



WHITEPAPER

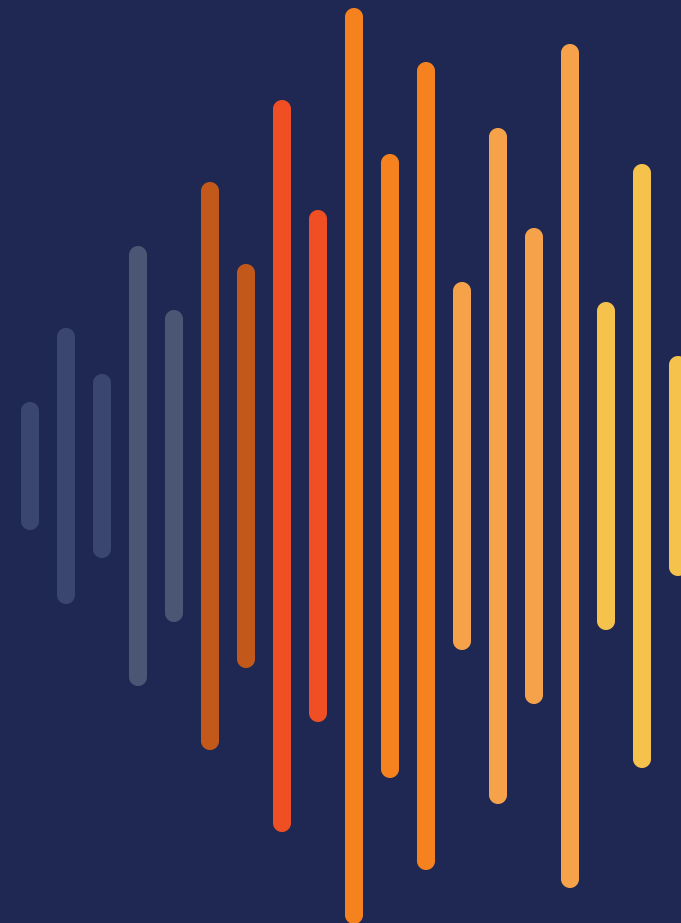
The AI Voice Agent Buyer's Guide

A practical framework for contact center managers and IT and UC leaders who are weighing an AI voice agent.

EVALUATE

COMPARE

DEPLOY



What is in this guide

01 Why teams are evaluating AI voice agents now

02 How an AI voice agent differs from your IVR

03 Where a voice agent earns its place

04 The evaluation checklist

+ Take it with you: the one-page checklist

05 Cost models and how to compare them

06 Containment versus resolution

07 Deployment realities and the first 90 days

08 How Variphy approaches this with Colleen

+ About Variphy & sources

You are not deciding whether software can answer a phone.

Contact center and IT leaders are looking at AI voice agents for reasons that have little to do with the technology itself. The work is getting harder to staff, call volume is unpredictable, and a large share of inbound calls ask the same handful of questions.

The real question is whether an AI voice agent fits your environment, your callers, and your tolerance for risk. This guide covers what these tools actually are, where to put one first, what to ask vendors, how to compare quotes, what to measure, and how to roll it out.

ANATOMY OF ONE AUTOMATED CALL				
RING	INTENT	WORK THE REQUEST	RESOLVE OR HAND OFF	TRANSCRIPT & SUMMARY
Does the agent front every call or take overflow?	In which languages, and what if the caller switches?	Which systems and directories can it read and write?	Right queue, with the context attached?	Does the data reach your reporting and prove resolution?

Why teams are evaluating AI voice agents now

Two pressures sit behind almost every evaluation: the work is hard to staff, and too much of the queue is calls that never needed a person.

THE STAFFING PICTURE

If you run a contact center or a help desk, the staffing math is familiar. Seats are hard to fill and harder to keep, and each agent who leaves costs weeks of hiring and training before the replacement is up to speed.

The Bureau of Labor Statistics expects employers to fill about 340,000 customer service openings a year through 2034, and nearly all of them go to backfilling people who leave the role rather than to growth.

The routine, repeatable part of your volume is the part a voice agent is built to take, which is why more teams are past the question of whether to automate and are working out which calls to hand over first.

