

Omilia Voice Agents

Powered by the Omilia Self-Learning Agentic CX Platform: Voice AI built for enterprise contact centres that is accurate, self-improving, and deployable in minutes.

A customer-first voice agent service

High call volumes, agent attrition, and growing automation expectations are placing sustained pressure on contact centre operations. The teams tasked with modernising CX face a compounding problem: legacy IVR systems cap automation at a fraction of call volume, while rebuilding them requires months of services work, specialist expertise, and ongoing manual tuning.

Omilia Voice Agents resolve more than **90%** of customer interactions without human involvement.

They deploy in minutes via natural language instruction, no coded flows, no labelled training sets needed, and they improve continuously from every call, whether handled by AI or escalated to a human agent.

The self-learning loop is what separates Omilia from static IVR and first-generation AI deployments. Agents improve with every interaction, including escalations and corrections made by live agents, without manual retraining cycles or additional services engagement. They continuously discover new tasks that can be automated, expanding containment without additional project work.



Provide Engaging Interactions

Voice Agents understand caller needs and disambiguate unclear requests effortlessly, ensuring every customer feels heard and that their issues are resolved quickly. With out-of-the-box, industry-specific models pre-trained on billions of real customer interactions, Omilia Voice Agents deliver 96%+ understanding accuracy from day one.



Reduce Wait Times

Voice Agents can recognise multiple action items within a single call and logically sequence and execute each task. With intelligent self-service handling high volumes of routine calls efficiently, you can significantly reduce wait times and increase first-call resolution rates across the board.



Personalize Customer Engagement

Dynamic and fully context-aware, Voice Agents understand customer sentiment, past interaction history, and individual preferences, adapting their responses in real time as the conversation evolves. Every interaction is tailored to create a personalised and engaging experience that is as unique as the customer themselves.

Proven Performance at Scale:

- More than 90% of calls resolved without human agent involvement
- 97% of calls queued and routed correctly
- 50% reduction in Average Handling Time
- <1s End-to-end voice latency vs. 2–3s industry avg.
- <3% Word error rate vs. 5–7% industry avg.
- 5B+ Interactions handled per year on OCP

What used to take weeks or months to configure and deploy can now be achieved in minutes

Business Outcomes

With Omilia's platform, non-technical users can create Agents in minutes and go live within a day, and consistently deliver against the metrics that matter most to contact centre operations:

Increased automation rates and better customer experience

Reduced Average Handling Time (AHT) across all call types

Improved First-Contact Resolution (FCR) rates

Faster time to value for new use cases and business requirements

Lower total cost of ownership compared to multi-vendor technology stacks

Demonstrable return on investment achievable within weeks of deployment

Reduced agent workload by up to 100% for qualifying call types

Improved CSAT, NPS, and customer retention scores

Service That Meets Your Business Goals

Omilia Voice Agents are not a one-size-fits-all solution. They are tailored to your specific workflows, brand voice and existing backend systems.

The Omilia self-learning loop draws on your actual interaction data, whether it is handled entirely by AI or supported by a human agent. It looks at resolution rates, corrections made by live agents during escalations, and direct customer feedback to drive continuous improvement. Every interaction feeds back into the system, linking outcomes, agent behaviour, and customer signals to drive a continuous loop of improvement.

The result is a customer experience that becomes measurably more accurate, more personalised, and faster to resolve issues over time, all without adding operational overhead or requiring constant manual intervention.

What are Voice Agents?

Voice Agents are AI-powered conversational systems that handle inbound calls through advanced speech recognition and our reasoning engine that delivers over **95%** understanding accuracy with no pre-built intent trees or dialog flows required.

Most AI solutions stall over time because they rely on static models, siloed tools, or manual tuning by expert teams. Omilia's native platform is fully integrated so it combines task agents, continuous self-learning, and automated optimization, so performance improves every single day without additional effort from your team.



Concierge Agents

Most routing failures happen before the conversation begins. Traditional intent-based systems depend on pre-trained, manually labelled intents. When a caller phrases something unexpectedly, mixes two requests, or changes their mind mid-call, the system breaks.

The Concierge Agent replaces that layer entirely. It uses real-time reasoning and contextual awareness to interpret caller intent and route to the right specialist Agent with zero-shot routing - no static queues or keyword matching. Nothing to label, nothing to retrain, no flows to author for new use cases.

The result is fewer misroutes, lower transfer rates, higher first-contact resolution, and a routing layer that adapts the day a new product launches.



Task Agents

Task Agents execute specific, multi-step workflows to completion - collecting and validating data, calling APIs, invoking backend systems, making decisions, and confirming outcomes without human handoff.

Key Features:

Three Autonomy Tiers One Architecture

Choose the level of AI reasoning that fits each task: deterministic flows for compliance-heavy work, agentic fallback to recover when flows fail, or fully autonomous execution for complex multi-system workflows.

Modernize incrementally, maintaining existing investments

Add Task Agents directly into existing applications without rebuilding what already works. Agentic Fallback augments deterministic flows where they fail today. Adding modernised AI capabilities builds upon existing infrastructure and investments, without replacing them.

Open-Standard Integration via MCP and REST

Connect to any enterprise system through Model Context Protocol or via REST APIs to CRMs, billing, scheduling, ticketing, and custom backends. No proprietary connectors required.

Built-In Sensitive Data Protection

PII, payment data, and credentials are masked and validated by default. Output fields can be marked Secret for automatic masking and Required for validation - no custom redaction logic needed.

