AVAYA

THE ACES GUIDE TO GENERATIVE AI IN THE CONTACT CENTER

CONVERSATIONAL AI

AVAYA
Customer Experience
Services

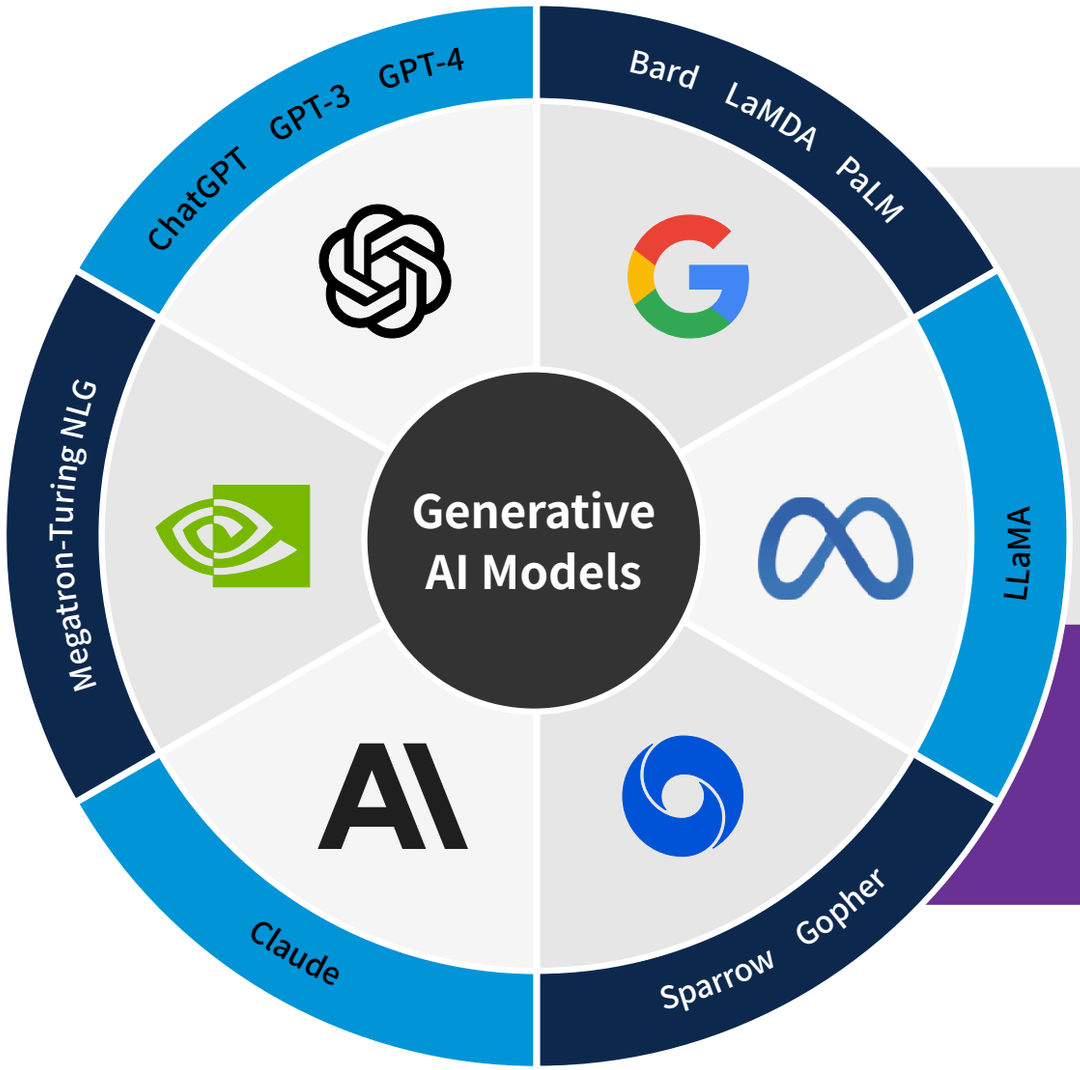
COGNIGY

TABLE OF CONTENTS

An Executive Summary of Generative AI

Generative Artificial Intelligence, which we'll refer to as Generative AI, is a computer program that can analyze and create new multimedia content, including text, images, video, audio, code, and more. It is not designed or ready for standalone use but is a powerful addition to an existing customer service solution.

Generative AI has two core capabilities: discrimination and generation. Discrimination refers to its ability to understand things like intent, recognize and extract entities from content, and analyze sentiment. Generation refers to ability to create content such as an image based on a description, translate into other languages, rephrase text, or summarize text.



