The Impact of Generative AI on the Conversational AI Market

Published 15 September 2023 - ID G00797697 - 20 min read

By Analyst(s): Gabriele Rigon, Bern Elliot, Arup Roy, Justin Tung, Adrian Lee, Stephen Emmott

The impact of generative AI on the conversational AI space is disruptive. Due to the dynamically evolving nature of this space, applications and software engineering leaders should revisit or reevaluate their CAI initiatives to discern the best approach to adoption.

Overview

Impacts

- Generative AI (GenAI) is transforming traditional conversational AI (CAI) use cases, revamping Q&A functionality, augmenting transactional virtual assistants (VAs) and paving the way for enterprisewide automation and human augmentation.
- As CAI capabilities can be found in numerous types of enterprise applications, the value proposition of conversational AI platforms (CAIPs) is challenged. We expect considerable market consolidation and product repositioning.
- Classical chatbot architectures are being augmented by GenAI techniques and services such as large language models (LLMs), and new ones have emerged, which anticipates the rise of radical transformations of the underlying CAI tools and approaches overall.
- In such an unstable and rapidly evolving space, leaders find it challenging to discern the pros and cons of their choice of adoption approaches for CAI initiatives.

Recommendations

- Improve the value of the CAI initiative and minimize risks by identifying enterprise-grade use cases beyond self-service chatbots and starting with low-risk GenAI-augmented implementations.
- Compare CAI capabilities offered by different types of enterprise applications by identifying customization and scalability options and ensuring your own domain models are loosely coupled with underlying vendors' technology.
- Prioritize the robustness and long-term resilience of the CAI implementation by assessing classical versus emerging chatbot architectures and selecting vendors with a solid GenAI roadmap.
- Choose adoption approaches strategically by defining the scope of the CAI initiative and carefully estimating the effort needed to move experimental LLM-based proofs of concept (POCs) to production.

Strategic Planning Assumption(s)

- By 2025, 80% of enterprise applications, such as productivity software, instant messaging platforms and BI applications, will embed a virtual conversational assistant to fulfill content fetching and content generation tasks.
- By 2025, 70% of the current CAIP vendors will be forced to reposition their product as, for example, contact center automation or CX/EX automation platforms.
- By 2025, GenAI will be embedded in 80% of CAI offerings, up from 20% in 2023.

Introduction

Beta Research

The following research is part of a new initiative Gartner is piloting to provide updates at a greater frequency. It is a work in progress that does not represent our final position. While we continue to monitor this topic, we invite you to provide constructive feedback. All relevant updates and feedback will be incorporated into the final research, which will undergo our standard review process.

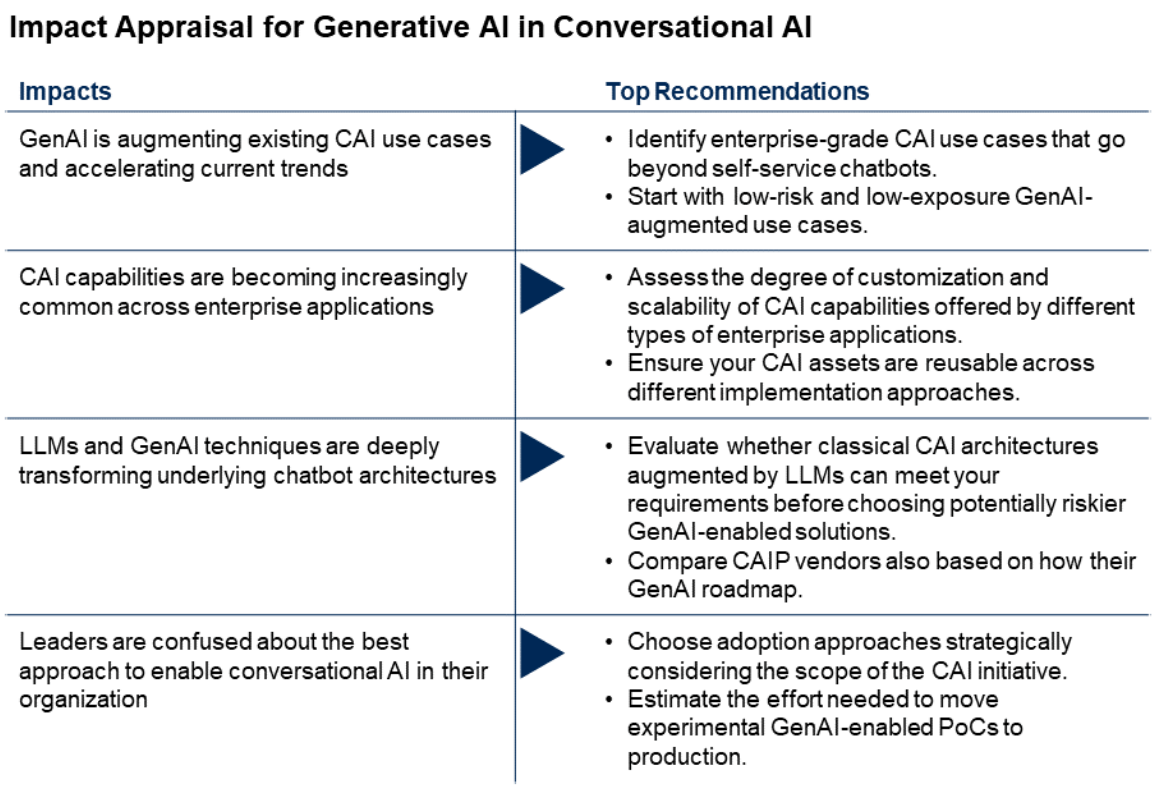
GenAI is disrupting the CAI space, while recasting the light on conversational user interfaces (CUIs) and reinvigorating the desire for human-to-machine interactions. Despite the excitement, the enterprise readiness of GenAI-driven approaches for CAI, in terms of costs and risks, has to be assessed depending on several factors, including the scope and the nature of the specific CAI use cases.

