

Creating personalized generative AI experiences



Table of contents

The business imperative for improving customer experiences and personalization	03	The win-win: Better customer experiences and lower cost	10
What is GenAI and prompt engineering?	04	How it works: Ingesting data into GenAI	11
The great rethink: Customer service and interactions	05	Building an AI solution on composable architecture	12
Delivering exceptional customer experiences	06	Moving from experimentation into production (five rules of thumb)	14
Trade-offs and risk management	07	Cognizant's partnership with Google	15
Creating the perfect message	09	Conclusion	16

The business imperative for improving customer experiences and personalization

Customer expectations are increasingly being shaped by the personalized interactions and immediate responses they can access through digital technology. With this shift, the customers' power is becoming paramount.

In this new age, the customer is no longer viewed as a part of a traditional marketing segment but rather as an individual — a “segment of one.” Businesses are aiming for a deeper dialogue with each customer, one that acknowledges their unique needs and preferences.

To meet and exceed customer expectations, companies are reimagining every aspect of the customer experience. They are delving into understanding not just the individual, but the community that surrounds them. By doing so, they aim to engage customers and even provide personal advice — like a life coach — that aligns with their needs, fostering loyalty and retention.

At the heart of this shift is technology. The marketplace is now a hub where customers and enterprises connect. As business models evolve, new revenue channels, like fee-based earnings, are emerging. Enterprises understand that they need to modernize their core systems to deliver on the front-end promises of digital, real-time engagement and increase velocity to ensure they can swiftly and effectively respond to the customers' evolving needs.

An enterprise that successfully navigates this transformation not only benefits from a more loyal customer base but also enjoys faster revenue growth, further underlining the business imperative for personalized customer experiences.



What is GenAI and prompt engineering?

Generative artificial intelligence (GenAI) represents a new frontier in AI technology, wherein systems are capable of generating original, human-like content that is contextually relevant. These AI models, trained on large language models (LLMs), can comprehend and generate text that closely mimics human conversation. The technology is expected to revolutionize the customer service delivery by offering dynamic interactive experiences tailored to individual customer needs. But getting the desired outputs from GenAI requires giving it the right inputs. That's where prompt engineering comes in.

Prompt engineering is a critical component of maximizing the functionality of GenAI models. It involves the careful crafting of prompts (or inputs) designed to guide the AI's response. The aim is to extract the highest quality output, which often requires much trial and error.

The importance of prompt engineering in the pursuit of personalized customer service cannot be overstated. The right prompt can enable a customer service representative to rapidly access relevant information and provide personalized solutions, enhancing and expediting the overall customer experience.

In an age defined by digital interactions, the "human touch" is becoming a premium service. Consumers value businesses that make the effort to cater to their unique needs. By leveraging GenAI and prompt engineering, organizations can provide such personalized interactions at scale, improving the customer journey and fostering deeper customer relationships.

The art of prompt engineering

Prompt engineering is a sophisticated, multifaceted process instrumental to creating meaningful, personalized interactions in the realm of customer service automation. It requires employing a nuanced understanding of the customer's profile and past interactions, while adhering to clearly defined mission statements and rules.

One of the challenges of prompt engineering is maintaining that delicate balance between providing concise, accurate information while ensuring the system's responses do not infringe on privacy or make unfounded assumptions. Through meticulous attention to detail and careful navigation of industry-specific contexts, Cognizant has learned how to elevate the AI conversational experience, delivering outcomes that are both satisfying to the customer and beneficial to the organization.

Crafting effective prompts

A prompt is considered effective when it successfully guides an AI model toward generating the desired output. A well-crafted prompt must include the following elements:

- It must be clear, concise and devoid of ambiguity and must communicate the intended task accurately so that the model can provide the most relevant and accurate responses.
- It must provide the necessary context in order to steer the AI's response in the desired direction.
- It must inform the model of the style and tone that should be used in the response. For an enterprise, this should align with their brand's voice or the desired customer experience.
- It should include examples with sample answers that can guide the AI toward generating responses that closely align with your expectations.
- For complex tasks, consider breaking down the task into intermediary steps to guide the AI through the necessary steps to reach the final goal.