Today, Generative AI can be used to:

Increase the speed and efficiency of customer and employee service.

Provide unparalleled natural and contextual experiences within your service processes.

Generative AI delivers innovative, technology-enabled experiences but does not replace service experiences with technology.

High Level Benefits in Customer Service

The two major areas where Generative AI benefits customer service are first speed and efficiency, and second personalization. Let's dig into each.

Speed & Efficiency

For Customers

- Empathetic, contextual responses in self-service
- Customers feel heard and understood
- More confident, accurate agent support
- More flexible interactions vs. rigid flows
- Customer input is better understood during self-service interactions

For Agents

- Contextual, Natural Suggestions during live interactions
- Lower error rates
- Automatic summarization during handovers and creating tickets
- Increased containment rate
- Faster bot-building and time to value
- Rapid testing with AI-simulated conversations

Killer Combo: Conversational AI & Generative AI

Generative AI alone is neither built nor suitable for standalone use in a service context for a variety of reasons. Yet, its groundbreaking ability to generate natural, contextual, and personalized responses cannot be ignored. Since we've already defined it, let's look at few dealbreakers for standalone use in customer service.

- Responses are inconsistent and unpredictable
- Cannot integrate with any 3rd party systems (e.g. CRM, CCaaS, etc.)
- Trained only on a static data set
- Output is not transparent and auditable
- Not designed for your specific use case
- Has no front-end for customers

	Conversational AI	Generative AI
Strengths	• Use case specific • Stays on task • System integrations • Channel integrations	• Very flexible • On the fly • Low effort • Human-like
Weaknesses	• Feels rigid • Pre-defined responses • Higher effort (manual) • Limited human-like responses	• Generic, not use-case specific • No system integrations • No channel integrations • Can go "off track" and hallucinate

Conversational AI in customer service is designed to model service processes and orchestrate actions by utilizing natural language processing technologies. It creates a conversational interface with integrated backend systems so customers can more efficiently and effectively self-serve. That means being able to speak to software, instead of using your mouse, keyboard and monitor.. It streamlines workflows and automates routine processes end to end resulting in faster and more efficient customer support and reduced burden on agents while enhancing the overall customer experience.

Yet, Conversational AI can feel rigid and impersonal at times as conversation designers have to model all possible conversation flows and scenarios. This is a result of it being use-case specific and built to stay on task.

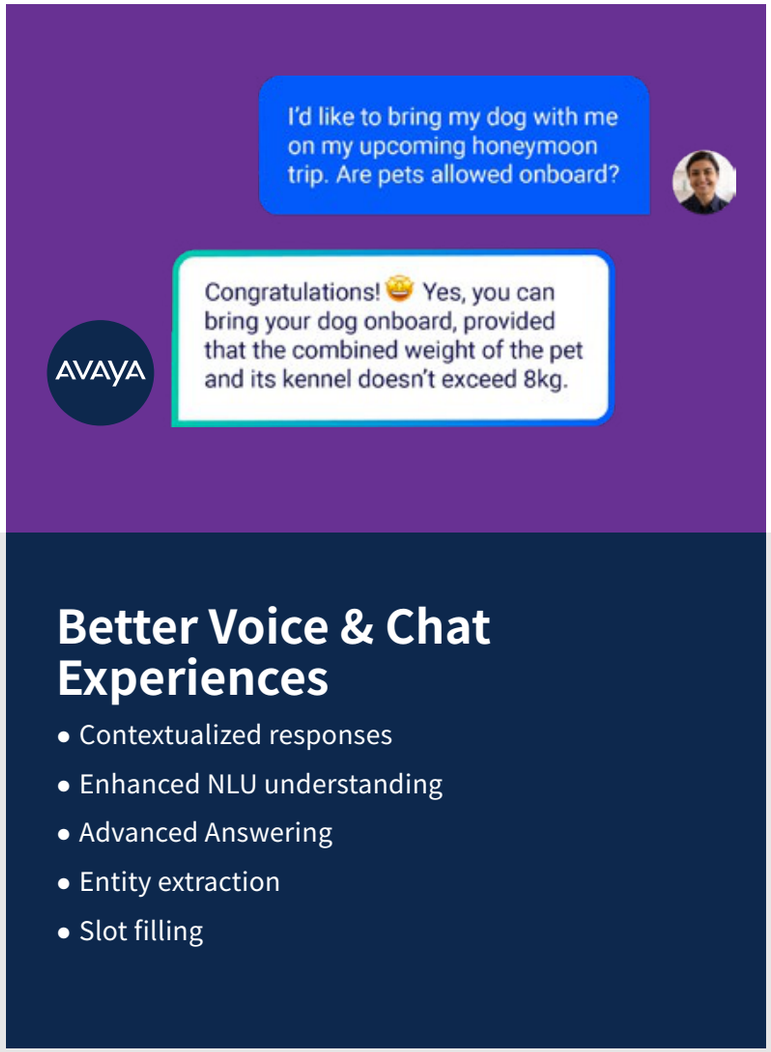
Luckily, the strengths and weaknesses of each solution perfectly complement each other.