We observe three main trends (see Figure 1):

- GenAI-enabled applications are pushing the automation of conversations between users (customers/employees) and software applications. This trend is likely to increase further.
- CUIs are beginning to complement (and may replace) traditional graphical user interfaces (GUIs) in a broad spectrum of business applications, such as search and insight engines, business intelligence platforms and productivity software (document and spreadsheet applications).
- Traditional approaches to CAI are being disrupted in the following ways:
  - CAI use cases are being augmented by LLMs and GenAI-specific design patterns.
  - GenAI-enabled CUIs are competing with traditional chatbots across a broad range of enterprise applications.
  - CAI architectures and approaches used to build and deploy chatbot applications are undergoing a radical transformation and need to be revisited to account for GenAI techniques

These shifts and rapid market evolution have created confusion and uncertainty. IT and business leaders in charge of exploring the CAI space and finding the best solution to support their use cases must navigate these dynamics in choosing the right approach to adopt CAI capabilities. Use this research to make sense of the multiplicity of options available, based on their benefits and risks.

Figure 1: Impact Appraisal for Generative AI in Conversational AI



Source: Gartner
797697

Impacts and Recommendations

GenAI Is Augmenting Existing CAI Use Cases and Accelerating Current Trends

GenAI techniques have accelerated existing trends in CAI use cases by enabling a higher degree of scalability and versatility, which is expected to translate into improved business outcomes. GenAI-enabled CAI use cases will allow opportunities to deploy more complex, multifunctional applications relatively more easily, therefore speeding up enterprisewide conversational-based automation but also posing new risks and challenges.

Figure 2: The Evolution of CAI Implementations

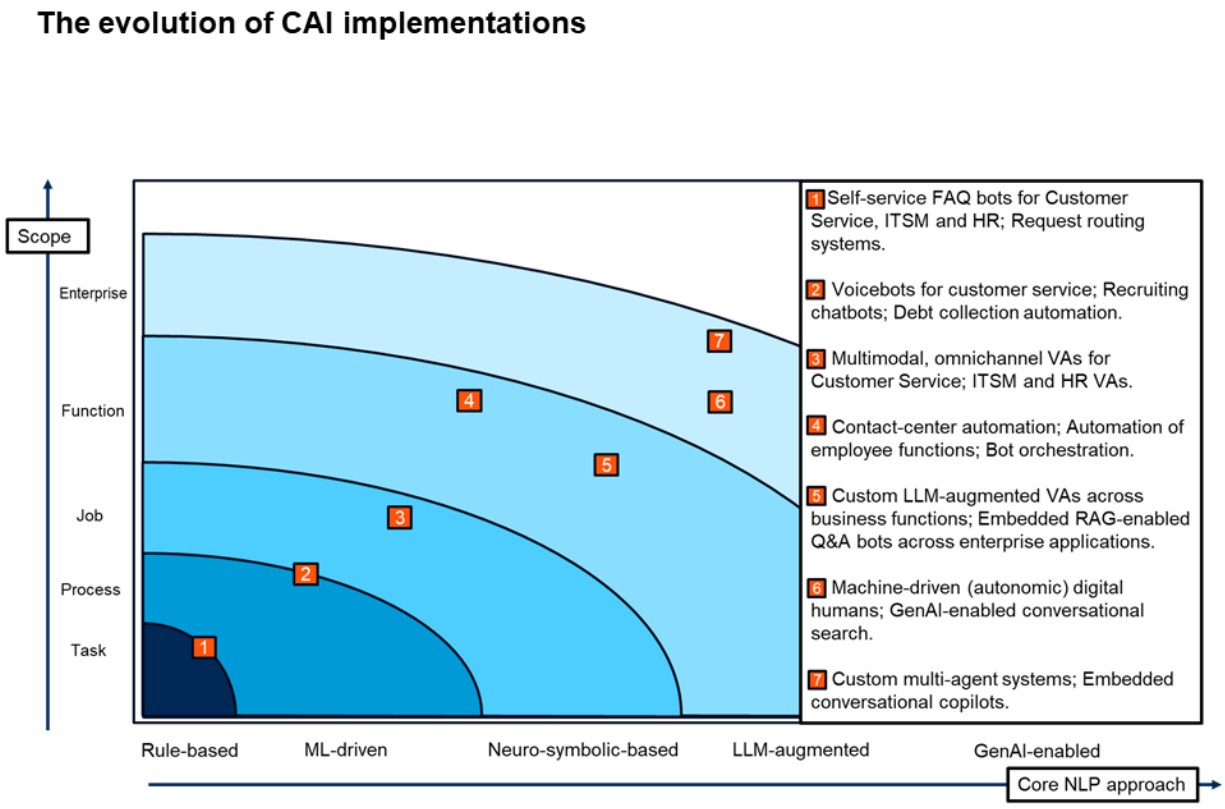


Figure 2 shows the evolution of CAI use cases as per two dimensions:

- **Core NLP approach** — This dimension is intended to offer an historical perspective of the underlying core NLP techniques. Rule-based Q&A chatbots represent the point of origin of the CAI tooling evolution. Such rudimentary implementations were gradually replaced by domain-specific VAs, whose sophistication was enabled by the introduction of ML-driven natural language processing (NLP) techniques and dedicated productized tooling, often provided in the form of CAIPs. “Neuro-symbolic” refers to composite AI approaches grounded on a hybridization of NLP techniques — specifically, rules and semantics (symbolic) combined with deep neural networks (DNNs). Such evolution is continuing with the hype around LLMs and retrieval-augmented generation (RAG) architectures, a new paradigm of approach.

