ASAPP

# Should you buy an AI agent or build your own?

The customer service leader’s guide to choosing which path is right for your business

ASAPP

# Contents

The promise of a generative AI agent in your contact center is clear – increased capacity, a better customer experience, and lower cost to serve. But the actual value you realize depends heavily on performance, safety, and flexibility. That makes choosing the right AI agent for your business critical. And because every business has its own unique needs and goals, it also opens the door for this question: **Should you buy an AI agent or build your own?**

For many businesses, building a solution is out of reach. But the build-or-buy equation is different if your enterprise has sufficient technical resources, maybe even your own AI models. If that describes your business, build-or-buy is a serious and important question. To answer it, you'll need to weigh a wide range of factors from technical considerations and operational realities to value calculations.

We want to be candid here. We want ASAPP to be the answer. We would love for you to choose GenerativeAgent® for your customer service – but only if that's the right choice for your business. With that in mind, here's our best advice on what to consider as you decide whether building your own solution is the best move.

GenerativeAgent is a fully autonomous virtual agent that safely resolves complex customer interactions over voice or chat.

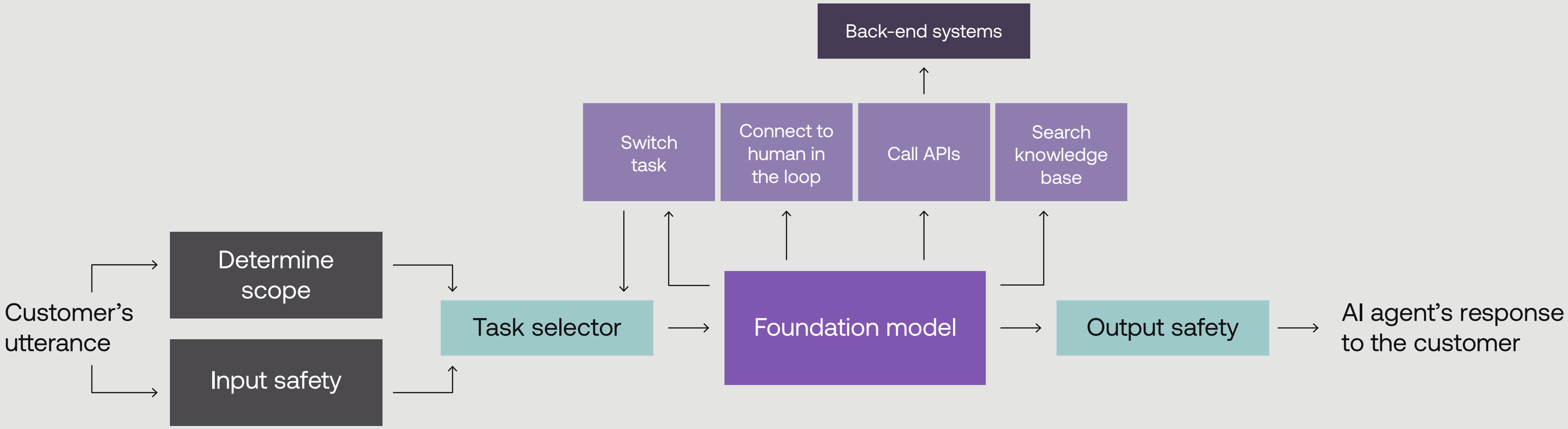
Learn how GenerativeAgent automates customer service

# What exactly would you need to build a robust solution?

A safe and reliable AI agent is much more than a large language model (LLM) that's grounded in your sources of truth. It's a complex orchestrated system that includes multiple LLMs, a variety of input and output safety mechanisms, automated QA for responses, secure connections to other systems and data sources, clear documentation of the agent's actions and reasoning, and no-code tools that allow business users to manage and optimize the AI agent's performance.

This complexity is necessary for safety, accuracy, and the level of performance an enterprise operation typically demands. It's also why there's a vast difference between building an impressive prototype and delivering a production-ready enterprise solution.

A safe and reliable AI agent requires the following components.



## Multiple LLMs

No one model will be best suited for every task, so it’s important to incorporate multiple models into your AI agent solution, using each one for what it does best. For example, you might choose one LLM to provide the primary reasoning for your AI agent and another to serve as a judge that evaluates responses for accuracy and relevance before they are sent to your customer. This real-time evaluation is important because LLMs are predisposed to prefer their own content, increasing the likelihood of inaccurate or irrelevant responses.

