deployment costs to decrease, long-tail use cases to become economically viable, and more at-scale automation to take place across a wider range of enterprise processes, employee experiences, and customer interfaces. This evolution will demand investing in strong AI trust and risk management practices and policies as well as platforms for managing and monitoring agent-based systems.

In this interview, McKinsey Digital's Barr Seitz speaks with senior partners Jorge Amar and Lari Hämäläinen and partner Nicolai von Bismarck to explore the evolution of gen AI agents and how companies can and should implement the technology, where the pools of value lie for the enterprise as a whole. They particularly explore what these developments mean for customer service. An edited transcript of the conversation follows.

Barr Seitz: What exactly is a gen AI agent?

Lari Hämäläinen: When we talk about gen AI agents, we mean software entities that can orchestrate complex workflows, coordinate activities among multiple agents, apply logic, and evaluate answers. These agents can help automate processes in organizations or augment workers and customers as they perform processes. This is valuable because it will not only help humans do their jobs better but also fully digitalize underlying processes and services.

For example, in customer services, recent developments in short- and long-term memory structures enable these agents to personalize interactions with external customers and internal users, and help human agents learn. All of this means that gen AI agents are getting much closer to becoming true virtual workers that can both augment and automate enterprise services in all areas of the business, from HR to finance to customer service. That means we're well on our way to automating a wide range of tasks in many service functions while also improving service quality.

Barr Seitz: Where do you see the greatest value from gen AI agents?

Jorge Amar: We have estimated that gen AI enterprise use cases could yield $2.6 trillion to $4.4 trillion annually in value across more than 60 use cases.[1] But how much of this value is realized as business growth and productivity will depend on how quickly enterprises can reimagine and truly transform work in priority domains—that is, user journeys, processes across an entire chain of activities, or a function.

Gen-AI-enabled agents hold the promise of accelerating the automation of a very long tail of workflows that would otherwise require inordinate amounts of resources to implement. And the potential extends even beyond these use cases: 60 to 70 percent of the work hours in today's global economy could theoretically be automated by applying a wide variety of existing technology capabilities, including generative AI, but doing so will require a lot in terms of solutions development and enterprise adoption.

Consider customer service. Currently, the value of gen AI agents in the customer service environment is going to come either from a volume reduction or a reduction in average handling times. For example, in work we published earlier this year, we looked at 5,000 customer service agents using gen AI and found that issue resolution increased by 14 percent an hour, while time spent handling issues went down 9 percent.[2]

[1] "The economic potential of generative AI: The next productivity frontier," McKinsey, June 14, 2023.
[2] Ibid.

The other area for value is agent training. Typically, we see that it takes somewhere between six to nine months for a new agent to perform at par with the level of more tenured peers. With this technology, we see that time come down to three months, in some cases, because new agents have at their disposal a vast library of interventions and scripts that have worked in other situations.

Over time, as gen AI agents become more proficient, I expect to see them improve customer satisfaction and generate revenue. By supporting human agents and working autonomously, for example, gen AI agents will be critical not just in helping customers with their immediate questions but also beyond, be that selling new services or addressing broader needs. As companies add more gen AI agents, costs are likely to come down, and this will open up a wider array of customer experience options for companies, such as offering more high-touch interactions with human agents as a premium service.

Barr Seitz: What are the opportunities you are already seeing with gen AI agents?

Jorge Amar: Customer care will be one of the first but definitely not the only function with at-scale AI agents. Over the past year, we have seen a lot of successful pilots with gen AI agents helping to improve customer service functions. For example, you could have a customer service agent who is on the phone with a customer and receives help in real time from a dedicated gen AI agent that is, for instance, recommending the best knowledge article to refer to or what the best next steps are for the conversation. The gen AI agent can also give coaching on behavioral elements, such as tone, empathy, and courtesy.

It used to be the case that dedicating an agent to an individual customer at each point of their sales journey was cost-prohibitive. But, as Lari noted, with the latest developments in gen AI agents, now you can do it.

