



# The State of Agentic CX



# Table of Contents

Introduction	3
• Executive Letter	
• Methodology	
• Key Findings	
Companies Don't Want to Talk to Their Customers	11
The Chat Experience is Broken	13
Voice Support is 30 Years Behind	20
Companies Are Not Ready for Personal AI Agents	23
Gaps Are Opportunities	25



# The State of Agentic CX in 2026

Enterprise customer experience remains focused on avoidance and deflection rather than deepening relationships

Customer expectations for support are only growing. As customers take advantage of the speed and effectiveness of AI in their everyday lives, they expect enterprises to innovate. AI agents have proven to reliably perform in customer experience settings, making it possible to deliver entirely new levels of immediate, personal, and effective service via voice and chat.

These are things we at Parloa know to be true. What we wanted to learn was the current state of customer experience, agentic or otherwise, across enterprises globally. Specifically, we wanted to baseline the gap between what customers expect and what the world's leading brands actually deliver.

The questions we wanted to answer were simple, yet telling:

- Can customers easily find a phone number or chat solution when they need help?
- Can chat conversations adequately solve customer problems?
- Are voice experiences conversational, seamless, and effective?
- Are enterprises ready for agent-to-agent customer experience?

Surveys are an excellent research mechanism, but we wanted to experience the experience. Our goal was to explore a wide variety of support systems across enterprise industries, geographies, and channels to create a stable benchmark that enterprises can use year over year to compare themselves against peers.



Hiring an army of human “mystery shoppers” to experience these journeys would be prohibitively expensive and time-consuming. Given that we build AI agents, it seemed only fitting, and frankly ingenious, to use our own technology to get the breadth and scale we needed. These discovery agents worked alongside our internal research team to study thousands of enterprise customer experience systems, creating the first-of-its kind, agent-led customer experience study.

Going in, we suspected there was an experience gap between customers and the companies. What we found was a gap far wider than any of us anticipated:

- **43% of enterprises** have no findable phone number or chat option.
- **9% of chat experiences** adequately resolve issues.
- **99% of voice experiences** rely on decades-old technology (or none at all).
- Hold times **exceed 90 minutes**.
- Agent-to-agent readiness — **not a chance**.

---

These statistics represent only a portion of the findings in the report. Find more detailed numbers on chat, voice, and overall website experience below.

### **While the gap is huge, this isn't a moment for despair.**

What these findings indicate is that the race is on, and the opportunity for enterprises is enormous. The fact that you're here, curious enough to read this, means you're already ahead.

Read the Parloa State of Agentic CX in 2026 below. Let us know what you think. We'd love to create a personalized report for your organization. Keep in mind, this is just the beginning. We have a lot more research underway.



*Latané Conant*

**Latané Conant**  
Chief Marketing Officer, Parloa



---

## METHODOLOGY

Parloa conducted an evaluation of enterprise chat and voice systems, including those of the Forbes Global 2000. The study largely represents the banking, financial services, insurance, travel, hospitality, retail, and health care industries. To perform the assessment, Parloa designed and deployed fully autonomous AI discovery agents. The foundational scan of these agents was customer service contact accessibility, which informed subsequent, deeper analysis. Ultimately, the study was divided into three phases:



# Customer service accessibility friction

In the first phase of the study, Parloa’s agents assessed websites for the accessibility of customer support channels. They tracked the visibility of a support channel on the website, how much scrolling or clicking was required to get to support, and which support channels were available.

Phase 1 agents applied a proprietary methodology for evaluation. Each company received a score on a five-point scale, from 1 (easiest) to 5 (most difficult). The score reflected the level of user effort required to locate phone or chat support.

<b>LEVEL 1</b>	<ul style="list-style-type: none"><li>• Contact options are immediately visible</li><li>• Phone or chat typically surfaced in the site header or footer</li><li>• No user action required</li></ul>
<b>LEVEL 2</b>	<ul style="list-style-type: none"><li>• Contact information available within one click</li><li>• Clearly labeled navigation link</li></ul>
<b>LEVEL 3</b>	<ul style="list-style-type: none"><li>• Moderate effort</li><li>• Requires two clicks, scrolling, or navigation to a secondary page</li></ul>
<b>LEVEL 4</b>	<ul style="list-style-type: none"><li>• Elevated friction</li><li>• Contact details buried in help centers or FAQs</li><li>• Requires three or more steps to find</li></ul>
<b>LEVEL 5</b>	<ul style="list-style-type: none"><li>• Very limited accessibility</li><li>• Contact information is difficult to locate, obscured, or absent</li></ul>



In parallel, the agents assessed contact availability, categorizing each site based on the support channels offered: phone only, chat only, both phone and chat, or neither. This classification enabled a more nuanced view not only of how easy it is to find contact options, but also of whether meaningful real-time support is available at all.

The discoverability data adds a critical layer to the study's broader findings. Subsequent sections detail what happens when customers do reach automated systems, and the results are poor. This section establishes that a substantial share of customers never even get to a system. The customer experience (CX) gap is not limited to broken chatbots and rigid interactive voice response (IVR) trees. It starts at the front door.

The research then proceeded down two tracks, voice and chat, each with a distinct methodology shaped by what the technology itself revealed in Phase One.

## Phase 2

# Chat journey evaluation

Parloa's discovery agents conducted 4,000 interactions across more than 800 enterprises, pressure-testing the systems that sit between customers and answers. The agents initiated scenario-based conversations focused on five common customer needs:



Booking requests



Routing inquiries



Transactional tasks



Language handling



Location-specific questions



Agents assessed the performance of enterprise chatbots using a structured four-part capability evaluation designed to distinguish modern LLM-based systems from traditional rule-based automation.

CAPABILITY	DEFINITION
Style adaptation	Measures if the system can modify language complexity or tone upon request
Instruction following	Tests adherence to explicit formatting requirements
Cross-turn memory	Evaluates recall of information introduced earlier in the conversation
Logical reasoning	Assesses the ability to respond coherently to non-trivial yes-or-no scenarios

Based on performance, each chatbot was classified along a four-tier scale ranging from rule-based to LLM-detected. Systems that passed three or four tests were classified as "LLM-detected." Those passing two were labeled "Likely-LLM." One test passed earned "Likely-rule-based LLM." Zero tests passed resulted in a "Rule-based" classification.







