Una caricatura de una persona

El contenido generado por IA puede ser incorrecto.UNIVERSIDAD DE BURGOS ESCUELA POLITÉCNICA SUPERIOR



Máster en Ingeniería en Informática

TFM del Máster en Ingeniería Informática

**Avatar**

**Estudio del Arte**



Presentado por Manuel C. León Rivera

en Universidad de Burgos — 20 de enero de 2021 Tutor: María Belén Vaquerizo García

**Índice General**

[**Plan de Proyecto Software** 1](#_Toc198030185)

[1.1 Introducción 1](#_Toc198030186)

[1.2 Planificación temporal 1](#_Toc198030187)

[**Estudio del Arte** 4](#_Toc198030188)

[**D-ID** 8](#_Toc198030189)

[1.1 Visión general 8](#_Toc198030190)

[1.2 Arquitectura general 8](#_Toc198030191)

[1.3 Funcionalidades principales 8](#_Toc198030192)

[1.4 Parámetros del modelo talk 9](#_Toc198030193)

[1.5 Características técnicas 9](#_Toc198030194)

[**Synthesia** 11](#_Toc198030195)

[1.1 Visión general 11](#_Toc198030196)

[1.2 Arquitectura general 11](#_Toc198030197)

[1.3 Funcionalidades principales 11](#_Toc198030198)

[1.4 Parámetros principales de creación de video 11](#_Toc198030199)

[1.5 Características técnicas 12](#_Toc198030200)

[**HeyGen** 14](#_Toc198030201)

[1.1 Visión general 14](#_Toc198030202)

[1.2 Arquitectura general 14](#_Toc198030203)

[1.3 Funcionalidades principales 14](#_Toc198030204)

[1.4 Parámetros principales de creación de video 14](#_Toc198030205)

[1.5 Características técnicas 15](#_Toc198030206)

[**Elay** 16](#_Toc198030207)

[1.1 Visión general 16](#_Toc198030208)

[1.2 Arquitectura general 16](#_Toc198030209)

[1.3 Funcionalidades principales 16](#_Toc198030210)

[1.4 Parámetros principales de creación de video 16](#_Toc198030211)

[1.5 Características técnicas 17](#_Toc198030212)

[**VEED** 18](#_Toc198030213)

[1.1 Visión general 18](#_Toc198030214)

[1.2 Arquitectura general 18](#_Toc198030215)

[**Colossyan** 19](#_Toc198030216)

[1.1 Visión general 19](#_Toc198030217)

[1.2 Arquitectura general 19](#_Toc198030218)

[1.3 Funcionalidades principales 19](#_Toc198030219)

[1.4 Parámetros principales de creación de video 19](#_Toc198030220)

[1.5 Características técnicas 20](#_Toc198030221)

[**Speechify** 22](#_Toc198030222)

[1.1 Visión general 22](#_Toc198030223)

[1.2 Arquitectura general 22](#_Toc198030224)

[1.3 Funcionalidades principales 22](#_Toc198030225)

[1.4 Parámetros principales de creación de audio 22](#_Toc198030226)

[1.5 Características técnicas 22](#_Toc198030227)

[**Tavus** 24](#_Toc198030228)

[1.1 Visión general 24](#_Toc198030229)

[1.2 Arquitectura general 24](#_Toc198030230)

[1.3 Funcionalidades principales 24](#_Toc198030231)

[1.4 Parámetros principales de creación de video 24](#_Toc198030232)

[1.5 Características técnicas 25](#_Toc198030233)

[**Runway** 26](#_Toc198030234)

[1.1 Visión general 26](#_Toc198030235)

[1.2 Arquitectura general 26](#_Toc198030236)

[1.3 Funcionalidades principales 26](#_Toc198030237)

[1.4 Parámetros principales de creación de video 26](#_Toc198030238)

[1.5 Características técnicas 27](#_Toc198030239)

[**A2E** 28](#_Toc198030240)

[1.1 Visión general 28](#_Toc198030241)

[1.2 Arquitectura general 28](#_Toc198030242)

[1.3 Funcionalidades principales 28](#_Toc198030243)

[1.4 Parámetros del modelo crear vídeo 29](#_Toc198030244)

[1.5 Características técnicas 29](#_Toc198030245)

[**AI Studios** 31](#_Toc198030246)

[1.1 Visión general 31](#_Toc198030247)

[1.2 Arquitectura general 31](#_Toc198030248)

[1.3 Funcionalidades principales 31](#_Toc198030249)

[1.4 Parámetros principales de creación de video 31](#_Toc198030250)

[1.5 Características técnicas 34](#_Toc198030251)

[**Akool** 35](#_Toc198030252)

[1.1 Visión general 35](#_Toc198030253)

[1.2 Arquitectura general 35](#_Toc198030254)

[1.3 Funcionalidades principales 35](#_Toc198030255)

[1.4 Parámetros principales de creación de video 36](#_Toc198030256)

[1.5 Características técnicas 37](#_Toc198030257)

[**Vidnoz** 38](#_Toc198030258)

[1.1 Visión general 38](#_Toc198030259)

[1.2 Arquitectura general 38](#_Toc198030260)

[1.3 Funcionalidades principales 38](#_Toc198030261)

[1.4 Parámetros principales de creación de video 38](#_Toc198030262)

[1.5 Características técnicas 39](#_Toc198030263)

[**Pipio** 40](#_Toc198030264)

[1.1 Visión general 40](#_Toc198030265)

[1.2 Arquitectura general 40](#_Toc198030266)

[1.3 Funcionalidades principales 40](#_Toc198030267)

[1.4 Parámetros principales de creación de video 40](#_Toc198030268)

[1.5 Características técnicas 41](#_Toc198030269)

[**Creatify** 42](#_Toc198030270)

[1.1 Visión general 42](#_Toc198030271)

[1.2 Arquitectura general 42](#_Toc198030272)

[1.3 Funcionalidades principales 42](#_Toc198030273)

[1.4 Parámetros principales de creación de video 42](#_Toc198030274)

[1.5 Características técnicas 43](#_Toc198030275)

[**Normalización de los datos** 44](#_Toc198030276)

[1.1 Normalización de columnas para el sistema CBR 44](#_Toc198030277)

[1.2 Cálculo de la Similitud Total en el Sistema CBR 49](#_Toc198030278)

[1.3 Elección de proveedor para el trabajo 50](#_Toc198030279)

[1.4 Proveedor obtenido mediante CBR 52](#_Toc198030280)

[**Proveedores fuera de estudio** 53](#_Toc198030281)

[**Especificación de Requisitos** 54](#_Toc198030282)

[1.1 Introducción 54](#_Toc198030283)

[1.2 Objetivos generales 54](#_Toc198030284)

[1.3 Catálogo de requisitos 54](#_Toc198030285)

[1.4 Casos de Uso 55](#_Toc198030286)

[**Bibliografía** 62](#_Toc198030287)

**Índice de Figuras**

[**Figura A1: Milestone Burndown - Sprint 1** 2](#_Toc194511389)

[**Figura A2: Issues - Sprint 1** 2](#_Toc194511390)

**Índice de Tablas**

[**Tabla A1: Tareas - Sprint 1** 3](#_Toc194511409)

# **Plan de Proyecto Software**

1. Introducción

En este primer apartado se detalla cómo se ha planificado el desarrollo del proyecto titulado “Avatar”.