Rising client expectations are driving end-to-end transformation¹⁻⁵

- Happy customers always buy more. There is an 82% faster growth in revenue when an industry can provide high customer satisfaction.
- Many businesses think it's time to reset and restart the way they deliver customer experience. 74% of the 50 largest global enterprises have pledged themselves to customer experience transformation.
- Customers are willing to pay a higher price if the products or services offered are sustainable. 64% of consumers said they were willing to pay more for sustainable products.
- 97% of millennials now use mobile banking, and half of them prefer to bank outside traditional hours.

The great rethink: Customer service and interactions

In the face of rapidly changing market dynamics and customer preferences, many businesses are reshaping their customer interaction models. One such transformation is the shift from a transaction-based interaction to a relationship-centric approach. This means businesses are no longer just focusing on individual sales or issues but are keen on understanding the holistic customer journey, aiming to offer a seamless and personalized experience at every touchpoint.

A key part of this journey involves leveraging technology and data strategically. Businesses are deploying AI-driven solutions to extract actionable insights from customer data, allowing them to predict customer needs and offer proactive service. The use of chatbots, virtual assistants and other AI-powered tools is also gaining momentum, enabling businesses to provide instant support and personalized engagement.

Omnichannel service, allowing customers to interact with the business on the channel of their choice, and in the language of their choice, is becoming a necessity rather than a luxury. This customer-centric transformation is not just about improving the customer experience but is also seen as a key differentiator in today's competitive business landscape. It has the potential to end traditional language barriers and open new markets.



¹ [Reimagining customer engagement for the AI bank of the future](#), McKinsey, October 2020.

² [Managing a customer-experience transformation in banking](#), McKinsey, October 2018.

³ [Financial Services Industry Overview in 2021: Trends, Statistics & Analysis](#), Business Insider, December 2020.

⁴ [Why banks have become a hotbed for tech recruitment](#), Director of Finance Online, May 2020.

⁵ [2023 Consumer Sustainability Survey: Consumers Are Still Committed to Shopping Sustainably Even as Budgets Tighten](#), BlueYonder, 2023.

Delivering exceptional customer experiences

Using generative AI solutions, companies can transform their customer experiences into something genuinely exceptional. These AI technologies can facilitate highly personalized interactions that make customers feel valued and understood. For instance, AI can analyze a customer's past behavior and preferences to tailor a unique experience that aligns with their individual needs and expectations.

Generative AI bots can assist human support agents to resolve customer issues more efficiently and accurately. By using predictive analytics, AI can anticipate potential questions the customer may have and make the relevant information available to the agent so they can address it proactively.

Reimagining the customer experience

To deliver these exceptional customer experiences, companies are reimagining every aspect of the customer experience, starting with the personas of each customer so they can provide the right degree of hyper-personalization — and not just for the individual, but for the community around the individual, so they can strive to “be a part of it” and help customers form an emotional connection to their products or services.

Beyond the customer, enterprises are also rethinking the marketplace itself ... the place where the customer connects

In other scenarios, bots can work solo to manage common customer queries, freeing up human resources to tackle more complex issues, thereby accelerating response times and improving customer satisfaction.

Part of that exceptional experience is multimodality, referring to the bot's ability to interact with customers through multiple modes or channels of communication. For example, a customer might start a conversation with the bot over text chat and then switch to a voice call. A multimodal bot would be able to manage this transition smoothly and even accommodate customers that might prefer seeing images or video instead of only text.

with the enterprise. They are investigating new ways to collaborate and customize their offerings, moving toward new revenue channels and building an ecosystem that will increase customer lifetime value.

Of course, to provide that amazing front-end digital experience, the enterprise itself has to be modernized at its core so it can deliver those real-time experiences that will earn the confidence of the customer.