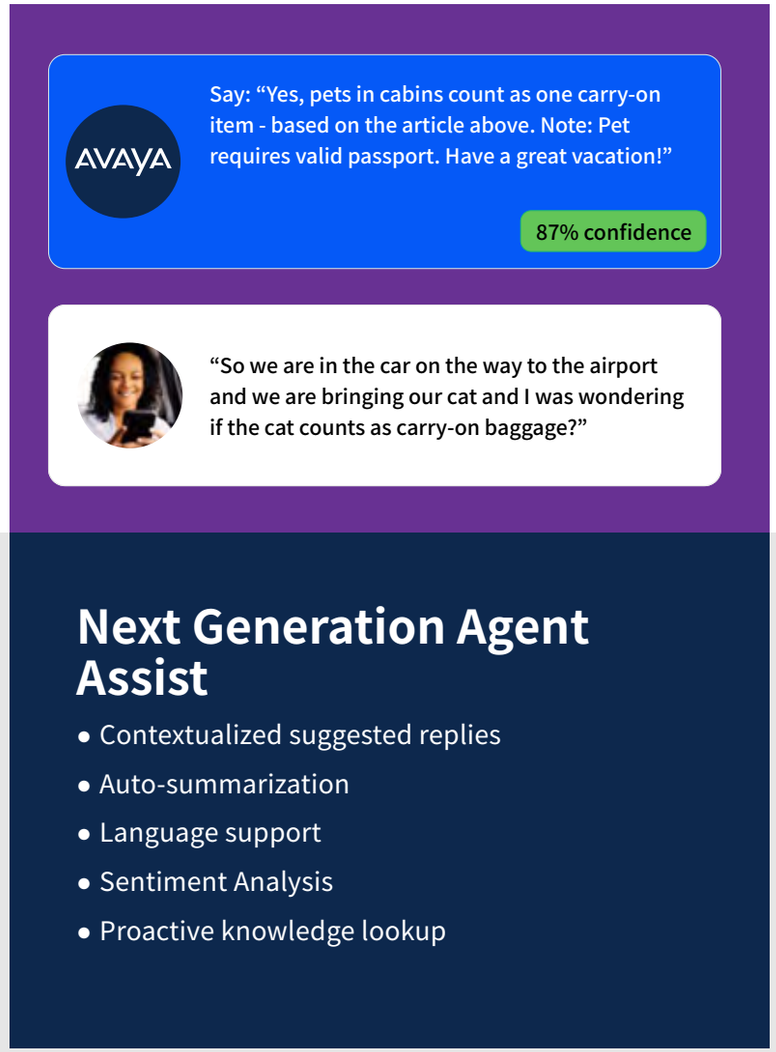
So, how can they be used together in a contact center or customer experience initiative?

Generative AI cannot reach into your Salesforce CRM to retrieve customer data, combine that with customer intent and reservation number to access your reservation system to make flight changes all while conversing via WhatsApp with a customer.

Conversational AI is designed to do just that and, when augmented with Generative AI, can provide customer and context specific personalization while staying on task and providing a fully automated self-service experience that feels natural and human-like.

Business Ready Use Cases





Better Voice & Chat Experiences

- Contextualized responses
- Enhanced NLU understanding
- Advanced Answering
- Entity extraction
- Slot filling

