2026

ASAPP

The agentic AI ecosystem for customer service

Whitepaper

Introduction

Customer service is undergoing its most consequential transformation in decades. The emergence of generative AI—and more specifically, the shift to multi-agent architectures—is redefining what it means to resolve a customer issue, assist a live agent, and orchestrate an enterprise workflow in real time.

For much of the past decade, the conversation in CX AI centered on containment: could a virtual agent handle enough of the inbound volume to justify the investment? The implicit assumption was that the agent was a gatekeeper—intercepting requests, deflecting what it could, and escalating the rest. Automation was a liability cap, not a growth driver.

That assumption is now obsolete. Generative AI agents can listen, reason, remember, and act. They can complete complex transactions—rebooking flights, processing claims, authenticating accounts, explaining policy in natural conversation—across voice and digital channels simultaneously, at any hour of the day. And crucially, they can do this not as isolated systems but as participants in a broader ecosystem of enterprise intelligence: reading from CRM agents, writing back to ticketing systems, delegating subtasks to specialized agents, and surfacing their own data for consumption by other enterprise systems like marketing, sales, and product analytics.

This white paper proposes a framework for thinking about AI agents in customer service across three dimensions: **Capabilities**, **Context Span,** and **Lifecycle**. It then applies that framework to ASAPP's **Customer Experience Platform (CXP)**—examining where the platform stands today, where the roadmap leads, and how CXP fits within a broader enterprise agentic architecture.

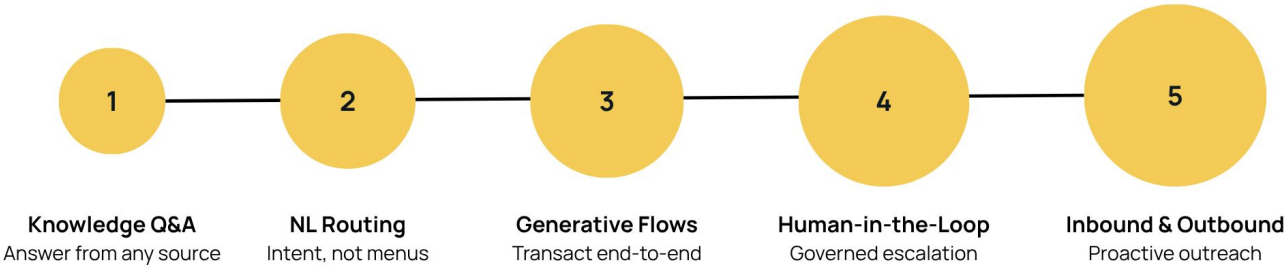
Nirmal Mukhi
Chief Architect, ASAPP

A framework for customer service agents

Before evaluating any specific platform, it helps to have a conceptual vocabulary for what a customer service agent actually is. Agents differ not just in quality—accuracy, naturalness, reliability—but in structure. What can they do? How much of the enterprise do they see? How long do they operate? These three questions define three dimensions that, taken together, characterize any agentic CX deployment.

Dimension 1: Capabilities

The first question is what kinds of tasks an agent can handle. Customer service interactions run a wide spectrum, and a mature platform needs to support the full range—not just the simple end.



Knowledge-Based Q&A. The foundational capability: answering questions grounded in product documentation, policy content, FAQs, and knowledge base articles. This is where virtually every CX AI deployment begins. The key differentiator is not just retrieval accuracy but the agent's ability to synthesize multiple KB sources, reason about policy nuance, and respond in a brand-appropriate voice.

Natural Language Routing. Rather than forcing callers through IVR menus or presenting chat users with fixed menu options, a capable agent understands intent from free-form language and routes directly to the right agentic workflow or human queue. This eliminates friction at the top of every interaction, reduces misroutes, and yields cleaner intent data than any IVR tree ever could.