Nicolai von Bismarck: It's worth emphasizing that gen AI agents not only automate processes but also support human agents. One thing that gen AI agents are so good at, for example, is in helping customer service representatives get personalized coaching not only from a hard-skill perspective but also in soft skills like understanding the context of what is being said. We estimate that applying generative AI to customer care functions could increase productivity by between 30 to 45 percent.[3]

Jorge Amar: Yes, and in other cases, gen AI agents assist the customer directly. A digital sales assistant can assist the customer at every point in their decision journey by, for example, retrieving information or providing product specs or cost comparisons—and then remembering the context if the customer visits, leaves, and returns. As those capabilities grow, we can expect these gen AI agents to generate revenue through upselling.

[For more on how companies are using gen AI agents, see the sidebar, "A closer look at gen AI agents: The Lenovo experience."]

Barr Seitz: Can you clarify why people should believe that gen AI agents are a real opportunity and not just another false technology promise?

Jorge Amar: These are still early days, of course, but the kinds of capabilities we're seeing from gen AI agents are simply unprecedented. Unlike past technologies, for example, gen AI not only can theoretically handle the hundreds of millions of interactions between employees and customers across various channels but also can generate much higher-quality interactions, such as delivering personalized content. And we know that personalized service is a key driver of better customer service. There is a big opportunity here because we found in a survey of customer care executives we ran that less than 10 percent of respondents in North America reported greater-than-expected satisfaction with their customer service performance.[4]

[3] Ibid.
[4] "Where is customer care in 2024?," McKinsey, March 12, 2024.

A closer look at gen AI agents: The Lenovo experience

Three leaders at Lenovo—Solutions and Services Group chief technology officer Arthur Hu, COO and head of strategy Linda Yao, and Digital Workplace Solutions general manager Raghav Raghunathan—discuss with McKinsey senior partner Lari Hämäläinen and McKinsey Digital's Barr Seitz how the company uses generative AI (gen AI) agents.

Barr Seitz: What existing gen AI agent applications has Lenovo been running and what sort of impact have you seen from them?

Arthur Hu: We've focused on two main areas. One is software engineering. It's the low-hanging fruit to help our people enhance speed and quality of code production. Our people are already getting 10 percent improvements, and we're seeing that increase to 15 percent as teams get better at using gen AI agents.

The second one is about support. We have hundreds of millions of interactions with our customers across online, chat, voice, and email. We're applying LLM [large language model]-enhanced bots to address customer issues across the entire customer journey and are seeing some great improvements already. We believe it's possible to address as much as 70 to 80 percent of all customer interactions without needing to pull in a human.

Linda Yao: With our gen AI agents helping support customer service, we're seeing double-digit productivity gains on call handling time. And we're seeing incredible gains in other places too. We're finding that marketing teams, for example, are cutting the time it takes to create a great pitch book by 90 percent and also saving on agency fees.

Barr Seitz: How are you getting ready for a world of gen AI agents?

Linda Yao: I was working with our marketing and sales training teams just this morning as part of a program to develop a learning curriculum for our organization, our partners, and our key customers. We're figuring out what learning should be at all levels of the business and for different roles.

Arthur Hu: On the tech side, employees need to understand what gen AI agents are and how they can help. It's critical to be able to build trust or they'll resist adopting it. In many ways, this is a demystification exercise.

Raghav Raghunathan: We see gen AI as a way to level the playing field in new areas. You don't need a huge talent base now to compete. We're investing in tools and workflows to allow us to deliver services with much lower labor intensity and better outcomes.

Barr Seitz: What sort of learning programs are you developing to upskill your people?

Linda Yao: The learning paths for managers, for example, focus on building up their technical acumen, understanding how to change their KPIs because team outputs are changing quickly. At the executive level, it's about helping leaders develop a strong understanding of the tech so they can determine what's a good use case to invest in, and which one isn't.

Arthur Hu: We've found that as our software engineers learn how to work with gen AI agents, they go from basically just chatting with them for code snippets to developing much broader thinking and focus. They start to think about changing the software workflow, such as working with gen AI agents on ideation and other parts of the value chain.

