boost·ai

The enterprise guide to Voice AI:

Transforming the contact center experience

Table of contents

3 Introduction

4 What is Voice AI?

5 The business case for Voice AI

6 How Voice AI works

9 The four levels of voice automation

13 Voice AI is not just “Chat read aloud”

15 A proven roadmap to Voice AI success

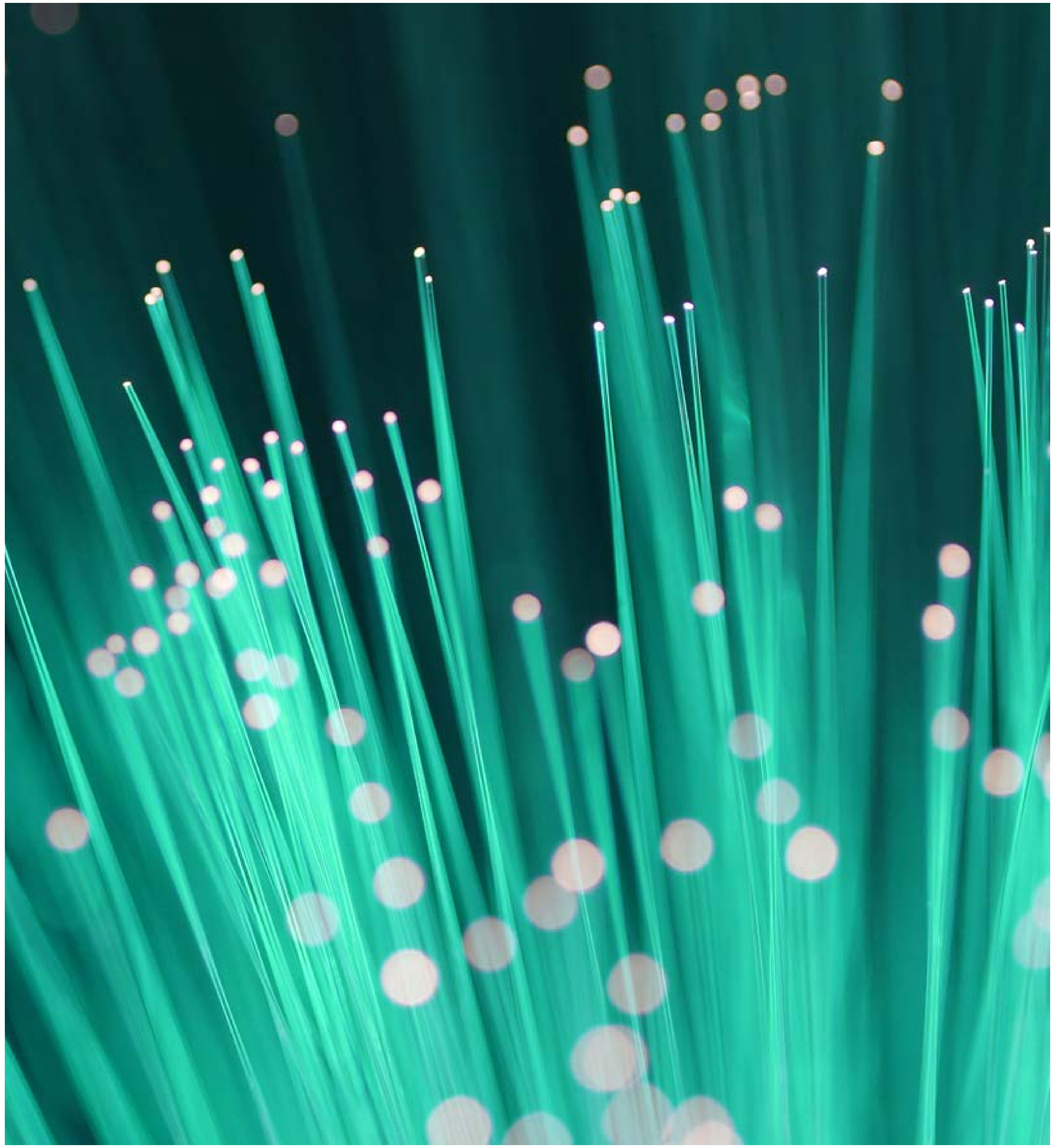
16 The future of voice is already here

Introduction

For modern enterprises, the financial downsides of customer hold times are a far greater expense than the cost of a new high-tech solution. When a customer picks up the phone, they are at their most vulnerable or most frustrated, and are seeking a human connection, likely having exhausted self-service options.

For decades, the legacy trap has defined this critical moment. Outstanding customer experience does not start with “Press 1 for mortgages” or “Press 2 for bank cards.” Traditional Interactive Voice Response (IVR) systems have become a primary source of friction, leading to dropped calls, degraded brand sentiment and a phenomenon where customers keypad mash or shout “agent” just to bypass a rigid system.

The consequences are clear: customers are no longer lost to competitors solely on price; they are lost to hold music and “we’re closed” messages. Voice remains the most heavily utilized and costly support channel in the world. However, a paradigm shift is underway. By modernizing the legacy stack with Voice AI, businesses can move beyond rigid, linear menus to create fluid, conversational experiences that scale, effectively transforming the contact center from a cost-heavy bottleneck into a hub of frictionless, 24/7 service.



What is Voice AI?

Voice AI is an advanced suite of technologies that allows automated systems to understand, process and respond to spoken language in real-time. Unlike traditional IVRs, which rely on DTMF (Dual-Tone Multi-Frequency) keypad inputs or simple keyword recognition, Voice AI interacts with users using the same natural language patterns humans use with one another. It represents the digital front door of the modern enterprise, capable of greeting every customer immediately and intelligently.

Beyond basic transcription

It is a common misconception that Voice AI is merely speech-to-text with a bot attached. While transcription is a fundamental component, the true enterprise value lies in Natural Language Understanding (NLU) and Generative AI. The technology can interpret nuance, recognize intent amidst rambility or fragmented sentences, and extract relevant entities (like complex account numbers or dates). This allows the system to not just listen, but cognitively process the request and execute the correct business logic.