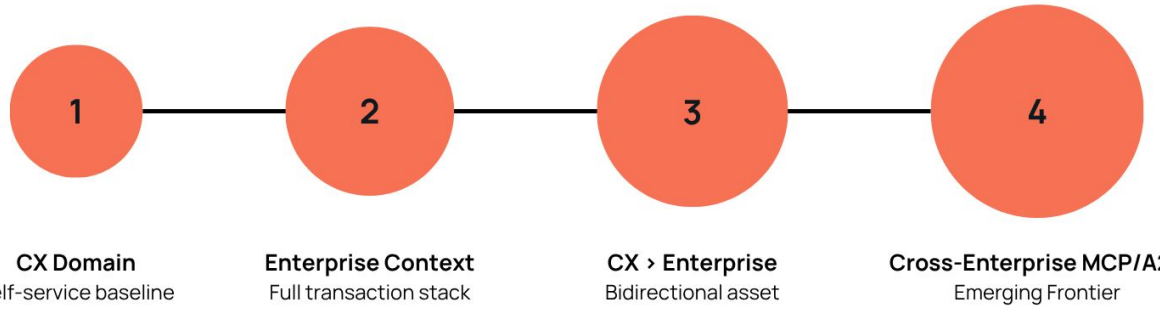
Generative Flows with Optional Determinism. This is where the real value unlocks. A generative agent can execute end-to-end transactional workflows—account changes, rebooking, authentication, cancellation, billing disputes—through natural multi-turn conversation, calling APIs at each step, and adapting based on what it learns. Critically, this generative flexibility can be combined with deterministic constraints: flows that enforce required sequences (e.g., mandatory disclosures, compliance scripts, data capture before handoff), flows that enforce step-by-step procedures - while still allowing the agent to handle variation and unexpected turns gracefully. The best CX agents blend structured control where it matters with generative flexibility everywhere else.

Collaborative Human Involvement. No matter how capable an AI agent becomes, there will always be moments that require human judgment—a policy exception, a sensitive situation, a high-value customer who needs personal attention. The question is not whether humans are in the loop but how efficiently and elegantly. A mature CX platform supports a collaborative workflow in which the AI agent remains in control of the conversation, reaches out to a human advisor behind the scenes, incorporates their input (approval, correction, additional context), and continues seamlessly. This is fundamentally different from a traditional escalation: the agent doesn't hand off, it consults. The result is higher containment with appropriate human oversight—a governance model, not a fallback.

Inbound and Outbound. Most CX AI today is inbound-reactive: the agent responds when a customer initiates contact. But a complete agentic capability set includes outbound: proactively contacting customers for appointment reminders, payment follow-ups, claim status updates, or re-engagement. Outbound agents powered by the same generative intelligence as inbound agents dramatically expand the surface area of automation—shifting customer service from a cost center that responds to problems, to a proactive function that prevents them.

Dimension 2: Context span

The second dimension is how much of the enterprise—and the world beyond it—an agent can see and influence. Context span determines whether an agent is a narrow specialist or a genuine orchestrator.



CX Domain Context. At minimum, a CX agent needs access to the information and systems relevant to its own domain: the knowledge base, customer account data, conversation history, and the tool set required to complete service transactions. This is the baseline for a functional self-service agent.

Enterprise Context. The more powerful version extends the agent's context across the enterprise. A CX agent that can read from a CRM agent—understanding purchase history, customer tier, prior interactions—delivers a fundamentally different experience than one that operates in isolation. Similarly, an agent that can access a billing system agent, an inventory agent, or a logistics agent can complete transactions end-to-end without requiring human intermediaries to bridge systems. This is the architecture of a genuinely capable AI agent: one that connects rather than silos.

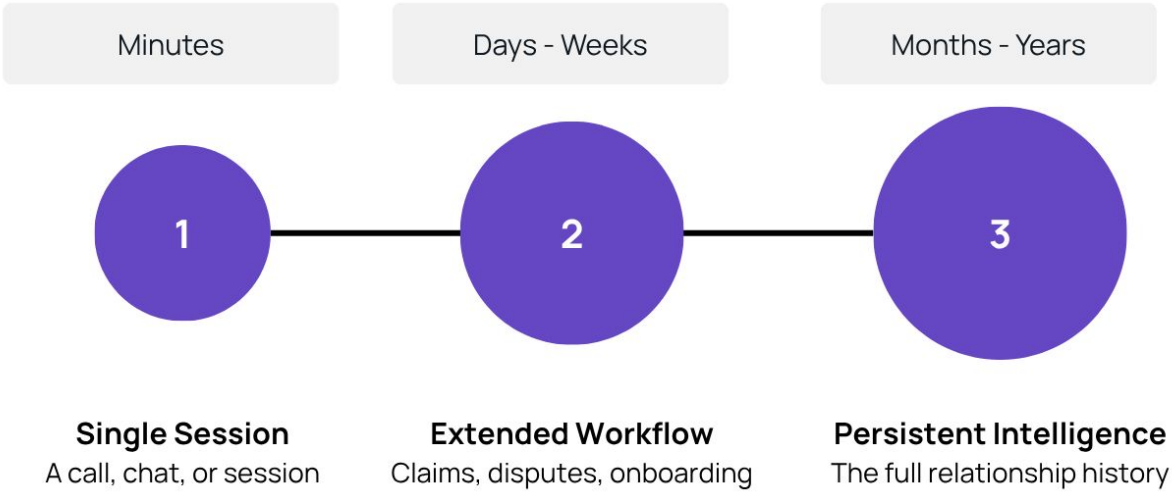
Making CX Data Available to the Enterprise. Context span is bidirectional. The data generated by customer service interactions—intent patterns, resolution outcomes, sentiment signals, structured conversation data—is enormously valuable to the rest of the enterprise. Marketing can use it to understand post-purchase friction. Product can use it to surface bugs and feature gaps. Sales can use it to identify at-risk customers or upsell moments. An enterprise that treats its CX data as a strategic asset—building an interaction intelligence repository that feeds dashboards, analytics models, and other enterprise AI systems—gains a compounding advantage over time.