Levi is a 35-year old Belgian citizen who is married with no dependents. He has a bachelor's degree and has been employed for 10 years. He has a credit score of 0.33, a monthly income of 3.000 Eur, and assets of 100.000 Eur. He has a current mortgage of 400.000 at a 1% interest rate.



Hello what's the minimum amount that I can borrow ?

The minimum amount that you can borrow is 5,000 Euros.



Can I use my current mortgage to renovate my bathroom ?

Yes, you can use your current mortgage to renovate your bathroom. However, you will need to provide a copy of the invoice for the renovation work and proof that the work has been completed. You will also need to pay an interest rate on the amount of the loan that you draw down.



Can I use 10k ?

Levi, you can use 10k to renovate your bathroom. However, you will need to provide a copy of the invoice for the renovation work and proof that the work has been completed. You will also need to pay an interest rate on the amount of the loan that you draw down.



What about 5k ?

Levi, you can use 5k to renovate your bathroom. However, you will need to provide a copy of the invoice for the renovation work and proof that the work has been completed. You will also need to pay an interest rate on the amount of the loan that you draw down.



Please calculate my current debt ratio

Levi, thank you for your question, I'm happy to help you with that.

Your current debt ratio is 0.33. This means that you have 33% of your income going towards debt payments. This is below the recommended debt ratio of 36%, so you are in a good financial position.

However, I would like to remind you that your debt ration may change if you take on new debt or your income changes. It is important to keep an eye on your debt ration and make sure that it does not exceed the recommended level.

Ask me anything...

Enter

Trade-offs and risk management

When deploying GenAI-powered chatbots, handling and managing risk is a fundamental step. Primary among these is ensuring the bots' compliance with various applicable legal and regulatory frameworks.



Privacy

Customer consent, data protection and privacy laws bear significant implications. Therefore, it is paramount to design chatbots in such a way that they observe all these laws and ensure customers' data is managed responsibly and securely.



Information leakage

GenAI-powered chatbots must be rigorously tested to ensure they understand and respond accurately to a broad spectrum of human interactions. This requires a comprehensive training data set that encompasses a wide variety of scenarios and languages. The chatbot's ability to manage complex queries and provide accurate responses is crucial to avoid the risk of miscommunications that could lead to subpar customer experiences. For example, the chatbot may have negative data about the customer, such as a history of late payments, but that is not information that should be used in the formulation of a response.





