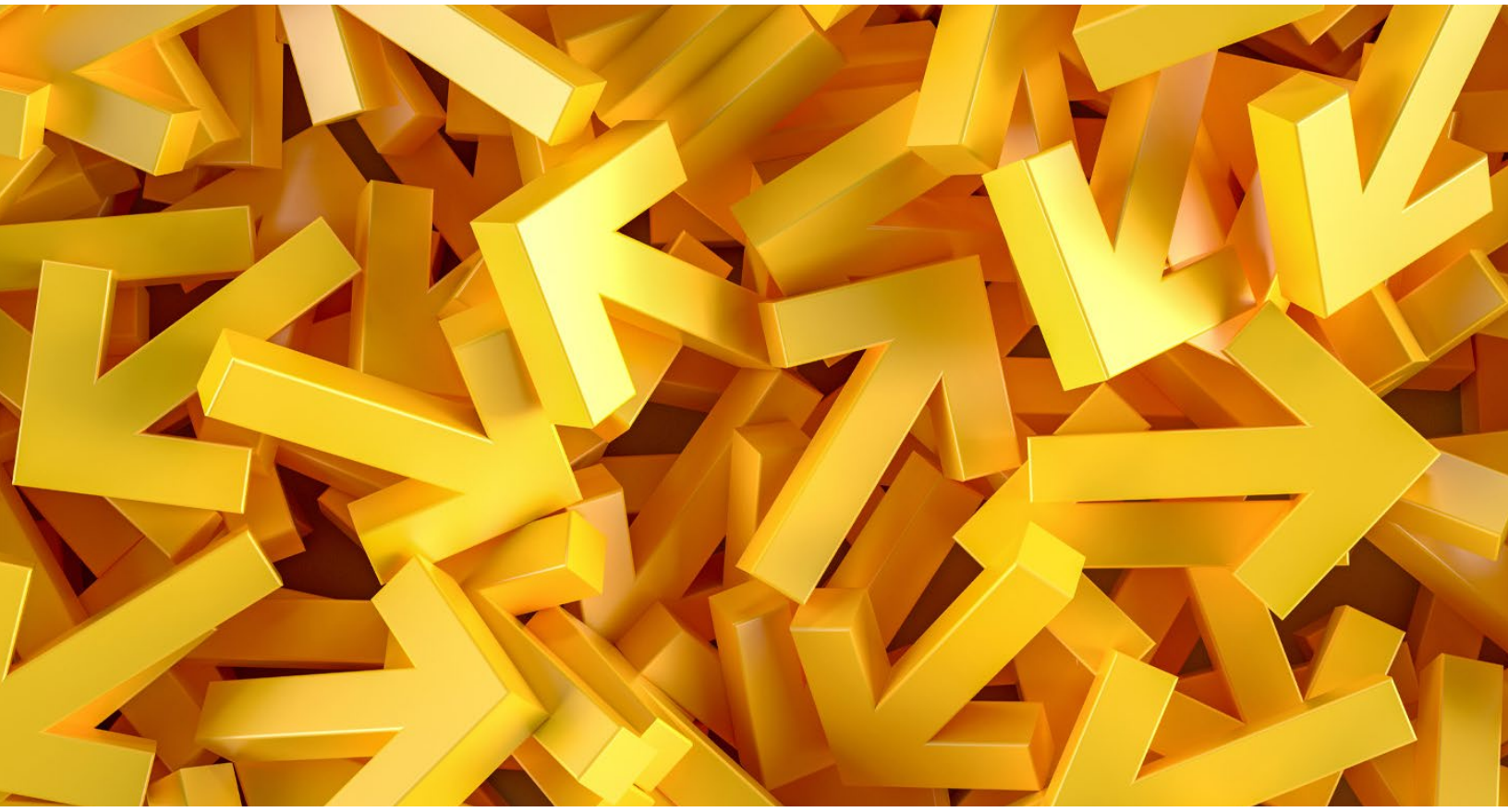
McKinsey & Company

McKinsey Direct

Gen AI in customer care: Early successes and challenges

The contact center emerged as an early use case for generative AI—with varied success. Organizations can learn from early adopters and think about their adoption strategies across three horizons.