Cross-Enterprise Context. At the frontier, agents operating in customer service may need to interact with agents in partner enterprises: airline systems to rebook flights, insurance networks to validate claims, logistics providers to track shipments. Emerging interoperability standards like Model Context Protocol (MCP) and Agent-to-Agent (A2A) protocols make this possible at scale—enabling enterprises to define governed, auditable interfaces through which their agents can exchange information and invoke actions across organizational boundaries.

Note that context has another type of span, across time. Transactional context differs from historical context (and there's an overlap with agent memory here). The definitions above encompass both the short-term (transactional) context and the long-term historical context.

Dimension 3: Lifecycle

The third dimension is time: how long does an agent operate, and what kind of continuity does it maintain?



Short-Lived Interactions. The canonical CX interaction is a bounded event: a phone call, a chat session, a digital conversation with a defined beginning, middle, and end.

Most deployed CX agents today operate at this granularity. Interaction-scoped memory, context, and state management are sufficient for the majority of use cases.

Extended Workflows. Some customer journeys span multiple interactions over days or weeks. A claim being processed. An installation being scheduled and completed. A dispute being investigated. An onboarding sequence being guided. For these cases, the agent needs to persist context across session boundaries—knowing what was promised, what has been completed, and what remains outstanding. Extended lifecycle agents operate more like case managers than conversation handlers, maintaining a thread of awareness that survives customer hang-ups, system restarts, and multi-day gaps.

Persistent Intelligence. Beyond individual workflows, the deepest form of lifecycle is long-term customer memory: an agent that knows not just this interaction but the entire history of the relationship. What the customer called about six months ago. What resolution they received. What product they use, and how they use it. This kind of persistent intelligence transforms the customer experience from a series of anonymous transactions to something that feels personal, contextual, and earned—the difference between a call center and a trusted advisor.

ASAPP CXP: Mapping the framework to reality

With the three-dimensional framework established, we can evaluate where **ASAPP's Customer Experience Platform** stands today and where it is heading.

GenerativeAgent® is ASAPP's primary customer-facing AI agent, designed from the ground up for enterprise-grade CX across voice and digital channels. It is the front door to the enterprise - the operational realization of the capabilities dimension described above.

On knowledge-based Q&A, GenerativeAgent retrieves from native knowledge bases, and KB integrations, synthesizing content in natural language rather than surfacing raw articles. Knowledge Guardrails allow builders to define explicit "do not answer" conditions for sensitive topics, ensuring the agent escalates rather than hallucinating in regulated domains.

Conversational Routing replaces IVR trees with intent-driven routing: the agent understands what the customer needs from natural speech or text, and routes directly to the appropriate workflow or human queue—no menus, no forced choices.

GenerativeAgent supports the full spectrum of automation depth, from simple FAQ responses to complex multi-step transactions involving API calls, conditional logic, and back-and-forth reasoning. Step-Based Workflows add structured control to this generative foundation—allowing builders to enforce required steps for compliance-sensitive flows while maintaining natural conversation. This blending of deterministic control and generative flexibility is one of the defining architectural choices that separates production-grade CX agents from demos.

HILA™ (Human-in-the-Loop Agent) is ASAPP's answer to the governance dimension of collaborative human involvement. Rather than treating human involvement as a failure mode, HILA frames it as a first-class workflow. GenerativeAgent can, mid-conversation, surface a request to a human advisor with full context, while maintaining the conversation with the customer as background work takes place. The customer never experiences a disruptive transfer.