Este proyecto tiene como objetivo realizar un estudio del estado del arte de las principales tecnologías existentes en el mercado que permiten generar avatares virtuales realistas a partir de texto o voz, así como desarrollar una aplicación funcional que, mediante razonamiento basado en casos (CBR), recomiende la API más adecuada según los requisitos del usuario.

La planificación incluye tanto una fase de investigación tecnológica como otra de desarrollo práctico en un entorno controlado, utilizando contenedores Docker para facilitar la portabilidad del proyecto, y herramientas de control de versiones para el seguimiento del desarrollo. Asimismo, se tendrá en cuenta la viabilidad técnica y económica, estableciendo estimaciones de tiempo, costes y beneficios que podría aportar la solución propuesta.

1. Planificación temporal

Durante las primeras reuniones de orientación, se identificaron los pasos iniciales necesarios para abordar correctamente el desarrollo del proyecto. Como metodología de trabajo, se optó por utilizar SCRUM, de forma que se establecen sprints iterativos y funcionales en los que se planifican tareas concretas con entregables parciales.

Desde el comienzo del proyecto se ha creado un repositorio en GitHub, donde se gestionará todo el código fuente, documentación y versiones del sistema. Para una organización visual y ágil de las tareas, se ha incorporado la extensión ZenHub, que permite definir Issues, agruparlas por Sprints y realizar un seguimiento del avance mediante su sistema de Boards.

Cada sprint se inicia con una reunión de planificación en la que se definen las funcionalidades a implementar. El enfoque busca que en cada iteración se obtenga un prototipo funcional —aunque incompleto—, asegurando así una entrega incremental y continua del producto, susceptible de mejora progresiva.

**SPRINT 1 – 25/03/2025 – 02/04/2025**

Durante esta primera iteración, más corta que el resto, se ha orientado a establecer un entorno de trabajo funcional, con herramientas acordes a las necesidades del proyecto tanto para la gestión como para el desarrollo del mismo.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Figura A1: Milestone Burndown - Sprint 1

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

El contenido generado por IA puede ser incorrecto.

Figura A2: Issues - Sprint 1

|  |  |  |
| --- | --- | --- |
| TAREA | NOMBRE DE LA TAREA | TAG |
| #1 | Elegir editor de texto para la memoria | Decisión |
| Investigación |
| #2 | Buscar trabajos relacionados | Investigación |
| #3 | Crear y configurar repositorio | Configuración |
| #4 | Instalación de Zotero | Configuración |
| Instalación |
| #5 | Uso de herramientas | Investigación |

Tabla A1: Tareas - Sprint 1

# **Estudio del Arte**

El presente trabajo se fundamenta en la metodología de Razonamiento Basado en Casos (CBR), que permite realizar comparaciones inteligentes entre tecnologías a partir de casos previos y criterios de decisión estructurados. En este contexto, cada proveedor de API se considera un caso, y se analiza bajo un conjunto común de características técnicas que sirven como parámetros de comparación.

Durante el proceso de investigación, se han identificado limitaciones comunes en la documentación disponible. Algunos proveedores no publican información de forma homogénea: ciertas funcionalidades son promocionadas pero no detalladas técnicamente, otras se encuentran restringidas a planes avanzados sin descripción explícita, y en muchos casos no existe acceso completo a la especificación de las APIs (como ocurre con plataformas comerciales como Synthesia o HeyGen). Esto genera una dificultad adicional para normalizar los datos y establecer criterios comparativos objetivos.

Por ello, se ha diseñado una plantilla de características técnicas, que guía la recopilación y análisis de la información. Estas características han sido elegidas por su relevancia en entornos reales de uso e integración técnica, y constituyen la base del estudio:

* **Precios (paquete anual, precio mensual)**: análisis del coste base mensual y anual según el plan mínimo disponible para desarrolladores o empresas. En ocasiones vendrán como créditos que podrán contratarse de forma opcional para aumentar los minutos de vídeo.
* **Tiempo de vídeo consumible al mes:** minutos de vídeo generables con avatares dentro del plan contratado.
* **Tiempo de streaming consumible al mes:** minutos disponibles para vídeo en tiempo real (si el proveedor lo permite).
* **Tiene streaming:** indica si la API ofrece generación de vídeo sincronizado en tiempo real.
* **Streaming con SDK:** La API ofrece unas herramientas de integración en el proyecto para Streaming.
* **Tiene traducción de vídeo:** evalúa si es posible traducir y re-generar vídeos manteniendo la sincronización labial.
* **Resolución del vídeo basado en foto:** calidad de salida cuando el avatar es generado a partir de una imagen fija (por ejemplo, rostro frontal JPG/PNG).
* **Resolución del vídeo basado en foto:** calidad de salida cuando el avatar es generado a partir de una imagen fija (por ejemplo, rostro frontal JPG/PNG).
* **Resolución máxima de vídeo:** resolución de exportación permitida en la API (720p, 1080p, 4K, etc.).
* **Tiempo máximo de vídeo:** duración máxima por cada solicitud.
* **Sincronización labial:** grado de precisión con el que el avatar mueve los labios en relación al audio (Baja, Media o Alta).
* **Número de avatares disponibles:** cantidad de presentadores o modelos humanos listos para usar.
* **Avatares personales:** posibilidad de crear avatares personalizados a partir de imagen o grabación.
* **Voces (proveedores):** número de voces disponibles y variedad de proveedores de TTS soportados.
* **Expresiones del avatar:** soporte para emociones visuales como alegría, neutralidad, tristeza o entusiasmo.
* **Clonado de voz:** funcionalidad para generar una voz IA personalizada a partir de muestras reales.
* **Idiomas soportados:** número de idiomas y dialectos disponibles para la generación de voz.
* **Entrada alternativa:** posibilidad de subir un archivo de audio propio en lugar de usar texto.
* **Licencia y uso comercial:** condiciones legales para utilizar los vídeos generados con fines empresariales, de marketing o distribución.
* **Calidad de la API (documentación):** valoración de la documentación oficial, ejemplos, claridad y recursos para desarrolladores.
* **Velocidad de generación de vídeo:** tiempo estimado entre el envío de la solicitud y la recepción del vídeo final.

Estas variables permiten construir una base de casos estructurada, con la que se podrá razonar de forma objetiva qué proveedor se adapta mejor a distintos escenarios definidos. A pesar de las dificultades para recopilar datos totalmente uniformes, el enfoque CBR (Moja, 2024) facilita la toma de decisiones al comparar soluciones existentes bajo un marco técnico y funcional común.

