

AI Voice Agent Pricing Explained: Per-Minute, Subscription and Usage Stacks

Most AI voice agents bill per minute of talk time. The total is a stack: a platform fee, the language model, speech-to-text, text-to-speech and telephony. Some vendors bundle the stack into one rate; others sell annual contracts.

Updated October 29, 2026

An AI voice agent bill has up to five parts. Each part has its own price. Some vendors show you every part. Other vendors bundle them into one rate per minute.

Prices change often. We checked every number below on 2026-10-29. Check the vendor page before you sign.

The five parts of the stack

| Part | What it does | Who usually bills it | |---|---|---| | Platform fee | Runs the agent, the call logic and the dashboard | The platform vendor | | Language model (LLM) | Decides what the agent says | The model provider, often passed through | | Speech-to-text (STT) | Turns the caller's speech into text | A speech vendor | | Text-to-speech (TTS) | Turns the reply into a voice | A voice vendor | | Telephony | Carries the phone call and owns the number | A carrier such as Twilio or Telnyx |

A speech-to-speech model can replace the STT, LLM and TTS parts with one model.

Three ways vendors price

1. Platform fee plus pass-through costs

You pay the platform a small fee per minute. You pay the other parts at cost or on your own accounts.

- **Vapi** lists \$0.05 per minute for hosting. It bills models "pass-through at cost". Its calculator estimates \$82 to \$129 a month for 1,000 minutes, with Deepgram, OpenAI and ElevenLabs as the providers.
- Vapi lists a HIPAA add-on at \$2,000 a month. It includes a BAA.
- **LiveKit** Cloud bills agent session time at \$0.01 per minute after the included minutes on its Ship and Scale plans. Its phone-call calculator example totals \$0.0479 per minute.

2. One bundled rate per minute

The vendor bills one rate. The rate includes the model and the voice.

- **Bland AI** lists \$0.14 per minute on its Start plan, with no platform fee. Its Build plan lists \$0.12 per minute plus a \$299 monthly platform fee. Bland says the rate covers the LLM, STT and TTS. Telephony is billed separately.

- **Retell AI** lists "\$0.07-\$0.31 / min" for voice agents. Its calculator example is \$0.11 per minute: \$0.055 voice infrastructure, \$0.04 LLM and \$0.015 TTS.

3. Annual contract

The vendor quotes a yearly price. The price depends on volume, concurrency and integrations.

- **Synthflow** states that "Enterprise contracts start at \$30,000 annually." Its page shows no per-minute rate.
- **Deepgram** Growth starts at "\$4,000+ / year" in prepaid credits.

Reference prices for each part

| Part | Example price | Source | |---|---|---| | Telephony, US local inbound | \$0.0085 per minute | Twilio | | Telephony, US local outbound | \$0.0140 per minute | Twilio | | US local phone number | \$1.15 per month | Twilio | | STT, Deepgram Nova-3 streaming | \$0.0077 per minute (pay as you go) | Deepgram | | TTS, ElevenLabs Flash/Turbo | \$0.04 per 1,000 characters | ElevenLabs | | TTS, ElevenLabs Multilingual v2 | \$0.08 per 1,000 characters | ElevenLabs | | Speech-to-speech, OpenAI gpt-realtime-2.1 | \$32 input and \$64 output per 1M audio tokens | OpenAI | | Speech-to-speech, OpenAI gpt-realtime-2.1-mini | \$10 input and \$20 output per 1M audio tokens | OpenAI | | Deepgram Voice Agent API, Standard | \$0.075 per minute (pay as you go) | Deepgram |

Twilio says its US page is "current as of August 2026."

Monthly fees outside the per-minute rate

Read the plan table, not only the headline rate. Look for these items:

- **Concurrency.** Vapi sells extra lines at \$10 per line per month. Retell sells extra concurrency at \$8 per month, with the first 20 free.
- **Phone numbers.** Retell lists \$2.00 per month per number. Twilio lists \$1.15 for a US local number.
- **Add-ons.** Retell lists PII removal at \$0.01 per minute and denoising at \$0.005 per minute.
- **Transfers.** Bland bills \$0.05 per transfer minute on Start and \$0.04 on Build.
- **Compliance.** A BAA or a HIPAA plan can cost extra. Vapi lists \$2,000 a month.

How to estimate your own cost

1. Count your expected minutes per month. Use call logs, not guesses.
2. Count your peak concurrent calls. This sets the plan you need.
3. Choose a model and a voice. These two choices move the rate most.
4. Add telephony and phone numbers.
5. Add the monthly fees and add-ons.
6. Ask each vendor for a written quote at your volume.

Example: 1,000 minutes on Bland AI's Start plan at \$0.14 per minute is \$140, before telephony.

Compare listings in the voice agent platforms and voice AI APIs categories to see which vendors publish their prices.

Frequently asked questions

What does an AI voice agent cost per minute?

On the public pricing pages we checked on 2026-10-29, all-in rates started near \$0.11 per minute in Retell AI's calculator example and \$0.12 to \$0.14 per minute on Bland AI's self-serve plans. Your model, voice and telephony choices move the number up or down.

Why do some vendors show two prices per minute?

Some vendors show only their own platform fee. Vapi, for example, lists \$0.05 per minute for hosting and passes model costs through at cost. You pay the model, voice and telephony costs on top.

Do enterprise contracts cost more than self-serve plans?

They usually add a commitment, not only a rate. Synthflow says its enterprise contracts start at \$30,000 a year. Deepgram's Growth plan starts at \$4,000 or more a year in prepaid credits.

Sources

[Vapi, Pricing \(fetched 2026-10-29\)](https://vapi.ai/pricing) — <https://vapi.ai/pricing>

[Retell AI, Pricing \(fetched 2026-10-29\)](https://www.retellai.com/pricing) — <https://www.retellai.com/pricing>

[Bland AI, Pricing \(fetched 2026-10-29\)](https://www.bland.ai/pricing) — <https://www.bland.ai/pricing>

[Synthflow, Pricing \(fetched 2026-10-29\)](https://synthflow.ai/pricing) — <https://synthflow.ai/pricing>

[Twilio, Programmable Voice pricing, United States \(fetched 2026-10-29\)](https://www.twilio.com/en-us/voice/pricing/us) — <https://www.twilio.com/en-us/voice/pricing/us>

[Deepgram, Pricing \(fetched 2026-10-29\)](https://deepgram.com/pricing) — <https://deepgram.com/pricing>

[ElevenLabs, API pricing \(fetched 2026-10-29\)](https://elevenlabs.io/pricing/api) — <https://elevenlabs.io/pricing/api>

[OpenAI, API pricing \(fetched 2026-10-29\)](https://developers.openai.com/api/docs/pricing) — <https://developers.openai.com/api/docs/pricing>

[LiveKit, Pricing \(fetched 2026-10-29\)](https://livekit.com/pricing) — <https://livekit.com/pricing>