Next Generation Agent Assist

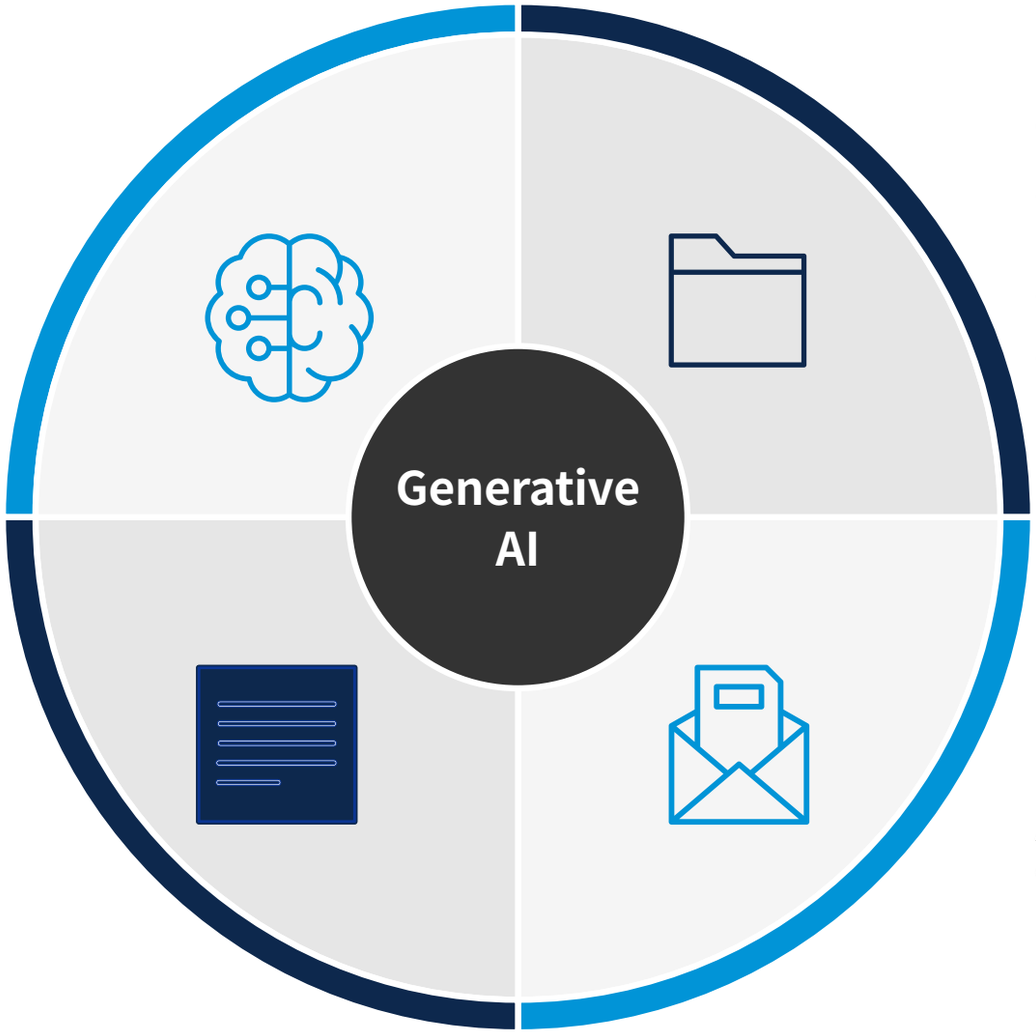
- Contextualized suggested replies
- Auto-summarization
- Language support
- Sentiment Analysis
- Proactive knowledge lookup

Potential Future Use Cases

The potential use cases for Generative AI are limited only by imagination at this point, and while there are plenty of questionable suggestions given the current lack of use in products and the early phase of the technology, it is easy to find several concrete ones in the customer service field we'll see within the next five years.

Agent Training

Generative AI bots simulating customers.



Case & Ticket Handling

- Automatically create case from self-service interaction
- Summarize and complete case data
- Case & Ticket routing

Conversational AI

Generate knowledge & training articles based on past cases, customer conversations, forums and more.

Email Drafting

Auto-generate email drafts to customers.

Business Impact by Role

We are in the initial stages of Generative AI technology, and it will clearly have a major impact on a wide range of industries, related technologies, and individual roles. Therefore, let's stick to changes based on capabilities that are currently production-ready in contact centers and customer service.

Frontline Agent

The integration of Generative AI in contact centers will transform the role and responsibilities of front-line contact center agents, initially via its use inside agent assistance technology. In the agent assist workspace

(whether browser tab, widget or other), Generative AI can be fed relevant customer information from your CRM including previous interactions, current customer subscriptions, purchases and the like. This enables it to produce individually relevant and contextual suggested responses and next best actions for each specific customer.

Yet, as self-service containment rate improves, agents will be freed up to focus on more complex, high-value interactions that require empathy and problem-solving skills. This shift in responsibilities is likely to lead to changes in KPIs, with less emphasis on metrics like AHT and more focus on CSAT, FCR and interaction quality.