Natural interaction

The fundamental shift is the removal of predefined, forced pathways. In a legacy system, the user is an operator navigating a maze, whereas with Voice AI, they are essentially one side of a conversation. Users can explain their needs using their own vocabulary, regional accents and unique speech patterns. This technology bridges the gap between machine efficiency and human-like understanding, allowing the system to adapt to the customer rather than forcing the customer to learn the system's rigid hierarchy.

The business case for Voice AI

The global contact center market represents approximately 17 million seats. Because voice is inherently more labor-intensive and difficult to manage than digital channels, it carries the highest operational cost per interaction. Shifting even a small fraction of this traffic to automated solutions represents more than just a marginal gain; it represents a $102 billion transformation opportunity for the enterprise to reallocate resources towards innovation and high-value growth.

The future outlook

The trajectory is untenable. Market analysts like Gartner predict that by 2028, 70% of customer service journeys will start and end with voice-based assistants. Organizations that fail to adopt these conversational layers risk being left behind in an era where instant is the expected standard of service.

Significant cost reductions: The math of voice automation is compelling. Automating just 1% of voice traffic can yield up to 19% more cost savings than shifting the same volume of email or chat traffic. This is due to the high baseline costs of telephony infrastructure and human labor required for voice.

Operational efficiency & FCR: Voice AI increases First Call Resolution (FCR) by ensuring calls are either resolved instantly or routed to the perfect specialist. This drastically reduces average handling times (AHT) and eliminates the “waiting room” experience.

Empowering human agents: Contact center attrition is a chronic industry challenge often driven by agent burnout. By allowing AI to handle high-volume, repetitive tasks – like password resets or status checks – live agents are freed to focus on complex, high-value and empathetic interactions where human judgement is preferred.

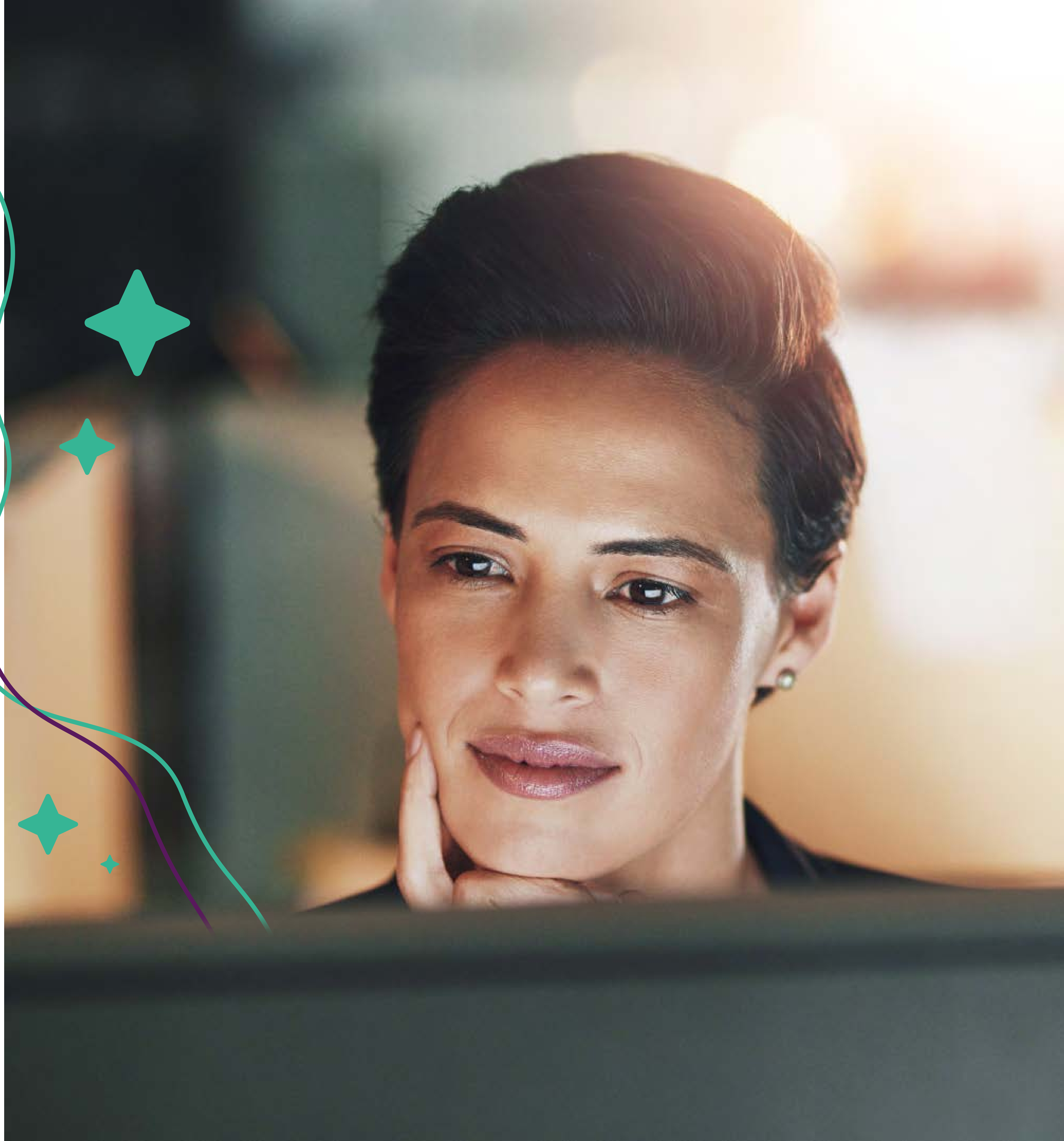
Always-on support: Voice AI provides accurate, 24/7 self-service. For global organizations, this means maintaining a consistent service level across time zones without the massive overhead of overnight staffing.

How Voice AI works

To demystify Voice AI, it is helpful to view it as a bridge between traditional telephony and digital intelligence. Modern enterprises typically choose between two architectural approaches, depending on their need for either speed or control.

The voice gateway

At the center of every setup is the Voice Gateway. This is the entry point where calls arrive. It connects the voice channel (telephony) to the AI platform and orchestrates the exchange of audio. Gateways connect to the outside world via SIP (Session Initiation Protocol) or WebRTC, allowing the AI to integrate directly with existing contact center infrastructure (CCaaS) or traditional phone networks (PSTN).