* **Recuperación (Retrieve):** el sistema analiza el nuevo caso de entrada (especificado por el usuario) y lo compara con todos los casos existentes (proveedores de API), utilizando una función de similitud basada en atributos normalizados.
* **Reutilización (Reuse):** se identifica el caso más similar y se toma como referencia para recomendar un proveedor.
* **Revisión (Revise):** esta fase no se aplica de forma operativa en el sistema desarrollado, ya que no existe una validación posterior de la solución ni retroalimentación basada en resultados. La selección se considera válida si cumple los criterios definidos por el usuario y obtiene la mayor similitud.
* **Retención (Retain):** si se desea, el sistema puede guardar el nuevo caso (por ejemplo, una configuración seleccionada) como parte del repositorio de experiencias para futuras consultas.

Para la identificación y selección de las APIs más relevantes de 2025 disponibles en el mercado actual, se ha realizado un proceso de exploración documental y análisis técnico a través de fuentes especializadas. Entre las referencias utilizadas destacan las siguientes:

* **13 best AI avatar generator to explore in 2025:** <https://www.heygen.com/blog/ai-avatar-generator>
* **The 15 best AI Avatar Generators of 2025**: <https://www.d-id.com/blog/best-ai-avatar-generators/>
* **Best AI Avatar Generator: Top 19 picks of 2025:** <https://www.synthesia.io/post/best-ai-avatar-generator>
* **7 Best AI Avatar Generators of 2025**: <https://www.synthesia.io/post/best-ai-avatar-generator>
* **Best AI Avatar Generators in 2025**: <https://www.vozo.ai/blogs/best-ai-avatar-generators>
* **5 Generadores de Vídeo con IA Gratuitos para Probar en 2025 – Aquí está mi experiencia honesta como creador de contenido**: <https://www.reddit.com/r/AIToolTesting/comments/1hm5gpr/5_free_ai_video_generators_to_try_in_2025_heres/>
* **Best AI Avatars for Video Marketing Compared in 2025:** <https://www.outmost.studio/post/best-ai-avatars-for-video-marketing-reviewed-in-2025>
* **The best avatar generators in 2025:** <https://www.producthunt.com/categories/avatar-generators?order=best_rated>
* **Top AI Avatar Creators You Need to Try in 2025:** <https://magicshot.ai/blog/top-ai-avatar-creators/>
* **¿Cuál es el mejor generador de avatares con IA?** **Nuestra 19 propuestas para 2025:** <https://www.synthesia.io/es/post/mejores-generadores-de-avatares-con-ia>
* **10 Best AI Avatar Generators (April 2025):** <https://www.unite.ai/best-ai-avatar-generators/>
* **The 18 Best AI Video Generatos (Free & Paid) to Try in 2025:** <https://www.synthesia.io/post/best-ai-video-generators>
* **11 Best AI Video Generatos in 2025:** <https://www.zebracat.ai/post/11-best-ai-video-generators-in-2025>

# **D-ID**

1. Visión general

La API de D-ID proporciona una interfaz RESTful orientada a la generación de contenido audiovisual mediante avatares virtuales animados a partir de una imagen estática (fotografía de rostro o torso) y una fuente de voz (texto o audio). Su arquitectura está diseñada para ofrecer servicios escalables de síntesis de vídeo con soporte multilingüe, personalización de voz y animación facial automática.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Basic Auth mediante API Key
* **API Referencia:** <https://docs.d-id.com/reference/get-started>
* **Base URL**: <https://api.d-id.com>
* **API versión**: 4.2.1
* **Formatos admitidos**: JSON en entrada y salida, imágenes JPEG/PNG, audio opcional

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /talks | Genera un video a partir de imagen + texto o audio |
| GET /talks/{id} | Recupera el estado o resultado de un video generado |
| POST /clips | Crea un vídeo de calidad a partir de avatares premium o personalizados. |
| POST /talks/streams | Genera animación en tiempo real (streaming interactivo) tomando como base una imagen. |
| POST /clips/streams | Genera animación en tiempo real (streaming interactivo) de alta calidad para asistentes virtuales (Premium). |
| POST /translations | Traduce un video a otros idiomas |
| GET /voices / GET /presets | Consulta voces TTS y modelos de avatares disponibles |

Tabla A1: Métodos D-ID Vídeos

1. Parámetros del modelo talk

**const** url **=** 'https://api.d-id.com/talks'**;**

**const** options **=** **{**

method**:** 'POST'**,**

headers**:** **{**

accept**:** 'application/json'**,**

'content-type'**:** 'application/json'**,**

authorization**:** 'Basic YzNCaGJXSmhaekl3TURCQVoyMWhhV3d1WTI5dDpibVdMdHVoVTN1WkJiYURsY2dCSXc='

**},**

body**:** JSON**.**stringify**({**

source\_url**:** 'https://d-id-public-bucket.s3.us-west-2.amazonaws.com/alice.jpg'**,**

script**:** **{**

type**:** 'text'**,**

subtitles**:** 'false'**,**

provider**:** **{**type**:** 'microsoft'**,** voice\_id**:** 'Sara'**},**

input**:** 'Making videos is easy with D-ID'**,**

ssml**:** 'false'

**},**

config**:** **{**fluent**:** 'false'**}**

**})**

**};**

fetch**(**url**,** options**)**

**.**then**(**res **=>** res**.**json**())**

**.**then**(**json **=>** console**.**log**(**json**))**

**.catch(**err **=>** console**.**error**(**err**));**

1. Características técnicas

|  |  |  |  |
| --- | --- | --- | --- |
| Parámetro | Build | Launch | Scale |
| Precios (paquete anual, precio mensual) | 12.63 € | 30.71 € | 121.60€ |
| Tiempo de vídeo consumible al mes | 16 minutos | 45 minutos | 200 minutos |
| Tiempo de Streaming consumible al mes | 32 minutos | 90 minutos | 400 minutos |
| Tiene Streaming | Sí | Sí | Sí |
| Streaming con SDK | Sí | Sí | Sí |
| Tiene traducción de video | Sí | Sí | Sí |
| Resolución videos de avatar basado en foto (D-ID Technical Support, 2024) | Depende de la resolución de la imagen subida. | Depende de la resolución de la imagen subida. | Depende de la resolución de la imagen subida. |
| Resolución vídeos avatar | 1080p | 1080p | 1080p |
| Tiempo máximo de video | 5 minutos | 5 minutos | 5 minutos |
| Sincronización labial | Normal | Normal | Normal |
| Número avatares | 0 (sólo avatares de foto) | 111 | 111 |
| Avatares Personales | 1 | 3 | 5 |
| Voces (Proveedores) | 127 voces (+idiomas) | 127 voces (+idiomas) | 127 voces (+idiomas) |
| Expresiones del avatar | Neutral, Feliz, Serio, Sorpresa | Neutral, Feliz, Serio, Sorpresa | Neutral, Feliz, Serio, Sorpresa |
| Clonado de voz | No | 1 voz | 3 voces |
| Idiomas soportados | 119 lenguajes | 119 lenguajes | 119 lenguajes |
| Entrada alternativa | Audio | Audio | Audio |
| Licencia y uso comercial | No comercial | Comercial | Comercial |
| Calidad API (Documentación) | Normal | Normal | Normal |
| Velocidad de generación de vídeo | Baja | Media | Alta |