THE CALL QUALITY PICTURE

You also know that not every call needs a person. A large part of the queue is the same short list of routine questions, alongside spam and nuisance calls that never had a reason to ring. That noise is growing. Spam and telemarketing volume rose about 20 percent in a single year, and the FTC logged more than 2.6 million Do Not Call complaints in fiscal 2025. Every one of those that reaches your queue is time an agent spends away from a caller who needs help.

341,700

Customer service openings to fill each year through 2034

Nearly all replacing people who leave

U.S. Bureau of Labor Statistics, Occupational Outlook Handbook: Customer Service Representatives, 2025.

2.6M

Do Not Call complaints logged by the FTC in FY2025

Nuisance traffic that still consumes queue time

Federal Trade Commission, Do Not Call Registry Data Book, FY2025.

MONTHLY AVERAGE SCAM & TELEMARKETING CALLS, U.S.



+20% in a single year, through September 2025 **57%** of robocalls classed as scam or telemarketing

U.S. PIRG Education Fund, Ringing in Our Fears 2025; call volume data via YouMail.

How an AI voice agent differs from your IVR

An AI voice agent works from what the caller says rather than from a fixed set of options, and that difference is the whole reason teams are moving past the IVR they already run.

ONE SENTENCE, TWO SYSTEMS

Caller: "I need to reschedule the appointment I booked last week."

Traditional IVR

"For appointments, press 3. For hours and locations, press 4."

AI voice agent

"I can move that. I see a visit on Thursday the 14th. What day works better?"

AI voice agents work from the caller's intent, while IVRs work from a menu.

AGAINST WHAT YOU RUN

	Traditional IVR	AI voice agent
Works from	A fixed menu tree it cannot deviate from	Caller intent, spoken in their own words
Reliable for	Simple, stable routing	Bounded call types handled end to end
Frustrating when	The caller's need does not fit a menu option	Its limits are not clearly drawn

Where a voice agent earns its place

Not every queue needs a voice agent. These are the situations where one pays for itself, and where the internal case for buying is easiest to make.

ROUTINE REQUESTS

Repeatable, high-volume calls

A large share of calls ask the same things: order status, hours, balances, appointment changes, password resets. A voice agent answers these in full and around the clock, which frees your people for the calls that actually need judgment.

PROTECTING AGENT TIME

Spam and nuisance calls

When a voice agent fronts the queue, the robocalls, misdials, and dead-air calls that would have tied up a person get handled by the agent instead. Nuisance traffic stops draining the hours you pay for.

COVERAGE

Staffing gaps and after-hours

When seats go unfilled, volume spikes, or the office closes, calls keep coming. A voice agent covers the overflow and the after-hours window, so callers reach help instead of a voicemail or a long hold.

If one or more of these describe your queue, a voice agent is worth evaluating.

The evaluation checklist

Seven questions buyers should ask, drawn from real evaluations.

WHAT TO ASK	WHY IT MATTERS	THE TRAP
01 Multilingual support Which languages does it support for understanding the caller and for speaking back, and what happens on a mid-call language change?	One of the first questions buyers raise and one of the most common gaps in competing tools. Mid-call switching, where a caller starts in one language and moves to another, is where many tools fall short.	A demo in English tells you nothing about the languages your callers actually use. Get the specific list and test it.
02 Call transfer and escalation Does it hand off to a live person cleanly, and to the right queue rather than a general line?	Escalation is the safety valve for every call that falls outside the agent's scope.	An agent that transfers but drops the context, forcing the caller to repeat everything. Confirm history and transcript travel with the transfer.
03 Call handling design Does it front every incoming call or take overflow only when the queue is full? Can your team edit the greeting and initial message?	Both models are valid; the right one depends on your volume and your comfort level.	A script you cannot change will not fit your brand or your compliance language.
04 Integrations How does it connect to the systems your work already lives in — ServiceNow, your CRM, your directory?	The agent is only as useful as the systems it can read from and write to.	More than one directory — common after mergers or across business units. Ask what happens when two disagree about the same person; that is where routing quietly breaks.