The employee experience may also improve as agents are empowered to tackle challenging issues and provide personalized support, which can lead to greater job satisfaction and a sense of accomplishment. Consequently, this may contribute to reducing employee attrition, as agents find their roles more engaging and rewarding in an AI-enhanced contact center environment.

Process, Handling and Wrap Up Automation

Generative AI can support agents via improved automation by:

- Using Speech-to-Text to understand customer speech, identify intents, search for knowledge base articles in real-time and rephrase and summarize them relevant to the context
- Drafting templates and suggested replies for chat, voice, and email responses
- Performing real-time sentiment analysis during voice and text interactions
- Summarizing conversation history within self-service for handover to an agent
- Summarizing live interactions and preparing the ticket or case during wrap-up (e.g. in Salesforce)

Augment and Maximize Limited Foreign Language Skills

In regions with multiple languages, you often have agents who can speak a second or third language but not necessarily at the level required for customer service. Generative AI can help by suggesting natural language in Spanish or French for example, to fill the gaps in an agent's ability while still enabling them to handle that interaction. This means not having to hire additional staff or offshore operations with the associated risks in quality and consistency. Of course, this comes in addition to Conversational AI like Cognigy's ability to do real-time bi-directional translation which is particularly useful in text-based channels and in multilingual regions like Europe.

Head of Conversational AI / Digital Assistants

For the Head of Conversational AI, the changes will be less impactful compared to others. Already tasked with managing AI implementation and balancing automation with human intervention, they will continue to focus on designing engaging conversation flows, establishing relevant KPIs, and managing costs.

They will continue to oversee the development, deployment, and continuous improvement of AI-powered customer service systems, while ensuring performance and alignment with organizational CX objectives and strategy. However, the role will require a deeper understanding of the rapidly changing Generative AI and Large Language Model landscape. This includes the technical differences, strengths and weaknesses of each model, costs and best practices in production. Finally, the benefits, costs and effort required for custom trained models and their value for the company based on size, contact volume, automation potential and so forth will also land with the Head of CAI.

Key job role changes include:

- Overseeing the development, deployment, and continuous improvement of AI-powered customer service systems
- Developing and implementing best practices for when and how to use Generative AI
- Ensuring conversation designs are natural, engaging, and tailored to customer needs
- Striking the right balance between automation and human intervention in customer interactions
- Establishing and monitoring KPIs for AI systems in addition to traditional human-centered ones
- Managing costs associated with AI system development, maintenance and API usages (e.g. to OpenAI) while maximizing ROI
- Staying up to date with advancements in Generative AI and LLM technology and its potential in contact center operations

Food for Thought:

While LLMs can deliver amazing human-like and contextual responses, they do incur additional costs per conversation. Is the additional cost of the LLM worthwhile for say every package tracking conversation? Probably not.

Contact Center Manager

The integration of Generative AI and Large Language Models in contact centers will reshape the future roles and responsibilities of contact center managers. As advanced AI automates and streamlines multiple areas of customer service operations, Contact Center Managers will need to adapt to this changing landscape.

Managing AI & People

Managers will need to shift from traditional workforce management tasks, such as scheduling and monitoring agent performance, to overseeing the performance and effectiveness of AI agents and their interactions. Additionally, they will be responsible for ensuring that human agents and AI systems collaborate effectively, harnessing the strengths of both to deliver exceptional customer experiences. While contact center managers won't need to be deeply technical, they'll still need to develop a fundamental understanding of AI technology similar to their people skills.

Labor and Software Costs

Cost management will also evolve for Contact Center Managers, as the implementation of Generative AI and Large Language Models can lead to both to significant cost savings but also a shift in costs from labor to software.

