

Voice Configuration

Provider: Azure

Voice:

- Isidora Multilingual es-es
- Ximena es-es
- Elias es-es
- Elvira es-es
- Beatriz es-mx
- Dalia es-mx

Provider: Cartesia

Voice: Young Spanish-speaking woman

Model: Sonic 2

Transcriber Configurations

Provider: OpenAI

Model: GPT-4o Transcribe

Language: ES

Provider: Deepgram

Model: Nova 2

Language: ES-419

Provider: Deepgram

Model: Nova 2 ó 3

Language: Multi

Provider: Google

Model: Gemini 2.0 Flash

Language: Multilingual

Provider: Cartesia

Model: Ink Whisper

Language: ES

Calidad de Proveedor para Transcribir

1. OpenAI – GPT-4o Transcribe – ES

- Ventaja: GPT-4o tiene un modelo de transcripción de última generación, optimizado para latencia muy baja y excelente precisión en español conversacional.
- Por qué está arriba:
 - Está afinado no solo para “transcribir”, sino para entender el contexto semántico.
 - Maneja bien distintos acentos (México, España, Latam).
 - Ideal para asistentes de voz en tiempo real.
Mejor balance entre velocidad + precisión + contexto.

2. Deepgram – Nova 2/3 – ES-419

- Ventaja: Deepgram Nova 2 y 3 son modelos especializados en speech-to-text con gran precisión en español latinoamericano (ES-419 = “español de LatAm”).
- Por qué aquí:
 - Tiene una tasa de error muy baja para acentos regionales (mexicano, argentino, colombiano).
 - Latencia muy baja.
 - Si lo configuras específicamente en ES-419 es más robusto que “Multi”.
La mejor opción si quieres priorizar velocidad y precisión bruta en español de LATAM.

3. Deepgram – Nova 2/3 – Multi

- Ventaja: reconoce varios idiomas a la vez.
- Por qué debajo:
 - Aunque útil si esperas que tus usuarios hablen en más de un idioma, pierde un poco de precisión frente a un modelo afinado solo en español.
Sirve si quieres asistentes bilingües, pero no es tan exacto como el ES-419.

4. Google – Gemini 2.0 Flash – Multilingual

- Ventaja: rápida y económica. Reconoce múltiples idiomas.
- Por qué abajo:
 - Gemini Flash es excelente en texto y visión, pero todavía no supera a Deepgram o OpenAI en ASR (Automatic Speech Recognition).
 - Puede equivocarse en acentos más marcados.
Bueno, si quieres algo generalista y barato, pero no es el mejor en español.

5. Cartesia – Ink Whisper – ES

- Ventaja: está basado en Whisper, que es muy sólido y open-source.
- Por qué último:
 - Aunque Whisper es preciso, suele ser más lento que Deepgram u OpenAI en tiempo real.
 - No está tan optimizado para baja latencia en asistentes conversacionales.
Fiable pero menos rápido; mejor para grabaciones que para voz en vivo.