Architecture A:

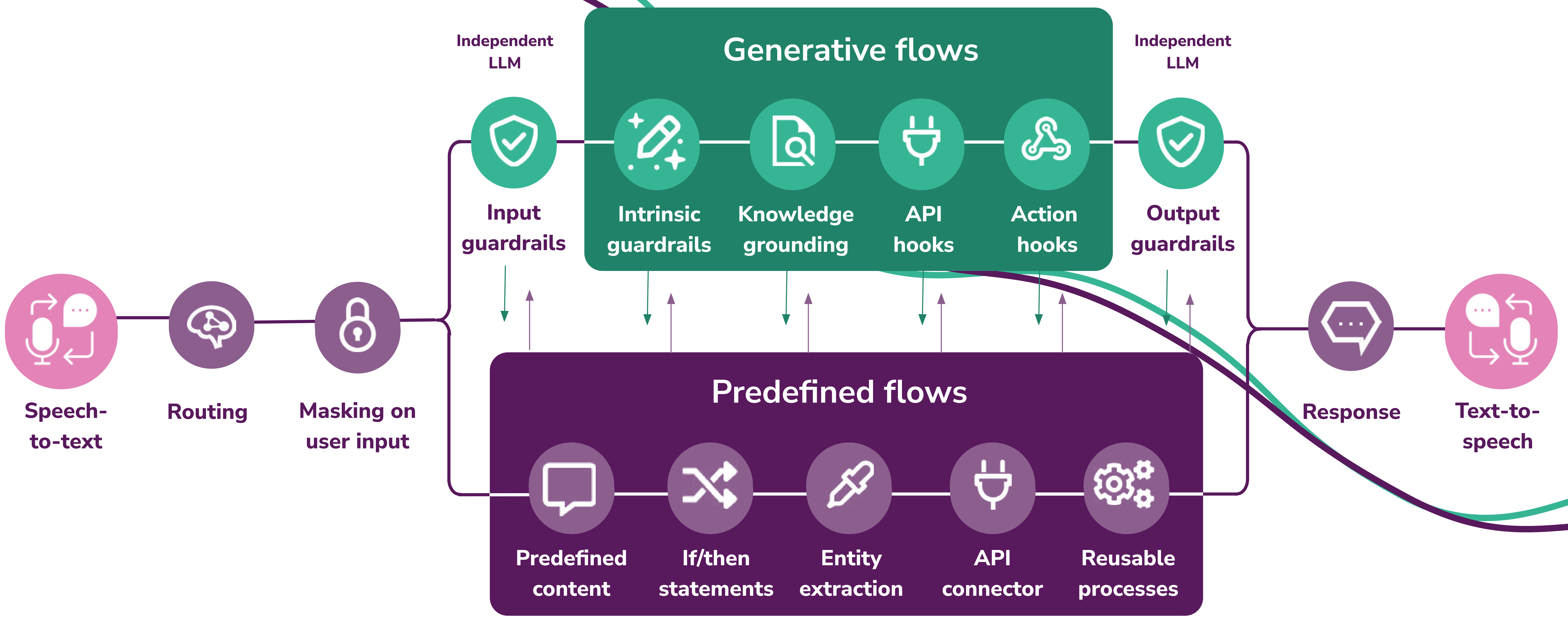
Multi-component voice pipeline

This is the traditional enterprise approach, favored by regulated industries for its modularity and high level of control. It breaks the process into distinct steps:

1 Speech-to-Text (STT): Transcribes the caller's audio into text.

2 An LLM handles routing, intelligently triggering Generative Flows (grounded in company knowledge) or Predefined Flows (for structured logic) as needed. An advantage of this approach is it allows for increased security, ensuring that sensitive data is filtered before it ever reaches a processing model.

3 Text-to-Speech (TTS): Converts the AI's textual response back into audio for the callers.



Architecture B: Speech-to-Speech

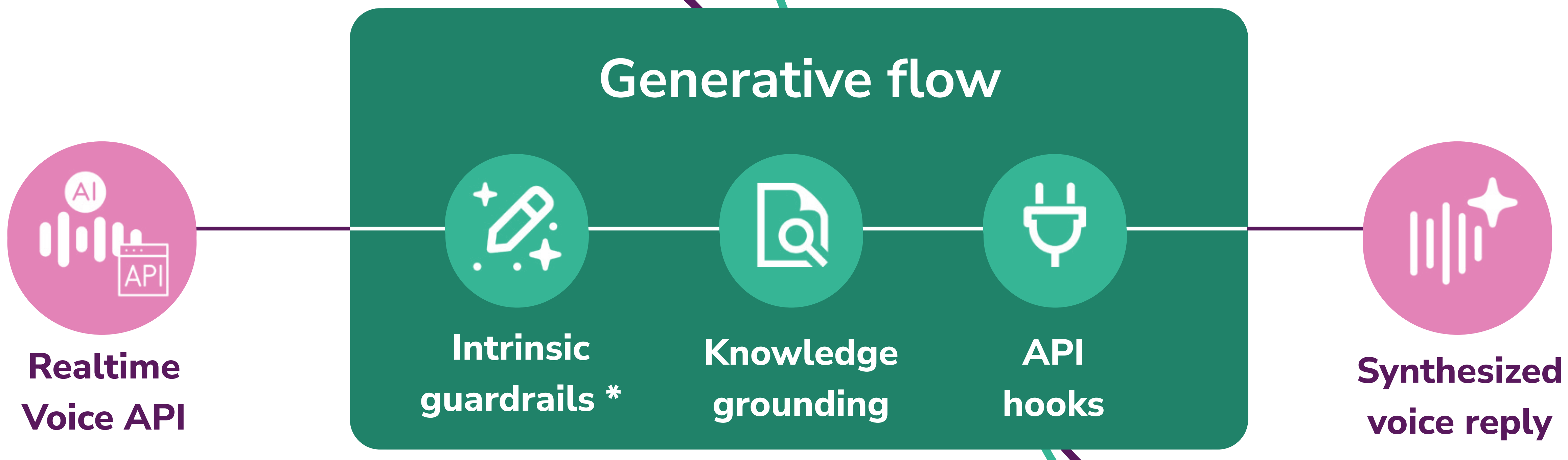
The cutting edge of the industry is moving toward a more direct model. In this setup, audio is sent directly to a multimodal Large Language Model (LLM) via a real-time API.