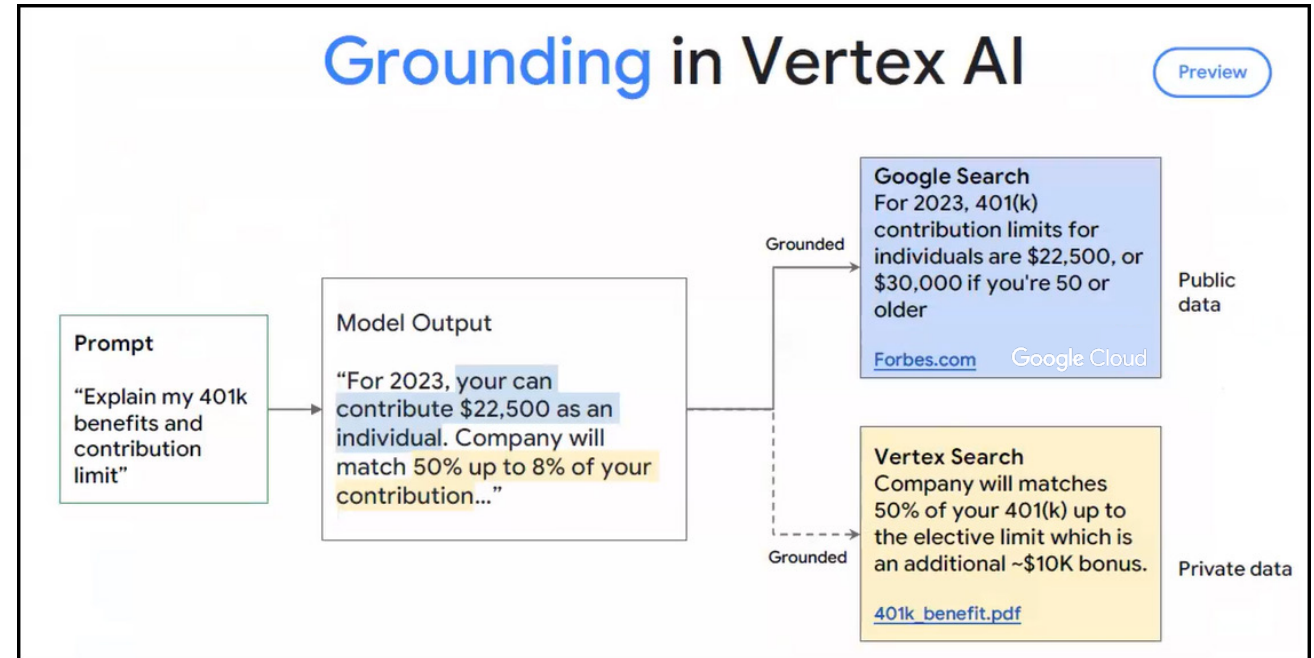
Hallucinations and grounding

Chatbots need to know when they've reached the limit of their capacity to answer a question and then seamlessly transition the interaction to a human agent. This minimizes the chance of "AI hallucinations" spreading misinformation and also ensures that more complex issues are always handled professionally.

In Google Cloud's Vertex AI machine learning (ML) platform, the term "grounding" is used in reference to a feature that helps connect and align the predictions made by LLMs with real-world concepts. It provides insights into how LLMs make their decisions.

By grounding the model's predictions in customer and/or external data, customers can better understand the reasoning behind the model's decisions, improving transparency and trust in the model's outputs. This image shows how the platform can use both public and private data to form a more complete response.

By implementing approaches and features like these, companies can confidently harness the benefits of GenAI-powered chatbots while effectively managing the potential risks.



Creating the perfect message

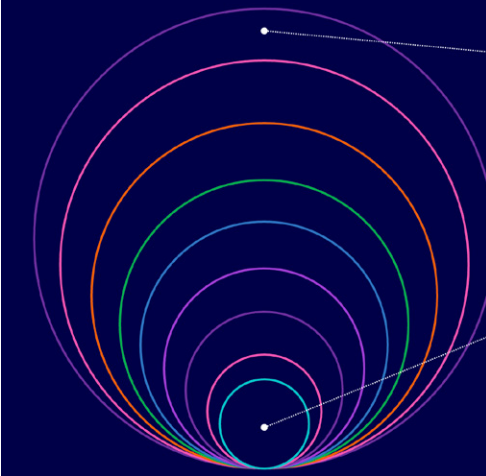
Prompt engineering in the context of generative AI is something all providers will need to master. Putting the right research questions in place to get to the core of the knowledge is a process we call “peeling the onion,” because you have to start off in a fairly general place and work your way through these distinct layers to get to a high-quality answer. And that’s really the art behind prompt engineering.

This is an ever-evolving space with new research and best practices coming out on an almost daily basis. Cognizant and Google collaborate closely together to ensure that we use the latest methodologies and share best practices to speed the path to the right customer outcomes. Critical to an effective prompt is getting the correct answer for the user only when that knowledge is grounded in the correct data. If not, it’s essential that the inquiry gets transferred to a human agent. Like our most effective colleagues, we want the models to say “I don’t know” rather than bluff.