LLM-detected	Passed three or four capability tests
Likely-LLM	Passed two of four capability tests
Likely-rule-based	Passed one of four capability tests
Rule-based	Passed zero capability tests
<b>LLM-powered</b>	<b>LLM-detected and Likely-LLM of classifiable</b>
<b>Rule-based</b>	<b>Rule-based and likely-rule-based of classifiable</b>



# Interactive voice response and voice testing

IVR and voice testing followed a different approach, as the majority of CX systems did not support agent-to-agent interaction. When Parloa deployed its generative AI voice agents to navigate enterprise phone trees, systems that depend on dual-tone multi-frequency (DTMF), rigid authentication gates, and unclear menu structures blocked them.

Therefore, Parloa's in-house human research team manually mapped the phone support architecture of more than 100 large enterprise accounts, calling main support lines and documenting menu depth, option load, authentication gating, callback availability, hold times, and use-case completion across six standard support scenarios:

 Paying a bill	 Booking a reservation	 Opening a new account
 Updating account information	 Answering an FAQ	 Requesting to speak to a live agent

This study establishes the first agent-conducted benchmark of agentic enterprise customer experience across chat and voice channels.



# Key Findings

**01**

## Companies don't want to talk to their customers

- **43.3%** of websites displayed no visible phone or chat option, and no obvious path to support.
- **56%** of websites required scrolling to locate support options.
- **58%** of enterprises presented at least some level of difficulty in accessing customer service.
- **37.8%** of enterprises displayed multiple phone numbers without a clear hierarchy.
- **2%** of all chatbots were accessible on support pages.

**02**

## The chat experience is broken

- **71.1%** of chatbots produced an actual conversation.
- **7.5%** of chatbots demonstrated genuine large language model (LLM) capabilities.
- **92.5%** of chatbots still run on traditional rule-based systems.
- **8.9%** of chatbot conversations achieved the customer's stated goal.
- **10.1%** of chatbots successfully routed customers to humans when prompted.

**03**

## Voice support is 30 years behind

- **96%** of dedicated support lines use legacy IVRs.
- **3-4** IVR menu options/button presses were required to reach a human in most industries.
- **6-8** path options were presented in the IVR menus for banking and insurance.
- Only **8%** offered a callback option.
- **90+** minutes was the longest observed hold time.

**04**

## Companies are not ready for personal AI agents

- Only **1%** of the enterprises demonstrated readiness for the next frontier: personal-agent-to-enterprise-agent interaction.



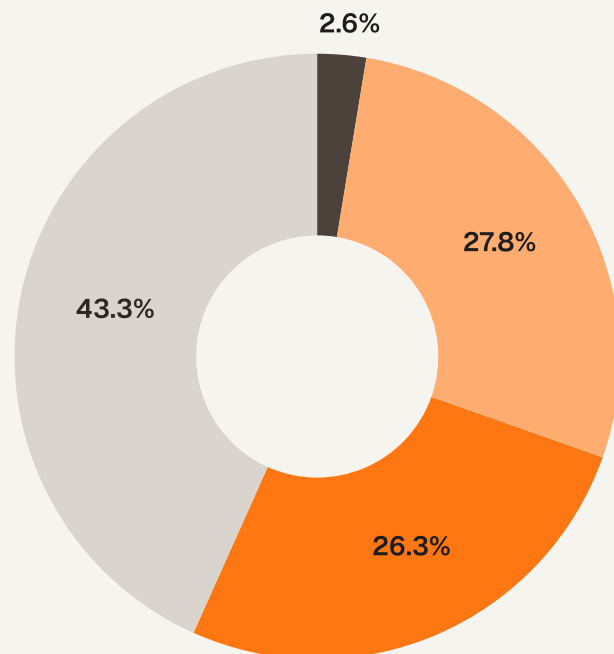
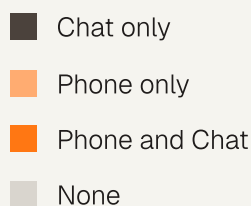
# Companies don't want to talk to their customers

The foundational scan by Parloa's AI discovery agents measured friction in customer service accessibility, seeking to determine how easily customers can find a way to get help. The results showed that nearly half of enterprise companies do not make it easy for their customers to talk to them.

For nearly half the world's largest companies, the first barrier to customer service is finding it at all.

**43.3% of the companies analyzed displayed neither a visible phone number nor a chat option. 58% presented at least some level of difficulty in accessing customer service.** This means that thousands of the world's largest enterprises are offering no obvious path to human or automated support. Customers coming to their websites with a problem need to hunt, guess, or give up entirely on solving it.

