ASAPP

Evaluating generative AI agents for your contact center

Uncover the truth beneath the marketing spin by asking the right questions

Every vendor who sells a generative AI agent for contact centers makes the same vague claims about what you can achieve with their product.

Increase efficiency, unleash your team's potential, and level up your CX. Supercharge everything from personalization to productivity—and realize value like never before.

Every vendor also assures us that they prioritize security, safety, and transparency. Trust us, they say.

There's nothing wrong with any of those claims. They're likely made in good faith. But without details to explain exactly what they will deliver—and how they'll do it—there's no way to know which solutions are right for your contact center. That's the problem with marketing language. It's often big on promises, but short on details. That makes a fair comparison of solutions and vendors more difficult than it should be.

We know that ASAPP is *yet another* AI vendor for the contact center making claims about how our generative AI agent can help your business. But we also recognize how similar many solutions sound based on their marketing content—and how that makes navigating this AI landscape nearly impossible.

So, if the marketing-spin machine has left you as dazed as a carnival ride, then let's stop the ride. It's time to get off. The best way to do that is to know which questions to ask and how they can help you gain the clarity you need to make an informed choice.

The following list of questions is not exhaustive. But it's a good start for rooting out the details you'll need to make a fair comparison of generative AI agents.

What *exactly* does your AI agent do?

Autonomous agents powered by generative AI are new to the contact center. And as technology vendors introduce new offerings, they are labeling a wide range of solutions as AI agents. In other words, all AI agents are not the same. Some simply automate specific processes or serve up information and other guidance to human agents, while others can operate independently to talk with customers, assess their needs, and take action to resolve their issues. So, it's important to dig into the details of what each solution actually does.

What to ask

In addition to providing responses, can your AI agent take action on behalf of customers?

What to look for in the answer

The current crop of AI agents varies widely when it comes to capabilities. Some are mostly talk with little action. They can listen, speak conversationally, and serve up relevant information—but they can't execute tasks independently to resolve customer issues.

When customers reach the contact center, they often need more than answers and information. They need to have specific problems resolved. For AI agents to be genuinely effective, they need to go beyond merely responding to questions. They must also be adept at taking action, such as interfacing with back-end systems to access customer data and handling requests like initiating returns for e-commerce platforms or booking flights for airlines.

Today, *very few* AI agents can fully automate interactions. Providers of AI solutions often promote their systems' capabilities, but those unable to operate autonomously frequently highlight the following features:

- Identifying customer intent to direct them to the appropriate service flow
- Serving up answers to questions using a knowledge base (or *retrieval augmented generation*)
- Enhancing specific parts of pre-existing interaction flows to make them more conversational, such as collecting user information like names or birthdates

While these features do enhance traditional IVR systems, using generative AI capabilities solely for these tasks is akin to using a bazooka to kill an ant—it's an overpowered solution for limited gains. Large language models present a technological shift that will enable the industry to deliver truly effective customer self-service experiences.

Look for a solution that maximizes the impact of generative AI by fully automating interactions.

How do you configure flows for your AI agent?

While different solutions might seem similar at first glance, there can be significant differences in how their flows are configured. Some AI agents are designed from the ground up for generative AI. Others are more like traditional bots that have learned a few new generative AI tricks.

The configuration of these systems often reveals their nature.

Traditional bots, which have integrated large language models (LLMs), operate through **predefined flows**. These flows can be enhanced by LLMs at specific nodes to create more natural conversational experiences. For instance, if a traditional flow requires a customer to enter their birthdate and the customer responds with "Today's my birthday!" a LLM can interpret this phrase and convert it into a date format—something that would've surely broken an IVR flow.

In contrast, AI agents developed explicitly for generative AI operate without such structured flows. They are designed to be goal-oriented, adhering to procedures and guardrails set out in natural language. For example, an AI agent tasked with scheduling appointments might access a calendar to book a slot for a customer. Rather than following a series of preset steps, the AI agent identifies the necessary information—such as date and time preferences—from the customer and processes it to fulfill the appointment request autonomously.

Traditional IVR and chatbot flows are notoriously inflexible, failing when a conversation deviates from the expected "happy path" or when the customer switches topics. Incorporating LLMs into these traditional systems doesn't address that inflexibility.