The evaluation checklist, continued

WHAT TO ASK	WHY IT MATTERS	THE TRAP
05 Transcript capture and analytics How do transcripts and call data get out, in what format, and do they feed the reporting you already use?	Usable transcripts are the raw material for improving the agent over time.	An agent that handles calls but locks up the data leaves you unable to prove its value or spot where it fails.
06 Deployment and routing Does it run in the cloud or on-premises, and can calls route through a local gateway?	Matters most with strict data-handling rules or on-premises telephony you are not ready to replace.	Assuming a cloud-only agent can be retrofitted into an on-premises call path later.
07 Security and compliance fit How is call data handled, how long is it retained, and where does it live?	Map the answers to your own policy rather than accepting a general assurance.	Healthcare, government, and finance buyers leaving compliance to the last conversation instead of the first.

Cost models and how to compare them

Agents are priced in a few different ways. Comparing quotes means understanding what each shape rewards and what it hides.

Per minute

Charges for the time the agent spends on calls.

Rewards quick, transactional calls

Penalizes anything that runs long

Per call

A flat rate for each call handled, regardless of length.

Rewards easy forecasting

Penalizes centers fielding many short calls

Concurrent call paths

Charges for how many calls the agent can handle at once.

Rewards steady high volume

Penalizes centers with occasional spikes

THE NUMBER NO QUOTE PRINTS — COST PER **RESOLVED** CALL

STICKER

RESOLUTION RATE

TRUE COST PER RESOLVED CALL

$$35\text{¢} \div 0.70 = \approx 50\text{¢} \text{ plus the human cost on the rest}$$

70% resolved by the agent — you pay 35¢

30% escalate — you pay 35¢ and the human

One prospect evaluating a competitor felt 30 to 40 cents per call was high. Whether it is high depends on a question the sticker price does not answer. Convert every quote, whatever its shape, into an estimated cost per resolved call using each vendor's honest resolution rate, then weigh that against the fully loaded cost of a human handling the same calls today. A quote that looks expensive per call can be the cheaper option per resolved call, and the reverse holds too. Build a simple worksheet with your own volumes, and ask each vendor to help you fill in the resolution figure rather than accepting the per-call rate at face value.

Measuring success: containment versus resolution

Two metrics dominate the conversation, and buyers often treat them as the same thing. They are not, and the gap between them is where a lot of disappointment hides.

WEAK HEADLINE METRIC

Containment

The share of calls the agent keeps without transferring to a human.

ANCHOR YOUR EVALUATION HERE

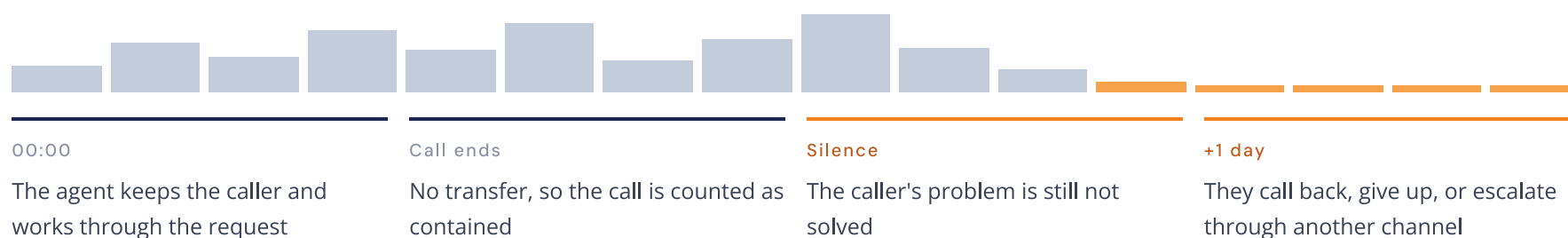
Resolution

The share of calls where the caller's actual need was met.

A caller asks to change an order, the agent walks them through it, and the call ends with no transfer. That counts as contained.

Whether it counts as resolved depends on one thing: did the order actually change? If it did not, the caller is back tomorrow, and containment still looks good even though nothing got fixed. That is why resolution is the number to hold vendors to.

HOW A CALL IS CONTAINED AND STILL UNRESOLVED



A containment rate reported on its own can look excellent while the experience underneath it is poor, which is why containment makes a weak headline metric.

What to track once the pilot is live

TRACK FROM DAY ONE OF THE PILOT

Resolution rate The number to build the evaluation around	Transfer & escalation rate Where the agent's scope ends in practice	Repeat-call rate Especially revealing, since two calls for one problem is not service	Caller satisfaction On the calls the agent handled
---	---	---	--

Ask every vendor for their Automated Resolution Rate, the share of calls that are both contained and actually resolved. That single number is harder to game than containment reported on its own.