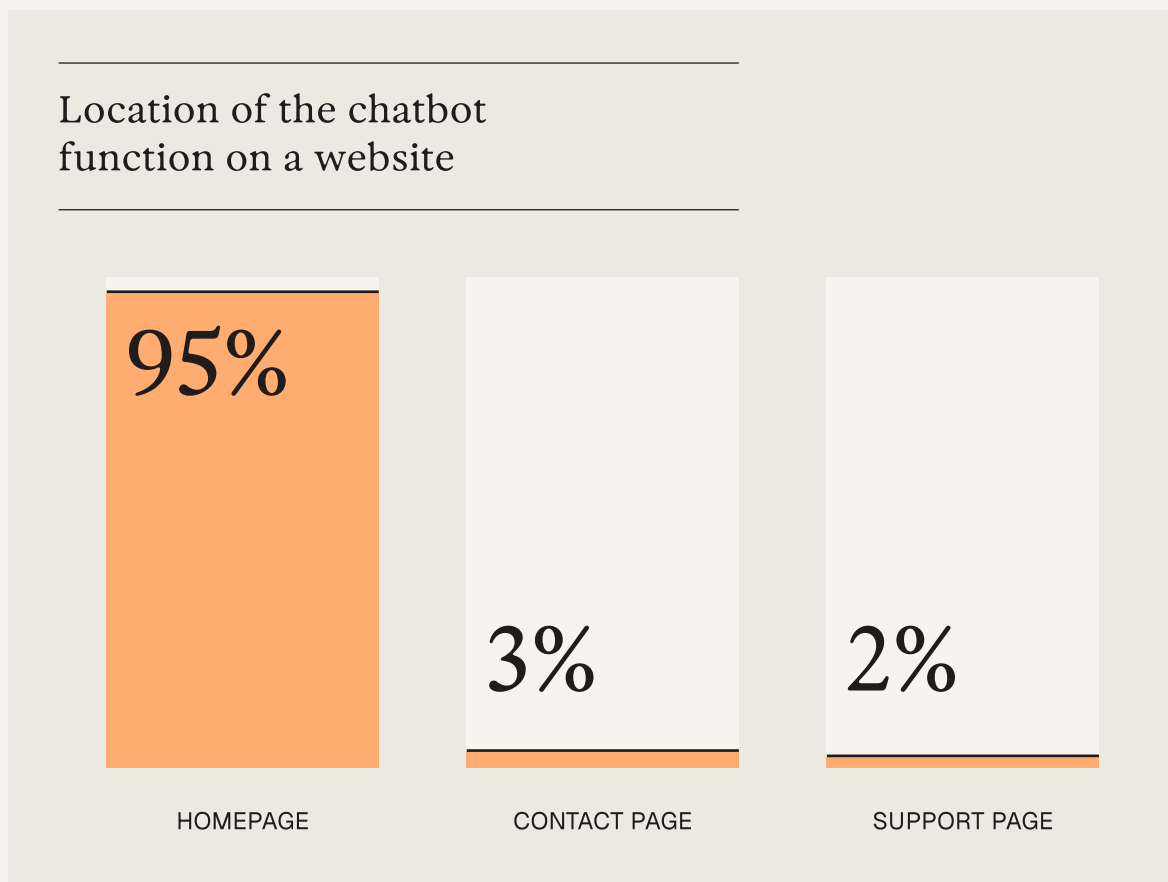
## Companies with chat or phone support





Among companies that did surface contact options, phone remained dominant. **Just over 54% of companies displayed a phone number, compared to 28.9% offering chat. While 26.3% of the sites provided both options,** only 2.6% of the sites analyzed delivered chat as their sole method of contact. These statistics acknowledge that enterprises are prioritizing voice as their customers' preferred channel, while suggesting that very few brands have fully committed to a digital-first service model.

Discoverability was uneven at best. **Nine out of 10 companies that displayed contact information placed it on the homepage, requiring zero clicks to find.** While that sounds encouraging, the scroll data tells a more disappointing story. **More than half of all sites (56%) required scrolling to locate support options.** Phone numbers appeared most often in the main content area, followed by headers and then footers. Placement in the footer, a location most users never reach, accounted for more than a third of visible phone numbers. Nearly all chatbots identified (95%) appeared only on the homepage, with just 3% on dedicated contact pages and less than 2% on support pages. This finding suggests that most enterprises treat chat as a homepage feature rather than embed it across the service journey.





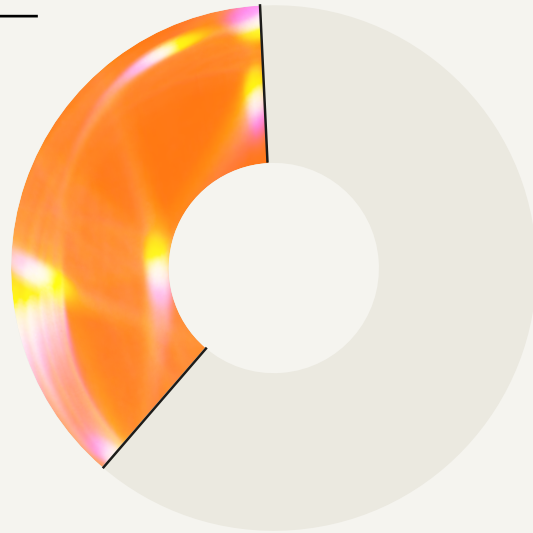
Of the enterprises that offered a phone number, **37.8% displayed multiple phone numbers without a clear hierarchy**, further highlighting that when the world's largest companies cannot present a single, unambiguous way to get help, the customer experience fails before any chatbot or IVR system has a chance to underperform.

---

# 37.8%

---

of enterprises displayed multiple phone numbers without a clear hierarchy.



Key finding 2

## The chat experience is broken

On the websites that offered chat support channels, Parloa's AI discovery agents evaluated the chat journeys to answer a more relevant question: do these systems perform effectively? In most cases, the agents found that they do not.

Among companies where chatbots were detected, **71.1% produced a measurable interaction**, defined by at least two non-empty messages or at least 30 characters of content. Of these 2,710 measurable interactions, only 241 met the test goal, yielding an **8.9% resolution rate**.



