

Plan:

☐ Actualizaciones:

- ☐ ChatGPT 5.1: Personalización, conectores.
- ☐ Gemini Deep Research: Archivos y fuentes.
- ☐ Gemini: Google Slides
- ☐ NotebookLM: Google Sheet (toma de decisiones)
- ☐ Kimi K2: <https://www.kimi.com/>

- ☐ **Generación de Audio con IA:** Texto a voz, clonación de voces, voz a texto, música.
- ☐ **Generación de Vídeo con IA:** De Hailuo a Veo 3.1
- ☐ Gemini, ChatGPt, Claude *html*
- ☐ **Google IA Studio “Build”**
- ☐ **Hiperparamétro - Modelos Parte I:** Google IA Studio PlayGround, HuggingChat, LMStudio

Generación de contenido multimodal



Fundamentos y Tipos de Generación

- ✓ Texto a Vídeo (Text-to-Video)
- ✓ Imagen a Vídeo (Image-to-Video)
- ✓ Vídeo a Vídeo (Video-to-Video)

Generación de Audio Con IA



Generación de Audio Con IA

Audio: Generación de efectos, transcripción y edición de podcasts por texto.

Música: Composición de canciones completas con voz e instrumentos desde un simple prompt.

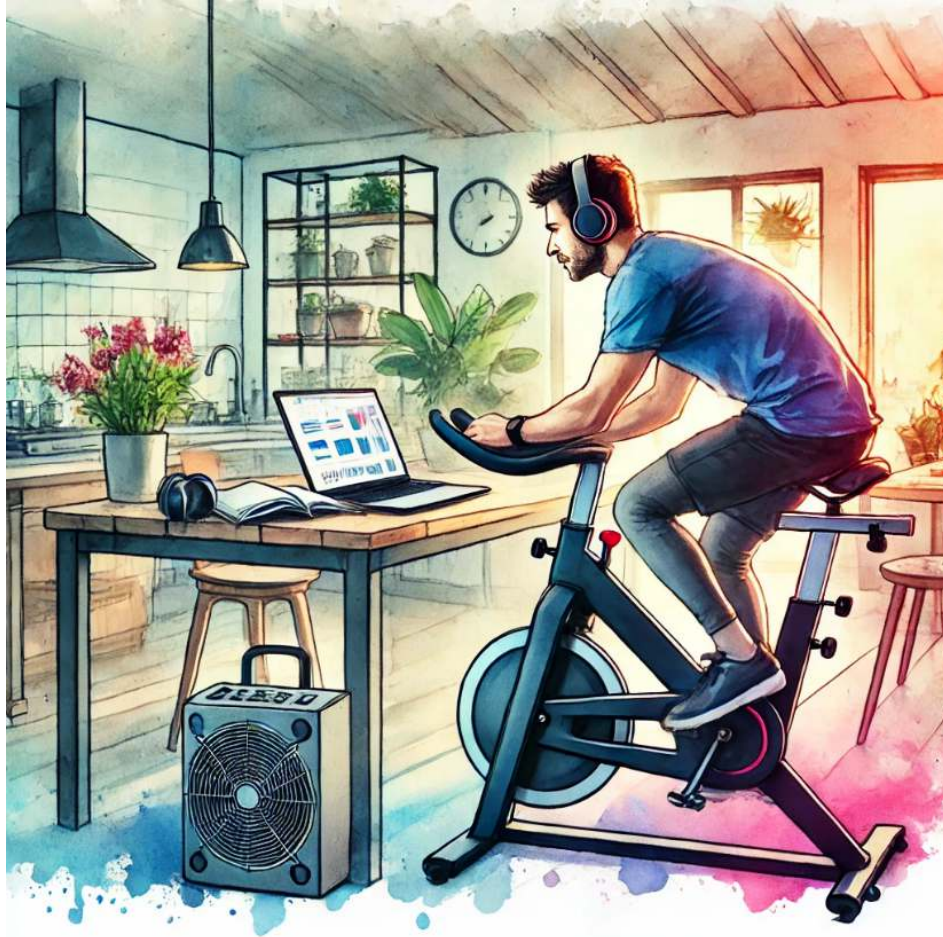
Voz: Síntesis y clonación de voz hiperrealista en docenas de idiomas.