The advantage: By removing the middleman of transcription, latency is significantly reduced, resulting in a more fluid, human-like conversation where the AI can respond almost instantly.

The trade-off: While faster, this approach is currently best suited for lower-risk use cases as it offers fewer granular guardrails than the multi-component pipeline.

Seamless handoffs

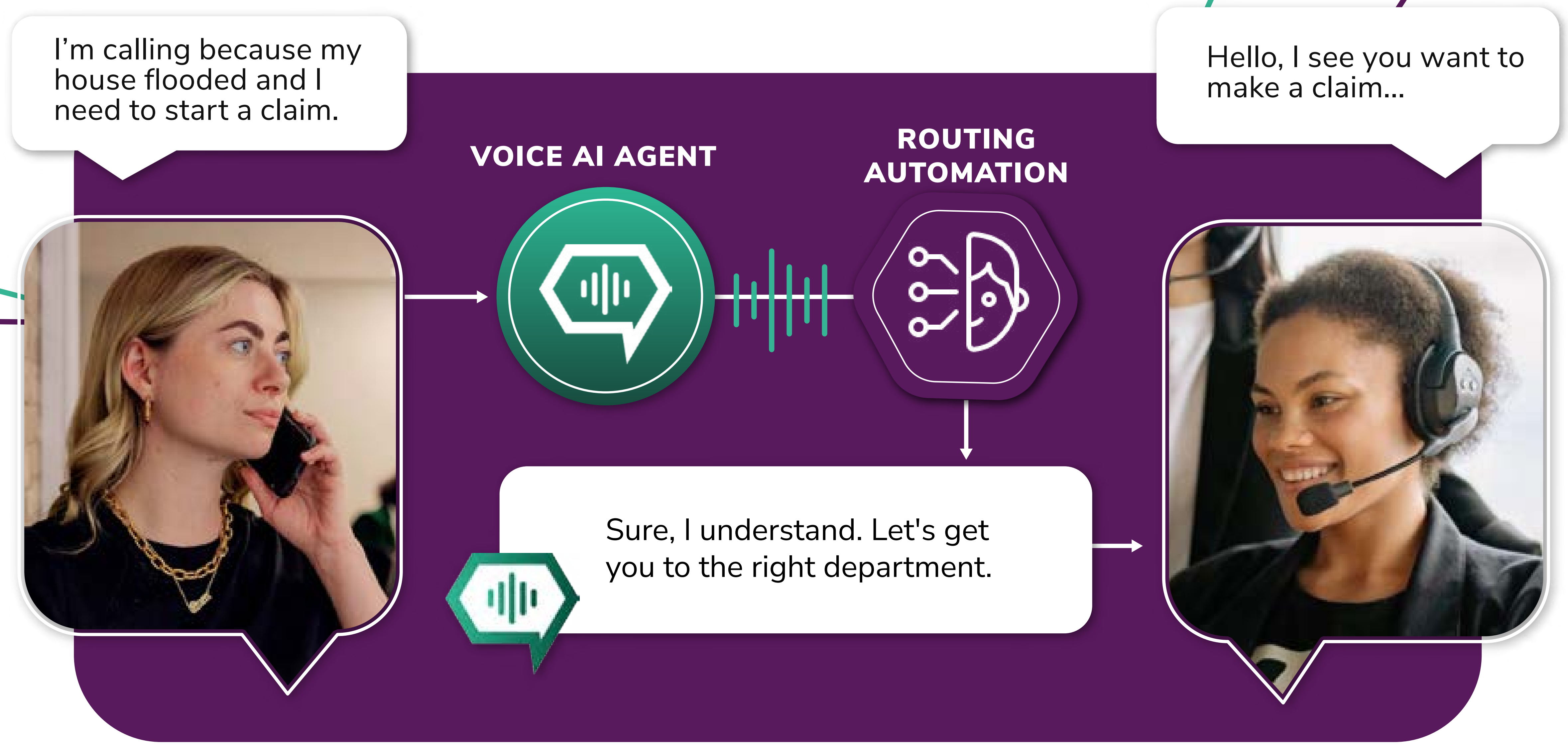
Regardless of the internal pipeline, a key technical requirement is the ability to hand off a call to a human agent when complexity arises. Using protocols like SIP REFER, the Voice AI can transfer the call to a live agent's queue along with a full conversation transcript. This ensures the human agent has the complete context before they even say "hello."



Combining architectures

Until recently, enterprises were forced to choose between the speed of Speech-to-Speech models and the security of traditional voice pipelines. The next evolution in the field is an adaptive voice architecture that removes this compromise.

In an adaptive model, the system intelligently applies the appropriate approach in real-time. It enables fast, free-flowing interactions when simplicity allows, but seamlessly switches to reinforced, structured control when the conversation shifts to sensitive data, complex compliance or high-precision regulation. This ensures the right experience for every moment in the customer journey.


I'm calling because my house flooded and I need to start a claim.
VOICE AI AGENT
ROUTING AUTOMATION
Hello, I see you want to make a claim...
Sure, I understand. Let's get you to the right department.

