

Brief sample - Voice API: S2S, TTS, STT, custom voices (oficial)

SKU: R\$ 50 - ROUND1 (sample gratuito - piloto)

Data: 2026-09-05 ~20:55 America/Sao_Paulo

Fontes oficiais (só estas):

- <https://x.ai/pricing> (consultado nesta data via fetch HTTP)
- <https://docs.x.ai/docs> - <https://docs.x.ai/docs/guides/voice>
- <https://docs.x.ai/developers/models> - <https://docs.x.ai/developers/pricing>
- <https://docs.x.ai/console/billing>

Produzido localmente (API credit US\$ 0.00; sem top-up). Sem inventar model IDs nem preços.

Ângulo 1 - Três produtos Voice (não misturar)

Fonte: <https://docs.x.ai/docs/guides/voice>

Capacidade | Endpoint / canal | Para quê | Notas docs

Speech to Speech (S2S) | WebSocket [wss://api.x.ai/v1/realtime](https://api.x.ai/v1/realtime) | Agente de voz em tempo real (turn-taking, tools) | Modelo alias `grok-voice-latest`; VAD server-side; tools ex. `web_search`

Text to Speech (TTS) | POST <https://api.x.ai/v1/tts> (+ WS streaming) | Ler texto em áudio | Voices embutidas + custom; tags de fala; MP3 / ?-law

Speech to Text (STT) | POST <https://api.x.ai/v1/stt> (+ WS streaming) | Transcrever áudio | 12 formatos; timestamps; diarization; 25 idiomas

Consumer Voice no chat (Free/SuperGrok) != Voice API prepaid. Assinatura não enche Credit balance.

Ângulo 2 - Preços Voice (API)

Fonte: <https://docs.x.ai/developers/models> - <https://docs.x.ai/developers/pricing>

Modo | Modelo / variante | Preço publicado

Speech to Speech | `grok-voice-think-fast-2.0` | US\$ 0.08 / min áudio (US\$ 4.80 / hr) + US\$ 0.004 / text input

Speech to Speech | `grok-voice-think-fast-1.0` | Deprecated - US\$ 0.05 / min (+ US\$ 0.004 / text input)

Speech to Text | REST | US\$ 0.10 / hr

Speech to Text | Streaming | US\$ 0.20 / hr

Text to Speech | - | US\$ 15.00 / 1M chars

Recomendação oficial na página de models: Voice -> Grok Voice API.

Ângulo 3 - Orçamento rápido (só números oficiais)

Exemplos ilustrativos com a tabela acima (sem inventar descontos):

Cenário | Conta oficial

30 min S2S 2.0 (só áudio) | 30 x US\$ 0.08 = US\$ 2.40 (+ text input à parte)

1 h STT REST | US\$ 0.10

1 h STT Streaming | US\$ 0.20

TTS 100k chars | (100k / 1M) x US\$ 15 = US\$ 1.50

Tools no S2S (ex. `web_search`): cobrança tokens do modelo + invocações (Web Search US\$ 5 / 1k calls) - ver <https://docs.x.ai/developers/pricing>. Não misturar com preço de minuto de áudio.

Ângulo 4 - Custom voices + segurança

Fonte: <https://docs.x.ai/docs/guides/voice>

- Clone: POST <https://api.x.ai/v1/custom-voices> com clip de referência (máx. 120s).
- Resposta traz voice_id reutilizável em TTS (REST/WS) e no S2S realtime.
- Voice default citada nos exemplos: eve.
- Compliance docs: SOC 2 Type II; HIPAA Eligible (BAA); GDPR; data residency; SSO/RBAC.
- Docs: áudio processado em tempo real; não armazenado nem usado para treino (texto da página Voice).

Ângulo 5 - Planos consumer (Voice no chat)

Fonte: <https://x.ai/pricing>

Plano | Preço | Voice / uso (página)

Free | US\$ 0/mês | Voice listado nos destaques Free

SuperGrok | US\$ 30/mês | limites maiores (Chat / Imagine / Voice)

SuperGrok Plus | US\$ 100/mês | uso bem maior em Chat / Imagine / Voice / Build

Lite / Heavy / Business / Enterprise aparecem na comparação; preços Lite/Heavy não firmados nesta consulta - reabrir pricing.

Ângulo 6 - Multi-modelo: quando Voice vs texto vs Imagine

Fonte: <https://docs.x.ai/developers/models> - <https://docs.x.ai/docs>

Precisa de... | Caminho oficial

Code / chat texto | Grok 4.6 (Responses / Chat)

Imagem | Imagine Image 2.0 (US\$ 0.04 / image)

Vídeo | Imagine Video 1.5 (US\$ 0.080 / sec)

Voz tempo real / TTS / STT | Grok Voice API (tabela Ângulo 2)

Bulk assíncrono texto/imagem/vídeo | Batch API (desconto 20% só em modelos listados; Imagine/vídeo = taxa standard)

Knowledge cut-off declarado Grok 4.6: 2026-02-01. Sem search tools, sem eventos em tempo real.

Checklist freela / orçamento Voice

1. Separar Voice no app Grok (assinatura) vs Voice API (prepaid).
2. Credit US\$ 0 -> sem demo API paga; sample = este brief. Sem top-up sem Luiz.
3. Escolher S2S vs TTS vs STT antes de cotar (unidades: min / hr / chars).
4. Se S2S + tools: somar invocações (pricing tools).
5. Custom voice: clip <=120s; guardar voice_id.
6. PIX BRL -> LUIZ; ledger com arquivo EMV (_4=R\$10, _2=R\$50, _0/_3=R\$100; R\$20 = round1/pix_r20_*).

Como comprar o brief pago (não este sample)

Ver OFERTA.md - Pages <https://ziuluiziul.github.io/round1-xai/> - PIX /workspace/pix-clean/out/ (texto CRC > PNG).