This article is a collaborative effort by Eric Buesing, Oana Cheta, Vinay Gupta, Paul Kline, Rohit Sood, and Sophie Weihmann, representing views from McKinsey's Operations Practice.



© Getty Images

April 2024

Generative AI (gen AI) arrived in the contact center in a big way in 2023—and with it came the potential for transformational improvements in agent efficiency and effectiveness, reduced operational costs, and improved customer experience. Since then, successful adoption has been uneven, with some contact centers deploying gen AI and AI technologies effectively while others are struggling to capture value.

This is a continuation of the transformation story seen in contact centers in recent years. Despite years of effort and a plethora of efficiency-enhancing technologies, customer care leaders still report that contact centers are burdened by increasing call volumes, consistent employee attrition, and talent shortages.[1]

As businesses across industries look to capture the benefits of gen AI—and some 80 percent of organizations are anticipating an increase in AI and gen AI investments—it may be time for customer care leaders to reflect on the lessons learned from one and a half years of gen AI adoption.[2] What are the barriers to capturing the value of gen AI technologies? And in the contact centers that have deployed gen AI, what were the results?

This article looks at how organizations are leveraging gen AI capabilities to improve their business operations. It shines a light on early case examples and successes, and explores how companies can think about their gen AI adoption plans to emerge from the sidelines and realize the technology's potential value.

Understanding the barriers to impactful adoption

For many contact centers, gen AI may not have been the cure-all solution to entrenched challenges. Five barriers have been shown to stand in the way of successful adoption.

Maintaining high-quality data: Gen AI outputs are only as valuable as the data sources they draw from. In the contact center, data sources are varied and potentially inconsistent and can include knowledge base articles, product manuals, standard operating procedures (SOPs), call recordings and transcripts, and customer details.

Understanding the business case: When considering gen AI technologies, focusing on the wrong use cases could lead to a low ROI—gen AI may not always be the appropriate tool to address specific pain points. For example, a contact center that mostly handles calls rather than chat may find that gen-AI-enabled transcription and synthesis cannot keep up with the pace of real-time customer conversations.

Designing with the end user in mind: Gen AI deployment may lose steam if the end users' needs and experiences are not considered. Workflow and user interface design can be as important as technical functionality.

Securing the right talent: Contact centers don't always have the in-house technical talent, such as data scientists and data engineers, to develop and deploy gen AI solutions at scale.

Prioritizing change management: It's never just tech. In some cases, adoption may be low, even if gen AI tools are well designed and an excellent fit for the contact center's context. Often, experienced agents may not see the need to augment their tasks, or the value of gen AI tools has not been communicated across the organization.

Three adoption approaches: Assessing your options

With these barriers in mind, should organizations go all-in and develop a proprietary foundation model from scratch? Or will purchasing an off-the-shelf gen AI tool do the trick? In the contact center context, it may be helpful to learn from the three common adoption archetypes that companies have explored—each requiring varying degrees of effort, internal capabilities, and investment, and each presenting different levels of risk and possible return.[3]

[1] "Where is customer care in 2024?," McKinsey, March 12, 2024.
[2] Ibid.
[3] "Technology's generational moment with generative AI: A CIO and CTO guide," McKinsey, July 11, 2023.

For each archetype, the total value at stake may differ depending on the technical maturity of the contact center. The estimations that follow assume medium maturity.

The **taker** approach uses existing models—off-the-shelf gen-AI-enabled products that internal IT teams can integrate into workflows and deploy relatively quickly. In contact centers, translating customer conversations and summarizing customer interactions are among the few use cases where ready-to-use models are appropriate.