The four levels of voice automation

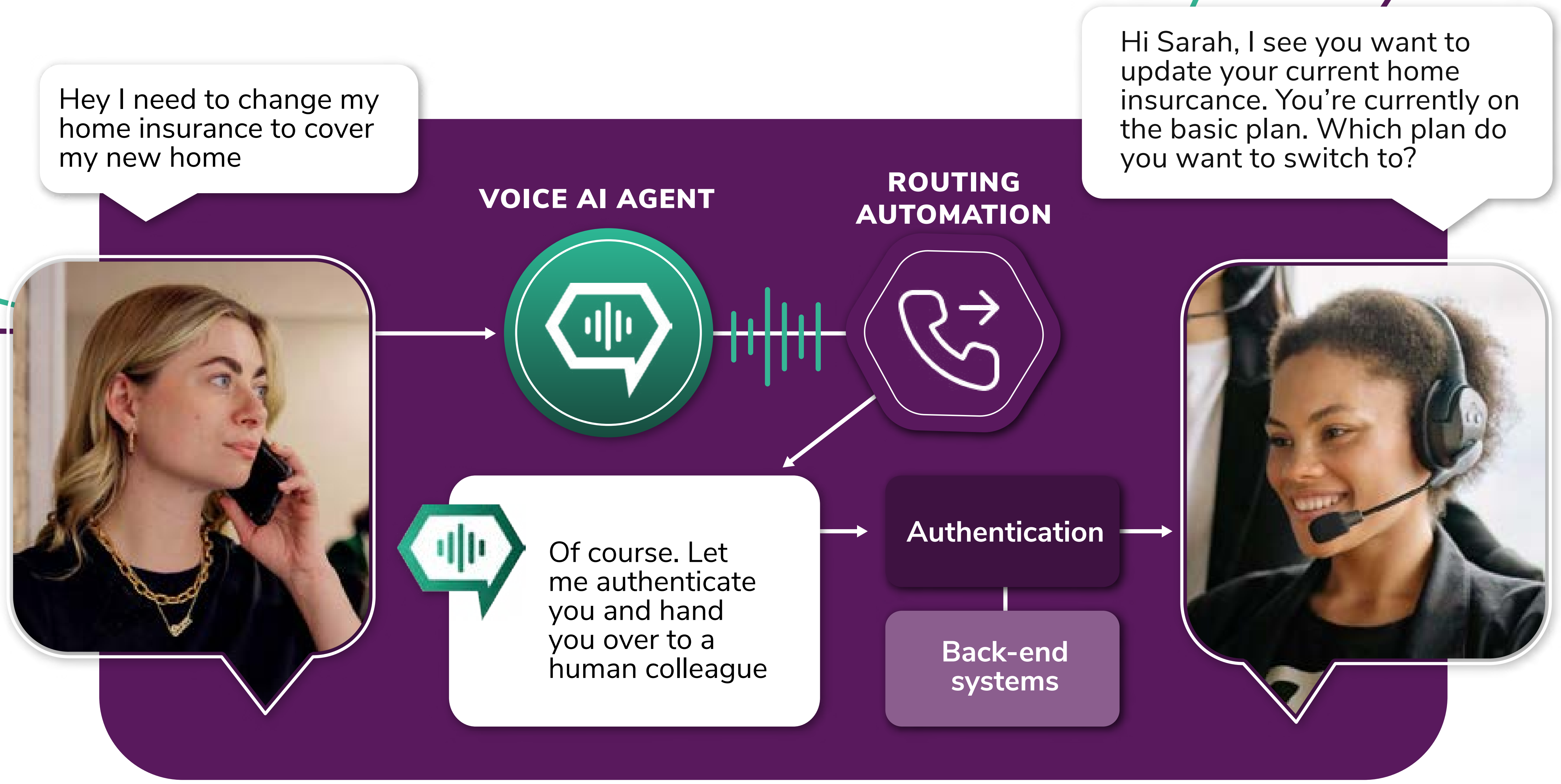
Understanding how Voice AI functions technically allows organizations to strategically deploy it across different tiers of customer experience. From initial triage to complex task resolution, Voice AI typically operates across four distinct levels of automation.

1. Routing automation

The first level of automation involves replacing the "Press 1" menu with a simple question: "How can I help you today?" Callers state their issue in natural language, and the AI identifies the correct department. This drastically reduces mis-routing – a common issue where agents spend their time manually transferring callers who chose the wrong menu option.

EXAMPLE

A customer says, "I'm calling because my house flooded and I need to start a claim." The AI recognizes the urgency and intent, bypasses the general queue and routes them directly to the Emergency Claims team.