Tabla A2: Características D-ID

# **Synthesia**

1. Visión general

La API de Synthesia proporciona una interfaz RESTful orientada a la generación de videos mediante avatares virtuales con apariencia humana. Estos videos se crean a partir de texto proporcionado por el usuario, que es interpretado mediante modelos de texto a voz (TTS) y renderizado por presentadores digitales predefinidos. La API está diseñada para integrarse en flujos de trabajo automatizados, permitiendo crear, gestionar y descargar videos de forma programática.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Bearer Token
* **API Referencia:** <https://docs.synthesia.io/reference/introduction>
* **Base URL**: <https://api.synthesia.io>
* **API versión**: V2.
* **Formatos admitidos**: JSON en entrada y salida, vídeo MP4 en salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /v2/videos | Crea un video con un avatar digital a partir de texto y parámetros de configuración |
| GET /v2/videos/{id} | Consulta el estado del video generado (pendiente, procesando, completado, error) |
| POST /v2/webhooks | Lista de Webhooks activos. Pare recibir notificaciones cuando ocurran ciertos eventos en Synthesia. |

1. Parámetros principales de creación de video

**const** url **=** 'https://api.synthesia.io/v2/videos'**;**

**const** options **=** **{**

method**:** 'POST'**,**

headers**:** **{**

accept**:** 'application/json'**,**

'content-type'**:** 'application/json'**,**

Authorization**:** 'AasdefeASFEAFE'

**},**

body**:** JSON**.**stringify**({**

test**:** **true,**

title**:** 'My first Synthetic video'**,**

visibility**:** 'private'**,**

aspectRatio**:** '16:9'**,**

input**:** **[**

**{**

scriptText**:** ' input: \'Hola, soy un avatar virtual creado con tecnología de inteligencia artificial. Mi propósito es ayudarte a comunicar de forma moderna, clara y profesional.\','**,**

avatar**:** '865f06e0-4199-4c5f-97bc-6b713544946e'**,**

avatarSettings**:** **{**horizontalAlign**:** 'center'**,** scale**:** 1**,** style**:** 'rectangular'**,** seamless**:** **false},**

background**:** 'green\_screen'**,**

backgroundSettings**:** **{**

videoSettings**:** **{**

shortBackgroundContentMatchMode**:** 'freeze'**,**

longBackgroundContentMatchMode**:** 'trim'

**}**

**},**

scriptLanguage**:** 'es'

**}**

**]**

**})**

**};**

fetch**(**url**,** options**)**

**.**then**(**res **=>** res**.**json**())**

**.**then**(**json **=>** console**.**log**(**json**))**

**.catch(**err **=>** console**.**error**(**err**));**

1. Características técnicas

|  |  |  |
| --- | --- | --- |
| Parámetro | Starter\* | Creator |
| Precios (paquete anual, precio mensual) | 16 € | 58 € |
| Tiempo de vídeo consumible al mes | 10 minutos | 30 minutos |
| Tiempo de Streaming consumible al mes | - | - |
| Tiene Streaming | No | No |
| Streaming con SDK | No | No |
| Tiene traducción de video | No | No |
| Resolución videos de avatar basado en foto | - | - |
| Resolución vídeos avatar | 720p | 1080p |
| Tiempo máximo de video | 5 minutos | 5 minutos |
| Sincronización labial | Alta | Alta |
| Número avatares | 125 | 180 |
| Avatares Personales | 3 | 5 |
| Voces (Proveedores) | 127 voces (+idiomas) | 127 voces (+idiomas) |
| Expresiones del avatar | No configurable | No configurable |
| Clonado de voz | No | No |
| Idiomas soportados | 140 | 140 |
| Entrada alternativa | Audio | Audio |
| Licencia y uso comercial | No comercial | No comercial |
| Calidad API (Documentación) | Baja | Baja |
| Velocidad de generación de vídeo | Media | Media |

\*El modo Starter no incluye integración con API

# **HeyGen**

1. Visión general

La API de HeyGen proporciona una interfaz RESTful diseñada para la generación de contenido audiovisual mediante avatares digitales realistas. Permite generar vídeos a partir de texto en múltiples idiomas utilizando avatares predefinidos o personalizados, con capacidades avanzadas de integración en flujos de trabajo web y corporativos. Además, ofrece funcionalidad de streaming con respuesta en tiempo real mediante WebRTC y SDKs.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON) + WebSocket/WebRTC (streaming, SDK)
* **Autenticación**: Bearer Token
* **API Referencia:** <https://docs.heygen.com/docs/quick-start>
* **Base URL**: <https://api.heygen.com>
* **API versión**: V4.0.8
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /v1/video/generate | Crea un avatar digital a partir de texto / audio y parámetros configurables |
| GET /v1/video\_status.get | Consulta el estado de un vídeo |
| POST /v2/videos/{id}/generate | Crea un vídeo a partir de una plantilla |
| POST /v2/video\_translate | Traduce a otro idioma un vídeo. |
| POST /v1/streaming.new | Inicia una sesión Streaming |
| POST /v2/photo\_avatar/photo/generate | Crea un avatar a partir de una foto |

1. Parámetros principales de creación de video

**import** movioApi from '@api/movio-api'**;**

movioApi**.**auth**(**'<your-api-key>'**);**

movioApi**.**uploadAsset**()**

**.**then**(({** data **})** **=>** console**.**log**(**data**))**

**.catch(**err **=>** console**.**error**(**err**));**

1. Características técnicas

|  |  |  |
| --- | --- | --- |
| Parámetro | Pro | Scale |
| Precios (paquete anual) | 87.65 € | 292.16 € |
| Tiempo de vídeo consumible al mes | 100 minutos | 660 minutos |
| Tiempo de Streaming consumible al mes | 500 minutos | 3300 minutos |
| Tiene Streaming | Sí | Sí |
| Streaming con SDK | Sí | Sí |
| Tiene traducción de video | No | Sí |
| Resolución videos de avatar basado en foto | Depende de la resolución de la imagen subida. | Depende de la resolución de la imagen subida. |
| Resolución vídeos avatar | 1080p (Streaming: 72op, 480p, 360p) | 4k (Streaming: 72op, 480p, 360p) |
| Tiempo máximo de video | 5 minutos | 30 minutos |
| Sincronización labial | Alta | Alta |
| Número avatares (consulta a través de API) | 1169 + 1699 (avatar foto) | 1169 + 1699 (avatar foto) |
| Avatares Personales (Group generation) | 3 | 10 |
| Voces (consulta a través de API) | 1855 | 1855 |
| Expresiones del avatar | Voz (Excited / Friendly / Serious / Smoothing / Broadcaster), Postura (Stable / Expressive), Expresión (default, happy) | Voz (Excited / Friendly / Serious / Smoothing / Broadcaster), Postura (Stable / Expressive), Expresión (default, happy) |
| Clonado de voz | No definido | No definido |
| Idiomas soportados | 175 | 175 |
| Entrada alternativa | Audio | Audio |
| Licencia y uso comercial | Comercial | Comercial |
| Calidad API (Documentación) | Alta | Alta |
| Velocidad de generación de vídeo | 25 segundos – 1 minuto.(*average time for a video to be created*, no date) | 25 segundos – 1 minuto.(*average time for a video to be created*, no date) |