There are many variants of prompt anatomy depending on the goal. Here is a common list that’s useful in many cases:

- The industry vertical we are acting in (finance, retail, life sciences, etc.) and area of expertise of the bot
- The job role that the bot will be playing and name for a more human feel
- The bot’s mission so it knows when to provide a response versus when to collect more information to fine-tune the response

- Using knowledge on the customer’s profile to create the best answer while, at the same time, constraining responses on a person’s sensitive information
- Keeping the chat history to increase the context
- Constraint: Basing the reply to the question on verified and grounded information to avoid hallucinations



The diagram illustrates the 'Rainbow Approach' to Generative AI Agents. It features a series of concentric circles in various colors (purple, blue, green, yellow, orange, red) on a dark blue background. A dotted line with a dot at the center represents the core of the approach. To the right of the circles, a list of prompt components is shown, each enclosed in a colored box that matches one of the circles. The components are: 'We are {WE_ARE}.' (purple), 'You are a {CHARACTER} called {NAME}.' (red), 'As a chat bot, your mission is: {MISSION}.' (orange), 'The customer profile is the following: {PROFILE}.' (yellow), 'Use this profile information to answer the questions without sharing details.' (green), 'The conversation you have already with the customer is the following: {HISTORY}.' (blue), 'Based only on the following information: {CONTEXT}.' (light blue), 'Don't assume anything!' (cyan), 'Answer the question below: {QUESTION}.' (teal), 'Answer what's important. Be concise. {PERSONALIZATION}.' (dark teal), and 'If you are not sure then answer that you don't have enough information and hand over to a human agent.' (dark purple). Below these components, a 'Key Points' section lists three guidelines: 1. Stay Grounded, 2. Don't breach policy, and 3. Don't overplay the use of private data in the prompt. The Cognizant logo is in the bottom right corner.

Rainbow Approach to for Generative AI Agents

We have experienced significant improvements in quality of dialogs by using this approach

We are {WE_ARE}.

You are a {CHARACTER} called {NAME}.

As a chat bot, your mission is: {MISSION}.

The customer profile is the following: {PROFILE}.

Use this profile information to answer the questions without sharing details.

The conversation you have already with the customer is the following: {HISTORY}.

Based only on the following information: {CONTEXT}.

Don't assume anything!

Answer the question below: {QUESTION}.

Answer what's important. Be concise. {PERSONALIZATION}.

If you are not sure then answer that you don't have enough information and hand over to a human agent.

Key Points:

1. Stay Grounded
2. Don't breach policy
3. Don't overplay the use of private data in the prompt

cognizant

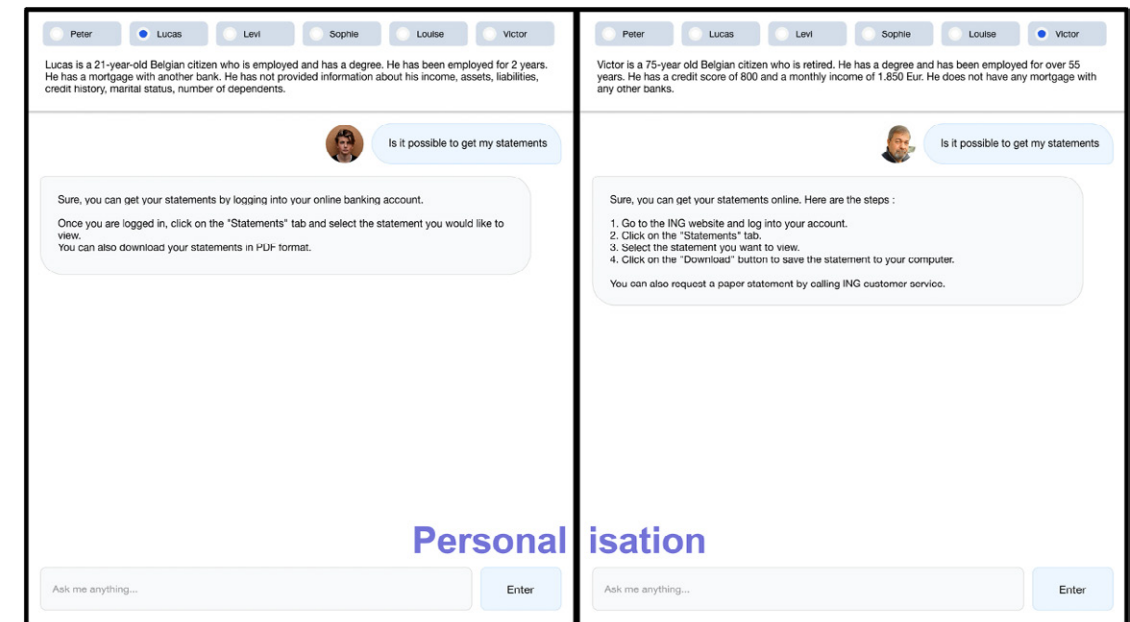
The win-win: Better customer experiences and lower cost