AI agents built on a core of generative AI are naturally more flexible. They can handle shifts in conversation topics, much like a human agent, accommodating the unpredictable flow of human interaction without breaking down.

As a general rule of thumb, if a vendor showcases a flowchart as part of their AI solution, it's likely that you're dealing with a traditional bot augmented with generative AI capabilities, rather than a true generative AI agent.

What channels does your AI agent support?

Most AI agent providers focus either only on chat or only on voice. Each channel presents unique challenges.

Voice is often a more popular, and costlier, channel. So, an AI agent that can provide a great customer experience and resolve issues on its own will often deliver greater value in a voice channel—but that depends on the channel mix and cost profile of your business.

So far, few solution providers have cracked the code for voice as a viable option with their AI agents. That's because creating safe and grounded AI agents requires orchestrating multiple models, which takes a few seconds. Voice AI agents that respond extremely quickly are often not grounded and will communicate inaccurate information. Ask the vendor how they balance accuracy with speed when designing a voice experience.

Look for a solution provider that already supports voice with an AI agent that's both accurate and provides a good conversational experience. They're leading the pack with more advanced technology and, long term, their track record of innovation will likely deliver greater value to your business.

How can I incorporate your AI agent into the infrastructure of my support channels?

Some solutions will integrate into existing platforms in such a way that the conversation never leaves the platform, leveraging the capabilities of an AI agent while maintaining the end-to-end visibility you get with your CX platform.

Exercise caution with solutions that divert calls or chats outside your existing technology ecosystem. If you are relying on a Contact Center as a Service (CCaaS) platform for comprehensive customer journey tracking, analytical insights, and call recordings, ensure that the AI provider integrates seamlessly within this framework. For voice calls, ask the AI provider whether they're doing a SIP transfer, which indicates the call is leaving your technology stack. If that's the case, you'll lose the end-to-end visibility for your calls. For chat, ask if they'll use your existing chat bubble on the website or on the app. Some vendors offer native connectors to CCaaS platforms, allowing you to leverage the AI agent without losing the advantages of your current system (although not all CCaaS platforms allow those connectors).

Can I keep my existing automation, or does your AI agent need to replace it?

There are generally two ways in which AI agents can be added to your workflows:

1. **AI agent first. Then a human agent.**

   Incoming chats or calls are initially routed to the AI agent. This approach bypasses your existing automation. If the AI agent cannot resolve the issue, it transfers the interaction to a human agent.

2. **Existing automation. Then an AI agent. Human if all else fails.**

   Calls or chats start with your existing automation, such as an IVR system or routing solution. Certain use cases are then handed off to the AI agent for resolution. If the AI agent is unable to assist, the matter is escalated to a human agent. Transitions here must be seamless, so it feels like a single experience to the customer.

If you have not developed robust self-service flows within your IVR or chatbots, piloting the AI agent to be the first point of contact for all interactions might be a good idea. On the other hand, if your existing automation can efficiently manage some queries, an AI agent can supplement those capabilities, handling specific use cases or stepping in when your automation reaches its limits. Both strategies are valid, so the right choice depends on your goals. You'll want to seek solutions offering the flexibility to tailor the AI agent to your needs.

Is there more to your solution than a LLM + RAG?

The introduction of retrieval augmented generation (RAG) dramatically improved the accuracy and relevance of generative AI output. By grounding a generative AI agent on an authoritative source, such as your knowledge base, RAG helps the solution overcome some of the natural tendencies of LLMs to produce erroneous or irrelevant responses.

The much-improved results that the addition of RAG has yielded begs the question—is RAG plus a foundational model all you need? The short answer is no. There's much more to a safe and high-performing generative AI agent, so you'll want to dig into how the solution is structured.

What to ask	What to look for in the answer
Which models (LLMs) does your solution use?	It's a basic question, but an important one. There's no single model you should look for in the answer. The point here is that the solution should be capable of using and orchestrating a wide variety of models.
Besides a LLM and RAG, what other technologies does your solution include? And how is it structured?	The best-performing generative AI agents don't rely on a single LLM. They use multiple models, each one for a specific purpose. With that approach, the solution provider can use each model for what it does best, choosing different models for different use cases. Some models will be included solely as quality assurance measures, reviewing output for accuracy, relevance, and appropriateness before it is communicated to your customer. Grounding the agent in the information sources you choose, such as your knowledge base, is critical. Using RAG, the solution should be able to refer to multiple authoritative sources as needed. Because the agent will need to retrieve data from your other systems using APIs, authentication and security measures are necessary to protect your data—and your customers' data.