On context span, ASAPP CXP currently integrates deeply within the CX domain—accessing conversation history, customer profile data from CRM connectors (Salesforce, HubSpot, Dynamics), knowledge systems, and enterprise data platforms (Snowflake, Databricks). MCP and A2A Support in the Connections component enables GenerativeAgent to connect to other agents, dramatically lowering the friction of extending context to enterprise systems that expose standard interfaces. CXP can reach across the enterprise and theoretically even across an ecosystem of enterprise partners to run complex workflows.

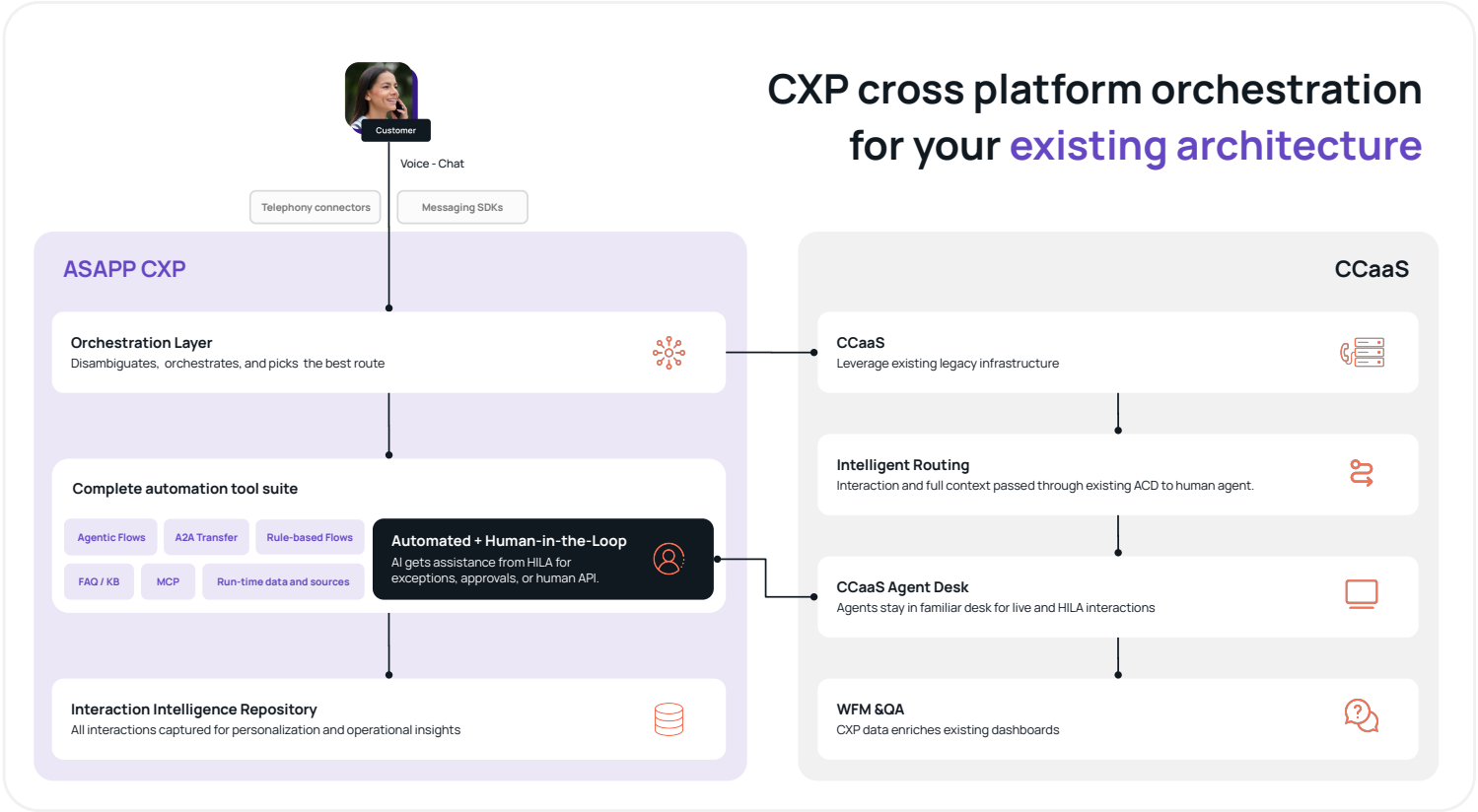
The Interaction Intelligence Repository is ASAPP's infrastructure for making CX data available to the broader enterprise. Every conversation is captured at high fidelity, structured with call summaries, sentiment signals, intent classifications, and custom-defined entities, and made available for consumption by BI tools, analytics platforms, and other enterprise AI systems. This transforms the contact center from a cost center into a data asset—generating the kind of behavioral intelligence that has historically been invisible to the enterprise.