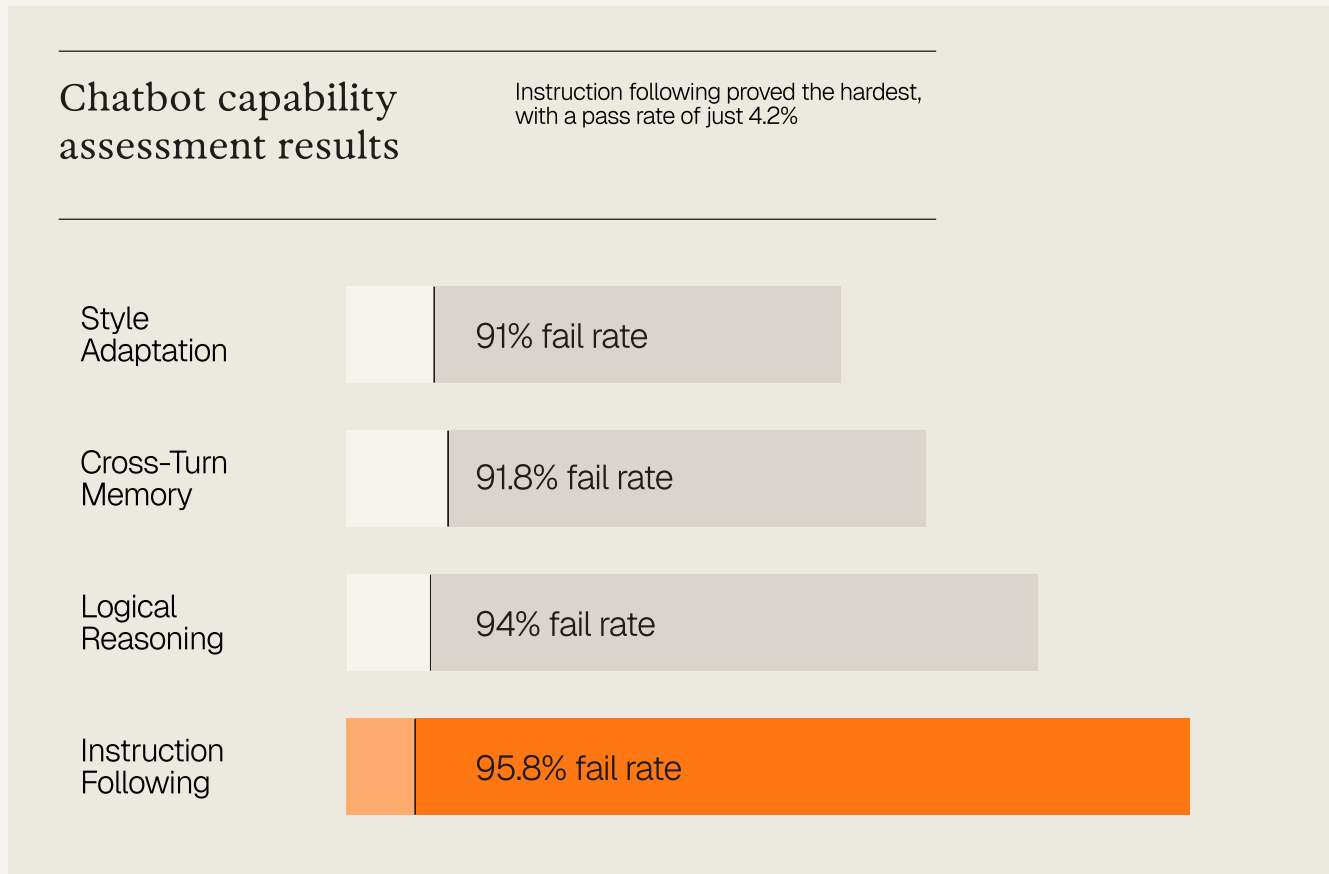
Based on these interactions, researchers collected sufficient data to classify hundreds of chatbots along a spectrum ranging from rule-based systems to LLM-powered implementations. The overall results were stark. Among classifiable systems, **3.7% met the criteria for LLM-detected classification**. Another 1.3% passed two tests and were classified as Likely-LLMs. 2.7% passed only one test and were categorized as Likely-rule-based. **The majority (58.7%) failed all four tests and were classified as Rule-based**. An additional 32.7% of chatbots produced no substantive responses, while the remaining 1% returned results that could not be conclusively classified.

LLM-detected	3.7%	Passed three or four capability tests
Likely-LLM	1.3%	Passed two of four capability tests
Likely-rule-based	2.7%	Passed one of four capability tests
Rule-based	58.7%	Passed zero capability tests
LLM-powered	7.5%	LLM-detected and likely-LLM of classifiable
Rule-based	92.5%	Rule-based and likely-rule-based of classifiable

Only 7.5% of the classified chatbots demonstrated LLM capabilities.

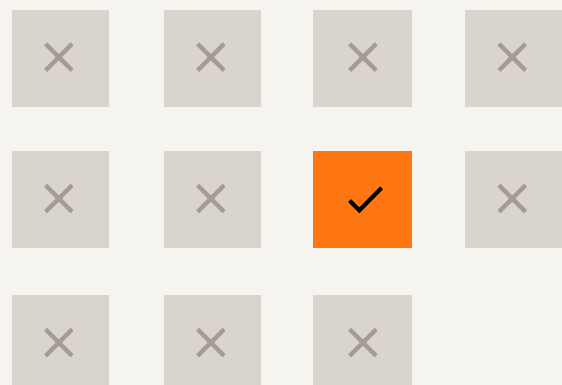


Performance across individual capability tests was consistently limited. Instruction following registered the lowest pass rate, at 4.2%. Only 4.9% of chatbots adhered to specific formatting instructions, while some 8.2% demonstrated cross-turn memory, indicating that **more than 90% could not recall information introduced earlier in the same conversation**. Logical reasoning passed at 6.0%, while style adaptation, the strongest category, reached 9.0%.



<1 in 11

interactions successfully resolved the customer's stated objective.





Furthermore, response speed did not correlate with effectiveness. **Average response time across all conversations was approximately 17 seconds**, though performance varied substantially by sector. Public-sector chatbots responded in roughly 10 seconds, while construction and engineering systems frequently required 30 to 45 seconds. Financial services averaged 15.6 seconds, a response time that appears competitive in isolation but led to fewer than 8% successfully-resolved interactions.

Speed without substance is not customer experience. It is just faster frustration.

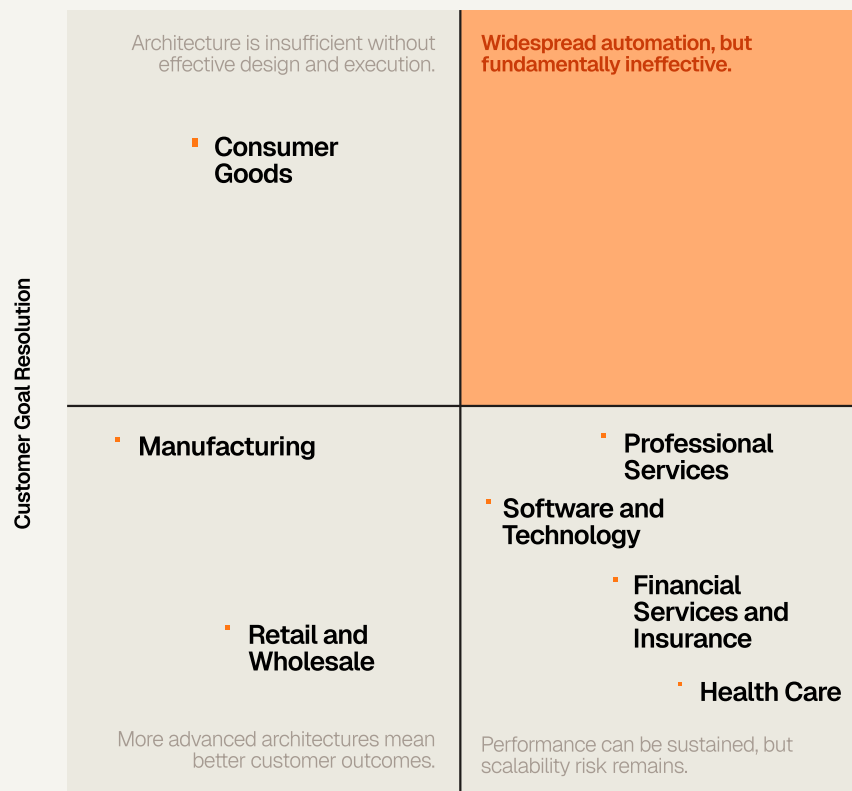
## Industry benchmarks for CX chatbots

The study establishes baseline performance metrics across all measured dimensions of automated CX, segmented by industry. The intent is twofold: to provide a standardized benchmark for the broader market and to offer individual enterprises a comparative view of their relative performance.

