

Análisis de costos BOTE3T

Usando los datos de uso obtenidos en el día que se realizaron las pruebas y encuesta pudimos realizar una tabla de costo aproximado por sesión. Los valores aproximados de cada API se recuperaron revisando el uso de cada API desde su plataforma correspondiente el día que se realizó la prueba.

Recurso (medida de uso)	Total usado	Promedio por usuario	Costo por unidad (USD)	Costo total (USD)
API GPT-4o input (tokens)	412.145	~41.215	2.5×10^{-6}	\$0.103
API GPT-4o output (tokens)	25.859	~2.586	1×10^{-5}	\$0.025
API Text-to- speech (caracteres)	~56.760	~5.676	16×10^{-5}	\$0.908
API Speech-to- text (segundos)	~860	~86	2.666×10^{-4}	\$0.022
API Embeddings (tokens)	1930	~193	1×10^{-7}	1.93×10^{-5}
API Pinecone (queries)	~1080	~108	9.6×10^{-8}	1.036×10^{-5}
Total				\$1.058

El costo aproximado de una sesión (usando ambos modos, usando TTS y STT) de 30-40 minutos por usuario es de USD \$1.058. Con la tasa de cambio del día 24 de jul. de 2025 el costo es de \$4.310 COP. A esto se le suma un costo fijo por uso del servidor virtual privado igual a USD \$5/mes, es decir \$20.368 COP.

Al ser la mayoría de los costos variables, permite que el sistema sea escalable, ya que el costo depende únicamente de su uso.

Precios recuperados de:

- GPT-4o y embeddings <https://platform.openai.com/docs/pricing?product=ER>
- Google TTS <https://cloud.google.com/text-to-speech/pricing?hl=en>
- Google STT <https://cloud.google.com/speech-to-text/pricing?hl=en>
- Pinecone <https://www.pinecone.io/pricing/>
- VPS <https://www.pythonanywhere.com/pricing/>