# **Elay**

1. Visión general

Elai es una plataforma de generación de vídeos mediante avatares digitales humanos que permite transformar texto en contenido audiovisual profesional. Su API está diseñada para automatizar la creación de vídeos, integrando funciones como selección de presentadores, idiomas, plantillas y configuración de música, todo a través de una interfaz RESTful. Está especialmente orientada a soluciones de aprendizaje corporativo para la generación de vídeos de avatares y la formación.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Bearer Token
* **API Referencia:** <https://elai.readme.io/reference/getting-started-with-your-api>
* **Base URL**: <https://api.elai.io>
* **API versión**: V1.0
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /api/v1/videos | Crea un vídeo |
| GET /api/v1/videos/{videoid} | Consulta el estado de un vídeo |
| POST /api/v1/streams | Crea un vídeo stream |
| POST /api/v1/videos/renderTemplate/{videoid} | Crea vídeos personalizados para clientes en base a un vídeo raíz. |
| POST /api/v1/videos/translate/{videoid} | Traduce un vídeo |
| POST /api/v1/avatars | Crea un avatar a partir de una foto |

1. Parámetros principales de creación de video

**import** elai from '@api/elai'**;**

elai**.**auth**(**'Bearer SECRET\_KEY'**);**

elai**.**createVideo**({**

name**:** 'Hello from API!'**,**

tags**:** **[**'test'**],**

**public:** **false,**

data**:** **{**

skipEmails**:** **false,**

subtitlesEnabled**:** 'false'**,**

format**:** '16\_9'**,**

musicUrl**:** 'https://elai-media.s3.eu-west-2.amazonaws.com/music/mixkit-driving-ambition-32.mp3'**,**

musicVolume**:** 0.17**,**

resolution**:** 'FullHD'

**}**

**})**

**.**then**(({** data **})** **=>** console**.**log**(**data**))**

**.catch(**err **=>** console**.**error**(**err**));**

1. Características técnicas

|  |  |  |
| --- | --- | --- |
| Parámetro | Basic | Advance |
| Precios (paquete mensual) | 20.22 € | 51.88 € |
| Tiempo de vídeo consumible al mes | 15 minutos (ampliable a 25, 40) | 100 minutos |
| Tiempo de Streaming consumible al mes | No se especifica. | No se especifica. |
| Tiene Streaming | No (Fase beta, sólo paquete Premium) | No (Fase beta, sólo paquete Premium) |
| Streaming con SDK | No | No |
| Tiene traducción de video | Sí | Sí |
| Resolución videos de avatar basado en foto | Depende de la resolución de la imagen subida. | Depende de la resolución de la imagen subida. |
| Resolución vídeos avatar | 1080p | 4k |
| Tiempo máximo de video | 5 minutos | 30 minutos |
| Sincronización labial | Normal | Normal |
| Número avatares | +80 | +80 |
| Avatares Personales | No (plugin aparte) | No (plugin aparte) |
| Voces | +450 | +450 |
| Expresiones del avatar | No | No |
| Clonado de voz | No definido | No definido |
| Idiomas soportados | +75 | +75 |
| Entrada alternativa | Audio | Audio |
| Licencia y uso comercial | Comercial | Comercial |
| Calidad API (Documentación) | Alta | Alta |
| Velocidad de generación de vídeo | 25 segundos – 1 minuto.(*average time for a video to be created*, no date) | 25 segundos – 1 minuto.(*average time for a video to be created*, no date) |

# **VEED**

1. Visión general

VEED.io es una plataforma SaaS centrada en la edición de vídeo directamente desde el navegador, que incluye la funcionalidad de generación de avatares parlantes (Talking Avatars). A diferencia de otros proveedores como D-ID, HeyGen o Elai, **VEED no ofrece una API pública**, lo que impide su integración en pipelines automatizados o en aplicaciones personalizadas mediante programación. Su propuesta de valor reside en la accesibilidad inmediata, una interfaz visual que no requiere conocimientos técnicos y herramientas complementarias como subtitulación automática, edición por texto y transcripción de audio.

1. Arquitectura general

* **Tipo de API**: No tiene. Se utiliza el estudio a través de navegador.
* **Autenticación**: -
* **API Referencia:** -
* **Base URL**: -
* **API versión**: -
* **Formatos admitidos**: Peticiones por navegador.

# **Colossyan**

1. Visión general

Colossyan se distingue por su enfoque especializado en formación corporativa interactiva. A diferencia de otros proveedores como D-ID o HeyGen, Colossyan permite crear conversaciones entre avatares, añadir cuestionarios y ramificaciones en los vídeos, y generar contenido compatible con SCORM para plataformas de e-learning. También destaca por ofrecer clonación de voz gratuita en todos los planes, algo inusual frente a soluciones que lo reservan para niveles Enterprise. Sin embargo, no cuenta con funcionalidades de streaming en tiempo real, y su API solo está disponible en planes superiores, lo que limita su acceso a integradores individuales o pequeños desarrollos.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Bearer Token
* **API Referencia:** <https://docs.colossyan.com/getting-started/quickstart>
* **Base URL**: <https://app.colossyan.com>
* **API versión**: V1.0
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /api/v1/video-generation-jobs/template-jobs | Crea un vídeo a partir de una plantilla |
| POST /api/v1/video-generation-jobs | Crea un vídeo manualmente |
| GET /api/v1/video-generation-jobs/{videoId} | Consulta el estado de un vídeo |
| GET /api/v1/generated-videos/{videoid} | Obtiene la información de un vídeo |

1. Parámetros principales de creación de video

**const** response **=** **await** fetch**(**'https://app.colossyan.com/api/v1/video-generation-jobs'**,** **{**

method**:** 'POST'**,**

headers**:** **{**

"Authorization"**:** "Bearer JWT"**,**

"Content-Type"**:** "application/json"