The five largest industries by sample size, which together represented more than 60% of the study, all exhibited rule-based chatbot rates above 50%, indicating that LLM adoption remains the exception rather than the norm.

### Customer outcomes by industry

More rule-based chatbots consistently deliver lower goal resolution.





Financial services and insurance accounted for the largest share of tested chatbot interactions. Despite the sector's scale and investment in technology, performance outcomes were mixed. While chatbot adoption stood at 64.2%, customer goal resolution stood at 7.4%. 65.7% of classifiable chatbots were categorized as rule-based.

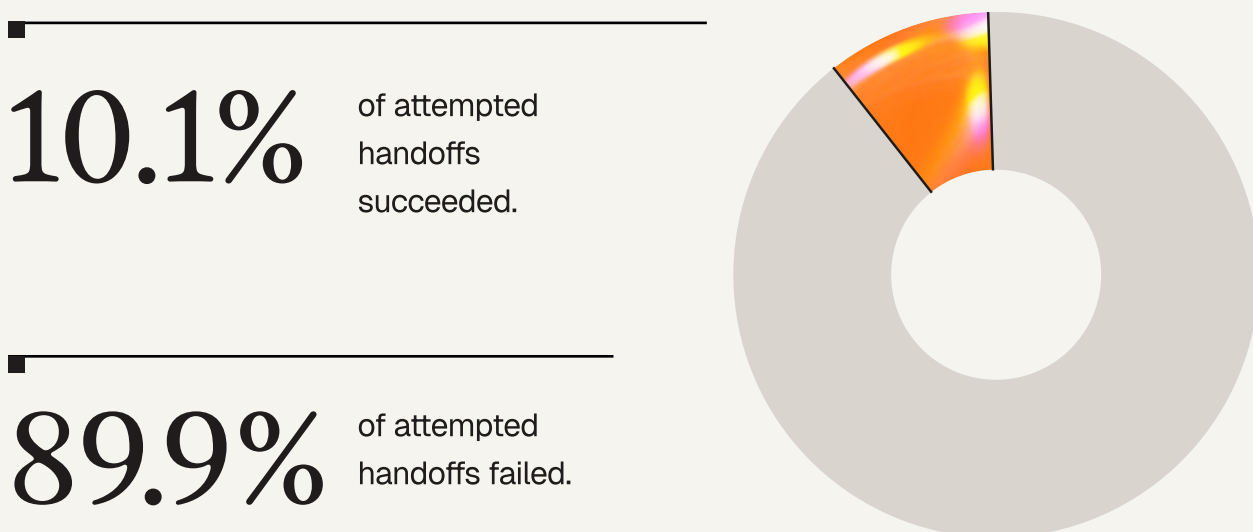
Software and technology companies demonstrated somewhat higher LLM adoption but still reported a 58.9% rule-based classification rate. Health care had one of the highest rates of rule-based systems (71.4%) and delivered above-average chat availability.

Taken together, these results show that even the strongest chatbot capabilities fall short of the cognitive performance required for reliable, end-to-end customer resolution.

## The handoff failure

In theory, when a chatbot cannot resolve a customer's issue, escalation to a human agent serves as the primary fallback. In practice, that escalation path rarely functions as designed.

Across 771 transfer attempts, only 78 resulted in a successful handoff to a human agent.





Handoff performance varied by industry but remained low across all sectors. Professional services recorded the highest success rate at 17.3%, while utilities performed the worst, at 3.0%. Even the strongest results reflect structural shortcomings rather than isolated execution failures.

The data indicate a readily-identifiable design pattern: Automation is deployed to handle straightforward requests, while unresolved or complex cases frequently terminate without escalation. This approach leaves a significant share of customer interactions unresolved and undermines the intended role of automation as a triage mechanism.

	GOAL RESOLUTION	AGENT HANDOFF
Financial Services and Insurance	↓ 7.4%	10.1%
Software and Technology	8.5%	10.1%
Health Care	↓ 6.5%	↓ 8.2%
Manufacturing	9.2%	11.5%
Professional Services	9.0%	↑ 17.3%
Retail and Wholesale	↓ 6.7%	↓ 7.5%
Consumer Goods	8.8%	10.0%
Hospitality	↑ 19.3%	12.0%
Education	10.2%	10.5%
Utilities	↓ 7.2%	↓ 3%

↑ ABOVE AVERAGE    ↓ BELOW AVERAGE



The implications also extend beyond the immediate interaction. Failed handoffs increase the likelihood of customer churn, negative reviews, and engagement with competitors.

In regulated industries, including financial services and health care, unsuccessful escalation also creates compliance risk. Customers who cannot reach a human agent to dispute a charge or resolve a benefits or claims issue may have grounds to legally escalate the matter.

These findings suggest that most enterprise CX systems treat escalation to a human as an exception rather than as part of the core workflow. Low resolution rates suggest the opposite approach is warranted. When automation accomplishes fewer than 10% of goals, escalation is not a secondary path. It is the primary one, and in its current form, it is failing.