On lifecycle, the current platform is primarily optimized for interaction-scoped, session-scoped operation and workflows spanning days to weeks. GenerativeAgent maintains full context within and across conversations, handles barge-in and interruption gracefully, and provides rich post-interaction summaries. Long-lived workflow capability - where the agent persists and autonomously manages long-running customer journeys that include a combination of inbound and outbound sessions - is an area of active development on the roadmap.

Reference architecture: CXP in the enterprise agentic ecosystem

Understanding ASAPP CXP in isolation is insufficient. The real value of the platform emerges when it is understood as a node in a broader enterprise agentic ecosystem.

CXP operates best as a single point of entry for the customer to interact with the enterprise, across multiple channels. It uses its automation capabilities to serve the customer, leveraging enterprise data sources, agents, and consulting with humans as needed.



How do you coordinate an ecosystem of agents to execute customer-facing workflows autonomously? The most typical topology we see is the “CXP as an Orchestrator” pattern. In this pattern, each agent within the enterprise manages a well-defined scope of operation within which it is capable of acting autonomously. These domains encompass traditional functional or departmental boundaries within the enterprise. Since many complex workflows that are initiated by customer interactions are cross-domain, CXP can act to work across these domains to serve the customer, across any channel. CXP would be the orchestrator, using tools it has access to, delegating work to other agents, and maintaining all the states needed for both real-time and long-running workflows with the customer.

With CXP as the orchestrator, where do you draw the boundaries between the responsibilities of the functional / departmental agents and the orchestrator? The answer is “it depends”.

The work of representing the brand, understanding the customer, and maintaining persistence belong in the orchestration layer: this would allow the orchestrator to anticipate what the customer may be calling or chatting about, and greeting them with that in mind using the brand voice. The orchestration layer also needs to plan out how to solve the problem, and which agents, tools, and resources to use - it needs to have enough context about the solution space to do that.

But a clear line is that the orchestrator isn’t connecting the systems or making the fine-grained decisions that go into fulfilling specific parts of the request, such as how to connect with the necessary backend services to file a new property damage claim, or when it is reasonable to provide a passenger with a delayed flight with a coupon or mileage credit as compensation. The policy minutiae and mechanics of performing actions within a functional domain lie within the specific domain of the sub-agents. The sub-agents themselves needn’t be capable of carrying on conversations with an end-customer or user using the brand voice; they would be internal experts within their domains.

It is also worth noting that an ‘orchestrator’ is typically not a single agent, but a family of agents that answer high-level questions, plan workflows, manage humans who may be involved in running parts of the workflow. Architecturally, this layer of agents can be thought of as supervisors, which perform all these tasks - and may themselves be organized based on the type of caller, or broad contact reason - in order to effectively serve the customer, plan, and leverage other agents in the system.

Toward the enterprise system of interaction

The framework presented here—Capabilities, Context Span, Lifecycle—is not merely a taxonomy of current technology. It is a map of the territory that enterprise AI in customer service needs to cover to deliver on its promise.

Today's deployments are largely concentrated in the lower-left corner of this space: narrow capabilities (FAQ, routing, handling simple intents), limited context (CX domain only), and short lifespans (single interactions). The value at that corner is real but bounded. The compounding value lies in expanding along all three dimensions simultaneously—toward agents that can handle the full spectrum of customer needs, that see and act across the enterprise, and that maintain awareness across the entire customer lifecycle.

ASAPP's CXP is designed with this expansion in mind. GenerativeAgent at the top of telephony and digital channels, a rich HILA workflow for governed human-AI collaboration, an Interaction Intelligence Repository that makes every conversation a data asset, and a growing library of enterprise integrations. Key to the system is the pervasive use of skilled AI agents that assist in each aspect of the lifecycle of creating and managing AI agents for the customer's business: an agent to discover opportunities for automation, an agent to develop agents, an agent to test and monitor agents, an agent to extract insights from use of the agents, and an agent to use these signals to optimize the agents. This creates a virtuous cycle of automation that drives outcomes: the feeling of having your own personal AI agent every time you interact with a brand.

The destination is what ASAPP calls **an enterprise system of interaction**—where every customer conversation drives execution, ties new intelligence to prior intelligence, and fuels continuous growth across the business. Enterprises that understand and act on this vision—investing in agents that grow in capability, deepen in context, and extend in lifecycle—will define the competitive standard for customer experience in the years ahead.

Learn more about ASAPP's CXP at asapp.com/cxp