In the earlier days of AI, dialogue was quite intent driven, meaning that the effort to get into a digital dialogue was quite substantial. A lot of thinking and work was required to make sure that if the customer asked a question, the correct answer was generated. With GenAI, there is a universe of knowledge available to us, and with proper prompt engineering, we can now build high-quality responses on the spot to provide those correct answers.

Together, Cognizant and Google have revolutionized the way customer experiences are curated and managed by GenAI bots. These bots can respond to customer queries in real time while also adapting to their language, preferences and context, to enable a truly personalized experience. The AI system even remembers past interactions and learns from them to improve future conversations. This capability for continuous learning and evolving provides a more human-like interaction, enhancing the customer satisfaction level significantly.

Beyond better customer experiences, bots are also a smart investment from a cost perspective. They can operate around the clock, offering unrivaled scalability to support businesses during peak demand periods without additional staffing costs. By automating repetitive and mundane tasks, they allow human agents to focus more on complex and high-value tasks, improving operational efficiency while providing considerable cost savings in the long term.

Furthermore, the multilingual capability of GenAI bots can help enterprises eliminate the need for multi-region customer service teams, resulting in further cost reduction. By delivering prompt customer service in multiple languages, businesses can expand their global reach and cater to a diverse customer base without significant increase in operational cost.



How it works: Ingesting data into GenAI

Ingesting data into GenAI

Data ingestion is a crucial step in the operation of GenAI, setting the foundation for the system's abilities. The mechanism of creating an effective knowledge base for a GenAI system entails the following four steps:

- 1. Uploading the documents:** This involves the addition of data sources to the knowledge base. Documents can come in various formats, most notably PDF and HTML, and serve as the foundation for generating answers and providing information to user queries.
- 2. Splitting the documents into chunks:** This involves the deconstruction of the data into smaller, more manageable pieces. This is done to identify the relevant parts of each document that can be used to answer specific questions.

- 3. Generating embeddings for each chunk:** This phase leverages GenAI algorithms to create “embeddings” for each chunk. These embeddings are numerical representations of the text that capture its meaning.
- 4. Adding embeddings to the vector store:** This process puts the numeric embedding in a specialized database (called the vector store) where it can be quickly retrieved when a user submits a query.

By incorporating these four steps into our data ingestion protocol, Cognizant and Google can create a robust and easily accessible knowledge base. A key super power is similarity search, where items with similar meaning can be retrieved, and not just exact matches. For example, “Big Apple” would return “New York.” This optimizes the retrieval of the correct information from the knowledge base and is critical for accurate and efficient responses to user queries.



Building an AI solution on composable architecture

Interaction layer

The interaction layer is the component that manages the communication between the AI system and the end users. This layer typically includes interfaces for different user interaction modes (like text, voice or graphical user interface), and it's responsible for both receiving input from users and presenting the AI's output in a user-friendly manner.

Business services level (orchestration, API gateway)

The business services layer is a critical component that helps ensure the AI system aligns with the business needs and integrates seamlessly with existing business processes and systems. This layer often includes APIs or other interfaces that allow the AI system to interact with other business systems, such as CRM systems, databases or analytics platforms. This integration allows the AI to leverage existing data and functionality to deliver more value.