¡Palabras que se oyen: IA convirtiendo texto en voz!

Existen varias herramientas de **texto a voz** que convierten texto escrito en voz hablada. Estas herramientas son útiles para personas con discapacidades visuales, dificultades de lectura, o para quienes prefieren escuchar contenido en lugar de leerlo.



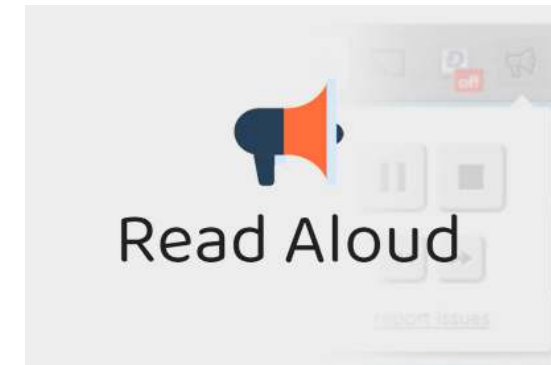


Convirtiendo texto en voz

Ahorra tiempo y aumenta tu productividad

- **Automatización de tareas:** Convierte documentos, artículos y libros en audio mientras realizas otras actividades.
- **Multitarea eficiente:** Escucha contenido mientras conduces, haces ejercicio o realizas tareas domésticas.
- **Mayor retención de información:** El audio facilita la memorización y comprensión de conceptos complejos.

Existen varias herramientas de **texto a voz** que convierten texto escrito en voz hablada, en particular **extensiones para el navegador**.














Available in the
Chrome Web Store

<https://chromewebstore.google.com/?hl=es>

*Generación de texto a
audio realista, voces,
clonación de voz,...*



| All | Open weights | Global Leaderboard | Personal Leaderboard | All models | To |
|-----|--|--------------------|----------------------|------------|-------|
| ↑↓ | Creator | ↑↓ | Model | ↑↓ | ELO |
| 1 |  MiniMax | | <u>Speech-02-HD</u> | | 1,133 |
| 2 |  MiniMax | | Speech-02-Turbo | | 1,122 |
| 3 |  OpenAI | | TTS-1 | | 1,120 |
| 4 |  ElevenLabs | | Multilingual v2 | | 1,115 |
| 5 |  ElevenLabs | | ElevenLabs v3 | | 1,114 |
| 6 |  ElevenLabs | | Turbo v2.5 | | 1,108 |
| 7 |  ElevenLabs | | Flash v2.5 | | 1,097 |
| 8 |  Inworld | | Inworld TTS 1 | | 1,096 |
| 9 |  Fish Audio | | OpenAudio S1 | | 1,092 |
| 10 |  MiniMax | | T2A-01-HD | | 1,064 |
| 11 |  Kokoro | | Kokoro 82M v1.0 | | 1,063 |

All

Open weights

Global Leaderboard

Personal Leaderboard

All models

↑↓

Creator

↑↓

Model

↑↓

ELO

↑↓

1

K

Kokoro

Kokoro 82M v1.0

1,063

2

Fish Audio

Fish Speech 1.5

1,031

3

Resemble AI

Chatterbox

1,026

4

Zyphra

Zonos-v0.1

1,000

5

Microsoft Azure

VibeVoice 7B

982

6

OpenVoice

OpenVoice v2

963

7

StepFun

Step TTS Mini

957

8

Coqui

XTTS v2

903

9

S

StyleTTS

StyleTTS 2

893

10

MetaVoice

MetaVoice v1

810

<https://artificialanalysis.ai/text-to-speech/>











<https://artificialanalysis.ai/>

<https://www.minimax.io/audio>

<https://inworld.ai/tts>

<https://elevenlabs.io/app/home>

Google IA Studio

| All Open weights Global Leaderboard Personal Leaderboard All models Top | | | |
|---|--|--------------------------|-------|
| ↕ | Creator ↕ | Model ↕ | ELO ↕ |
| 1 |  Inworld | <u>Inworld TTS 1 Max</u> | 1,216 |
| 2 |  MiniMax | Speech-02-HD | 1,134 |
| 3 |  MiniMax | Speech-02-Turbo | 1,124 |
| 4 |  OpenAI | TTS-1 | 1,117 |
| 5 |  ElevenLabs | ElevenLabs v3 | 1,117 |
| 6 |  ElevenLabs | Multilingual v2 | 1,115 |
| 7 |  ElevenLabs | Turbo v2.5 | 1,109 |
| 8 |  Inworld | Inworld TTS 1 | 1,101 |
| 9 |  ElevenLabs | Flash v2.5 | 1,099 |
| 10 |  Fish Audio | OpenAudio S1 | 1,088 |