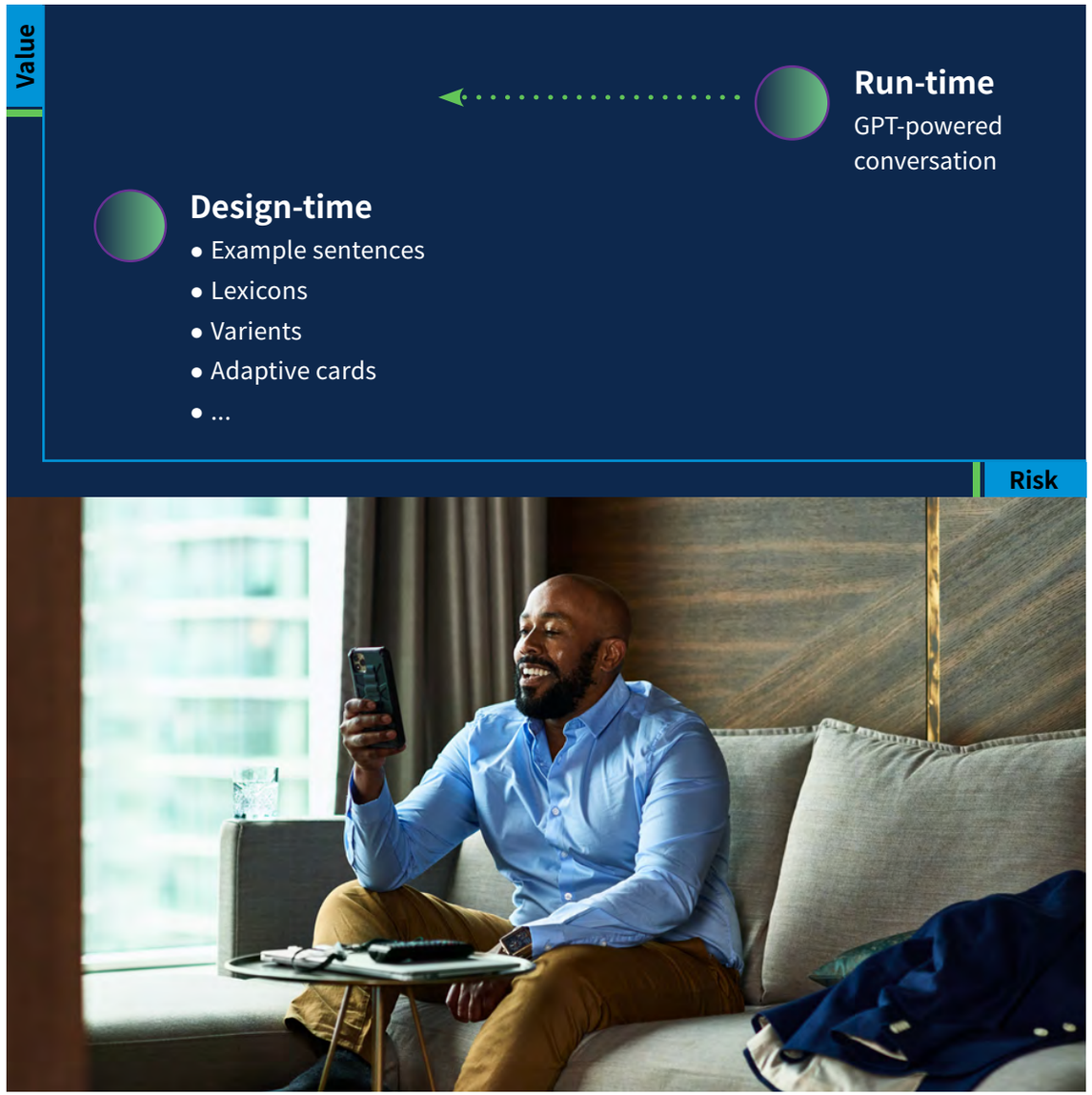
As AI-driven systems handle routine tasks and queries, the need for large teams of human agents may decrease, potentially reducing labor costs. However, the initial investment in AI technology, ongoing maintenance, and the need for skilled staff to manage and train the AI systems will become essential factors in cost management. These costs must be balanced with the expected benefits, including improved efficiency, customer satisfaction, and potential revenue growth.

From Human Focused to AI KPIs

AI's increasing share of the contact center will also impact the technology stack and KPIs for contact centers. Managers will need to familiarize themselves with the new tools and platforms used to develop, deploy, and monitor AI-driven customer service solutions. They will also need to reassess traditional KPIs, as some may become less relevant or even obsolete. New KPIs may arise, focusing on the performance of AI systems, such as NLU accuracy, response time, and customer satisfaction in AI-driven interactions. Contact Center Managers will need to establish these new KPIs in conjunction with management and the overall **Customer Access and CX strategies** of the organization.

Challenges, Risks and Questions

With new technology that lacks years of experience in the field, comes many questions, challenges, and potential risks. Weighing added value versus risk is key for any new solution and Generative AI is no different. Let's address the most common ones in the context of customer service.