Consider one contact center that handles a high volume of multichannel customer queries. Service agents typically spent 5 to 10 percent of the average handle time capturing notes after a call. The ticket management technology they use recently launched a gen-AI-powered summarizing feature. When the organization subscribed to and activated this add-on, agents could instantly summarize text and voice interactions, reducing contact time by ten to 20 seconds per interaction.

The post-call note-taking task, however, is not completely automated; beyond the live customer conversation, agents still need to consider all the information they can see in the system and manually add additional context and nuance to their notes. So, while the ease of implementation in the taker approach might be high and the investment relatively low-risk, use cases are limited and the value that can be captured is modest—the total value at stake represents only around 3 to 5 percent of the entire customer operation.

The second archetype, **shaper,** specializes an existing gen AI model with internal data and systems to fit the organization's specific use case. This might require creating a customized set of prompts that will surface context-appropriate responses, formatting internal data and incorporating it into the large language model (LLM), and setting up appropriate internal infrastructure. The total value at stake of all the use cases in this category represents 10 to 15 percent of the entire customer operation. In the future, these gains could be higher in contact centers that spend a lot of time resolving simple calls; however, since gen AI models currently take a few seconds to transcribe speech into text, chatbots and chat-based copilots may have more potential in the near term than voice copilots.

For instance, one healthcare company used the GPT 3.5 model to streamline claims processing and enable frontline agents to quickly retrieve claims-related information. Implementation involved creating a set of refined prompts, standardizing the format of internal data, and testing the system with agents until the technology worked (see sidebar, "From proof of concept to pilot—A healthcare claims processing case study").

In another example, a technical help desk for a machinery distributor adopted a smart-assist gen AI tool that processed over 13,000 knowledge base resources and equipment manuals and could diagnose issues and recommend solutions. The contact center saw a 10 percent increase in first-time-right resolution and a reduction in task completion time to one minute from 15 minutes. Customers, too, benefited from the cost savings of shorter machinery downtimes.

Makers, on the other hand, build and train proprietary foundation models. An organization developing a gen-AI-enabled voice- or chatbot, for example, may want its LLM or, more likely its small language model (SLM), to respond in a tone and manner matching the company's voice, adopt a range of context-appropriate personas, and disengage if customers cross predefined boundaries. While these solutions can be incredibly powerful, training and running a bespoke LLM or SLM requires significant upfront and recurring investment, highly specialized skills, enormous high-quality data sets, and powerful computers. The size of the investment means the risk of not realizing favorable returns could also be higher for solutions that do not have a solid business case.

Frontline learning is another potential use case for these proprietary foundation models. One technical help desk built a simulation tool—a model that ingested existing call recordings and

internal knowledge resources such as SOPs. The AI could generate multiple simulations that mirrored customer interactions across a variety of realistic issues and customer personas. By interacting with the simulation tool over time, agents could practice content and soft skills and improve their simulation scores—improvements that were reflected in real-time customer calls. Early results from this proof of concept suggest that 80 percent of ongoing training could be replaced by AI-driven training, and lead to a 33 percent improvement in customer interactions.

Technology solution providers are also beginning to scale their LLMs into product offerings, providing models pretrained on industry-specific customer care data.

For those organizations that take the riskier and more complex maker route, the rewards can be significant. The value at stake could be 20 to 30 percent of the entire customer operation—potentially higher depending on organizational readiness and maturity.

Move quickly from assessment to deployment

To determine which archetype is most suitable, organizations can look at both their intended use cases within the contact center, and their current maturity and state of organizational readiness.

There is intense hype around gen AI, and organizations are at risk of being weighed down by excessive proofs of concept and pilots. To bring clarity and focus to this process, readiness assessments are critical. Organizations can then move beyond readiness assessments and begin building out key tools to enable successful gen AI deployment—including LLM decisions, governance models, and talent plans—as well as the enabling and mitigating systems and processes required to manage enterprise-level disruption.