Resolution also connects back to cost: the cost per resolved call in section 5 only holds up when you measure resolution honestly.

Deployment realities and the first 90 days

An AI voice agent is not a switch you flip. Most of the value comes from tuning it after it goes live, so treat the first 90 days as a structured rollout rather than a launch.

Day 1	30	60	90 · expand
DAYS 1–30	DAYS 31–60	DAYS 61–90	
Start narrow	Review and correct	Measure, then expand	
• One high-volume, well-understood, low-risk call type — that alone • Confirm routing and the CRM and directory integrations • Check behavior when a directory returns an unexpected result • Set the greeting and initial message in your own words	• Read the transcripts; find the intents it misreads and fix them • Tighten escalation paths so transfers land in the right queue with context • This is where a good deployment separates from a mediocre one	• Compare resolution and repeat-call rate against your baseline • If the numbers hold, add the next call type — one at a time • If they do not, you learned it on a narrow slice of traffic	

SURPRISES COMMON ENOUGH TO PLAN FOR

More than one directory produces conflicting lookups

Transfers land in the wrong queue when escalation paths are loosely mapped

Transcript data is harder to extract than the demo suggested

Language coverage proves narrower in production than in the sales call

How Variphy approaches this with Colleen

Variphy builds communications intelligence for unified communications environments, with analytics, reporting, and provisioning across Cisco CUCM, Webex Calling, Webex Contact Center, Microsoft Teams Phone, and Zoom Phone. Its AI voice work sits under that same line: AI, automation, and analytics applied to the calls your organization already handles.

Colleen AI Voice Agent

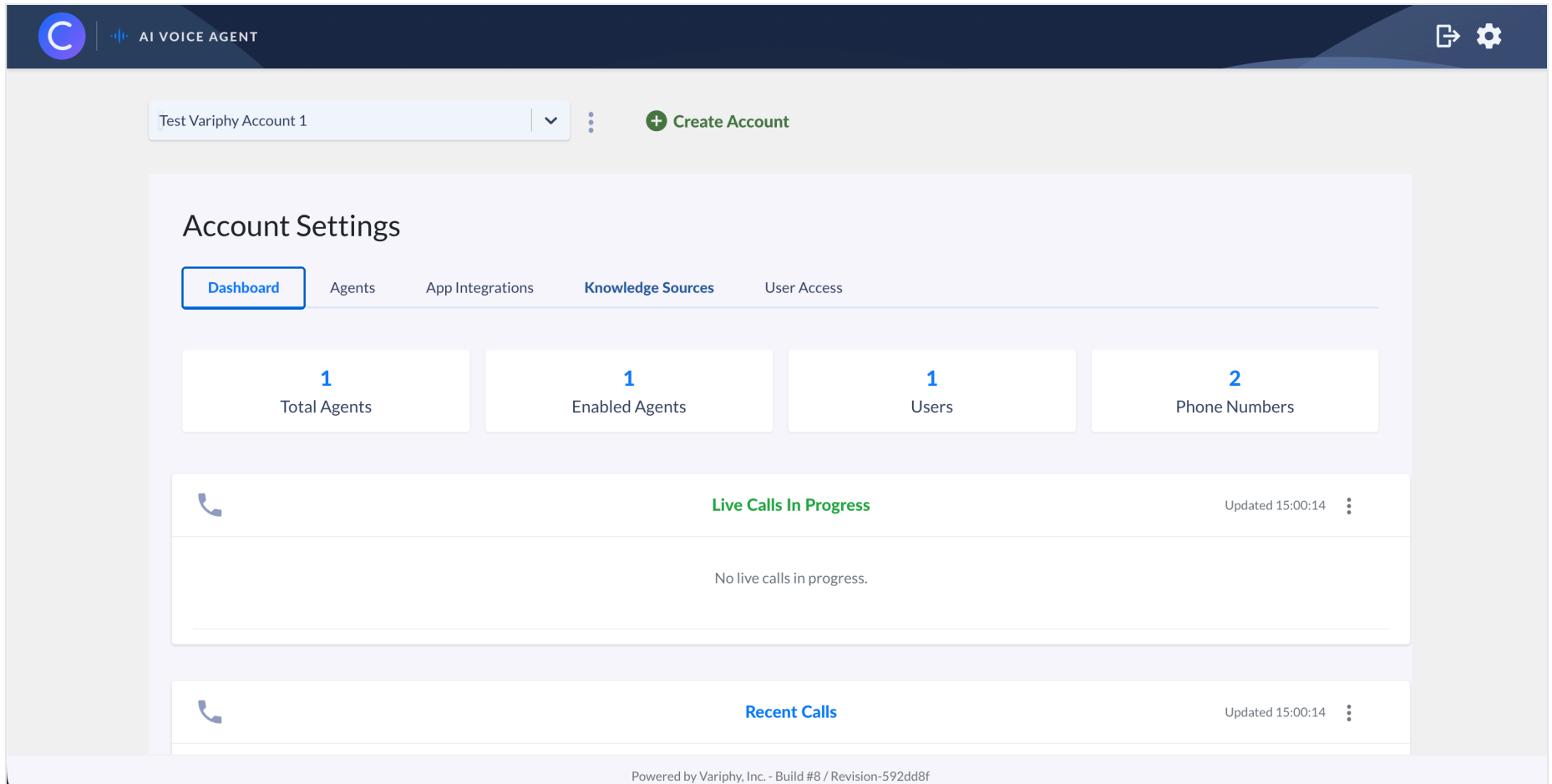
Colleen AI Voice Agent handles inbound calls automatically using natural language, without the menus, decision trees, or intent training traditional systems require. It deploys in front of any hunt group or call queue, with no CCaaS needed and no infrastructure changes.

