

AI rates

What the AI costs on every call, per minute of conversation. No per-channel fees and no minimums: you pick the three engines and pay for the ones you use.

YOU PICK ALL THREE ENGINES 27 models to choose from for listening (STT), thinking (LLM) and speaking (TTS). The cheapest and the priciest combination are 3.2× apart: \$0.051 to \$0.162 per minute. You set the rate — we don't impose one.	YOU PAY FOR THE MINUTE SPOKEN Billed per actual minute of AI conversation — not per call, channel or licence. Time the agent spends querying your calendar or CRM (tools) is not billed.	CHEAPER AS YOU GROW The discount applies to ALL AI usage, not just some models: -10% on Business, -20% on Scale. No renegotiation and nothing to request: it comes off automatically as you consume.
--	--	--

100 credits = \$1.00

A credit is worth **\$0.01** everywhere: **100 credits = \$1.00**. Any figure on this sheet divided by 100 is dollars per minute. Every plan includes a monthly credit allowance; if you need more, you top up separately and that balance **never expires**. Subscriptions are billed in euros in Europe and dollars in the Americas, but an AI credit costs the same in both. Prices exclude tax.

What a call costs

Three real combinations from the catalogue below, at the base rate. The example call runs 3 minutes.

CHEAPEST \$0.051 PER MINUTE GPT-4o mini transcribe + GPT-5 Nano + WaveNet/Standard \$0.15 · 3-min call \$0.041 · on Scale (-20%)	MIDDLE GROUND \$0.095 PER MINUTE Deepgram Nova-3 + GPT-5.4 Nano + Aura-2 \$0.29 · 3-min call \$0.076 · on Scale (-20%)	MOST CAPABLE \$0.162 PER MINUTE Google STT (telephony) + Gemini 3.6 Flash + Cartesia Sonic \$0.49 · 3-min call \$0.130 · on Scale (-20%)
---	---	---

None of these models train on your calls' audio or text.

01 Listen · STT

6 MODELS · \$0.025 TO \$0.035

MODEL	\$/MIN	BASE	BUSINESS -10%	SCALE -20%
ASSEMBLYAI				
Universal-Streaming ▲	\$0.025	2.5	2.25	2
OPENAI				
GPT-4o mini transcribe	\$0.025	2.5	2.25	2
SPEECHMATICS				
Speechmatics RT Standard	\$0.025	2.5	2.25	2
DEEPGRAM				
Nova-3 (MIP) *	\$0.026	2.6	2.34	2.08
Deepgram Nova-3	\$0.031	3.1	2.79	2.48

MODEL		\$/MIN	BASE	BUSINESS -10%	SCALE -20%
GOOGLE					
Google STT (telephony)		\$0.035	3.5	3.15	2.8

02

Think · LLM

14 MODELS · \$0.018 TO \$0.065

MODEL		\$/MIN	BASE	BUSINESS -10%	SCALE -20%
MISTRAL					
Ministral 3B ▲		\$0.018	1.8	1.62	1.44
Ministral 8B ▲		\$0.019	1.9	1.71	1.52
Mistral Small (latest) ▲		\$0.020	2	1.8	1.6
GOOGLE					
Gemini 2.5 Flash-Lite		\$0.021	2.1	1.89	1.68
Gemini 3.1 Flash-Lite		\$0.022	2.2	1.98	1.76
Gemini 3.5 Flash-Lite		\$0.022	2.2	1.98	1.76
Gemini 2.5 Flash		\$0.031	3.1	2.79	2.48
Gemini 3.5 Flash		\$0.065	6.5	5.85	5.2
Gemini 3.6 Flash		\$0.065	6.5	5.85	5.2
OPENAI					
GPT-5 Nano		\$0.021	2.1	1.89	1.68
GPT-OSS 120B		\$0.023	2.3	2.07	1.84
GPT-5.4 Nano		\$0.027	2.7	2.43	2.16
GPT-5.4 Mini		\$0.042	4.2	3.78	3.36
ANTHROPIC					
Claude Haiku 4.5		\$0.044	4.4	3.96	3.52

03

Speak · TTS

7 MODELS · \$0.005 TO \$0.062

MODEL		\$/MIN	BASE	BUSINESS -10%	SCALE -20%
GOOGLE					
WaveNet/Standard		\$0.005	0.5	0.45	0.4
Neural2		\$0.020	1.98	1.78	1.58
Chirp3 HD		\$0.037	3.73	3.36	2.98
OPENAI					
OpenAI TTS		\$0.019	1.86	1.67	1.49

MODEL		\$/MIN	BASE	BUSINESS -10%	SCALE -20%
MISTRAL					
Mistral Voxtral 		\$0.035	3.51	3.16	2.81
DEEPGRAM					
Aura-2		\$0.037	3.73	3.36	2.98
CARTESIA					
Cartesia Sonic		\$0.062	6.21	5.59	4.97

04

Text chatbots • per 1,000 tokens

17 MODELS • \$0.0001 TO \$0.0108

MODEL		\$/1,000 TOKENS	BASE	BUSINESS -10%	SCALE -20%
ANTHROPIC					
Claude Haiku 4.5		\$0.0054	0.54	0.49	0.43
Claude Sonnet 5		\$0.0108	1.08	0.97	0.86
GOOGLE					
Gemini 2.5 Flash-Lite		\$0.0005	0.05	0.04	0.04
Gemini 3.1 Flash-Lite		\$0.0015	0.15	0.14	0.12
Gemini 2.5 Flash		\$0.0022	0.22	0.2	0.18
Gemini 3.5 Flash-Lite		\$0.0022	0.22	0.2	0.18
Gemini 3.6 Flash		\$0.0081	0.81	0.73	0.65
Gemini 3.5 Flash		\$0.0090	0.9	0.81	0.72
MISTRAL					
Ministral 3B		\$0.0001	0.01	0.01	0.01
Ministral 8B		\$0.0002	0.02	0.02	0.02
Mistral Small (latest)		\$0.0007	0.07	0.06	0.06
Mistral Large 3		\$0.0021	0.21	0.19	0.17
OPENAI					
GPT-5 Nano		\$0.0004	0.04	0.03	0.03
GPT-OSS 120B		\$0.0007	0.07	0.06	0.06
GPT-5.6 Luna		\$0.0012	0.12	0.11	0.1
GPT-5.4 Nano		\$0.0012	0.12	0.11	0.1
GPT-5.4 Mini		\$0.0045	0.45	0.41	0.36

* MIP: Deepgram's Model Improvement Program. A cheaper rate in exchange for letting the provider use the audio to improve its models. It is in the catalogue in case the trade is worth it to you; plain Nova-3, right below it, does not.

 The provider trains on the audio or text of calls made with that model. The rest do not.

Rates in credits per minute (100 credits = \$1.00), indicative and subject to the AI provider's own pricing and to change. Every call adds one model from each block: STT + LLM + TTS. Choosing models with EU data residency may change the rate. · Alpine