Organizations can take six key considerations into account as they undertake readiness assessments and build tools and capabilities:

Data. Before getting started, it's fundamentally important to understand current data access and the quality of the data available. One organization on this journey found that call recordings were not easy to access and format, which limited its ability to adopt the taker approach.

Servicing strategy. It's important to align the contact center's gen AI strategy with the company's overall technology aspirations. Organizations new to AI and gen AI may find trusted, off-the-shelf solutions with substantial vendor support the most attractive to implement in the short term. There are rarely one-size-fits-all solutions on the market, however; organizations could rather focus on use cases that address opportunities and pain points specific to their contact center, identifying gen AI interventions that drive real business value where it matters most. The taker approach often has less value at stake, for example, compared to the more resource-intensive shaper or maker use cases.

Operating model. As with any transformation program, a carefully planned gen AI initiative brings together business, operations, and technology. This can ensure that the resources required to select, develop, implement, maintain, and adapt the AI and gen AI tools are readily available to support business objectives. It is also important to establish risk guardrails early on, particularly for companies that deploy gen AI in customer service operations first before scaling to other areas of the organization.

Talent. People are central to the success of gen AI adoption, so leaders need to ensure they have the right skills and capabilities to innovate and execute on their strategy. Keeping ahead of the emerging gen AI skills gap and talent demand may be especially critical in shaper or maker use cases.

Technology. The gen AI solution ecosystem is constantly expanding, with an almost overwhelming number of options to choose from. Leaders can adopt an agile innovation approach and revisit their adoption strategies to keep up with technological changes. For instance, they can seek to understand which AI solutions are offered

by current telephony and customer relationship management (CRM) vendors, and where specific use cases may require LLM solutions to be built or modified in the short term. For the "agent assist" use case to work, for example, the organization would need to take the cost of physical infrastructure and other licenses into account, along with the process of customizing the model with internal data and refined prompts.

Adoption and change management. Defining the path to adoption and scaling early could set the customer care team up for success. Effective communication—keeping the operations team in the loop, continuously providing feedback, and enabling precise communication mechanisms—can make a difference. For instance, an insurance company built a gen-AI-powered agent assist tool, but found that adoption was low. Agents and management were not made aware of the potential benefits, and top-performing agents had already memorized the steps it took to complete their tasks without the agent assist. Only after a few agents started using the tool did the benefits come to light. This group of early adopters was encouraged to talk to their colleagues about the advantages of the new technology. Other agents followed their lead, adoption spread, and the value of the technology was amplified.

From proof of concept to pilot—A healthcare claims processing case study

A healthcare company introduced a gen-AI-enabled app to streamline claims processing—a six-week proof of concept that laid the foundation for a pilot and, potentially, a contact-center-wide rollout.

To process one customer's claim, claims adjudicators typically consulted at least three additional data sources, including an internal knowledge base system and the company's member portal. To speed up this manual task, the company imagined an application with an intuitive chat interface where adjudicators could ask plan-specific questions pertinent to customer inquiries or claims, and receive responses synthesized from internal data sources.

The company deployed a functioning gen AI application based on GPT 3.5 that targeted the steps in the claims process where the technology would add the most value. The model drew on the company's high-quality internal data sources, and customized prompts were created and refined. These prompts or questions could provide answers such as whether a particular service or service type is covered for the customer and how much, the details of the customer's plan, and the maximum eligibility age.

To demonstrate the potential impact of the gen AI intervention, the company conducted rigorous testing. Ten testers, among them team leaders and select working-level specialists, used the technology to process live calls and recorded calls. The application's outputs were continuously verified, and testers shared feedback during debriefing sessions.

The proof of concept demonstrated substantial potential benefits. For instance, 15 to 20 percent of total time could be saved, depending on adoption and uptake. Reduced handling time could lead to improved customer satisfaction. And an elevated employee experience could be anticipated, too, thanks to more efficient tools and processes.