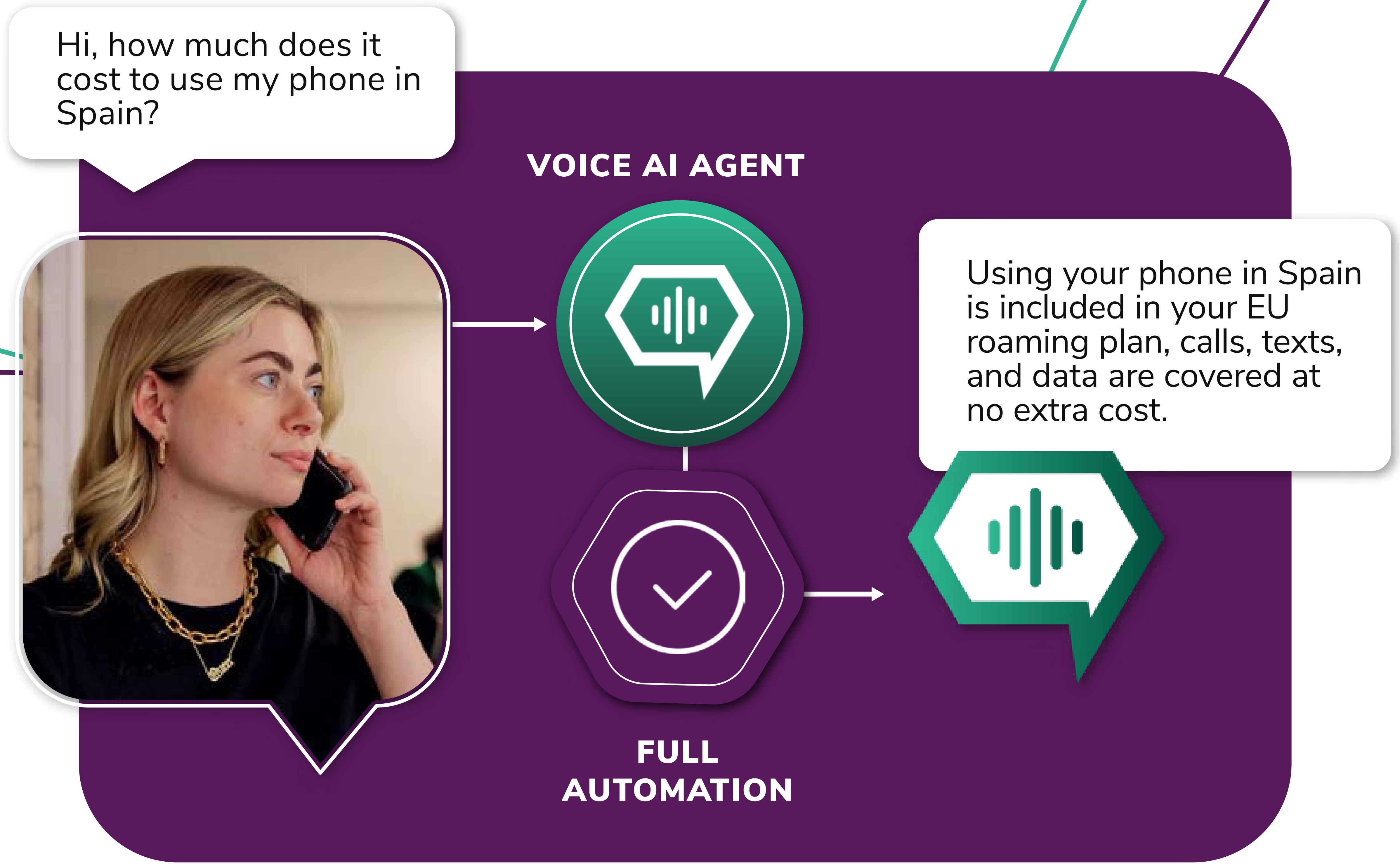

Hey I need to change my home insurance to cover my new home
VOICE AI AGENT
ROUTING AUTOMATION
Hi Sarah, I see you want to update your current home insurcance. You're currently on the basic plan. Which plan do you want to switch to?
Of course. Let me authenticate you and hand you over to a human colleague
Authentication
Back-end systems

2. Partial automation

Partial automation focuses on offloading routine, repetitive steps before a seamless handoff to a human colleague. The AI handles the "prep work"—such as identity checks, account lookups, or providing answers to basic FAQs—effectively shortening call duration and increasing the number of calls agents can handle per hour. This frees human staff to focus on the high-value, complex conversations where their empathy and judgment matter most.

EXAMPLE

"I need to change my home insurance to cover my new home." After capturing this, the AI authenticates the user and hands the call to an agent with the context already



Hi, how much does it cost to use my phone in Spain?

VOICE AI AGENT

Using your phone in Spain is included in your EU roaming plan, calls, texts, and data are covered at no extra cost.

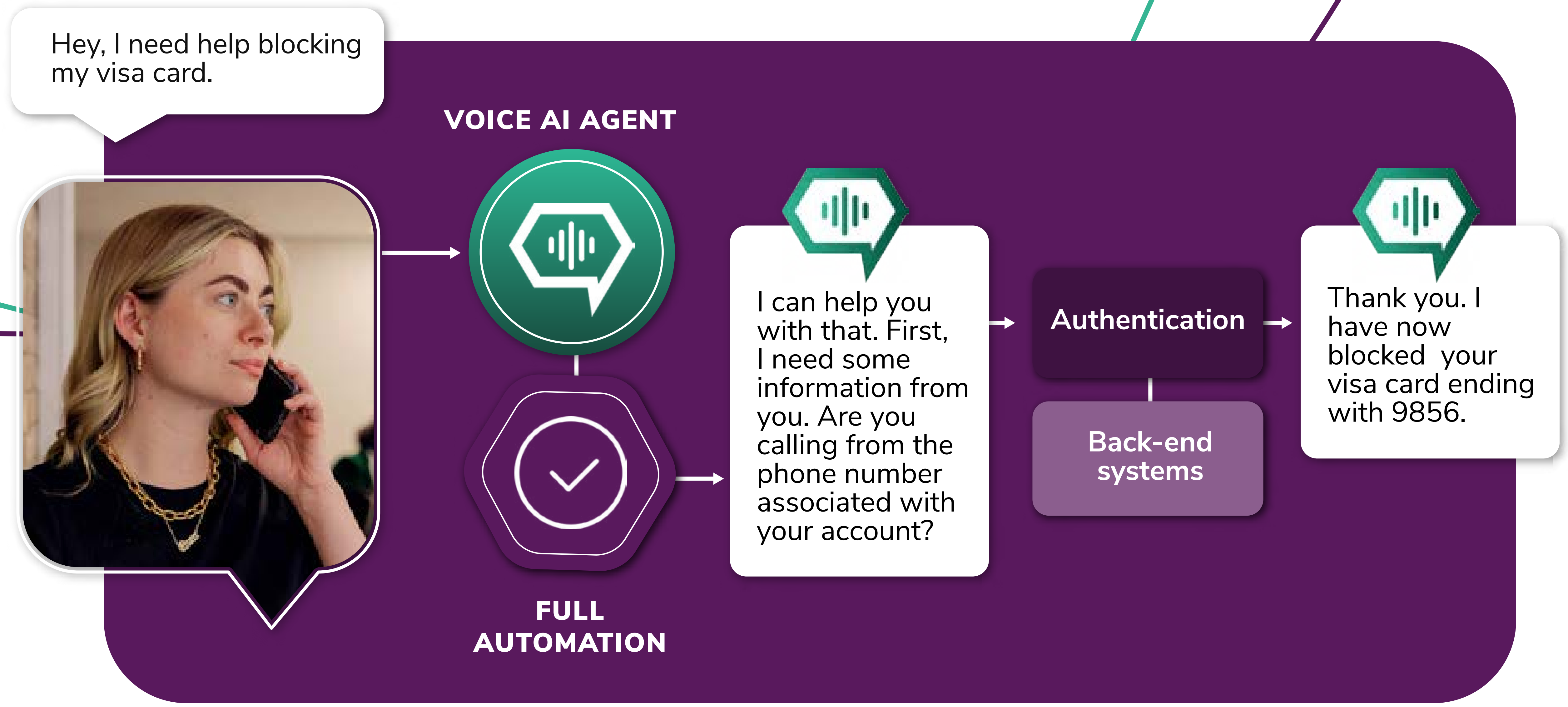
FULL AUTOMATION

3. Full automation: informational

At this level, the AI provides instant resolution for common questions with no human agent required. This is essential for deflecting high-volume inquiries like opening hours, service coverage, or login issues. By providing accurate self-service at scale, organizations can extend their support capabilities beyond traditional business hours to offer 24/7 availability.