Orchestrated or Standalone Deployment

Deploy as orchestrated agents under a Concierge for multi-service phone numbers, or as standalone Autonomous Task Agents for single-purpose entry points (password reset, balance check, claim status). Same agent, two deployment models, fits any contact-center architecture.

Advantages of Task Agents

1. **Dynamic planning and execution at runtime.** The agent decides the path during the conversation rather than following a pre-built reactive workflow.
2. **Faster time to value for complex use cases.** Skip the build cost of a deterministic flow when the behavior space is wide or poorly defined.
3. **Native integration with enterprise systems via MCP and WebService tools.** Built for cross-system orchestration (CRM, billing, ticketing).
4. **Reasoning over extracted data unlocks new use case classes.** Behaviors like “where did I spend all my money last month” – which require synthesis, not just lookup – become feasible.
5. **Fits MCP-first customers.** For organizations that already expose data via MCP, a High Autonomy Task Agent slots in with ease.



Efficiency and value at the speed of voice

As Voice Agents learn from every interaction across your contact centre, you get the combined benefit of AI precision and human agent expertise. You can deploy, improve, and scale AI Agents in minutes, and with human oversight built into the process, you retain full control over what is automated and when it goes live. Voice Agents continuously discover new resolution patterns, improve automation performance, and reduce operational cost, without touching the live interaction path.



Higher Task Resolution

Customers can speak to Voice Agents naturally, using their own words and phrasing. Even if a customer switches topics mid-conversation or interrupts the flow of dialogue, the agent understands the intent behind what is being said and successfully resolves the call. This frees your human agents to focus their time and expertise on more complex, nuanced tasks where human judgement genuinely adds value.



Infinite Scale

Omilia Voice Agents are architected to scale without limits. Unlike many competing solutions that degrade in performance as call volumes increase, Omilia’s platform is designed to handle thousands of concurrent calls simultaneously, maintaining consistent quality and response times regardless of demand spikes.



One Fully native stack

Unlike platforms assembled from third-party ASR, LLM, and TTS APIs, OCP is a vertically integrated stack. Omilia's proprietary models are fine-tuned for enterprise voice – not adapted from general-purpose frontier models. This means sub-second latency that external API chains cannot achieve, and a certified security perimeter that zero-dependency architecture uniquely enables.



Unit Economics & Cost Control

Full stack ownership eliminates the token burn problem. One price per resolved interaction, regardless of agentic steps taken. Self-learning codifies patterns offline, reducing LLM invocations in runtime over time.

- Outcome-based pricing: pay per resolved interaction, not consumed compute
- No third-party LLM dependencies – no token pricing volatility exposure
- Budget certainty: volume scales proportionally, not exponentially
- Self-learning reduces runtime LLM calls as patterns are codified deterministically



Security, compliance, and sovereignty

- Walled-garden architecture with zero third-party data sharing
- Certified PCI DSS Level 1, SOC 2 Type II, ISO 27001, HIPAA, GDPR, FedRAMP-Ready,
- Cyber Essentials, Responsible AI
- Deployment choice – fully managed SaaS or on-premise for Tier 1 financial / government
- data-residency mandates
- No utterance data sent to third-party ASR, LLM, or TTS providers



Deployment & time-to-value

- 340+ pre-built NLU intents for financial services
- 95%+ Day-1 intent accuracy via AI Bootstrapping from existing transcripts, recordings, SOPs, and API specs
- Pre-production quality validation and synthetic QA before any change reaches production
- 23+ years building for regulated industries

Advantages of Omilia Agentic Voice Agents

	Legacy Approach	Omilia Agentic Voice
Deployment time	Weeks to months With intent design, training sets and flow diagrams required before go-live, deployment is longer	Deploy and go live on day one Deploy in minutes by instructing the platform in natural language to create an agent, no coding required.
Ease of build	Intents, flows, training data Hand-built intents, dialog trees and labelled training sets per use case.	Natural-language instruction Describe what the agent should do; ingest existing knowledge sources and APIs in seconds.
Learning model	Manual retraining cycles Periodic tuning and expensive services to maintain accuracy as use cases evolve.	Continuous self-learning Learns from every interaction (AI and human) and improves itself safely under governance.
Understanding & reasoning	Rigid NLU on determined paths Fixed dialogue flows can struggle on edge cases, ambiguity or mid-call topic switches.	95%+ understanding accuracy Free-flowing conversation with topic switching and human-like comprehension out of the box.
Performance improvement over time	Containment hits a ceiling Static logic caps automation; further gains need more services, long projects, more investment.	Compounds every day Breaks the automation ceiling – 90%+ task completion that improves with every call.

About Omilia

Omilia is the global standard for AI-driven customer service transformation. Our native Self-Learning Agentic CX platform revolutionizes how enterprises engage with customers, automating interactions with precision, empowering agents in real time, and delivering seamless, personalized experiences across all channels.

Powered by deep expertise in developing proprietary Agentic AI technology, and multi-layered anti-fraud capabilities, we enable enterprises to move decisively and safely into the era of AI-first contact centers. Omilia's Self-Learning Agentic CX learns from across the entire customer journey – from self-service to live agent interactions – unlocking continuous improvement and breaking the "glass ceiling" of containment that legacy siloed models can't achieve.

Omilia is trusted by the world's most demanding enterprises across all industries. Built on over two decades of AI innovation, Omilia delivers measurable outcomes: lower costs, higher efficiency, and unmatched customer satisfaction – all while preserving the human touch where it matters most.