The solution should also include a variety of safety mechanisms to prevent harmful input, such as prompt injection, as well as inaccurate or harmful output, such as hallucinations or offensive language.

AI agent systems are usually controlled through an orchestration engine.

The bottom line is that a LLM and RAG are not enough. To ensure safe, consistent performance that you can trust to represent your brand, these other components are critical.

Will I get locked into using a specific LLM forever? Or is your solution flexible enough to allow changes as models evolve?

The competitive landscape with LLMs is constantly changing. Today's best-performing models are already very different from what was state-of-the-art a few months ago. Even channel choice is a factor. As model developers improve accuracy, reduce latency, and invest in multi-modality, their models become a more attractive option for voice interactions.

You'll want to choose a solution provider that maintains flexibility regarding model selection. One that has a track record of successfully switching models as their relative benefits change will be your best bet for long-term flexibility.

Ask the provider whether they have experience migrating models in a system that's been deployed to production, how they reap the advantages of better models, and how they manage that process.

How will your solution protect our data (and our customers' data)?

Security is a top concern with all technology solutions, and AI is no exception. In fact, it adds some new risks into the mix. A good example is prompt injection, in which a bad actor attempts to disguise malicious input as a legitimate prompt to manipulate the AI into leaking sensitive data, granting access to restricted systems, or saying something it shouldn't. Any AI vendor worth considering should be able to explain their approach to security boundaries in detail.

What to ask	What to look for in the answer
How do you ensure that the AI agent cannot be exploited by a bad actor to gain unauthorized access to data or systems?	The solution should be able to detect when someone is trying to exploit the system by asking it to do something it should not. It should include mechanisms that detect and block malicious inputs, including code and phrases that would effectively redirect the AI agent with new orders that override its intended purpose. Of course, robust security boundaries are crucial to successfully limit the AI agent's access to *only* the data it needs to resolve the customer's issue. They also help prevent bad actors from gaining access to the data.
How do you ensure that the AI agent cannot retrieve data it is not authorized to use?	AI agents access other systems using APIs. An API is a set of rules that enables software applications to communicate with each other to exchange data or activate functionality. Security and authentication in this API layer are critical to ensuring that the AI agent can access data only for the customer it's actively interacting with and only for the task it's currently performing, and that it can't retrieve data it is not authorized to use.

How does your solution maintain data privacy during customer interactions?

During customer interactions, personal identifiable information (PII) is sometimes part of the conversation and is often necessary to resolve the customer's issue. To maintain data privacy, this PII should always be redacted before the data is stored. Ask the provider what mechanisms they have for redacting PII and how accurate that redaction engine is.

Will your solution send any of our data or our customers' data to third parties?

Many vendors include third-party components within their solutions. Often, that means that your data, including customers' PII, could be sent to those third parties.

Look for vendors who can guarantee that your sensitive data will stay on their infrastructure, and if they share data with any subprocessors, that any sensitive data will be redacted prior to that data leaving the vendor's infrastructure. Also ensure that the vendor can meet your organization's data retention and deletion needs for historical data.

Ask vendors for an approved list of subprocessors, which should be published and accessible. Be wary of long lists of subprocessors, particularly if data is leaving the cloud and isn't redacted, as that means your customers' data will be exposed to more points of failure.

How do you keep your AI agent from ticking off my customers or damaging my brand?

Some of the biggest risks with generative AI aren't related to data security. Because generative AI creates responses on the fly, it can generate output that is inaccurate, biased, or otherwise harmful. We've all heard stories of bots that spouted offensive language, agreed to sell pricey products for a pittance, or encouraged people to do unsafe things. Solution providers that are worth considering should have robust safety mechanisms built in to ensure that the AI agent stays on task, produces accurate information, and operates ethically.

What to ask

How do you mitigate and manage hallucinations?

How do you prevent the agent from sharing misinformation with our customers?

What to look for in the answer

Generative AI agents are a lot like many of your new hires—eager to please and hesitant to say *I don't know*. That tendency to sometimes answer even when it doesn't have accurate information is at the heart of genAI's hallucinations.