Inteligencia Artificial Conversión de Audio a Texto con IA

Transcripciones



La Revolución del Audio: Un mundo mejor

Mayor accesibilidad:

Subtitulado automático para personas con discapacidad auditiva, generación de transcripciones para personas con dificultades de aprendizaje



Desafíos y Limitaciones

- **Calidad del audio:** Ruido, eco, diferentes ambientes.
- **Idiomas y dialectos:** Dificultades con acentos y lenguas poco comunes.
- **Contexto y comprensión:** Limitaciones en la interpretación de sarcasmo, ironía y lenguaje figurado.
- **Privacidad:** Protección de datos personales en las transcripciones.



TurboScribe es una poderosa herramienta de transcripción que convierte audio y video en texto de manera rápida y precisa. Algunas de sus principales funcionalidades incluyen:


1. **Transcripción Rápida y Precisa:** TurboScribe ofrece tres modos de transcripción (Cheetah, Dolphin, Whale) para adaptarse a tus necesidades, siendo Whale el más preciso.
2. **Reconocimiento de Voces:** Identifica y etiqueta a diferentes hablantes en la transcripción, ideal para entrevistas, podcasts, y reuniones.
3. **Soporte Multiformato:** Puedes subir archivos de audio y video en varios formatos o incluso enlaces de YouTube, Dropbox, Google Drive, y más.
4. **Exportación Versátil:** Una vez que la transcripción está lista, puedes exportarla en múltiples formatos como TXT, DOCX, PDF, y formatos de subtítulos como SRT y VTT.
5. **Opciones Gratuitas y de Pago:** Puedes transcribir hasta 3 archivos por día de hasta 30 minutos cada uno de manera gratuita. Si necesitas más, puedes optar por un plan ilimitado.
6. **Interfaz Intuitiva:** La plataforma es fácil de usar, permitiéndote gestionar tus transcripciones de manera eficiente desde un panel de control.





• <https://turboscribe.ai/>

TurboScribe Free

Free
100% Free

 **3 Transcripts Daily**
Transcribe 3 files for free every day.


 **30 Minute Uploads**
Each file can be up to 30 minutes long. Upload 1 file at a time.


 **Lower Priority**
Wait longer before your files are transcribed.


CURRENT PLAN


TurboScribe Unlimited

10 US\$ / month
120 US\$ billed yearly **SAVE 50 %**
≈ 502.360,94 COP billed yearly ⓘ

 **Unlimited Transcriptions**
Unlimited transcriptions for one person.

 **10 Hour Uploads**
Each file can be up to 10 hours long / 5 GB.
Upload 50 files at a time.

 **All Features**
Translation to 134+ languages. Bulk exports.
All transcription modes. Unlimited storage.

 **Highest Priority**
We'll always transcribe your files ASAP with the highest priority.

GO UNLIMITED



Modelos de Interacciones de Voz en IAs

¿Qué son los modelos de interacciones de voz?

- Procesan audio de entrada (*speech-to-text*) y generan respuestas habladas (*text-to-speech*).
- Beneficios: Accesibilidad, multitarea, inmersión natural.
- Evolución: De asistentes como Siri a conversaciones fluidas con IA generativa.
- Ejemplos: **Grok, Perplexity, ChatGPT y Claude.**

Hacia Interacciones Más Humanas

- Las IAs de voz transforman la interacción: De texto a diálogo natural.
- Tendencias: Mayor integración multimodal, accesibilidad global y personalización.