AI models are evolving quickly. The best model for a task today might not be the best model for that task in the future. **Your solution should allow you to easily swap out models over time as better-performing models are developed.** Keep in mind that swapping out a model in your solution will require both development and testing resources. It can also affect the effectiveness of your AI agent. Without sufficient testing and optimization, you are likely to see a temporary degradation in performance.

### How ASAPP delivers

GenerativeAgent incorporates multiple LLMs, each dedicated to a different task. The architecture of GenerativeAgent is designed so that we can swap out the backend LLM as needed whenever a more advanced or suitable model becomes available. Because ASAPP is focused on AI research and development, we stay on top of model differences, and we manage all testing and QA for seamless maintenance, even when a model is swapped. That reduces your development costs, minimizes your risk, and produces consistently high performance.

## Secure access to other systems

As your AI agent works to resolve customer issues, it will need to access other systems, such as your CRM, to retrieve data and to take action on behalf of the customer. The AI agent will rely on APIs to access these systems. In many cases, relevant APIs likely already exist in your technology ecosystem. However, they may not return data in a form that your AI agent can understand or use. Your technical teams will either need to modify or wrap those APIs, or develop another method for translating their output so your AI agent can use it.

### How ASAPP delivers

ASAPP has developed an API transformation layer that simplifies the retrieval and passing of information to GenerativeAgent to ensure high performance while at the same time filtering out data that should not be exposed to an LLM. **As a result, there’s no need to rewrite your existing APIs.** This reduces costs for both initial deployment and later expansion to additional use cases.

## Model orchestration

As powerful as the current crop of foundational models are, there are still serious limits to what they can accomplish on their own. They sometimes struggle with multi-step processes and applying contextual information in a meaningful way. Using multiple models, each for a specific purpose, can help overcome these limitations – but only with nuanced orchestration. Model orchestration manages the complex workflow across components and processes, including the interaction between the various LLMs, API calls, data retrievals, prompt management, and more. **This orchestration layer is actually the backbone of a robust AI agent solution.**

### How ASAPP delivers

GenerativeAgent is built on a sophisticated approach to model orchestration. In addition to a foundational LLM, the system uses more than ten models that each perform discrete tasks. The orchestration layer manages their complex, simultaneous interactions to deliver superior accuracy and safety.

## Data security

Because your AI agent will access other systems using APIs, security and authentication in the API layer are critical to ensure 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 cannot retrieve data it is not authorized to use.

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. If you use third-party components in your solution, you'll need to be able to guarantee that if any data is shared with a third party, it's redacted before leaving your infrastructure.

### How ASAPP delivers

**Authentication in ASAPP's API layer only permits the retrieval of data for the authenticated user.** This approach ensures that GenerativeAgent cannot make API calls for other users or systems, which enhances security and data privacy

ASAPP redacts all data before storage, including when working with third-party providers (e.g. LLM model providers).

## Input safety mechanisms

Your 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 that would effectively redirect the AI agent with new orders that override its intended purpose. Given the range of techniques for exploiting AI, you'll need to 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
- Exploiting the system to perform abusive actions
- Engaging in out-of-scope topics that could be damaging to your brand

**How ASAPP delivers**

GenerativeAgent is a multi-layered solution that employs automated input safety evaluators to filter malicious inputs and guard against a wide range of jailbreaking techniques. These mechanisms include prompt filtering, content filtering, models to detect harmful language, and mechanisms to keep the AI within scope.

## Output safety mechanisms

Simply grounding your AI agent on your sources of truth is not enough to ensure that its actions and responses are safe and reliable. Even with sufficient grounding, AI models sometimes hallucinate, generating inaccurate information. You'll need safety mechanisms to review the model's output before it's shared with the customer. And you'll need to make some decisions about what happens when the output is incorrect. You could choose to have the foundational model try again, generating a new response, or you could have the AI agent transfer the interaction to a human.

**How ASAPP delivers**

GenerativeAgent addresses and significantly reduces the risk of AI hallucinations through several measures, including the use of guardrails and safety agents that analyze traffic and outputs to identify and prevent the generation of inaccurate or misleading information. In addition, both automated quality assurance testing and manual red teaming focus on identifying hallucinations. A human-in-the-loop escalation process provides the opportunity for your customer service team to intervene and rectify any AI-generated inaccuracies. We also ensure that responses are firmly grounded in your data, policies, and knowledge base.

For actions that alter customer data, GenerativeAgent supports a confirmation workflow outside of the LLM that requires explicit user confirmation before the action is executed.