- **Scope** — The "scope" of CAI initiatives spans:
  - **Tasks** — Task-specific elements within a process (e.g., finding answers to specific FAQs), selective workflows (e.g., routing conversations)
  - **Processes** — Complete workflows and end-to-end processes (e.g., appointment booking, recruitment automation, contextual assistance for online shopping, debt collection automation)
  - **Jobs** — Function-specific automation and human augmentation (e.g., customer service chatbots, recruitment chatbots)
  - **Business units** — Business-unit-wide automation and human augmentation (e.g., contact center automation, ITSM automation)
  - **The enterprise** — Cross-function automation and human augmentation (e.g., enterprisewide GenAI-enabled conversational search)

As of now, GenAI techniques are augmenting function-specific VAs and new types of LLM-driven Q&A bots across enterprise applications. We expect low-risk and low-exposure use cases (involving employee-facing VAs in HR and ITSM, or agent assist products) to be the first to be significantly augmented by the appetite for GenAI. CAI vendors are adjusting workflows and improving risk mitigation controls, such as those needed to reduce hallucinations. By 2028, GenAI tools will become critical for the enablement of enterprisewide CAI use cases such as machine-driven digital humans, conversational search, custom multiagent systems and embedded conversational copilots. Privacy and data security, but also accuracy and reliability of generated content, will be pivotal to adoption and differentiation.

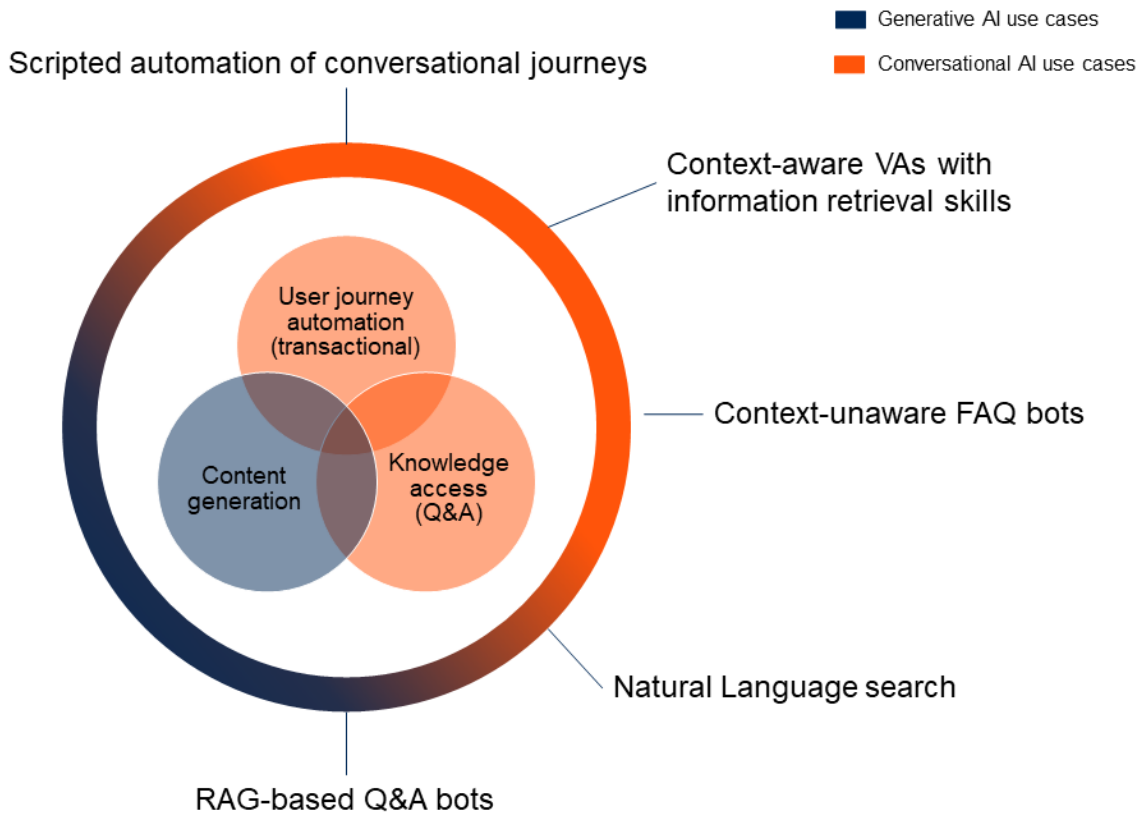
CAI Use Cases Types and Objectives

The key objectives of CAI use cases have also been impacted by GenAI. Figure 3 shows the overlap among three use case categories, which require a CUI to enable user interactions, but whose goal is very different. It also includes some representative implementation examples, from scripted automation of conversational journeys to RAG-based Q&A bots.

The core CAI use cases belong to the “user journey automation” and the “knowledge access” groups. Use cases whose primary objective is to enable content generation are not in the scope of CAI, but rather fall under the “GenAI umbrella.” However, some degree of overlap is expected because these categories converge into one another. IT and business leaders should decide where their use cases belong to in this high-level picture before navigating any relevant markets any further.

Figure 3: Implementation Examples Through the Lens for Conversational AI Use Cases

Implementation examples through the lens for Conversational AI use cases.



Source: Gartner
797697

Gartner

Recommendations

- Identify enterprise-grade use cases that go beyond self-service chatbots and depend on a broader spectrum of capabilities.
- Start with low-risk and low-exposure GenAI-augmented use cases, and prioritize human augmentation versus full automation.

CAI Capabilities Are Becoming Increasingly Common Across Enterprise Applications

CAI capabilities and CUIs are spreading across all types of enterprise applications. This is due to the availability of GenAI techniques and the expectations they are raising across users. The value proposition of CAI-specific products is then challenged on the range of use cases that shifts toward knowledge access, Q&A, contextual assistance, summarization and prompt-based content generation.