THREE ROLES THE VOICE AGENT PLAYS

Virtual receptionist Routes callers by name or department using your existing directory (Webex, Microsoft Teams, Active Directory, or a spreadsheet) and handles multiple matches intelligently.	General knowledge Q&A Answers caller questions from a configured knowledge base and delivers a full summary and transcript when the call is done.	Support desk Detects urgent issues, transfers calls to the right person automatically, and creates tickets in ConnectWise or ServiceNow, so routine calls stay automated and critical ones never get missed.
--	---	--

Every call lands in the analytics platform

Live call monitoring shows caller, agent, and call state in real time. Completed calls carry an AI-generated summary and a full searchable transcript, alongside call volume analytics and historical records across every Colleen-handled call.



The screenshot displays the Colleen AI Voice Agent Dashboard. At the top, there's a dark blue header with the Colleen logo (a blue circle with a white 'C') and the text "AI VOICE AGENT". To the right of the header are icons for a document and a gear. Below the header, there's a light blue bar with a dropdown menu showing "Test Variphy Account 1" and a green button labeled "+ Create Account". The main content area is titled "Account Settings" and has four tabs: "Dashboard" (selected), "Agents", "App Integrations", "Knowledge Sources", and "User Access". Below the tabs, there are four white cards with blue numbers and labels: "1 Total Agents", "1 Enabled Agents", "1 Users", and "2 Phone Numbers". Below these cards, there are two sections: "Live Calls In Progress" and "Recent Calls". The "Live Calls In Progress" section shows a green status bar with a phone icon, the text "Live Calls In Progress", and "Updated 15:00:14". Below this, it says "No live calls in progress." The "Recent Calls" section shows a blue status bar with a phone icon, the text "Recent Calls", and "Updated 15:00:14". At the bottom of the dashboard, there's a footer that says "Powered by Variphy, Inc. - Build #8 / Revision-592dd8f".

Colleen AI Voice Agent Dashboard: Agents, users, and numbers at a glance, with live calls in progress and recent call activity underneath.

Recordings and transcripts, call by call


AI VOICE AGENT



Test Variphy Account 1

+ Create Account



Live Calls In Progress
Updated 11:12:55

No live calls in progress.



Recent Calls
Updated 11:12:57

Start Date/Time	Calling Party	Duration	Agent Name	Agent Phone Number	Result	Transferred To
Jun 11, 2026, 11:12:39	+16478027505	00:01:53	Variphy Support Agent	+19016573034	Answered	Hide

Summary

The caller inquired about Variphy's services, specifically the AI voice agent, its integrations, and use cases. The agent explained the AI voice agent's purpose (handling incoming calls, replacing/reducing IVR, answering FAQs, routing calls, qualifying calls, collecting information pre-transfer, and providing 24/7 support) and noted that integrations can be tailored to specific workflows. The caller then requested to be transferred to Simran Kanda for deeper technical integration details. The agent agreed and initiated the transfer. No bookings or service commitments were made on the call.