It's important to understand that because of the way genAI functions, it's not possible to *completely* eliminate hallucinations. Solution providers should be candid about that. If a solution provider isn't clear and unequivocal on that point, that should be a red flag. But just because hallucinations can't be totally prevented, that doesn't mean there's nothing you can do about them.

Vendors should *measure* and *reduce* their occurrence and have mechanisms in place to prevent the solution from communicating inaccurate information to your customers when hallucinations do occur.

Trustworthy solution providers will explain how they ground the AI agent in information specific to your business, such as your knowledge base. You might hear them refer to RAG, or retrieval-augmented generation, a technique that improves the output of LLMs with relevant reference data. This approach reduces inaccurate output (hallucinations).

They should also be able to explain how the solution catches hallucinations with QA mechanisms specifically designed for that purpose. These mechanisms often incorporate additional AI models to evaluate output, spot hallucinations, and require the agent to try again—before responding to your customer. And if an accurate response cannot be generated and human assistance is necessary, the solution should be able to hand off the interaction to a human co-worker.

How do you prevent jailbreaking?

Trustworthy generative AI agents are multi-layered solutions that incorporate a variety of guardrails to protect against people who want to exploit them. Jailbreaking refers to a number of techniques that bad actors use to get around those guardrails. It's a little like a prisoner who convinces a guard to help them escape or maybe just "forget" to lock a door.

The techniques for jailbreaking run the gamut from complex and highly technical to very simple and surprisingly human. People with malicious intent might try to inject code through a prompt that would override the agent's instructions or force it to bypass guardrails. Or they might just sweet-talk the AI agent into breaking the rules. That's not as far-fetched as it might sound. Generative AI can be very literal in its reasoning, and it lacks the human capacity for nuanced judgments about things like another person's motives. That makes it a little gullible and easy to persuade.

But the LLMs at the core of a trustworthy generative AI agent are just one component of the complete solution. Those models should be wrapped in safety mechanisms.

Given the range of techniques for exploiting AI, solution providers must take a layered approach to detecting and mitigating actors who are engaging in bad behavior, including:

- **Leaking prompts from the system**
- **Injecting instructions into the system**
- **Using foul language**
- **Engaging in out-of-scope topics that could be damaging to your brand, such as jokes about politics or religion**

How hard will the solution be to use and maintain?

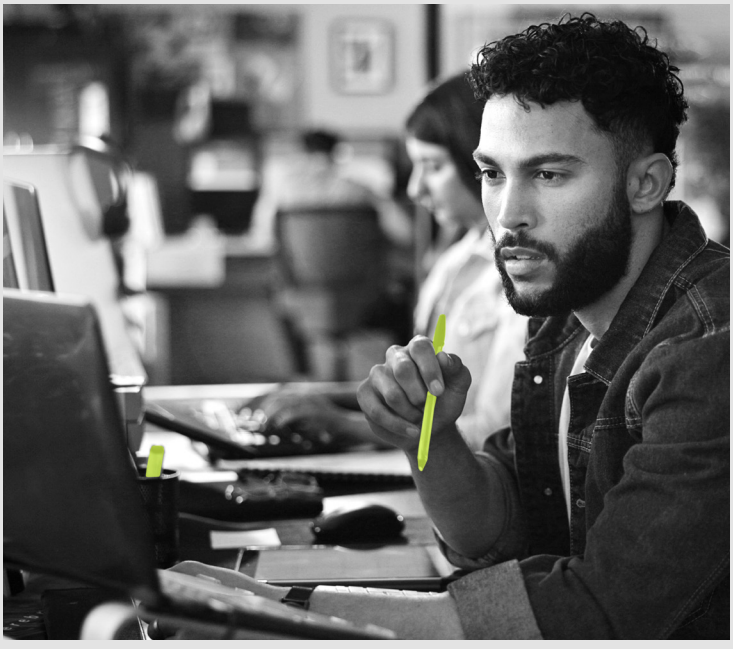
Conditions in a contact center can change quickly. Product updates, new service policies, modified workflows, revised knowledge base content, and even shifts in customer behavior can require your agents to adapt—including your AI agents. That means your team will need a flexible solution that empowers them to handle simple tasks like policy updates on their own, without waiting on technical resources.