## Mechanisms to keep a human in the loop

Few enterprises are ready to turn over too much of their customer service to AI without clearly defined processes for human intervention and oversight. In other words, there's broad agreement that it's important to keep a human in the loop. **The question is, what do those processes look like, and what exactly do these humans in the loop do?** You have a range of options to consider:

- Treat the humans as an escalation point to take over an interaction when the AI agent fails, much like they do with traditional deterministic bots
- Require human oversight and approval for every action the AI agent takes, which severely limits the impact the AI agent can make on your contact center's capacity
- Define a middle ground between complete AI independence and total human oversight by creating a solution that can work collaboratively with human agents

Whatever path you choose, your solution will need to enable the human-AI relationship you're envisioning.

See why Forrester says the ASAPP human-in-the-loop capability enables the transition to AI-led customer service by capturing the tacit knowledge of human agents.

### How ASAPP delivers

GenerativeAgent supports multiple approaches for integrating humans into the workflow. It can often avoid a hard escalation by asking a human-in-the-loop agent for the help it needs to continue resolving a customer's issue on its own. It's smart enough to know when it needs help, and it knows how to ask for what it needs. Through real-time collaboration with a human in the loop, GenerativeAgent can continue serving the customer without handing off the interaction. Because it was designed and built with this collaboration in mind, GenerativeAgent enables the human-in-the-loop agent with user-friendly tools in an intuitive interface for an efficient and streamlined workflow.

You can also configure GenerativeAgent to ask humans for help at certain critical points, or even more frequently.

This advanced capability creates new possibilities for how you can incorporate GenerativeAgent into your customer service workflows. With this innovative approach to the human in the loop, you can expand automation opportunities safely, improve the performance of GenerativeAgent through self-learning and continuous optimization, and increase your contact center capacity without shortchanging the customer experience.

## Intuitive controls for business users

Customer service operations are fluid and must remain agile to respond quickly as business needs, market trends, and customer behavior change. For your AI agent, that can mean following new processes, taking on new intents, or simply optimizing performance. If adapting to new realities depends heavily on your developers or other technical resources, you're likely to experience bottlenecks that degrade both the customer experience and the efficiency of your contact center.

To avoid those bottlenecks, your AI agent solution will need no-code tools that enable non-technical business users to enact policy changes, shift brand expression, and deploy new flows without significant developer support.

### How ASAPP delivers

GenerativeAgent was built for business users to operationalize quickly with fewer labor hours and at lower operational cost. No-code tooling enables non-technical business users to handle regular optimization and adjustments to policies on their own. Over time, as you expand your use of GenerativeAgent, you can add intents without heavy reliance on IT.

## Production monitoring

One of the most significant ways a generative AI agent differs from traditional IT projects is the non-determinism of the responses. Because of its inherent probabilistic nature, you cannot simply configure your AI agent, provide some tools, put in guardrails, and then trust that it will perform as expected in production. This kind of system can do many things – and as it interacts with more customers, it will sometimes do unexpected (and potentially undesirable) things. You'll need to create tools that allow you to monitor your AI agent in real time and catch problems before they get too big.

### How ASAPP delivers

ASAPP provides robust production monitoring of 100% of GenerativeAgent responses, and has an ongoing quality improvement process (including human annotation) to identify and reduce problems that happen at scale. And, if something does go wrong, ASAPP can alert you early so that you can get ahead of any potential issues while they are still small.

# Common challenges in building an AI agent

Building and perpetually supporting a safe, reliable AI agent comes with some novel challenges. The nature of generative AI, the speed with which LLMs are evolving, and the lingering limitations of the technology make this effort quite different from other software development and infrastructure projects. Here are a few of the challenges you'll need to keep in mind.

## Balancing safety and performance

One of the toughest challenges in building an AI agent is finding the optimal balance between safety and performance. It's tempting to simply say that safety is non-negotiable, and on some level, that's true. You don't want to end up with an AI agent that does damage to your brand and your business by providing misinformation, revealing confidential data, or offering responses to out-of-scope questions.

But you also can't afford a solution that fails to resolve a high percentage of customer issues or takes so long to act that your customers will do anything to avoid it, including choosing a competitor instead. Automated safety mechanisms that evaluate validity and relevance throughout the AI agent's workflow take time to run. When a customer is waiting on the phone, every fraction of a second matters. A long wait for a resolution, or worse, a transfer to a human agent does more harm than good.

ASAPP has spent years calibrating the balance between safety and performance. As a result, GenerativeAgent effectively automates customer interactions with fast resolutions and personalized service experiences without compromising safety and security.