Prompt Injection Attacks

An increasing number of jailbreaking methods or prompt injection attacks have been discovered and shared widely in the news using OpenAI or Bing AI. One particularly amusing example comes from a recruitment startup who released a Twitter bot. It automatically responded to any mentions of "remote work" using GPT-3 so a clever user decided to trick it as you can see in the screenshot. Clearly, a company's worst nightmare would be someone tricking their customer service bot into giving incorrect answers or making controversial statements.

The reason the example and similar methods work is because the input is being passed directly to the Large Language Model with no filter, prompt template or software layer between the public and the model. Thus, anything is possible.

Luckily, when using Generative AI in combination with Cognigy, this becomes nearly impossible. Conversational AI serves as the user-facing software and will initially receive and analyze the user input.

So, if someone tried a similar trick, it would not be matched against an intent (unless you created an intent to identify tricksters) and the bot would simply say it was unable to understand and ask again. CAI provides a protective layer that only passes information to the LLM when and how you define, including the exact instructions given to the LLM.

Negligible risk in customer service when using CAI augmented by an LLM. When allowing users to interact directly with an LLM, which is strongly recommended against, then the risk is high.

Hallucinations

An increasing number of jailbreaking methods or prompt injection attacks have been discovered Hallucination is the technical term (believe it or not), for when Large Language Models confidently give completely incorrect responses, or as one person joked truth versus made-up truth. Here, it's important to remember that LLMs are actually prediction models designed to generate human-like and realistic text. They are not designed to generate or recite facts, but create language that's natural by predicting the most likely word or phrase that comes next.

The LLM does not understand English or any language in the way humans think, it's looking at mathematical patterns, how words and phrases are related to each other and the frequency with which they occur together in different contexts. Thus, as far as the LLM goes, it's successfully completed its task as instructed.

Hallucinations can occur when the user is interacting directly with an LLM, with no filter, or instructions in the middle to create the necessary guardrails.

Negligible risk in customer service when using CAI augmented by an LLM. When allowing users to interact directly with an LLM (which is not recommended in customer service) the risk is high.

Discrimination, Hate Speech and Social Stereotypes

One of the most shared forms of dangers in the news and social media is that of a large language model generating speech that is discriminatory, hateful, or promotes stereotypes. Again, it's critical to understand that an LLM is trained on massive volumes of text from all over the internet. Since the internet is filled with a wide range of speech, an LLM whose training data includes such content can potentially regurgitate it. Again, this is not to say the LLM believes, promotes, or even understands what it is generating; it's simply working on a prediction model of language.

Every major tech company has experienced this, from Meta to Microsoft. Even Microsoft's recent integration of OpenAI into Bing search quickly led to issues of the bot going off script.

Negligible risk in customer service when using CAI augmented by an LLM. When allowing users to interact directly with an LLM, which is strongly recommended against, then the risk is high.

Increased Cost

As discussed earlier in terms of how Conversation Designers' role will change, indiscriminate use of LLMs present a risk of reducing or eliminating the ROI of using them. Here are the costs of two different OpenAI models at the time of writing:

Model	Cost per 1,000 tokens (~750 words)
Gpt-3.5-turbo	$0.002
GPT-4 with 8K context	$0.03

Now 3 cents per 750 words may not sound like much and depending on your situation and usage frequency it may not be. However, if you're handling millions of conversations a year, it's crucial to look at your analytics and get an idea of the average conversation length to estimate the additional cost. Consider categorizing conversation by type (e.g. password reset, refund, item return, etc.) and then look at the metrics for each category. With that you can begin to estimate costs and begin to consider in which types of scenarios, the additional customer value is worth the cost.

LLM Prompts

Final Comments on Risks

It should be clear by now that there's a single common theme among the risks (except cost):

direct unfiltered interaction with an LLM. When using Cognigy.AI, there multiple ways to use LLM technology. Only one node enables direct user to LLM interaction: The "Complete Text node." Thus, it should be used with great caution and is not recommended for direct customer interactions at the moment.

Near Future: Looking over the Generative AI Horizon

As the saying goes, making predictions is hard, especially about the future. While we've focused on the concrete contact center impact thus far, let's take a brief look just over the horizon with the most likely near-future developments and outcomes in the field of Generative AI, and how they may affect customer experience.

Custom Models (Released Sep 2023)

The cost and computing power of training custom LLM models may get to a level at which it's financially worthwhile to do. Yet, the desired result is a conversational agent that can communicate at near human level, but also has access to up-to-date and accurate information. This is in contrast to ChatGPT whose data ends around September 2021 and is known to hallucinate. That being the goal, custom models may be unnecessary.