Transcription

Variphy Support Agent: Thanks for calling Verify! How can I help you today?

Caller: Uh, hey, I was just wondering a little bit more about your guys' services.

Variphy Support Agent: Variphy provides a comprehensive suite of analytics and reporting tools for unified communications platforms like Cisco, Webex, and Zoom. We specialize in call accounting, billing, and

Powered by Variphy, Inc. - Build #13 / Revision-41a5742

Recent calls with result and transfer target per call, expanded to the recording and the full transcription of the conversation.

How Colleen stands up to the eval

CHECKLIST ITEM	HOW COLLEEN ANSWERS IT
Transfer and escalation to the right queue	Detects urgent issues and transfers to the right person automatically; the result and transfer target are recorded per call.
Integrations, including more than one directory	Routes from your existing directory (Webex, Microsoft Teams, Active Directory, or a spreadsheet) and handles multiple matches intelligently. Creates tickets in ConnectWise or ServiceNow.
Transcript capture and analytics	AI-generated summaries, full searchable transcripts, live call monitoring, and call volume analytics inside Variphy.
Deployment and routing	Deploys in front of any hunt group or call queue on any phone system, in any environment, with no infrastructure changes and no CCaaS required.

CHECKLIST ITEM	HOW COLLEEN ANSWERS IT
Call handling design	No menus, decision trees, or intent training to maintain.
Multilingual support	Supports a wide range of languages, including regional variants you can specify by country, such as Mexican Spanish or Spain Spanish. The most requested are Spanish, Italian, and Portuguese, alongside English. It also follows a language change mid-call, so a caller who starts in Spanish and switches to English is answered in English.
Security, compliance, and pricing model	Call data is encrypted and stored securely, with HIPAA-compliant handling available when required. Pricing is volume-based, structured in call tiers.

Test these questions against your own environment

If you are evaluating an AI voice agent, Variphy can walk the checklist through with you, on your call types, your directories, and your volumes.

[Talk with Variphy about a scoped pilot ›](#)

The one-page evaluation checklist

Bring this into the vendor call. Any question you cannot get a specific answer to — a list, a number, a demo — is the one worth pressing on.

- | | |
|--|--|
| ☐ Which languages, for listening and for speaking, and what happens mid-call? | ☐ Does a transfer reach the right queue with the context attached? |
| ☐ Front every call or overflow only — and can we edit the greeting? | ☐ How does it connect to ServiceNow, the CRM, and every directory we run? |
| ☐ What happens when two directories disagree about the same person? | ☐ How do transcripts and call data get out, and do they feed our reporting? |
| ☐ Cloud, on-premises, or routed through a local gateway? | ☐ How is call data handled and retained, and where does it live? |
| ☐ What is the honest resolution rate, and the cost per resolved call? | ☐ Which call type do we start on, and what is the baseline today? |

Take the checklist into your next vendor call.

Then bring the answers back to one number: cost per resolved call, measured against the fully loaded cost of a human handling the same calls today.

[Request a scoped pilot ›](#)

Two decades of turning communications data into answers.

Variphy is headquartered in California's Silicon Valley and has delivered unified communications analytics and tooling for more than 20 years. Our mission is to build the most powerful and complete Communications Intelligence platform available, with a personalized level of service for every enterprise and partner we work with.

2,000+

CUSTOMERS

10M+

PHONES AND USERS

35+

COUNTRIES

20+

YEARS IN UC ANALYTICS

KEY BENEFITS

- Unlimited data retention
- Centralized platform support
- Advanced search and filtering
- AI-powered intelligence
- Risk-free proof of concept
- Role-based user access
- Automated report delivery
- Fully customizable reports
- Standardized provisioning
- Variphy deployment options

SUPPORTED PLATFORMS

On-premises UCM

Webex Calling

Webex Contact Center

UCCX

CUBE

Zoom Phone

Microsoft Teams

[Start a free trial ›](#)

[Request a demo ›](#)