Inteligencia Artificial **Conversión de Audio a Audio con IA**

Traducciones al instante

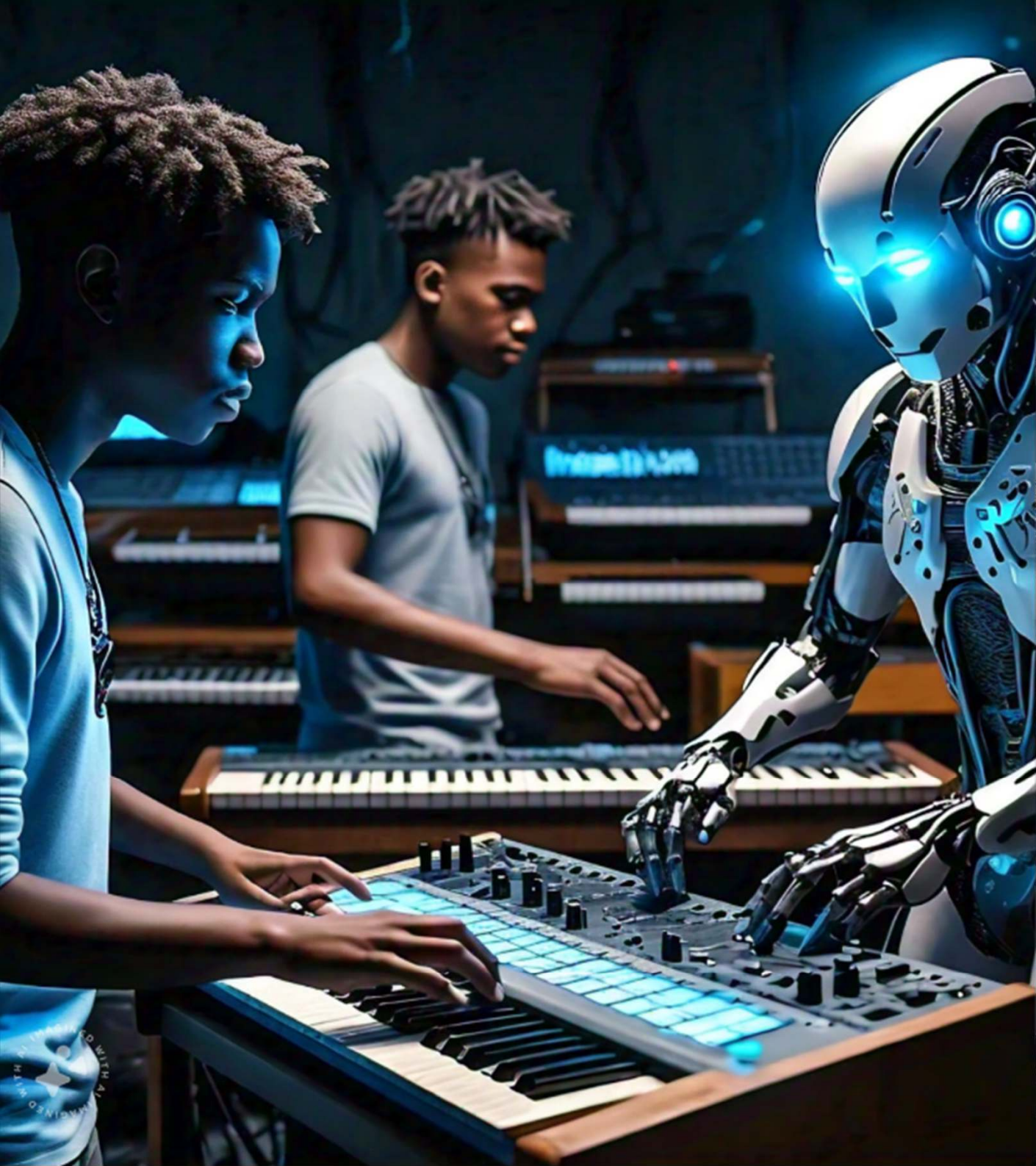


Inteligencia Artificial

Conversión de Texto a Música con IA

¿Alguna vez te has preguntado si
una máquina podría componer
una sinfonía?