**},**

body**:** JSON**.**stringify**({**

"dynamicVariables"**:** **{},**

"callback"**:** "text"**,**

"callbackPayload"**:** **{},**

"videoCreative"**:** **{**

"settings"**:** **{**

"videoSize"**:** **{**

"width"**:** 1**,**

"height"**:** 1

**},**

"name"**:** "text"**,**

"alphaChannel"**:** **true**

**},**

"scenes"**:** **[**

**{**

"name"**:** "text"**,**

"duration"**:** 1**,**

"intrinsicDurationTrackReference"**:** "text"**,**

"transition"**:** **{**

"type"**:** "fade"**,**

"duration"**:** 1

**},**

"tracks"**:** **[**

**{**

"type"**:** "actor"**,**

"variant"**:** "full\_body"**,**

"view"**:** "front"**,**

"actor"**:** "text"**,**

"audioUrl"**:** "text"**,**

"text"**:** "text"**,**

"speakerId"**:** "text"**,**

"position"**:** **{**

"x"**:** 1**,**

"y"**:** 1

**},**

"size"**:** **{**

"width"**:** 1**,**

"height"**:** 1

**},**

"rotation"**:** 1**,**

"referenceId"**:** "text"

**}**

**]**

**}**

**]**

**}**

**})**

**});**

1. Características técnicas

|  |  |
| --- | --- |
| Parámetro | Business (único paquete con API) |
| Precios (paquete anual) | 61.15 € |
| Tiempo de vídeo consumible al mes | Ilimitado |
| Tiempo de Streaming consumible al mes | - |
| Tiene Streaming | No |
| Streaming con SDK | No |
| Tiene traducción de video | Sí |
| Resolución videos de avatar basado en foto | Depende de la resolución de la imagen subida. |
| Resolución vídeos avatar | 1080p |
| Tiempo máximo de video | 10 minutos |
| Sincronización labial | Normal |
| Número avatares | +170 |
| Avatares Personales | 10 |
| Voces | 70 |
| Expresiones del avatar | No |
| Clonado de voz | 2 |
| Idiomas soportados | 70 |
| Entrada alternativa | Audio |
| Licencia y uso comercial | Comercial |
| Calidad API (Documentación) | Bajo |
| Velocidad de generación de vídeo | No definido |

# **Speechify**

1. Visión general

Speechify es una plataforma centrada en la conversión de texto a voz (TTS), que ofrece una API RESTful diseñada para integrarse en aplicaciones, servicios y flujos automatizados. Su enfoque principal es la generación de voz realista y personalizada, con soporte para más de 60 idiomas, control por SSML y clonación de voz. A diferencia de proveedores como Elai, D-ID o HeyGen, Speechify no genera vídeos ni avatares, sino que se especializa en producción de audio con alta calidad y personalización, siendo ideal para aplicaciones centradas en accesibilidad, lectura de contenido o asistentes virtuales.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Bearer Token
* **API Referencia:** <https://docs.sws.speechify.com/v1/api-reference/api-reference/introduction>
* **Base URL**: <https://api.sws.speechify.com>
* **API versión**: V1
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /v1/audio/speech | Crea un audio a partir de texto |
| POST /v1/voices | Clona una voz |

1. Parámetros principales de creación de audio

from speechify **import** Speechify

client **=** Speechify**(**

token**=**"YOUR\_TOKEN"**,**

**)**

client**.**tts**.**audio**.**speech**(**

input**=**"input"**,**

voice\_id**=**"voice\_id"**,**

**)**

1. Características técnicas

|  |  |
| --- | --- |
| Parámetro | Plan Genérico para API |
| Precios | No especificado |
| Tiempo de vídeo consumible al mes | - |
| Tiempo de Streaming consumible al mes | - |
| Tiene Streaming | No |
| Streaming con SDK |  |
| Tiene traducción de video | No |
| Resolución videos de avatar basado en foto | - |
| Resolución vídeos avatar | - |
| Tiempo máximo de video | - |
| Sincronización labial | - |
| Número avatares | - |
| Avatares Personales | - |
| Voces | +1000 |
| Expresiones del avatar | No |
| Clonado de voz | Sí |
| Idiomas soportados | +50 |
| Entrada alternativa | - |
| Licencia y uso comercial | Comercial |
| Calidad API (Documentación) | Bajo |
| Velocidad de generación de audio | 300ms latencia |

# **Tavus**

1. Visión general

Tavus es una plataforma especializada en la generación automatizada y personalizada de vídeos mediante inteligencia artificial. Su propuesta se centra en la creación de vídeos personalizados a escala, donde el avatar del usuario puede ser generado a partir de una grabación original y replicar múltiples mensajes de forma dinámica. Permite generar un avatar personalizado del usuario para crear miles de vídeos con nombres, productos o segmentos variables, lo que la hace ideal para marketing automatizado y ventas personalizadas B2B.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON) + WebSocket/WebRTC (streaming)
* **Autenticación**: Bearer Token
* **API Referencia:** <https://www.postman.com/tavus-api/tavus/collection/393k3hs/tavus-platform-api-developer-beta>
* **Base URL**: <https://tavusapi.com>
* **API versión**: V2
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /v2/avatars/train | Crea un avatar a mediante un vídeo |
| POST /v2/requests | Crea un vídeo con un Avatar |
| POST /v1/requests | Crea un vídeo con el input dinámico |

1. Parámetros principales de creación de video

var settings = {

"url": "https://tavusapi.com/v2/requests",

"method": "POST",

"timeout": 0,

"data": "{\n \"avatar\_id\": 8252,\n \"script\": \"Hi Ari, how are you doing today? How is working on phoenix going?\",\n \"callback\_url\": \"https://webhook.site/dacb6637-3d5e-4fb3-b705-acc3a25c0ac1\"\n}",

};

$.ajax(settings).done(function (response) {

console.log(response);

});

1. Características técnicas

|  |  |  |
| --- | --- | --- |
| Parámetro | Starter | Growth |
| Precios | 51.75 € | 348.25 € |
| Tiempo de vídeo consumible al mes | 10 minutos | 100 minutos |
| Tiempo de Streaming consumible al mes | 100 minutos | 500 minutos |
| Tiene Streaming | Sí | Sí |
| Streaming con SDK | No | No |
| Tiene traducción de video | No | No |
| Resolución videos de avatar basado en foto | - | - |
| Resolución vídeos avatar | 1080p | 1080p |
| Tiempo máximo de video | 20 minutos | 60 minutos |
| Sincronización labial | Alta | Alta |
| Número avatares (consulta a través de API) | 25 | +100 |
| Avatares Personales (Group generation) | 3 | 7 |
| Voces (consulta a través de API) | 25 | +100 |
| Expresiones del avatar | No | No |
| Clonado de voz | Sí | Sí |
| Idiomas soportados | +30 | +30 |
| Entrada alternativa | No | No |
| Licencia y uso comercial | Comercial | Comercial |
| Calidad API (Documentación) | Normal | Normal |
| Velocidad de generación de vídeo | Rápida (1 segundo latencia) | Rápida (1 segundo latencia) |

# **Runway**

1. Visión general

Runway se diferencia de otras soluciones al centrarse en tareas avanzadas de edición y manipulación visual mediante inteligencia artificial, más allá de la simple generación de avatares parlantes. Su API permite crear vídeos completos a partir de texto (modelos Gen-3), eliminar fondos en tiempo real, aplicar estilos visuales y realizar inpainting inteligente en vídeo. A diferencia de APIs como D-ID o Tavus, que están orientadas a la personalización facial o avatar, Runway ofrece herramientas de postproducción creativa basadas en IA que permiten construir contenido desde cero o transformar material ya existente con un alto grado de control artístico y visual.