CAI-Specific Products

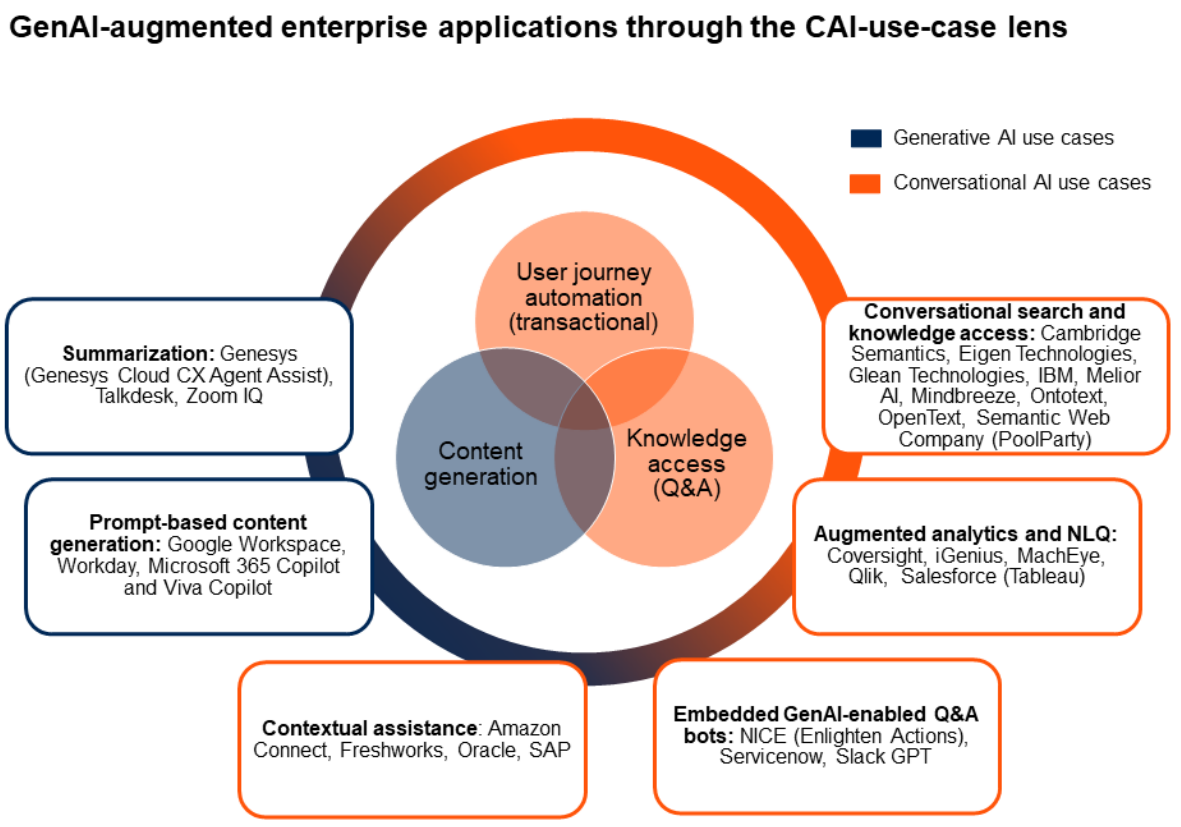
CAIPs typically meet the needs of both advanced and simpler initiatives across the whole spectrum of core CAI use cases. As per a trend we observed in the latest Magic Quadrant for Enterprise Conversational AI Platforms, several enterprise CAIP vendors are expanding their capabilities to support broader use cases beyond self-service chatbots, such as contact center automation and automation/augmentation of employee functions. With the rise of several competing implementation approaches for the enablement of CAI capabilities, we expect such trends to accelerate, as CAIP vendors seek differentiation by repositioning their offering in the broader automation space.

RAG-enabled CAI products are a new entry in this space. They are often stand-alone, low-code or no-code products offered as targeted services by non-core-AI application vendors or systems integrators. They are also offered as separate, dedicated modules in a productized fashion by established CAIP vendors. They offer varying degrees of control to mitigate GenAI risks in Q&A use cases. But from the perspective of reliability and robustness for supporting enterprise-grade user journey automation, RAG-enabled chatbots still lag behind traditional CAI architectures enabled by proven tooling offered by CAIPs.

Other Enterprise Applications

Figure 4 shows some examples of GenAI-augmented enterprise applications and how such newly enabled functionalities can be categorized as per our CAI use-case lens. The list of vendors and products included in the graphic is not exhaustive.

Figure 4: GenAI-Augmented Enterprise Applications Through the CAI-Use-Case lens



Source: Gartner
797697

Gartner

- **Conversational search and knowledge access:** GenAI-enabled CUIs will become a customer expectation across enterprise search and insight engines products to drive adoption by nontechnical users (e.g., Glean, IBM, Mindbreeze, OpenText). In intelligent document processing (IDP), functionalities have emerged that allow users to “interrogate documents” in natural language (e.g., Eigen Technologies, Melior AI). Semantic platforms, which provide ontology or knowledge-graph-based approaches for semantic modeling, are instead experimenting with composite AI approaches that leverage semantic reasoning with the GenAI capabilities of LLMs to build more explainable conversational applications for search and knowledge access (e.g., Cambridge Semantics, Ontotext, Semantic Web Company’s PoolParty).