Instead, we'll see a combination of NLU, vector search and LLMs which will achieve the same result while being just as fast and actually cost-effective. Vector search enables natural language search based on the relation between words and ideas instead of the legacy keyword model.

Simply put, your knowledge will be indexed by a vector database. Natural language understanding will detect the user's intent and search the vector database for the relevant information. The LLM will receive the relevant passages, add context and output a natural human-like response. That makes it more than just a bot, virtual agent or search. It's a real-time answer machine straight out of Sci-Fi movies.

Travel Assistant

Can I bring my emotional support elephant on the flight?

I'm sorry, but emotional support animals other than dogs are not allowed on the flight according to U.S. Department of Transportation regulations.

In this example, we took the support information of a major airline, indexed it and then connected OpenAI. We asked it "Can I bring my emotional support elephant in the cabin?" Although elephants are never discussed in the knowledge base, it understands the request, references the relevant information and answers. Compare this with an empty "0 results found" search results page.

Contextual AI

Since the public release of ChatGPT in November 2022, there has been rapid development of both OpenAI's Large Language Models as well as those of competitors. As they continue to mature and all face fierce competition, costs will be driven down as their features improve. As a result, businesses and contact centers will increasingly face rising demands and expectations from customers to incorporate Generative AI into their service while costs and accessibility are driven down.

This will lead to the accelerated democratization of truly powerful AI that will be within the reach of small and medium sized contact centers. Ultimately, this cycle of innovation and cost reduction will lead to a more efficient and customer-centric future for contact centers and businesses alike.

Custom Models

Another benefit of the fierce competitive landscape in Generative AI will be the rise of custom language models tailored to specific businesses, industries and even narrow use cases. These bespoke AI models will be trained on industry-specific and even individual company-specific data, including terminology, customer interactions, and unique business requirements, enabling them to deliver more accurate, relevant, and efficient responses to customer queries. By integrating custom language models into their contact centers, businesses will be able to provide a higher level of personalized support, ultimately enhancing customer satisfaction and loyalty.

With the ability to fine-tune their own AI models, organizations can ensure seamless alignment with their brand voice, compliance requirements, and desired customer experience. As the technology behind generative AI and large language models continues to advance, we can expect to see an increasing number of businesses harnessing the power of custom language models to drive innovation, efficiency, and success in their contact centers.

Finally, we will likely see tech giants such as Microsoft, OpenAI, Meta and Google beginning to offer pre-built and trained models in different verticals as well as LLMs as a Service.

Artificial General Intelligence (AGI)

Artificial General Intelligence (AGI) represents a significant leap forward in the development of AI, as it refers to machines that possess the ability to understand, learn, and apply knowledge across a wide range of tasks, much like humans. As Generative AI and Large Language Models continue to evolve, they may eventually contribute to the realization of AGI, transforming the way customer service and contact centers operate. This shift could lead to unparalleled efficiency, automation, and personalization in customer interactions, as AGI-powered systems would be capable of handling complex, multi-domain queries and tasks with ease.

In the context of customer service and contact centers, AGI could lead to an even more human-like understanding of customer needs and preferences. This would enable AI systems to provide highly tailored and personalized support, anticipating customer concerns, and offering proactive solutions. Moreover, the integration of AGI in contact centers could enhance collaboration between human agents and AI, empowering agents to focus on more complex and high-value or edge cases while AI systems handle routine queries and issues. This symbiotic relationship could lead to improved overall customer experience, higher agent job satisfaction, and more efficient contact center operations, ultimately driving business growth and reducing costs.

Conclusion

The marriage of Generative AI and Conversational AI means the era of human-like and truly useful virtual agents is finally here. By harnessing the power of these twin AI technologies, contact centers can achieve greater speed and efficiency, as well as deliver dynamic contextual personalization to improve customer experiences. Business-ready use cases already exist with many more on the horizon. By understanding the benefits and risks today, organizations can quickly and responsibly adopt Generative AI solutions driving contact center success and fostering exceptional customer experiences that were impossible less than a year ago.

Stay Connected:

About Avaya

Businesses are built by the experiences they provide, and every day, millions of those experiences are delivered by Avaya. Organizations trust Avaya to provide innovative solutions for some of their most important ambitions and challenges, giving them the freedom to engage their customers and employees in ways that deliver the greatest business benefits.

Avaya contact center and communications solutions power immersive, personalized, and unforgettable customer experiences that drive business momentum. With the freedom to choose their journey, there's no limit to the experiences Avaya customers can create. Learn more at **www.avaya.com.**

12/23 • CC15941EN-03