EXAMPLE

A customer asks about roaming costs while traveling. The AI instantly replies: “Using your phone in Spain is included in your EU roaming plan. Calls, texts, and data are covered at no extra cost.”



Hey, I need help blocking my visa card.

VOICE AI AGENT

I can help you with that. First, I need some information from you. Are you calling from the phone number associated with your account?

Authentication

Back-end systems

Thank you. I have now blocked your visa card ending with 9856.

FULL AUTOMATION

4. Full automation: transactional

The highest level of maturity is secure, transactional automation at scale. By integrating directly with back-end systems, the AI can fetch customer information and execute actions autonomously. This offers full resolution on the first contact without human involvement, handling high-value tasks like account changes, policy updates, bookings or SIM activations.

EXAMPLE

A caller needs to block a lost visa card. The AI verifies the phone number, fetches the account, confirms the specific card, and completes the block instantly: All without involving a human agent in the interaction.

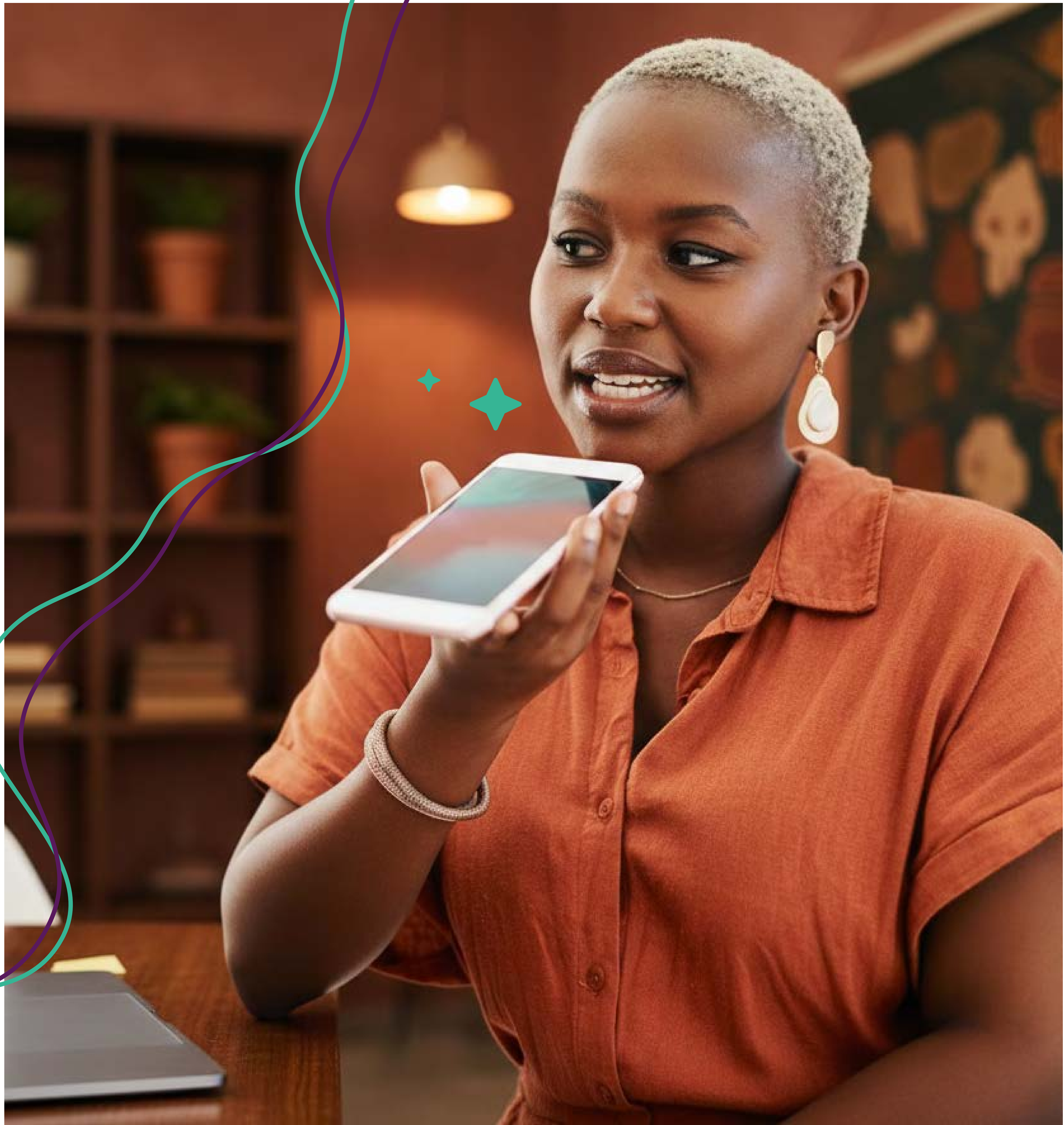
Voice AI is not just “Chat read aloud”