- **Augmented analytics and NLQ:** BI and augmented BI platforms are augmenting their natural language queries (NLQ) capabilities to allow users to send requests and query BI data in natural language (e.g., ConverSight, iGenius, MachEye, Qlik, Salesforce's Tableau).
- **Embedded GenAI-enabled Q&A bots:** Such conversational modules are being announced by vendors in the area of collaboration and instant messaging applications, ITSM platforms, ERP platforms, CRM platforms, BI platforms, CCaaS platforms and productivity software offerings (e.g., NICE's Enlighten Actions, ServiceNow, Slack GPT).
- **Contextual assistance:** Real-time in-app contextual assistance functionalities are currently found across multiple BI platforms, ERP platforms, CRM platforms and productivity software offerings (e.g., Amazon Connect, Freshworks, Oracle, SAP).
- **Prompt-based content generation:** Several enterprise application vendors are leveraging LLMs for enabling prompt-based content generation. Although those are noncore CAI use cases, such functionalities typically require CUIs with some degree of conversational skills (e.g., Google Workspace, Workday, Microsoft 365 Copilot and Copilot in Microsoft Viva).
- **Summarization:** Summarization is a core GenAI use case, which we mention here for the sake of completeness, and that is being offered by several collaboration platforms, instant messaging applications, CRM platforms, productivity software offerings, BI platforms and CCaaS products (e.g., Genesys within Genesys Cloud CX Agent Assist, Talkdesk, Zoom IQ).

We expect the sophistication of the conversational experience per se to vary notably across these enterprise applications and to be inferior in terms of control over dialogue processes compared to custom CAI implementations. Integrations to back-end systems are also expected to be platform-specific, and hence, narrow in scope and not scalable.

Many of the enterprise applications in question had traditionally been regarded as back-end systems VAs were integrating to pull out data or fulfill some request, or as channels users were interacting through for sending their prompts to the underlying bot. Such status will not change any time soon.

Some of the platform vendors in the CRM, ERP or ITSM space, for instance, have also been offering dedicated modules to develop custom chatbots in the last few years — those in previous research we called "targeted service." It is not clear whether such products will be leveraged and augmented or retired. From application vendors' standpoint, they certainly will go in the direction of augmentation as they are best positioned to enrich the leveraged LLMs with enterprise data repositories as these applications house a rich repository of good, structured, functional and/or vertical-specific data. From buyers' standpoint, they are unsure whether to leverage such application vendors' product or use another generic product and do the same in terms of data enrichment.

Recommendations

- Assess the degree of customization and scalability entailed by CAI solutions offered by different types of enterprise applications.
- As the CAI-product space is consolidating, ensure your CAI assets are reusable across different implementation approaches and that your own domain models are loosely coupled with underlying vendors' technology so you can swap straightforwardly.

LLMs and GenAI Techniques Are Deeply Transforming Underlying Chatbot Architectures

Augmentation of Classical Chatbot Architectures

Prior to the hype around ChatGPT and GenAI, the high-level architecture of chatbots was regarded as largely standardized. [1] Exceptions were found in products such as those we covered in Cool Vendors in Conversational and Natural Language Technology. The key differentiator of such approaches is their ability to convert unstructured information into structured knowledge (for example, knowledge graphs) in a largely automated way. Such knowledge is then consumed by overlying chatbots in an intentless fashion — that is, without platform users having to explicitly define natural language understanding (NLU) intents by hand.

Several CAIP vendors have been enhancing the building blocks of CAI implementations by integrating LLMs in their pipelines. Examples include the following use cases for LLMs in CAIs:

- **NLU** — Synthetic data generation, generation of entity lists, augmentation of intent classification (also in multilingual scenarios), augmentation of entity extraction (including PII), prompt modification and completion. Data exploration for topic and intent discovery is also found among tasks vendors are augmenting via LLM-based approaches.
- **Dialogue management** — Automatic identification and creation of dialogue flows.
- **Integration** — Code and query generation.
- **Response generation** — Batch-response generation, real-time response generation/suggestion, dynamic personalization of responses, response translation.

Evaluate LLMs' utility and the effectiveness in these scenarios case by case, and compare LLM-augmented functionalities to classical ones for performance, accuracy and cost-effectiveness. Composite AI approaches, which leverage a concatenation of layers based on different techniques, should normally be preferred to improve the overall accuracy and transparency of specific tasks, such as NLU. CAIP vendors normally leverage LLMs on a selective basis in their pipelines, only when strictly needed and intended. This is to minimize costs and risks related to unreliability of outputs and data privacy and security when disclosing information with any third party that may provide the LLM.

CAIP vendors are also starting to offer their own domain-specific pretrained LLMs, often a fine-tuned version of proprietary or open-source models, or they allow clients to bring their own.

The Rise of LLM-Enabled RAG Architectures

GenAI techniques have enabled new chatbot architectures, visualized in Figure 5. RAG approaches use a search engine to locate and retrieve any information necessary to construct a prompt including proprietary information. More specifically:

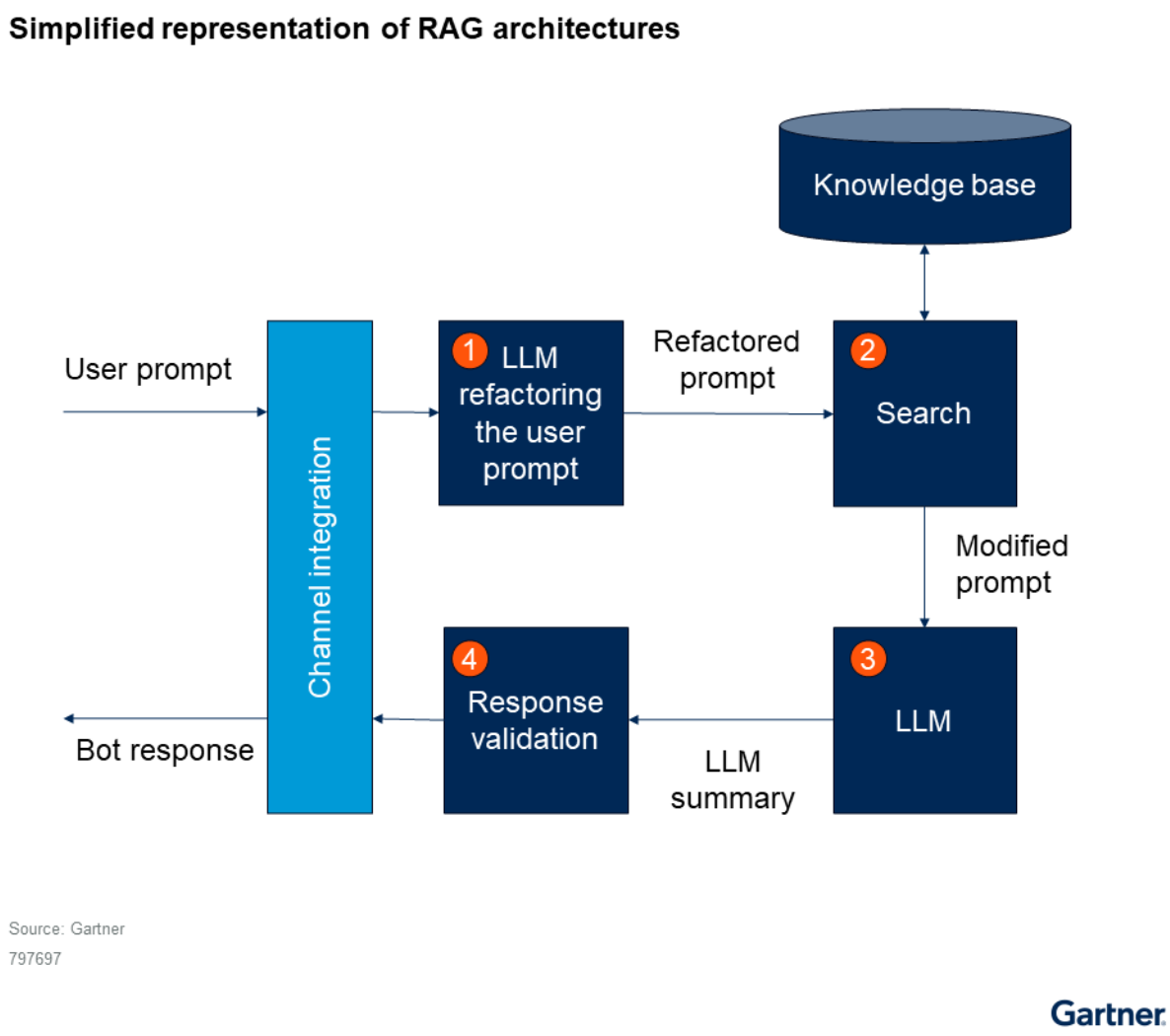
1. An LLM may be employed to rephrase/refactor the original user request in a way it is compatible with the format required by the search engine that applies in the subsequent step.

2. A search based on the refactored prompt is run. Such search can be grounded on any knowledge representation or approach of choice, including knowledge graphs and vector databases.

3. An LLM is used to summarize the information included in the modified prompt and generate a fluent, human-like response.

4. A "response validation/grounding" layer monitors and checks responses generated by LLMs for accuracy, security and such before they are displayed in front of final users.

Figure 5: Simplified Representation of RAG Architectures



The range of use cases where RAG architectures are employed are primarily Q&A chatbots. The LLM layers are stateless, meaning they do not have any contextual awareness of the previous steps of the conversation, unless the prompt itself informs them about it at each turn. Conversational processes are then bound to this additional modification of the prompt, which, unlike the case of scripted dialogue flows, is not guaranteed to be robust and accurate all the time. Additionally, it may turn out to have a serious impact on costs in the long term if a third-party LLM priced by tokens is used.

Future Convergence

RAG-like approaches are already made available by several vendors in the CAIP space. Such products are offered in addition to classical chatbot implementation approaches, often in beta. Additionally, vendors are also experimenting with intentless approaches driven by LLMs, in a way that is not conceptually different from capabilities offered by the Cool Vendors in Conversational and Natural Language Technology. Such approaches appear to be promising to effortlessly expand chatbots' knowledge beyond the intents that CAI developers normally manually craft and maintain.

CAIP vendors are then rethinking the basic building blocks of chatbot architectures, and new standards are being defined. In the midterm, we expect CAI application architectures to converge to new design patterns that leverage both the control and the transparency enabled by classical CAI techniques, and the fluency and the scalability from GenAI-enabled tools.

Recommendations

- Evaluate whether classical CAI architectures augmented by LLMs can meet your requirements before choosing potentially riskier solutions such as pure RAG-enabled chatbots.
- Compare CAIP vendors also based on how they are incorporating LLMs in their product pipelines (for example, in terms of how they manage context awareness of LLMs and answers' feasibility risks) and their GenAI roadmap.

Leaders Are Confused About the Best Approach to Enable CAI in Their Organization

Due to the impact GenAI is having on the CAI space, leaders are confused when it comes to selecting the best approach for adopting GenAI-enabled VAs/CAI offerings into the organization. We identified five adoption approaches:

1. **Buy from CAIP vendors and customize** — Buy productized tools and functionalities from a CAIP vendor and use them to custom-build the CAI application.
2. **Buy from CAIP vendors and customize, bringing your own fine-tuned models** — Same as option one, but the organization brings one or more LLMs that were customized (fine-tuned) elsewhere.

3. **Buy from non-CAIP vendors and customize** — Buy GenAI-enabled conversational functionalities provided by non-CAIP vendors and customize the CAI application/functionality (for example, integrating with/bringing your own data and documents).

4. **Consume** — Use noncustomizable CUIs embedded in enterprise applications. Data dependencies and integrations with underlying data sources and knowledge bases are not intended to be customizable in this option.

5. **Custom-build from underlying tooling** — Build RAG-like CAI applications from the ground up using your internal tech and expertise.

Table 1 covers the pros and cons of such approaches.