1. Arquitectura general

* **Tipo de API**: Integración
* **Autenticación**: Bearer Token
* **API Referencia:** <https://docs.dev.runwayml.com/>
* **Base URL**: <https://tavusapi.com>
* **API versión**: V2024-11-06
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

Esta API no está orientada a la generación de avatares reales, sino a la modificación de videos con avatares. No entra dentro del ámbito de las necesidades del proyecto.

1. Parámetros principales de creación de video

**import** RunwayML from '@runwayml/sdk'**;**

**const** client **=** **new** RunwayML**({**

apiKey**:** process**.**env**[**'RUNWAYML\_API\_SECRET'**],** // This is the default and can be omitted

**});**

**async** **function** main**()** **{**

**const** imageToVideo **=** **await** client**.**imageToVideo**.**create**({**

model**:** 'gen4\_turbo'**,**

promptImage**:** 'https://example.com/assets/bunny.jpg'**,**

promptText**:** 'The bunny is eating a carrot'**,**

**});**

console**.**log**(**imageToVideo**.**id**);**

**}**

main**();**

1. Características técnicas

|  |  |
| --- | --- |
| Parámetro | API Usage |
| Precios | 0.22 € / 5 segundos |
| Tiempo de vídeo consumible al mes | Sin límite |
| Tiempo de Streaming consumible al mes | - |
| Tiene Streaming | No |
| Streaming con SDK | Sí |
| Tiene traducción de video | No |
| Resolución videos de avatar basado en foto | - |
| Resolución vídeos avatar | 720p |
| Tiempo máximo de video | 5 segundos |
| Sincronización labial | - |
| Número avatares (consulta a través de API) | No |
| Avatares Personales (Group generation) | No |
| Voces (consulta a través de API) | No |
| Expresiones del avatar | No |
| Clonado de voz | No |
| Idiomas soportados | No |
| Entrada alternativa | Imagen, vídeo |
| Licencia y uso comercial | Comercial |
| Calidad API (Documentación) | Normal |
| Velocidad de generación de vídeo | No especifica |

# **A2E**

1. Visión general

A2E (Avatar to Emotion) es una plataforma que combina generación de vídeo con control emocional avanzado sobre avatares digitales. A diferencia de otras soluciones como D-ID o Tavus que se centran en la síntesis facial o voz, A2E destaca por permitir al usuario elegir un avatar y aplicar emociones realistas aumentando el tiempo de procesado de vídeo, obteniendo como salida un vídeo renderizado directamente desde su API. Esto lo hace especialmente útil para contenidos donde la expresividad facial del avatar es un factor determinante. Es una plataforma que está más orientada a avatares personales que los que pueda ofrecer el servicio.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Basic Auth mediante API Key
* **API Referencia:** <https://api.a2e.ai/>
* **Base URL**: <https://video.a2e.ai>
* **API versión**: V1
* **Formatos admitidos**: JSON en entrada y salida

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /api/v1/video/generate | Genera un video a partir de un avatar seleccionado. Si se utiliza el parámetro **isSkipRs** el procesado tarda más, pero la sincronización labial es mejor. |
| POST /api/v1/userVideoTwin/startTraining | Crea un avatar personal a partir de un vídeo o una imagen. |
| POST /api/v1/userVideoTwin/continueTraining | Continúa el entrenamiento de un avatar para la sincronización labial. |
| POST /api/v1/userImage2Video/start | A partir de una imagen crea un vídeo de 5 segundos con una calidad de 720p. Normalmente tarda 10 minutos en procesar. |
| POST /api/v1/streaming-avatar/set-qa-context | Crea un contexto para un avatar que tomará como base de conocimiento en un stream. |
| POST /api/v1/streaming-avatar/agora-token | Devuelve la información de un Avatar en Stream para conectarse a su canal |

Tabla A1: Métodos D-ID Vídeos

1. Parámetros del modelo crear vídeo

**var** settings **=** **{**

"url"**:** "https://video.a2e.ai/api/v1/video/generate"**,**

"method"**:** "POST"**,**

"timeout"**:** 0**,**

"headers"**:** **{**

"Content-Type"**:** "application/json"

**},**

"data"**:** JSON**.**stringify**({**

"title"**:** "My first AI video"**,**

"anchor\_id"**:** "660556b3cd492168141926d3"**,**

"anchor\_type"**:** 1**,**

"back\_id"**:** "62a6d3e56c209514b155b12b"**,**

"audioSrc"**:** "https://speech\_cdn.ai-anchor.com/dubbing\_stable/pyK1dvqcDKpKJo1IcEPm\_9714\_fr.mp3"**,**

"resolution"**:** 1080**,**

"web\_bg\_width"**:** 853**,**

"web\_bg\_height"**:** 480**,**

"web\_people\_width"**:** 270**,**

"web\_people\_height"**:** 480**,**

"web\_people\_x"**:** 292**,**

"web\_people\_y"**:** 0**,**

"isSkipRs"**:** **true,**

"isCaptionEnabled"**:** **true,**

"captionAlign"**:** **{**

"language"**:** "en-US"**,**

"PrimaryColour"**:** "rgba(247, 89, 171, 1)"**,**

"OutlineColour"**:** "rgba(0, 0, 0, 1)"**,**

"BorderStyle"**:** 4**,**

"BackColour"**:** "rgba(146, 84, 222, 1)"**,**

"FontName"**:** "AaFengKuangYuanShiRen"**,**

"Fontsize"**:** 50**,**

"subtitle\_position"**:** 0.3

**}**

**}),**

**};**

$**.**ajax**(**settings**).**done**(function** **(**response**)** **{**

console**.**log**(**response**);**

**});**

1. Características técnicas

|  |  |  |
| --- | --- | --- |
| Parámetro | Pay as You Go | Dedicated Line |
| Precios (precio mensual) | 8.77 € | 526.05 € |
| Tiempo de vídeo consumible al mes | 10 minutos | 694 minutos |
| Tiempo de Streaming consumible al mes | 40 minutos | 2777 minutos |
| Tiene Streaming | Sí | Sí |
| Streaming con SDK | No | No |
| Tiene traducción de video | No | No |
| Resolución videos de avatar basado en foto | Depende de la resolución de la imagen subida (recomendado 1080p). | Depende de la resolución de la imagen subida (recomendado 1080p). |
| Resolución vídeos avatar | 1080p | 1080p |
| Tiempo máximo de video | 5 minutos | 5 minutos |
| Sincronización labial | Alta | Alta |
| Número avatares | No especificado | No especificado |
| Avatares Personales | 1 | 3 |
| Voces (Proveedores) | No especificado | No especificado |
| Expresiones del avatar | Alta | Alta |
| Clonado de voz | Sí | Sí |
| Idiomas soportados | No especificado | No especificado |
| Entrada alternativa | Audio | Audio |
| Licencia y uso comercial | No comercial | Comercial |
| Calidad API (Documentación) | Buena | Buena |
| Velocidad de generación de vídeo | 1 minuto video = 5 minutos procesado | Media |