## The Chatbot Provider Landscape

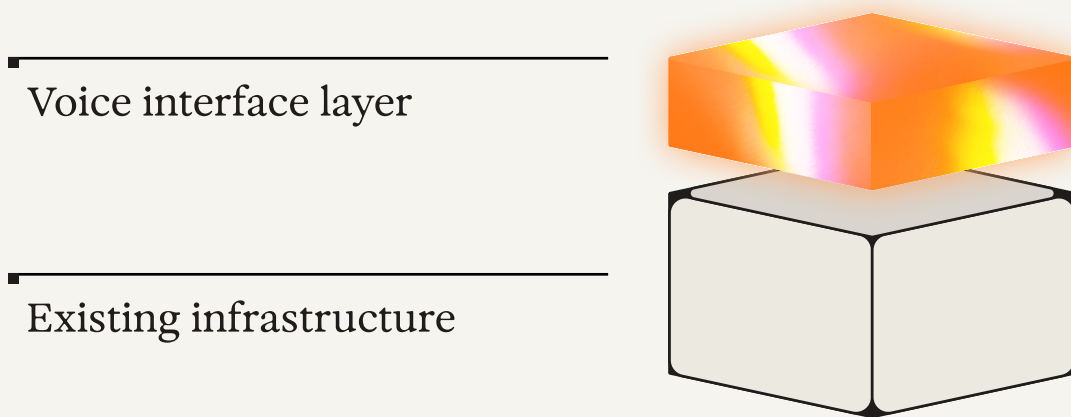
The chatbot provider landscape proved fragmented and opaque. Of the detected chat widgets, nearly 39% could not be attributed to a recognized vendor. HubSpot led the identifiable providers, followed by Salesforce, Zendesk, LiveChat, Genesys, and Intercom.

Outcomes varied by chatbot platform. Intercom achieved a 93.3% conversation viability rate while HubSpot, despite being 2.5 times more prevalent in the sample, produced a conversation in only 68% of cases. The difference suggests that platform selection influences not only resolution outcomes but also customer engagement.



# Voice support is 30 years behind

While voice AI capabilities have been deployed broadly across enterprise service environments, their functional contribution remains limited. The study found that voice interfaces are most often layered onto existing infrastructure without materially improving automation, decision making, or task completion.



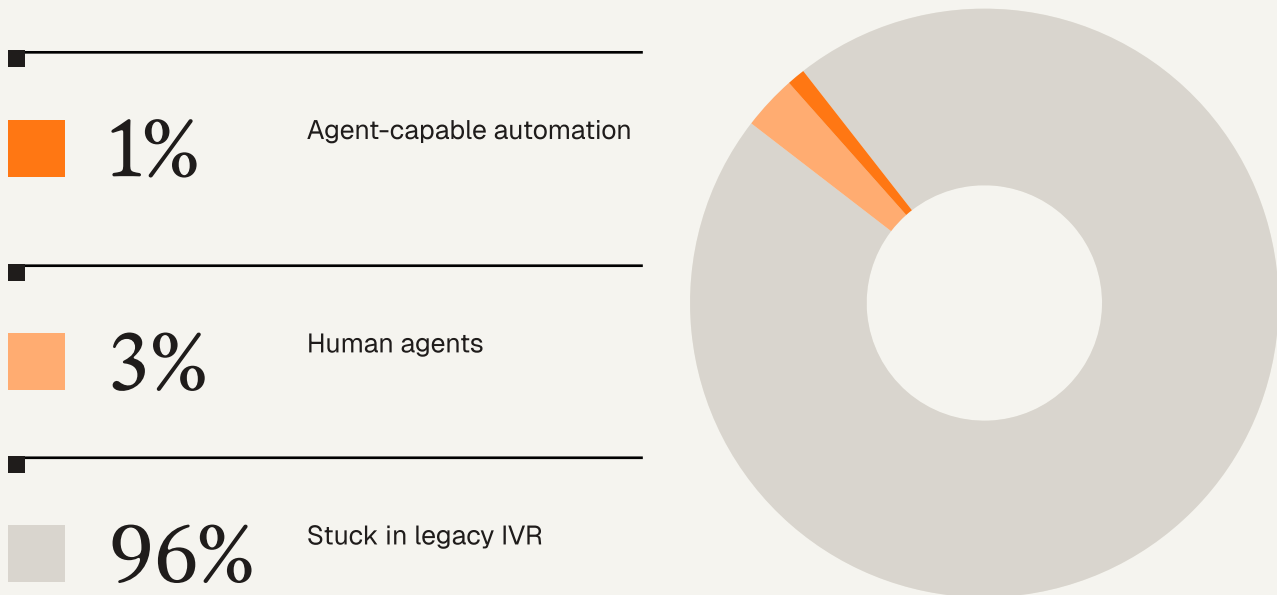
In practical terms, this structure means that while a caller might hear a natural-sounding voice prompt say, “Tell me what you need help with,” they will soon get routed into the same rigid menu tree that existed before the AI layer was added. The conversational interface screens intent but does not act on it. Additionally, the system still fails to preserve what the caller said during the transfer to a human, and the voice AI rarely resolves issues on its own.

Voice AI was frequently added on top of legacy DTMF routing rather than replacing it, providing a conversational interface without reducing structural complexity, preserving context, or resolving issues autonomously.



This pattern reflects a broader industry tendency to adopt new technology as a wrapper around old processes. The underlying architecture does not change. The customer journey does not shorten. What changes is the first 15 seconds, which now sound more modern. Everything after that remains the same.

The study's approach to voice channel assessment evolved in response to these findings. Researchers initially utilized a purpose-built voice AI agent, which sought to measure a representative customer's effort and resolution outcomes across voice interactions. Across the more than 5,000 enterprise phone contacts identified in Phase One, a global average of 96% of the calls made by Parloa's voice agent were contained by legacy IVR systems, 3% by human agents, and only 1% by automation capable of progressing conversations initiated by customer agents. The prevalence of DTMF requirements, live agent intervention, and identity verification made such measurement impractical in many cases. As a result, the focus shifted from performance evaluation to architectural analysis and task completion.



Parloa's research personnel evaluated 100 large-enterprise voice support lines globally, mapping the IVR architectures that sit between incoming calls and resolution. The patterns observed point to systems optimized primarily for operational control rather than for customer convenience.



Across insurance, health care, and travel, reaching a human agent typically required navigating three to four menu levels. Banking and insurance systems frequently presented six to eight options per menu, requiring callers to navigate complex decision trees before they could describe their issue. Greater menu depth and segmentation did not correspond to faster resolution. Instead, they correlated with longer call durations and higher abandonment rates.

The absence of callback functionality was one of the most consistent findings across industries. Only 8% of enterprises offered a callback as an alternative to holding, even when estimated wait times extended to 15, 30, or 90 minutes. Callers were also rarely informed of their position in the queue. This void highlights the limited adoption of intelligent callback orchestration, a technically mature capability that remains underutilized in practice.