Table 1: Pros and Cons of Currently Available Approaches to Adopt CAI Capabilities

(Enlarged table in Appendix)

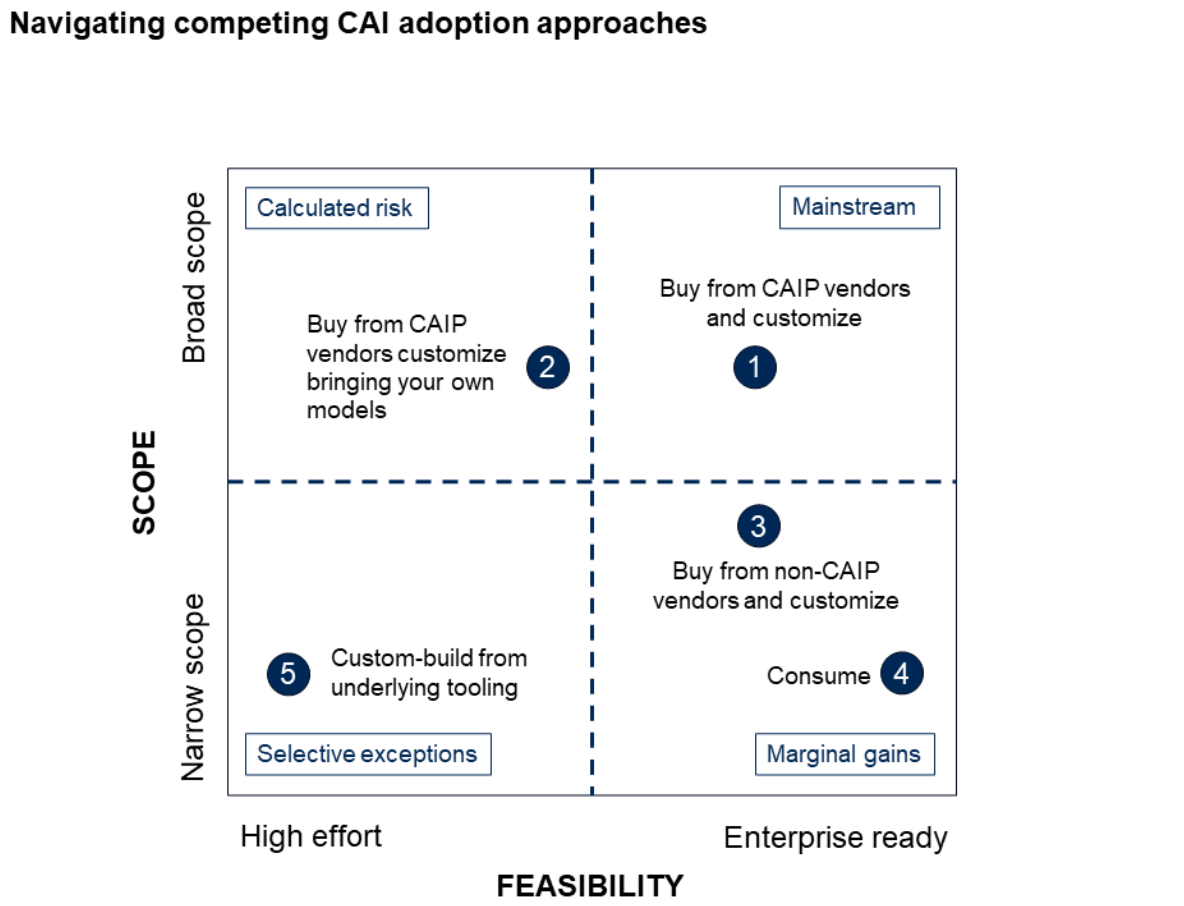
Approach	Pros	Cons
1. Buy from CAIP vendors and customize	**Flexibility and scalability:** Support for diverse use cases with varying scope and integration needs. **Enterprise readiness of GenAI-enabled capabilities:** Technical service providers (TSPs) incorporated LLMs in their pipelines accounting for GenAI-specific risks [2] so LLMs are used in a conservative way when they provide value and within the context of more extensive, composite AI pipelines. **Robustness:** Proven tooling and vendors' expertise in the CAI domain.	**Application customization effort:** Building a CAI application is intrinsically complex, despite low-code tooling offered by CAIPs, and the learning curve may vary depending on the complexity of the implementation.
2. Buy from CAIP vendors and customize, bringing your own fine-tuned models	**Flexibility, scalability and robustness:** Same as option one. **Reusability of assets:** A custom fine-tuned LLM represents an asset that, if loosely coupled to the vendor's platform, can be preserved and reused across different initiatives.	**Application customization effort:** Same as option one. **Additional model customization effort:** LLM fine-tuning requires advanced skills and significant effort. Availability of pretrained domain-specific models should be considered.
3. Buy from non-CAIP vendors and customize	**Non-CAI capabilities:** Additional capabilities such non-CAI vendors may provide in their area of expertise (from IDP to enterprise search).	**Limited CAI-specific functionalities and scalability:** Additional CAI-specific capabilities offered in such products are likely to be limited, which could reduce the overall scalability of the implementation.
4. Consume	**Enterprise readiness and virtually no implementation effort:** No customization effort is required, and CAI capabilities are immediately available.	**Tactical and narrow scope:** The scope of the use cases/chatbot application is probably going to be limited and nonscalable.
5. Custom-build from underlying tooling	**Flexibility:** The customization possibilities are endless.	**Effort:** Building enterprise-grade risk mitigation measures and additional capabilities needed to build and maintain a robust CAI initiative entails significant skills and effort. Organizations are at risk of reinventing the wheel in several scenarios. **Scope and prod deployment:** Organizations may end up with narrow-scope POCs that are unlikely to make it to production.