Raghav Raghunathan: Gen AI provides an experiential learning capability that's much more effective. They can prepare sales people for customer interactions or guide them during sales calls. This approach is having a much greater impact than previous learning approaches. It gives them a safe space to learn. They can practice their pitches ahead of time and learn through feedback in live situations.

Barr Seitz: How do you see the future of gen AI agents evolving?

Linda Yao: In our use cases to date, we've refined gen AI agents so they act as a good assistant. As we start improving the technology, gen AI agents will become more like deputies that human agents can deploy to do tasks. We're hoping to see productivity improvements, but we expect this to be a big improvement for the employee experience. These are tasks people don't want to do.

Arthur Hu: There are lots of opportunities, but one area we're exploring is how to use gen AI to capture discussions and interactions, and feed the insights and outputs into our development pipeline. There are dozens of points in the customer interaction journey, which means we have tons of data to mine to understand complex intent and even autogenerate new knowledge to address issues.

Arthur Hu is chief technology officer of Lenovo's Solutions and Services Group. **Linda Yao** is COO and head of strategy at Lenovo. **Raghav Raghunathan** is general manager of Lenovo's Digital Workplace Solutions. **Lari Hämäläinen** is a senior partner in McKinsey's Seattle office; **Barr Seitz** is director of global publishing for McKinsey Digital and is based in the New York office.

Lari Hämäläinen: Let me take the technology view. This is the first time where we have a technology that is fitted to the way humans interact and can be deployed at enterprise scale. Take, for example, the IVR [interactive voice response] experiences we've all suffered through on calls. That's not how humans interact. Humans interact in an unstructured way, often with unspoken intent. And if you think about LLMs [large language models], they were basically created from their inception to handle unstructured data and interactions. In a sense, all the technologies we applied so far to places like customer service worked on the premise that the customer is calling with a very structured set of thoughts that fit predefined conceptions.

Barr Seitz: How has the gen AI agent landscape changed in the past 12 months?

Lari Hämäläinen: The development of gen AI has been extremely fast. In the early days of LLMs, some of their shortcomings, like hallucinations and relatively high processing costs, meant that models were used to generate pretty basic outputs, like providing expertise to humans or generating images. More complex options weren't viable. For example, consider that in the case of an LLM with just 80 percent accuracy applied to a task with ten related steps, the cumulative accuracy rate would be just 11 percent.

Today, LLMs can be applied to a wider variety of use cases and more complex workflows because of multiple recent innovations. These include advances in the LLMs themselves in terms of their accuracy and capabilities, innovations in short- and long-term memory structures, developments in logic structures and answer evaluation, and frameworks to apply agents and models to complex workflows. LLMs can evaluate and correct "wrong" answers so that you can have much higher accuracy. With an experienced human in the loop to handle cases that are identified as tricky, then the joint human-plus-machine outcome can generate great quality and great productivity.

Finally, it's worth mentioning that a lot of gen AI applications beyond chat have been custom-built in the past year by bringing different components together. What we are now seeing is the standardization and industrialization of frameworks to become closer to "packaged software." This will speed up implementation and improve cost efficiency, making real-world applications even more viable, including addressing the long-tail use cases in enterprises.

Barr Seitz: What sorts of hurdles are you seeing in adopting the gen AI agent technology for customer service?

Nicolai von Bismarck: One big hurdle we're seeing is building trust across the organization in gen AI agents. At one bank, for example, they knew they needed to cut down on wrong answers to build trust. So they created an architecture that checks for hallucinations. Only when the check confirms that the answer is correct is it released. And if the answer isn't right, the chatbot would say that it cannot answer this question and try to rephrase it. The customer is then able to either get an answer to their question quickly or decide that they want to talk to a live agent. That's really valuable, as we find that customers across all age groups—even Gen Z—still prefer live phone conversations for customer help and support..