Parálisis del sueño 🤤



mar-IA

 Follow

Pop

Parálisis del sueño 🤤

13 de septiembre de 2025, 7:10 p. m. v4.5+

▶ 103



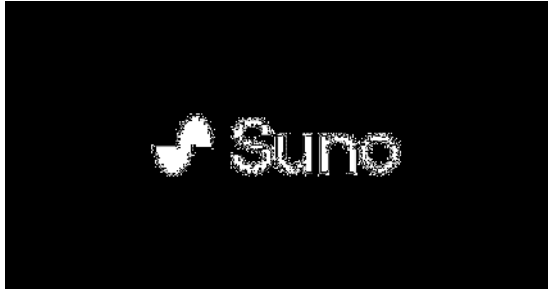
👍 37












Aplicaciones








- **Educación**
 - **Composición personalizada:** Creación de ejercicios musicales adaptados a cada estudiante.
 - **Aprendizaje interactivo:** Gamificación del aprendizaje musical.
 - **Accesibilidad:** Herramientas para personas con discapacidades.
- **Empresas y entretenimiento**
 - **Creación de jingles y bandas sonoras:** Música personalizada para campañas publicitarias.
 - **Generación de música de fondo:** Ambientes de trabajo más productivos.
 - **Personalización de experiencias:** Música adaptada a las preferencias de los clientes.
 - **Videojuegos:** Bandas sonoras adaptadas a la acción en tiempo real.
 - **Cine:** Creación de partituras originales para películas.



Instrumental

| ↑↓ | Creator ↑↓ | Model ↑↓ |
|----|---|-------------------|
| 1 |  Suno | Suno V4.5 |
| 2 |  ElevenLabs | Eleven Music |
| 3 |  Producer.ai | FUZZ-1.1 Pro |
| 4 |  Producer.ai | FUZZ-2.0 Raw |
| 5 |  Producer.ai | FUZZ-2.0 |
| 6 |  Google | Lyria 2 |
| 7 |  Udio | Udio v1.5 Allegro |

Vocals

| ↑↓ | Creator ↑↓ | Model ↑↓ |
|----|---|-------------------|
| 1 |  Suno | Suno V4.5 |
| 2 |  Producer.ai | FUZZ-2.0 |
| 3 |  Producer.ai | FUZZ-2.0 Raw |
| 4 |  Producer.ai | FUZZ-1.1 Pro |
| 5 |  ElevenLabs | Eleven Music |
| 6 |  Udio | Udio v1.5 Allegro |
| 7 |  Sonauto | Sonauto V2.1 |

Licencias y Derechos en Suno : El plan que tenías al momento de crear la canción define todos tus derechos. **Planes y Propiedad:**

- **Plan Pro o Premier:** Tú eres el dueño. Incluye la licencia comercial para distribuir y monetizar tu música donde quieras (Spotify, YouTube, etc.).
- **Plan Básico (Gratis):** Suno es el dueño. Tu uso se limita a fines *no comerciales*.

Puntos Clave de la Licencia

- **¿Si cancelo mi plan Pro?** Sigues siendo el dueño y mantienes todos los derechos comerciales de las canciones que creaste *durante* tu suscripción.
- **¿Es retroactivo?** No. Suscribirte a un plan Pro *después* de crear una canción en el plan gratuito **no** te da derechos comerciales sobre esa canción.

La Diferencia: Propiedad vs. Copyright

- **Propiedad (Con Plan Pro):** La canción es tuya y tienes la licencia exclusiva para monetizarla.
- **Copyright (Registro Legal):** Es complicado. Las leyes (especialmente en EE. UU.) protegen creaciones humanas. Una canción 100% generada por IA (música y letra) probablemente no se pueda registrar.
- **PERO:** Si tú escribiste la letra, esa letra es tuya. Aunque el *copyright* formal sea un tema legal gris, tu plan Pro te da la *propiedad* y licencia necesarias para ser el único que pueda monetizarla.

Conversión de Texto a Video con IA





Las aplicaciones de esta tecnología son extensas y abarcan múltiples sectores:

- ✓ **Marketing y Publicidad:** Creación rápida de material promocional, anuncios personalizados y contenido para redes sociales.
- ✓ **Educación y Formación:** Desarrollo de recursos educativos interactivos, simulaciones y material didáctico personalizado.
- ✓ **Entretenimiento y Medios:** Facilitación de la previsualización, generación de efectos visuales y conceptualización de narrativas.
- ✓ **Comunicaciones Corporativas:** Producción de vídeos internos, presentaciones y material de formación a escala.