Source: Gartner (September 2023)

The scope and feasibility of these five options are evaluated in Figure 6:

- **Scope** — The scope of the use case that the option allows to support, from task-specific to enterprisewide. Criteria that contribute to options to score higher in this dimension are:
  - Built-in capabilities beyond those strictly needed to build a self-service chatbot (for example, voice automation and document ingestion capabilities)
  - Predefined integrations to back-end systems and channels, analytics and agent-assist products
  - The ability to build a customized implementation across various use-case types (from Q&A bots to transactional voicebots for process automation)
- **Feasibility** — The skills and effort needed to implement a production-grade custom CAI application, but also any prebuilt tools and guardrails designed to ensure enterprise-grade risk mitigation.

Figure 6: Navigating Competing CAI Adoption Approaches



Recommendations

- Choose adoption approaches strategically, considering the scope of the CAI initiative and the cross-function automation and human augmentation needs it is intended to meet.
- When experimenting building CAI applications from underlying tooling and LLMs, estimate the effort, additional capabilities and costs entailed by moving such POCs to production.

Acronym Key and Glossary Terms

CAI	Conversational AI
CAIP	Conversational AI Platform
CUI	Conversational User Interface
DNN	Deep Neural Network
GenAI	Generative AI
GUI	Graphical User Interface
LLM	Large Language Model
NLG	Natural Language Generation
NLP	Natural Language Processing
NLQ	Natural Language Query
NLTs	Natural Language Technologies
NLU	Natural Language Understanding
Q&A	Question and Answer
RAG	Retrieval-Augmented Generation
TSP	Technical Service Provider
VA	Virtual Assistant

Evidence

This research is grounded on interactions (inquiries and briefings) with vendors and clients about conversational AI and generative AI. It is also based on previous Gartner research in these areas and public information on the topic (including news articles, vendors' press releases, academic papers and technical documents).

Endnotes

[1] The traditional core building blocks that still apply to the majority of CAIPs today are:

- **Channel integrations** — The ability of the CAI application to connect to a variety of different channels, use specific rich features of different channels and operate multiple channels centrally.
- **Natural language understanding** — The ability to understand the user's natural language, which includes intent matching as well as entity recognition.
- **Dialogue management** — The ability to handle complex and sophisticated dialogue that may span different types, such as question and answer, query, transactional, negotiation and review.
- **Response generation** — The ability to create the output message and transmit it to the channel integration module for delivery to the user channel.
- **Back-end integrations** — Back-end integrations are meant to enable communication with back-end systems, such as common enterprise applications like CRM, ERP and ITSM platforms, and data sources.

[2] Risks posed by GenAI and LLMs are normally related to unreliability of outputs, which includes inaccuracies, hallucinations, bias and copyright violations, but also privacy and data security risks, as well as cybersecurity risks. See Table 2 in How to Pilot Generative AI and Microsoft Azure OpenAI vs. OpenAI: Comparing GenAI Trust, Risk and Security for an exhaustive list. Additional concerns leaders should have when operationalizing LLM-based CAI implementations are related to the limited visibility on long-term costs associated with LLM usage, as well as their performance in specific use cases where low latency is key, such as voicebots.

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Magic Quadrant for Enterprise Conversational AI Platforms

Critical Capabilities for Enterprise Conversational AI Platforms

Quick Answer: What Impact Will Generative AI Have on Search?

Best Practices for the Responsible Use of Natural Language Technologies

AI Design Patterns for Large Language Models

AI Design Patterns for Knowledge Graphs and Generative AI

How Can Generative AI Be Used to Improve Customer Service and Support?

How to Pilot Generative AI

Microsoft Azure OpenAI vs. OpenAI: Comparing GenAI Trust, Risk and Security

How to Choose an Approach for Deploying Generative AI

Market Guide for Augmented Analytics

Table 1: Pros and Cons of Currently Available Approaches to Adopt CAI Capabilities

Approach	Pros	Cons
1. Buy from CAIP vendors and customize	**Flexibility and scalability:** Support for diverse use cases with varying scope and integration needs. **Enterprise readiness of GenAI-enabled capabilities:** Technical service providers (TSPs) incorporated LLMs in their pipelines accounting for GenAI-specific risks [2] so LLMs are used in a conservative way when they provide value and within the context of more extensive, composite AI pipelines. **Robustness:** Proven tooling and vendors' expertise in the CAI domain.	**Application customization effort:** Building a CAI application is intrinsically complex, despite low-code tooling offered by CAIPs, and the learning curve may vary depending on the complexity of the implementation.
2. Buy from CAIP vendors and customize, bringing your own fine-tuned models	**Flexibility, scalability and robustness:** Same as option one. **Reusability of assets:** A custom fine-tuned LLM represents an asset that, if loosely coupled to the vendor's platform, can be preserved and reused across different initiatives.	**Application customization effort:** Same as option one. **Additional model customization effort:** LLM fine-tuning requires advanced skills and significant effort. Availability of pretrained domain-specific models should be considered.
3. Buy from non-CAIP vendors and customize	**Non-CAI capabilities:** Additional capabilities such non-CAI vendors may provide in their area of expertise (from IDP to enterprise search).	**Limited CAI-specific functionalities and scalability:** Additional CAI-specific capabilities offered in such products are likely to be limited, which could reduce the overall scalability of the implementation.

4. Consume	**Enterprise readiness and virtually no implementation effort:** No customization effort is required, and CAI capabilities are immediately available.	**Tactical and narrow scope:** The scope of the use cases/chatbot application is probably going to be limited and nonscalable.
5. Custom-build from underlying tooling	**Flexibility:** The customization possibilities are endless.	**Effort:** Building enterprise-grade risk mitigation measures and additional capabilities needed to build and maintain a robust CAI initiative entails significant skills and effort. Organizations are at risk of reinventing the wheel in several scenarios. **Scope and prod deployment:** Organizations may end up with narrow-scope POCs that are unlikely to make it to production.

Source: Gartner (September 2023)