Time to reach a human agent ranged from 20 seconds to more than 90 minutes. Callback options were uncommon, and most systems provided no queue transparency.

#### Industry-specific patterns reflected these broader trends:

Banking systems	combined extensive authentication gating with high option density per menu.
Insurance	relied on deep segmentation across multiple levels.
Health care systems	provided broad routing options but layered them across complex structures.
Travel systems	paired extended wait times with conversational screening.
Retail	generally showed shallower menu depth but limited callback availability.
Telecommunications	often deployed hybrid systems affected by intent-recognition errors.



Authentication requirements further increased friction, particularly in financial services. Banks and insurers commonly required identity verification before allowing access to a human agent. In some cases, calls from prospective customers were terminated entirely when authentication could not be completed. In these scenarios, security mechanisms functioned less as safeguards and more as barriers to entry.

Together, these patterns describe service architectures designed around internal process control rather than customer resolution. The result is a consistent and measurable increase in effort required to reach assistance, even before the effectiveness of that assistance can be evaluated.

#### Key finding 4

# Companies are not ready for personal AI agents

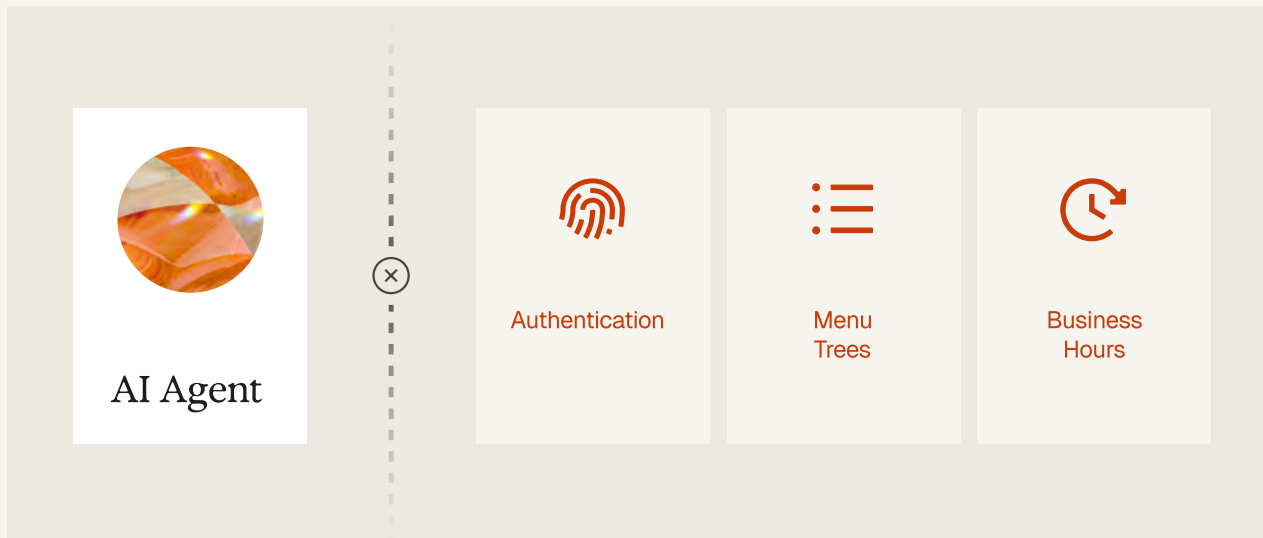
The voice AI findings introduced a new central question of the study: How ready are current enterprise voice infrastructures to support the next generation of automated and agent-to-agent interactions? **Only 1% of tested enterprises demonstrated readiness.**

In agent-to-agent interactions, a customer's personal AI agent engages directly with an enterprise service agent to negotiate, transact, and resolve issues without human involvement. Early implementations already exist, making this a near-term rather than speculative shift.

If best-in-class agentic AI cannot reach a live agent, customers have little chance of faring better.



Barriers to agent-to-agent interaction are primarily structural. Legacy IVR systems built around DTMF input cannot process natural-language requests from AI agents. Authentication mechanisms designed exclusively for human callers break down when the caller is software. Menu trees that depend on button presses are inaccessible to voice-based agents. In addition, systems that terminate calls outside of business hours or after failed authentication offer no pathway for asynchronous resolution.



Readiness for agent-based customer service requires more than incremental technology upgrades. It requires reconfiguring the service architecture to accommodate machine-initiated interactions. Enterprises need systems capable of receiving, interpreting, and acting on requests from other AI systems. Authentication frameworks must support both human and nonhuman callers. APIs and protocols must enable negotiation and transactions between agents. Service workflows must also evolve from synchronous, session-based interactions to persistent, outcome-oriented processes.

The limited readiness observed is not a reflection of insufficient ambition. It is a reflection of architectural constraints. Enterprises that address these constraints early are likely to gain a durable competitive advantage as agent-to-agent service models become more prevalent.



# Gaps Represent Opportunities

The findings of this report reveal a significant gap between customer expectations and industry reality. In 2026, customers expect fast, reliable support. Instead, they are being forced to jump through hoops only to reach dead ends.



In 2026, the technology is available to close this gap, and those who leverage it will quickly become industry leaders. Enterprises that redesign customer experience platforms to support autonomous resolution, reliable escalation, persistent context, and machine-to-machine interaction are likely to differentiate themselves on both cost and quality.

The findings of this report point to four strategic initiatives enterprise leaders should prioritize in order to close the gap:

## Put the customer at the center of service

For years, CX has been perceived as a "cost center" focused on deflection and containment. By leveraging AI that can reason and engage with customers, CX becomes a revenue-driving opportunity.

## Invest in end-to-end experiences

Just as shoppers should be able to easily buy products, active customers should be able to easily connect with support. Do not bury your support channels, rather, make them easy for your customers to access.

## Adopt a native voice AI approach

Adding voice AI on top of an existing IVR product only modernizes the first 15 seconds of a support conversation. Today's customers want to have end-to-end conversations, not be pushed through menu options.

## Make "agent-to-agent" interactions your North Star

While enterprises are operating behind rigid authentication gates and DTMF-only systems, customers are building personal AI agents to handle their chores. Enterprises must innovate to meet customers where they are, even if that is as an agent.