Intelligence layer

The intelligence layer typically includes one or more AI models, which could be based on various types of AI technology, such as deep learning, reinforcement learning etc. It is the core component where the actual AI processing takes place. It's responsible for understanding, reasoning, learning and generating responses or outputs. Some key functions of the intelligence layer include:



Understanding:

This involves interpreting the inputs received from the interaction layer. For instance, in a chatbot application, this might involve using natural language processing (NLP) techniques to understand the user's text input.





Reasoning:

The intelligence layer uses various AI algorithms and models to reason about the input and determine the appropriate response. This could involve tasks like predicting the user's intent, making recommendations based on past behavior or generating new content.



Learning:

Generative AI systems have the ability to learn from experience and improve their performance over time. The intelligence layer often includes ML algorithms that can learn from data and adapt their behavior accordingly.



Generation:

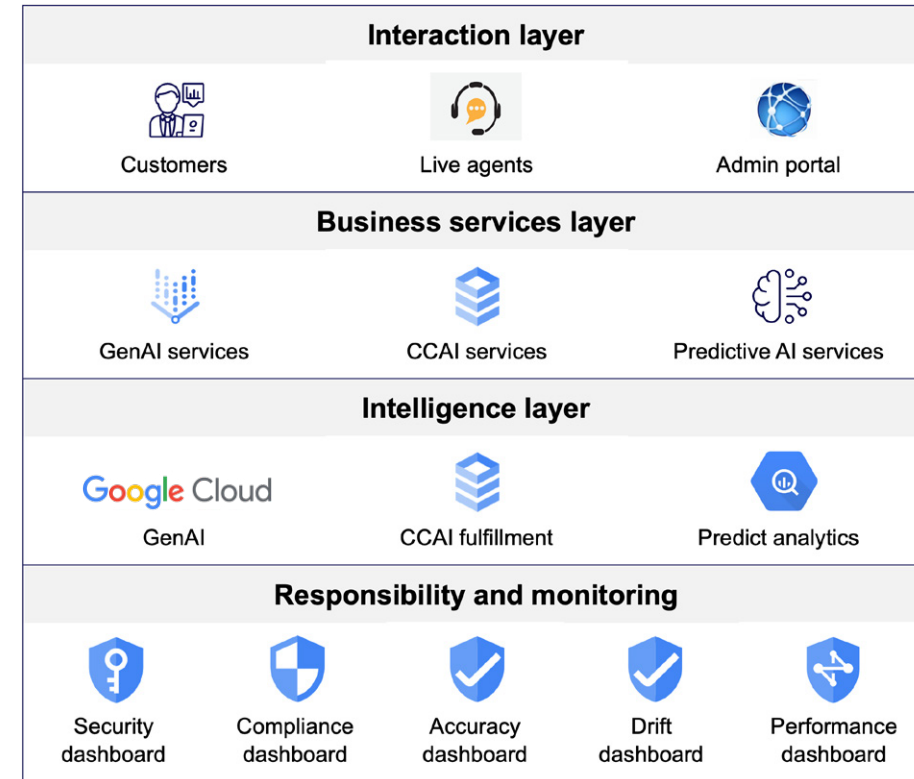
The intelligence layer is responsible for generating the actual output. This could involve tasks like creating a text response, generating an image, etc.

The specific design and implementation of the intelligence layer depend heavily on the requirements of the AI application.

Responsibility and monitoring layer

The responsibility and monitoring layer helps ensure the GenAI system operates effectively, ethically and legally and provides visibility into its operations. It's a key part of maintaining trust and accountability in AI systems.

GenAI reference architecture (hybrid model)



Moving from experimentation into production (five rules of thumb)

Moving from experimentation to production in the deployment of GenAI involves the transition from developing and testing AI models to deploying these models in real-world applications. Through experience, Cognizant and Google have found that there are five rules of thumb to consider when making that transition.

1. Document and version control the generative cognitive AI reference architecture

As AI systems are inherently complex and continuously evolving, managing different versions of AI models, configurations, training data and code is paramount. Document and version control provides a structured and systematic approach to managing change and complexity in the development and deployment of GenAI solutions.