The proof-of-concept program worked well for this healthcare company, likely because there was high engagement from workers during development and because infrastructure and data systems were in a good state. As the application was refined, the accuracy of answers went up to 90 percent from 60 percent, and the proof of concept was ready to scale to pilot.

Build momentum and demonstrate impact quickly

Organizations eager to deploy gen AI in the contact center can think about this transformation through three horizons, with the dual goal of building momentum and demonstrating impact quickly. While the three adoption archetypes outlined earlier describe how companies are approaching gen AI adoption, the three horizons illustrate how, practically, these approaches may be deployed and rolled out over time in the contact center environment.

These horizons enable organizations to: first, build momentum with simpler AI and lower-risk gen AI; second, unlock new potential with gen AI tools to perform coaching and enable consolidated access across different knowledge bases; and third, leverage more complex intelligent agents with hyperpersonalization or the automation of selected tasks.

1. Build internal AI muscle

Early AI implementations can help establish the fundamentals and provide a basis for further growth. Here organizations could focus on building capabilities with off-the-shelf solutions, deploy pilot groups to test and learn from their experiences, and establish principles for responsible AI use that endure as gen AI adoption scales.

Examples include the taker archetype use case of synthesizing customer interactions to reduce the administrative burden on customer service agents. Similarly, analytical insights can be drawn from customer interactions, which can be used to identify patterns in customer queries and pinpoint coaching opportunities to upskill agents. For instance, one bank's contact center handled many queries related to money transfers and could use gen AI tools to understand the granular reasons for these calls, including customers' questions about international transfer fees and how money transfer methods differ.

Translation from English to other languages, or vice versa, is another good taker example that can help build internal muscle with gen AI. Such tools can support chat agents who can only write in one language, meaning companies do not need to hire bilingual agents.

2. Scale up internal use cases

Once pilot programs are deployed and early successes captured, proofs of concept can be scaled to the entire contact center to capture potential savings. These proof points can help build confidence to invest in more complex AI use cases and realize even higher value. Momentum is further maintained as agents, analytical teams, IT teams, and management can see the potential of the technology firsthand and could be more likely to embrace it further.

Such "scale up" examples might include a gen AI agent copilot tool. Here organizations could build on the interaction synthesis from early use cases to support agents with knowledge extraction and summarization capabilities, personalized conversational dialogue, auto-generated content, and memory retention.

A telco that deployed a gen AI agent copilot has seen a 40 percent decrease in handle time for chat queries and a 25 percent decrease in costs, alongside a 14 percent increase in agent productivity (as defined by number of cases resolved per hour). For lower-skilled workers, this improvement was as high as 35 percent.

3. Scale to external-facing use cases

Finally, those organizations that have built AI deployment capabilities at scale using internal use cases may be best positioned to embrace the customer-facing opportunities, particularly to supplement agent support during surge conditions.

This third horizon of gen AI adoption in the contact center might include well-designed empathetic and intelligent voice- and chatbots that customers interact with directly. While such custom-built LLMs may have the highest value at stake, they demand considerable customization and expertise.

Until now, the contact center environment provided an important organizational test bed for generative AI, as organizations looked to leverage this technology to drive efficiency and improve customer experience. Now, equipped with hard-learned lessons, customer care leaders can set loftier aspirations and embrace gen AI transformation with focus.

As companies determine the pace and scale at which they enter the world of gen AI, thinking about their adoption journey over three horizons can help them capture learnings, build momentum, and harness value—for employees and customers.

Eric Buesing is a partner in McKinsey's Charlotte office; **Oana Cheta** is a partner in the Chicago office; **Vinay Gupta** is a senior expert in the Boston office; **Paul Kline** is a senior expert in the Dallas office; **Rohit Sood** is a senior partner in the Toronto office; and **Sophie Weihmann** is a consultant in the New Jersey office.

The authors wish to thank Cherisse Diaram and John Larson for their contributions to this article.

Copyright © 2024 McKinsey & Company. All rights reserved.