# To achieve these strategic initiatives, the study highlights five technical recommendations:

---

01

## Transition from rigid trees to AI-optimized intent routing

Multi-level IVR trees were built for the touch-tone era. In the AI era, enterprises must move toward voice infrastructure that captures intent in a single turn. By allowing CX architects, not just IT, to manage these flows through prompt-based platforms, businesses can iterate on routing logic as fast as end users' needs change.

---

02

## Orchestrate reliability through intelligent callbacks

The study found that the absence of callback options is a near-universal failure across the Global 2000. Intelligent callback should not be perceived as an add-on feature. It should be a continuity and reliability driver that is integrated across the entire agent lifecycle.



---

# 03

## Re-engineer authentication as a service

Currently, authentication acts as a gatekeeper that often terminates interactions before they begin. To survive in a global regulated market, enterprises need security by design, contextual verification that adjusts based on the task's sensitivity and systems that are tuned to verbal alphanumeric authentication. Security by design allows for seamless global scaling that satisfies both EU-style regulations and customers' desire for frictionless experiences.

---

# 04

## Enable & stabilize context across the journey

This study validated that context is often the barrier to successful customer experiences. Brittle architectures with several potential points-of-failure make this worse. By adopting a composable, modular platform, enterprises can build resilient infrastructure that can be triaged and more effectively maintained. With this type of platform, every piece of data, from a chat widget to a voice agent, is preserved and passed to the human agent, preventing the repetitive restatement of issues that defines 89.9% of failing chat handoffs.

---

# 05

## Shift from screening layers to reliability-driven resolution

Many of today's AI deployments are "wrappers" that screen intent but do not act on it. While the next phase of agentic CX focuses on autonomous resolution, in order to succeed, AI agents need to be trusted by end users. This transition requires focus on reliability throughout the agent lifecycle, not just algorithms that determine whether an enterprise can be charged for a successful outcome.



Today's customers want support experiences only AI can deliver. This study highlights the reality that while many companies say they have AI and support solutions, few can successfully validate that statement. Enterprises that redesign their CX experiences to connect, not deflect, customers will be the ones who stop managing costs and start winning the market.



# Methodology Notes and Limitations

The Parloa State of Agentic CX in 2026 establishes a baseline. Future iterations will measure whether enterprises narrow or widen the gap between what customers expect and what enterprises deliver.

## How Parloa AI agents conducted this study

Phases I (Customer Service Accessibility Friction) and II (Chat Journey Evaluation) of the study involved more than 10,000 enterprise online offerings and nearly 4,000 interactions across more than 800 companies in 27 industries, all conducted within a single week. Executing this effort using human researchers alone would have required hundreds of trained testers operating simultaneously across time zones, languages, and service channels. To manage consistency, quality control, and cost at that scale would have posed significant constraints.

Autonomous AI agents enabled the study to be conducted within days rather than months.

Parloa designed the research to test not only enterprise CX systems but also the feasibility of agent-led evaluation at scale. The AI agents navigated chatbot interfaces, adapted to differing interaction patterns, assessed system behavior against predefined criteria, and documented outcomes consistently across thousands of interactions.

The AI agents performed the same functions that enterprises increasingly expect from customer-facing AI systems. They initiated interactions, interpreted responses, maintained consistency across channels, and reached structured conclusions based on observed behavior. The methodology illustrates how agentic AI can operate with sufficient reliability, adaptability, and judgment to support complex, real-world enterprise use cases.



## PHASE I – Customer service accessibility friction

**AI-led sample composition.** Phase I covered some 10,000 unique companies across banking, financial services, insurance, travel, hospitality, retail, and health care sectors. Some industries had small sample sizes and have been grouped for analytical purposes.

## PHASE II – Chat journey evaluation

**AI-led sample composition.** Parloa’s AI discovery agents performed 3,958 interactions with 800 unique companies in 27 industries. Financial services and insurance represented the largest segment, followed by software and technology, health care, manufacturing, and professional services.

---

**Classification funnel.** Of the chatbot instances detected, 71.1% produced an actual conversation. Of those, 345 were classifiable on the LLM-to-rule-based spectrum. The remainder dropped out due to empty transcripts, timeouts, insufficient data, or inconclusive results. This attrition is itself a finding, as companies whose chatbots failed to engage represent a distinct tier of CX performance.

---

**Testing window.** All interactions occurred February 18 through 23, 2026. A single-week window introduces potential bias due to seasonal variation, staffing patterns, and system maintenance. Future waves of the study will test across multiple windows to strengthen longitudinal claims.

---

**Enterprise scope.** By design, this study covers the largest enterprises globally. These companies arguably have more resources and more mature technology stacks than the broader market. Performance among mid-market and small-to-medium-sized companies may differ and, in some cases, be worse.



**The guardrail caveat.** Nearly 60% of chatbots that self-identified as AI failed every capability test. Some of these systems may rely on LLMs operating under extensive guardrails that limit conversational engagement. If so, actual LLM deployment could exceed the measured 7.5%, though such implementations remain functionally indistinguishable from rule-based systems for an end user.

---

**Data discrepancies.** Minor inconsistencies exist across source documents regarding the total company counts (777, 779 and 842). These reflect different counting methods: unique companies, unique domains, unique chatbot instances and total test runs. The appendix standardizes 800 unique companies and 345 classifiable chatbots for all headline statistics.

## **PHASE III – Interactive voice response and voice testing**

**Voice sample.** IVR and voice testing covered approximately 100 enterprise support lines. This is a materially smaller sample than the chat component, and readers should note the difference in scale when evaluating voice findings.



The State of Agentic Customer Experience in 2026

Research conducted February 2026

Parloa GmbH © 2026

#### Disclaimer

This report is provided for informational purposes only and does not constitute advice, assurance, certification, or compliance determination. To the maximum extent permitted by applicable law, the authors and publishers disclaim all warranties, representations, and conditions, express or implied, including warranties of accuracy, completeness, merchantability, fitness for a particular purpose, and non-infringement. Neither the authors nor the publishers shall be liable for any direct, indirect, incidental, consequential, special, or punitive damages arising from or related to the use of, reliance on, or interpretation of this report. Nothing in this report is intended to create, nor should it be construed as creating, any contractual obligation, duty of care, or legal responsibility to any party in any jurisdiction.