Beneficios Clave de la Adopción de IA en Video












Beneficios Operativos

- **Menos costos:** elimina gastos de rodaje, locaciones, equipo y postproducción.
- **Más velocidad:** producción en horas o minutos en lugar de días o semanas.
- **Escalable:** genera múltiples versiones (formatos, estilos, idiomas) automáticamente.
- **Localización inteligente:** avatares y sincronización labial en decenas de idiomas.

Beneficios Creativos y Estratégicos

- **Creatividad democratizada:** cualquiera puede producir contenido cinematográfico sin experiencia técnica.
- **Más enfoque estratégico:** la IA asume tareas repetitivas; el creador se centra en narrativa y distribución.
- **Personalización masiva:** contenido adaptado a cada audiencia o campaña.

| # | Modelo | Fortalezas |
|----|---------------------------------|---|
| 1 | Kling 2.5 Turbo 1080p | Realismo fotorealista en movimientos complejos y física Sci-Fi; alta adherencia a prompts detallados . |
| 2 | Veo 3 (No Audio) | Cinematografía precisa con control de cámara (pan/zoom); consistencia temporal en narrativas multi-objeto . |
| 3 | Veo 3.1 Preview (No Audio) | Fidelidad en previsualizaciones creativas; edición granular para estilos variados (3D/fantasy) . |
| 4 | Veo 3.1 Fast Preview (No Audio) | Velocidad en drafts iterativos; movimientos fluidos y escalabilidad para marketing educativo . |
| 5 | Sora 2 Pro (No Audio) | Narrativa emocional con stitching multi-escena; física realista para simulaciones y cameos . |
| 6 | Ray 3 | Consistencia lumínica y foco dinámico; realismo ambiental en vida silvestre o arquitectura . |
| 7 | Hailuo 02 Standard | Fluidez en simulación humana y física (fluidos); rápida para social media y estilos anime . |
| 8 | Vidu Q2 | Expresiones faciales sutiles y acting humano; control cinematográfico para moda/emociones . |
| 9 | Veo 3 Fast Preview (No Audio) | Iteraciones rápidas con transiciones suaves; accesible para principiantes en diseño . |
| 10 | Waver 1.0 | Coherencia espacio-temporal; motion compleja (deportes) e imagen-vídeo inmersivo . |
| 11 | Wan 2.5 Preview | Precisión semántica en prompts largos; open-weights para abstracciones artísticas e investigación . |

| | | | |
|--|--|---------------------------------|-------|
| <div>Current modelsAll modelsAllOpen weights</div> | | | |
| ↕ Creator ↕ | | Model ↕ | ELO ↕ |
| 1 |  Kuaishou KlingAI | Kling 2.5 Turbo 1080p | 1,233 |
| 2 |  Google | Veo 3 (No Audio) | 1,228 |
| 3 |  Google | Veo 3.1 Preview (No Audio) | 1,215 |
| 4 |  Google | Veo 3.1 Fast Preview (No Audio) | 1,214 |
| 5 |  OpenAI | Sora 2 Pro (No Audio) | 1,214 |
| 6 |  Luma Labs | Ray 3 | 1,212 |
| 7 |  MiniMax | Hailuo 02 Standard | 1,201 |
| 8 |  Vidu | Vidu Q2 | 1,187 |
| 9 |  Google | Veo 3 Fast Preview (No Audio) | 1,184 |
| 10 |  Bytedance | Waver 1.0 | 1,184 |
| 11 |  Alibaba | Wan 2.5 Preview | 1,182 |

<https://artificialanalysis.ai/text-to-video/arena?tab=leaderboard-text&format=all>

Artificial Analysis Text to Video Leaderboard

Style

- All
- Abstract
- Cartoon & Anime
- Fantasy
- Photorealistic
- Sci-Fi
- 3D Animation

Subject

- All
- Action
- Animals
- Buildings
- Fashion
- Food
- Indoor
- Nature
- Outdoor
- People
- Physics
- Screens
- Specific Lighting
- Specific Location or Era
- Sports
- Technology
- Text
- Transport
- Water
- Weather and Effects
- Show less

Format

- All
- Long Prompt
- Moving Camera
- Multi Scene
- Short Prompt

<https://artificialanalysis.ai/text-to-video/arena?tab=leaderboard-text&format=all>