What to ask

What kinds of changes and updates can our contact center team make to the solution without pulling in developers or other technical resources?

What to look for in the answer

Solutions with robust no-code tooling empower non-technical business users to manage changes and updates on their own. Look for a solution with an easy-to-understand console that offers a substantial range of options.



How will we know what the AI is doing—and why?

When a human agent performs exceptionally well—or makes a mistake—you can ask them to explain their reasoning. That's often the first step in improving performance and ensuring they're aligned with your business goals. A supervisor can then listen to the call to provide targeted coaching.

It's equally important to understand how an AI agent is making decisions. But many solutions remain a black box, offering little or no insight into how the AI is reasoning and what specific data it's using to generate responses. It's crucial to prioritize solutions that shine a light into that box to provide a high degree of *transparency* and *explainability*.

What to ask	What to look for in the answer
How will we know what specific tools and data the AI agent is using for each customer interaction?	Generative AI agents access other systems to retrieve data and information and perform tasks to help the customer. All of these actions should be clearly documented in a readily available audit trail. Look for a solution that provides an interface where you can see what actions the AI agent is taking beneath the hood. This solution should provide aggregate metrics and reporting that allow your data teams to understand what's happening at scale.
In what ways do you surface information about how the AI agent is reasoning and making decisions?	In addition to listing the systems and information sources it accesses, a generative AI agent should explain its reasoning for each action it takes. This record of the AI agent's decision-making should be retained for later analysis.

What do you mean by human in the loop?

You probably won't find a solution provider who doesn't acknowledge the importance of having a human in the loop with a generative AI agent. But that doesn't mean they all agree on what that human should be doing or how the solution should enable human involvement. For some, human in the loop is little more than a general assurance for CX leaders that their team can provide oversight. Others mean solutions in which AI agents support human agents but don't ever interact with customers. And then there are the pioneers who are elevating the concept of the human in the loop with AI agents that serve customers directly and rely on humans as advisors, as needed. So, you'll want to assess exactly what each vendor means when they say *human in the loop*.

What to ask

What roles do the humans in the loop play? Are they involved primarily during deployment and training, or are they also involved during customer interactions?

What to look for in the answer

Typically, solution providers involve members of your team during deployment and initial training to provide input and feedback that will improve the AI agent's performance. And many are developing consoles that will enable ongoing oversight of the AI agent with methods for continued optimization.

But for some vendors, that's largely where the humans' role ends. When it comes to customer interactions, your human agents are simply escalation points for when the AI agent fails, and the experience is a lot like what happens when a traditional bot fails. Some vendors still refer to that scenario as a human in the loop, mainly because your customer gets handed off to a human.

A better approach is to enable human agents to work directly with AI agents much as they would with a new hire. Inexperienced agents typically ask a supervisor or more experienced colleague for help when they get stuck. Look for a solution in which the AI agent can do the same thing. In this type of solution, the AI agent works independently and alerts the human in the loop when it needs help. That human agent can then provide what the AI agent needs to continue handling the interaction: guidance, a decision, or authorization for actions the AI agent is prohibited from doing on its own. This type of solution also allows the human agent to step in to speak with the customer directly as needed.

When and how does your gen AI agent interact with a human agent?

The most important thing to listen for here is flexibility. The solution should be able to enlist the help of a human whenever it:

- Needs to access a system it cannot access on its own
- Cannot resolve the customer's issue on its own
- Requires a decision or authorization by policy

And it should be configurable so *you* can set the ground rules for when the AI agent acts independently and when it relies on its human coworker. You might even want to limit some of the AI agent's actions with the initial deployment until you and your team have confidence in its performance. The solution should make it easy to expand what the AI agent can do over time.

When the AI agent does ask for help or hand off a customer to a human coworker, it should clearly communicate what's needed and what it is unable to do.

Can the AI agent ask the human agent for the input it needs to resolve the customer's issue without handing over the interaction to the human?

Some solutions treat the human in the loop mostly as a failsafe. They automate some processes and then escalate the interaction to a human when the AI reaches a point of failure.

But the most effective and efficient solutions create a more cooperative relationship. Look for a provider that strives for an ideal balance between humans and AI, relying on each for what they do best. With this type of solution, instead of completely handing off an interaction, the AI agent asks the human agent for input, authorization, or a decision about what it should do next. With that guidance, it continues interacting with the customer and resolving their issue. And interactions that require human skills, such as building trust or selling services, are routed directly to human agents, who have more time to focus on them now that the AI agent is handling the more repetitive interactions.