## Testing a probabilistic solution

Unlike most software development where the goal is to create a deterministic system with the same, expected outcome every single time, generative AI generates new information and content based on patterns, often producing varied responses that can be difficult to predict. **This variability makes testing an AI agent exceptionally difficult, but it is a crucial component in delivering an enterprise-grade solution.** You'll need a robust testing process that combines multiple techniques, including simulation-based testing from production data.

GenerativeAgent comes with easy-to-use testing and simulation tools to help you ensure that it performs as you expect, even at scale. These tools include mock APIs, mock data, and simulation testing interfaces.

## Keeping pace with AI innovation

As quickly as the technology is evolving, keeping up with every model improvement and new technique for applying AI in the real world is a full-time project. If you're planning to build your own solution, that burden will fall on your technical teams.

ASAPP is an AI company with dozens of leading experts in AI, data science, and solution development. Our commitment to AI research is evident in the 60+ patents we've been granted so far, with more pending. We've published dozens of research papers on AI, and we are continuing to push the boundaries of AI solution development. That means GenerativeAgent is backed by a team that is totally focused on AI innovation.

# Important considerations for value realization

In the end, the question of whether to buy an AI agent solution or build your own isn't really about technical complexity. It's about which option delivers the best value for your business. Some of the factors that drive the relative value of each option aren't necessarily obvious. In addition to the standard cost/benefit analysis, you'll want to consider the following factors.

## The cost of missed opportunities

Development of a robust AI solution is costly. You'll need infrastructure and expertise to support the project. Given the expertise required, you might need to hire additional resources. But even if you already have the personnel you need, there are opportunity costs to consider. Every hour spent developing a custom AI agent is an hour not spent on other programs and initiatives.

## Delayed deployment = delayed value

Building a safe, reliable AI agent takes time. Successfully deploying it into production the first time is also a lengthy and cumbersome process. The longer it takes to get an AI agent live in your customer service operation, the longer it will take to realize the value it can offer. In addition to the costs you would rack up for development, you'll need to consider the cost of delayed deployment with an in-house solution vs. an enterprise-ready product that can be deployed now. As fast as the technology landscape is changing, lost time can be especially costly if your competitors are already deploying and realizing value with an AI agent.

## Long-term development and maintenance burdens

The technical burden doesn't end when you deploy your AI agent. Maintenance requires work, too. Generative AI models require continuous monitoring, training, and refinement. In addition, every change in your products, processes, or policies could require an update to your AI agent or to the systems and information it depends on. Troubleshooting and fixing problems that arise can be time-consuming and costly. And as AI models continue to evolve, you will eventually need to replace the models you originally used. That all adds up to an ongoing burden for development and testing.

## The cost of falling short of best-in-class performance

Some enterprises have the technical resources to build a solid AI agent solution. But solid is not best-in-class, and there's a cost to falling short. Let's say you identify a use case for your AI agent with a volume of 20,000 interactions per month. Your AI agent successfully resolves 70% of those interactions. That keeps 6,000 interactions out of your human agents' queues. Considered on its own, that's a good result, and you might be happy with it.

But if GenerativeAgent could successfully resolve 85% of those interactions, that's another 3,000 interactions it could have kept out of your agents' queues. When you consider the cost of an interaction with a human agent, the difference matters. And when you consider expanding the use of your AI agent to more use cases, the difference between a solid solution and a best-in-class solution gets much costlier.

# So what's the answer – build or buy?

The right path for your business depends on a lot of factors. In the end, it boils down to these key questions:

- Do you have the internal expertise, infrastructure, and resources to build a safe, reliable AI agent?
- Are you prepared for both the short-term and long-term investment in development and maintenance that building will require?
- Will a DIY approach deliver the value you need on a timetable that works for your business?

If you can answer a resounding yes to all three questions, then building is at least a viable option. And if not, we'd love the chance to show you how GenerativeAgent can fit into your customer service operation and deliver on your long-term business goals.

# Get in touch with us

☎ +1 (646) 386-8639

✉ hello@asapp.com

🌐 www.ASAPP.com

### About ASAPP

ASAPP is an artificial intelligence solution provider committed to solving the toughest problems in customer service. Because we automate what was previously impossible to automate, our AI-native® solutions deliver more than efficiency gains. They redefine the role of AI in the contact center and lay the groundwork for businesses to reimagine their customer experience delivery for the age of AI. Leading enterprises rely on ASAPP's generative and agentic AI solutions to dramatically expand contact center capacity and transform their contact centers from cost centers into value drivers. To learn more about ASAPP, visit www.ASAPP.com.