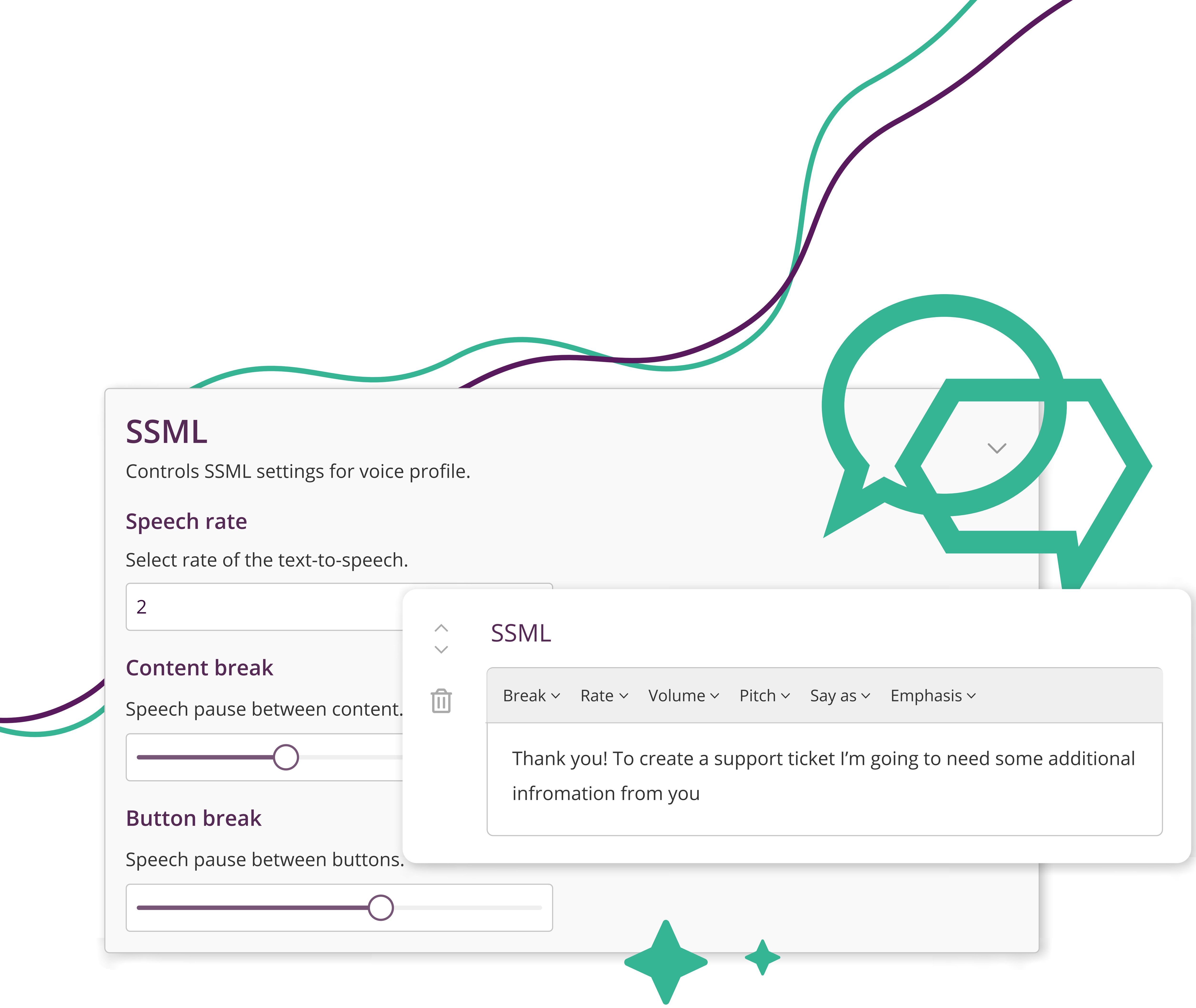
Successful Voice AI requires a dedicated conversational design philosophy. A script that works in a chat window will often fail over the phone because the user interface is purely auditory.

The latency challenge

In digital chat, a five-second wait for a response is acceptable. On the phone, three seconds of dead air feels like a technical failure. Industry research indicates that transparency is the most effective cure for wait-time frustration. Studies show that a significant majority of consumers are willing to tolerate processing delays as long as the system actively communicates the status and reason for the wait.

Best practice: Certain systems can incorporate filler phrases or background/keyboard noises to mask the time the AI takes to query a database.





SSML

Controls SSML settings for voice profile.

Speech rate

Select rate of the text-to-speech.

2

Content break

Speech pause between content.

Button break

Speech pause between buttons.

SSML

Thank you! To create a support ticket I'm going to need some additional infromation from you

Designing for the ear

Brevity and clarity. Humans have limited auditory working memory. Voice AI responses must be significantly shorter than text responses. Information should be front-loaded so the most important details come first.

SSML (Speech Synthesis Markup Language). Designers use SSML to adjust pacing, volume and emphasis in AI responses. This ensures that the AI doesn't stumble over brand names, correctly reads phone numbers and uses appropriate pauses for natural flow.

Handling interruptions and background noise. Real conversations involve interruptions. A sophisticated Voice AI must support "Barge-in", allowing the user to speak over the Voice AI Agent to provide an answer or change the subject. It also requires repeat prompts that can rephrase information if the caller sounds confused or if there is background noise.

Safety and guardrails

In regulated industries like finance or healthcare, compliance is the highest priority. Enterprise-grade Voice AI must include PII (Personally Identifiable Information) masking to ensure sensitive data is stripped from logs. The system must also employ strict output guardrails to prevent hallucinations – ensuring the AI only provides information based on verified data.

A proven roadmap to Voice AI success

Voice AI implementation is a strategic journey. At boost.ai, we outline a phased approach that allows the organizations we work with to learn, iterate and prove value without disrupting core operations.

Phase 1:Pilot

1

Launch a “Smart Routing” pilot on a single, low-risk entry point. Focus on identifying the top 10 reasons people call and automate routing for those interactions.

Impact:

- 10%+ of traffic automated
- Easy validation & buy-in
- Fast time to value.

Phase 2: Implement

2

Expand the routing to 100% of inbound traffic and introduce “Informational automation.” At this stage, the AI handles high-frequency, low-complexity FAQs.

Impact:

- 20%+ of traffic automated
- 90%+ routing accuracy
- Reduced wait times.

Phase 3: Optimize

3

Introduce authentication and transactional automation. By integrating with the CRM, the AI begins resolving tasks end-to-end. This is where “Self-service rates” become the primary KPI.

Impact:

- 40%+ of traffic automated
- 90%+ routing accuracy
- CSAT improvement.

Phase 4: Transform

4

The voice channel becomes a fully integrated, proactive service hub. Organizations can automate complex customer journeys end-to-end, deploy proactive use cases and deliver personalized, real-time service.

Impact:

- 60%+ of traffic fully automated
- 90%+ routing accuracy
- Further CSAT improvement.

The future of voice is already here

Customers are already choosing speed and clarity over the friction of legacy systems. With the majority of customer journeys expected to involve AI voice agents by 2028, the transition from rigid menus to fluid conversations is no longer a luxury for the few. By starting now, even with a small and focused pilot, you ensure your organization is ready for this shift before hold times become a permanent barrier to scaling your business. The technology to bridge the gap between machine efficiency and human empathy exists today. While the market continues to evolve, the cost of waiting is often higher than the cost of beginning the journey. At boost.ai, we are ready to help you take that first step toward a more conversational future.

Try out our voice experience at [voice.boost.ai.](voice.boost.ai)