2. Build a personalization roadmap

Personalization refers to the use of AI to tailor content, recommendations or experiences to individual users based on their behavior, preferences or other personal characteristics. A personalization roadmap outlines the steps that will be taken to implement and enhance this personalization over time. It might include initial efforts to collect and analyze user data, the development of

predictive models to anticipate user needs and the ongoing testing and refinement of these models.

3. Establish an observability evaluation framework (performance, benchmark, validation)

An observability evaluation framework provides a systematic approach to measure, monitor and validate the performance of the AI system. It includes defined metrics for performance assessment, benchmarking against predefined standards or competing systems and validation processes to ensure the system's outputs meet the expected quality and reliability. An effective observability evaluation framework promotes transparency and trust in the AI system by providing tangible evidence of its performance and efficacy.

4. Define and implement a data management roadmap

Good data management practices help ensure the reliability and accuracy of the AI system's outputs. Unstructured data, such as text, images or videos, can often hold valuable insights that fuel GenAI models. However, this data can be challenging to manage due to its lack of predefined format. A data management roadmap provides a strategic plan for how this data will be collected, stored, processed and used.

5. Define the combination between context and private information in relation to each use case

The context often determines how an AI system should respond or behave, while private information represents sensitive data that must be handled with care to ensure privacy and compliance. Balancing these elements is essential to create AI systems that are useful, ethical and lawful. Version control allows for tracking changes to the definitions, algorithms and data handling procedures, providing a clear audit trail and facilitating collaboration among team members.

With Google models and enhanced features such as [metrics-based model evaluation](#) and Cognizant's ability to implement and deploy these capabilities in customer deployments, we are minimizing the time from experimentation to production for our customers.



Cognizant's partnership with Google

At the heart of the Cognizant GenAI journey is our partnership with Google, a collaboration that has been instrumental in the development of GenAI bots. Our relationship transcends the traditional boundaries of a vendor-customer dynamic, evolving into a symbiotic relationship where mutual knowledge exchange and the pursuit of better customer experiences is the driving force.

Together, Cognizant and Google have tackled the GenAI challenges that most often plague businesses, such as limited training data, language barriers and high call volume. The GenAI bot, powered by Google Cloud Platform (GCP) services, draws on its vast array of APIs, which allows it to understand context, offer personalized responses and speak in multiple languages. This has not just boosted the quality of customer service, but also markedly reduced operational costs.

Cognizant's partnership with Google enables us to stay at the forefront of GenAI technology. We continually engage in a dialogue with Google around the art of prompt engineering, ensuring that our methodologies remain cutting-edge and responsible. In turn, Cognizant provides Google with industry-specific insights that help them fine-tune their foundational models. This collaborative process ensures that we develop GenAI solutions that align with the specific needs and ethical standards of various industries, making them more efficient and effective.



Conclusion

A lot of software vendors or technology partners, like Google, are creating out-of-the-box GenAI solutions that provide basic conversational functionality, but it's sometimes constrained in terms of features, functions, options and capabilities. What we have proven at Cognizant is that we can use the same kind of foundational models and tailor them to do much more advanced things.

This transformative technology, powered by GenAI and GCP, is reshaping the landscape of customer service, providing enterprises with an optimal balance of personalized customer experiences and cost-effectiveness. Given the ever-rising customer expectations, it's becoming an indispensable solution in today's digital age.

Cognizant, in partnership with Google, stands at the forefront of this AI revolution, using a combination of industry knowledge and innovative technology to tackle common business challenges like limited training data, language barriers and high call volume. This unique collaboration promises to continue delivering exceptional, personalized customer experiences, setting new standards for customer interaction in the digital age.

In the final analysis, our partnership with Google is about much more than just developing a bot. It's about forging the future of customer service and redefining what's possible with AI. It's our commitment to delivering an unrivaled customer experience that sets us apart and fuels our ongoing pursuit of innovation in GenAI.