Why should we trust *your* team?

Trust depends on a number of factors, but it starts with expertise in both AI and in the real-world challenges of the enterprise contact center. What you need to know is whether a vendor has the expertise to deliver a reliable solution that solves real CX problems now—and continue improving it in the future. Generative AI is evolving at lightning speed. What's cutting edge today will be out of date shockingly soon. So, you need to be sure that your solution provider can keep up with the pace of innovation. If they depend on a third-party model provider for a significant portion of their solution, that's a strong reason for caution. But you should dig deeper than that one question.

What to ask	What to look for in the answer
Have you successfully deployed your solution in a company like ours?	Any serious vendor can provide an impressive demo of an AI agent. But there's much more to a successful deployment than the solution's core functionality. That's why many enterprise deals fall through in the procurement, security, privacy, and governance reviews. Look for AI vendors who have successfully navigated those requirements in companies like yours. Whether you're a growing technology company, or a large enterprise in a regulated industry, look for a vendor who has a positive track record and understands what it takes to navigate these processes successfully.
What components of your solution were developed in-house vs. acquired from third parties?	It's not necessary for *all* components of a solution to be built in-house, but exercise caution with any provider that lacks experience with in-house development of generative AI solutions. They might not be well positioned to continue expanding the product's capabilities as AI technologies evolve. That could leave you tied to a solution that falls farther and farther behind the rest of the market over time. It can also be helpful to discuss the vendor's product development roadmap. Just remember that a roadmap is just a plan. The vendor's track record of innovation, development, and deployment is a better gauge of what they are capable of delivering.

What kind of validation can you share from third parties?

Every vendor will point to happy customers. Some will even let you talk to a few. And they will all share performance metrics that demonstrate the impact of their solution. All of this information is helpful, but it's tightly controlled and strategically selected by the vendor.

Look for additional validation from third parties, such as reviews and articles by industry experts and reports by top-tier analysts.

Can you point me to your team's research publications and patents?

Solution providers with a deep bench of experts who are actively pushing the boundaries of what's possible with AI through published research will be ahead of the curve on key issues that affect speed, accuracy, and safety. Patented innovations are a strong sign that they have moved beyond theory to practical application.

Your demo looks great. How close is that to the real thing?

Every technology vendor cherry-picks the information they share to make the best impression. Sometimes, that approach carries through to the product demo. It looks fantastic. But how much is a real representation of the product? How much is smoke and mirrors? A highly controlled and scripted presentation can make a product look much better than it actually is. And that can lead to a disappointing deployment that fails to deliver on your goals.

What to ask	What to look for in the answer
Can I provide some of the customer inputs and answers instead of your scripted ones?	The more you can interact with the demo yourself, the more confident you'll feel that the product will perform well. If the demo cannot handle your off-the-cuff inputs and responses, it's hard to be sure it will be ready to interact with your customers.
Is this a live instance of your solution or a prototype?	If the demo is just a prototype, that shouldn't necessarily disqualify the vendor. But it should give you a good reason to dig deeper into the vendor's ability to deliver a viable product. If it is a live instance of the product, be sure to ask how this instance differs from the solution they will implement for your contact center.

Better questions lead to better information

Every solution provider wants to leave a good impression. They play up their strengths and focus on the big-picture impact they believe their solution will make for your business. The result is a market full of messaging that all sounds the same, even when the products are very different.

That makes it hard to tease out the distinctions that matter. And with new technology like generative AI agents, it's not always clear which differences are most important.

One thing is clear: **asking the right questions leads to better information —and that equips you to make a fair comparison of solutions and choose the one that's right for your business.**

About ASAPP

ASAPP is an artificial intelligence cloud provider committed to solving how enterprises and their customers engage. Inspired by large, complex, and data-rich problems, ASAPP creates state-of-the-art AI technology that covers all facets of the contact center. Leading businesses rely on ASAPP's AI Cloud applications and services to multiply agent productivity, operationalize real-time intelligence, and delight every customer.

To learn more about ASAPP innovations, visit **www.asapp.com**

LEARN MORE