Jorge Amar: We are seeing very promising results, but these are in controlled environments with a small group of customers or agents. To scale these results, change management will be critical. That's a big hurdle for organizations. It's much broader than simply rolling out a new set of tools. Companies are going to need to rewire how functions work so they can get the full value from gen AI agents.

Take data, which needs to be in the right format and place for gen AI technologies to use them effectively. Almost 20 percent of most organizations, in fact, see data as the biggest challenge to capturing value with gen AI.[5] One example of this

[5] "The state of AI in 2023: Generative AI's breakout year," McKinsey, August 1, 2023.

kind of issue could be a chatbot sourcing outdated information, like a policy that was used during COVID-19, in delivering an answer. The content might be right, but it's hopelessly out of date. Companies are going to need to invest in cleaning and organizing their data.

In addition, companies need a real commitment to building AI trust and governance capabilities. These are the principles, policies, processes, and platforms that assure companies are not just compliant with fast-evolving regulations—as seen in the recent EU AI law and similar actions in many countries—but also able to keep the kinds of commitments that they make to customers and employees in terms of fairness and lack of bias. This will also require new learning, new levels of collaboration with legal and risk teams, and new technology to manage and monitor systems at scale.

Change needs to happen in other areas as well. Businesses will need to build extensive and tailored learning curricula for all levels of the customer service function—from managers who will need to create new KPIs and performance management protocols to frontline agents who will need to understand different ways to engage with both customers and gen AI agents.

The technology will need to evolve to be more flexible and develop a stronger life cycle capability to support gen AI tools, what we'd call MLOps [machine learning operations] or, increasingly, gen AIOps [gen AI operations]. The operating model will need to support small teams working iteratively on new service capabilities. And adoption will require sustained effort and new incentives so that people learn to trust the tools and realize the benefits. This is particularly true with more tenured agents, who believe their own skills cannot be augmented or improved on with gen AI agents. For customer operations alone, we're talking about a broad effort here, but with more than $400 billion of potential value from gen AI at stake, it's worth it.[6]

Barr Seitz: Staying with customer service, how will gen AI agents help enterprises?

Jorge Amar: This is a great question, because we believe the immediate impact comes from augmenting the work that humans do even as broader automation happens. My belief is that gen AI agents can and will transform various corporate services and workflows. It will help us automate a lot of tasks that were not adding value while creating a better experience for both employees and customers. For example, corporate service centers will become more productive and have better outcomes and deliver better experiences.

In fact, we're seeing this new technology help reduce employee attrition. As gen AI becomes more pervasive, we may see an emergence of more specialization in service work. Some companies and functions will lead adoption and become fully automated, and some may differentiate by building more high-touch interactions.

Nicolai von Bismarck: As an example, we're seeing this idea in practice at one German company, which is implementing an AI-based learning and coaching engine. And it's already seeing a significant improvement in the employee experience as measured while it's rolling this out, both from a supervisor and employee perspective, because the employees feel that they're finally getting feedback that is relevant to them. They're feeling valued, they're progressing in their careers, and they're also learning new skills. For instance, instead of taking just retention calls, they can now take sales calls. This experience is providing more variety in the work that people do and less dull repetition.

Lari Hämäläinen: Let me take a broader view. We had earlier modeled a midpoint scenario when 50 percent of today's work activities could be automated to occur around 2055. But the technology is evolving so much more quickly than anyone had expected—just look at the capabilities

[6] "The economic potential of generative AI: The next productivity frontier," McKinsey, June 14, 2023.

of some LLMs that are approaching, and even surpassing, in certain cases, average human levels of proficiency. The innovations in gen AI have helped accelerate that midpoint scenario by about a decade. And it's going to keep getting faster, so we can expect the adoption timeline to shrink even further. That's a crucial development that every executive needs to understand.

Jorge Amar is a senior partner in McKinsey's Miami office, **Lari Hämäläinen** is a senior partner in the Seattle office, and **Nicolai von Bismarck** is a partner in the Boston office. **Barr Seitz** is director of global publishing for McKinsey Digital and is based in the New York office.

Designed by McKinsey Global Publishing