Tabla A2: Características D-ID



https://www.a2e.ai/volume-price-of-a2e-api/

**Plan Pay as You Go**: 8.77€ = 600 coins = 600 /60 = 10 minutos de vídeo = 600 coins/15 -> 40 minutos de streaming.

**Dedicated Line**: 5026.05€ = 41666 coins /mes = 41666/60 -> 694 minutos apróx de vídeo = 41666/15 -> 2777 minutos apróx de streaming.

# **AI Studios**

1. Visión general

AI Studios destaca por ofrecer avatares de alta fidelidad que simulan presentadores humanos reales, con enfoque en formación empresarial, anuncios corporativos y producción de contenido institucional. Permite crear vídeos a partir de texto con sincronización labial precisa y control total sobre idioma, voz, formato de salida y fondo.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON), integrada en el proyecto.
* **Autenticación**: Bearer Token
* **API Referencia:**  <https://v2.aistudios.com>
* **Base URL**: <https://aistudios.com>
* **Planes:** <https://www.aistudios.com/pricing>
* **API versión**: V3
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /api/odin/v3/simple/video | Crea un vídeo con un avatar y unas ropas definidas |
| GET /api/odin/v3/editor/progress/{projectKey} | Consulta un vídeo |
| POST /api/odin/v3/editor/Project/{projectKey} | Edita un vídeo |

Para Streaming, se usa un SDK integrado en la aplicación.

1. Parámetros principales de creación de video

curl https**:**//app.aistudios.com/api/odin/v3/editor/project \

-H "Authorization: ${API KEY}" \

-H "Content-Type: application/json" \

-X POST \

-d '{

"name"**:** "Test Project Using Api"**,**

"orientation"**:** "landscape"**,**

"scenes"**:** **[{**

"background" **:** **{**

"id" **:** "background"**,**

"type" **:** "background"**,**

"source\_type" **:** "image"**,**

"source\_url" **:** "/images/background/bg\_blue\_gradient.png"**,**

"source\_color" **:** "rgb(54,188,37)"

**},**

"watermark" **:** **false,**

"clips" **:** **[**

**{**

"id" **:** "aiModel-1h4ij5h8e87"**,**

"type" **:** "aiModel"**,**

"layer" **:** 1**,**

"top" **:** 144.5979042385424**,**

"left" **:** 630.2493927359487**,**

"script" **:** **{**

"org" **:** "<p>Hello, this is test video using Api.</p>"

**},**

"effects" **:** **[**

**{**

"type" **:** "head-only"

**}**

**],**

"height" **:** 2229**,**

"width" **:** 679**,**

"model" **:** **{**

"ai\_name" **:** "M000045058"**,**

"emotion" **:** "BG00002320"**,**

"language" **:** "en"**,**

"source\_url" **:** "https://cdn.aistudios.com/ai/ai\_mov\_thm/tight\_ai\_mov\_thm\_M000045058\_BG00002320.png"**,**

"editor" **:** **{**

"headCenterX" **:** 613.3333333333334**,**

"headCenterY" **:** 290**,**

"headWidth" **:** 182**,**

"headHeight" **:** 185**,**

"modelTightX" **:** 367.33333333333337**,**

"modelTightY" **:** 168.16666666666669**,**

"modelTightS" **:** 1**,**

"modelTightW" **:** 679**,**

"modelTightH" **:** 2229**,**

"modelOriginW" **:** 1374**,**

"modelOriginH" **:** 2444**,**

"scale" **:** 0.3**,**

"adjustX" **:** **-**0.016860747210092203**,**

"adjustY" **:** **-**0.024822695035461**,**

"spaceB" **:** 46.833333333333314**,**

"spaceT" **:** 168.16666666666669**,**

"spaceL" **:** 367.33333333333337**,**

"spaceR" **:** 327.66666666666663**,**

"top" **:** 168.16666666666669**,**

"left" **:** 367.33333333333337**,**

"height" **:** 2229**,**

"width" **:** 679

**},**

"origin" **:** **{**

"height" **:** 2444**,**

"width" **:** 1374

**},**

"deployImage" **:** **{**

"themb\_src" **:** "https://cdn.aistudios.com/ai/model-introduce/thumbnails/M000045058\_BG00002320.png"**,**

"themb\_width" **:** 384**,**

"themb\_height" **:** 240**,**

"org\_src" **:** "https://cdn.aistudios.com/ai/ai\_mov\_thm/tight\_ai\_mov\_thm\_M000045058\_BG00002320\_org.png"**,**

"org\_width" **:** 1374**,**

"org\_height" **:** 2444**,**

"edit\_src" **:** "https://cdn.aistudios.com/ai/ai\_mov\_thm/tight\_ai\_mov\_thm\_M000045058\_BG00002320.png"**,**

"edit\_width" **:** 692**,**

"edit\_height" **:** 2277

**},**

"deploySize" **:** **{**

"org\_width" **:** 1374**,**

"org\_height" **:** 2444**,**

"edit\_width" **:** 679**,**

"edit\_height" **:** 2229

**},**

"editorValue" **:** **{**

"headCenterX" **:** 613.3333333333334**,**

"headCenterY" **:** 290**,**

"headWidth" **:** 182**,**

"headHeight" **:** 185**,**

"modelTightX" **:** 367.33333333333337**,**

"modelTightY" **:** 168.16666666666669**,**

"modelTightS" **:** 1**,**

"modelTightW" **:** 679**,**

"modelTightH" **:** 2229**,**

"modelOriginW" **:** 1374**,**

"modelOriginH" **:** 2444**,**

"scale" **:** 0.3**,**

"adjustX" **:** **-**0.016860747210092203**,**

"adjustY" **:** **-**0.024822695035461**,**

"spaceB" **:** 46.833333333333314**,**

"spaceT" **:** 168.16666666666669**,**

"spaceL" **:** 367.33333333333337**,**

"spaceR" **:** 327.66666666666663

**},**

"maskFile" **:** "M000045058\_BG00002320H\_alpha\_INV.mp4"

**},**

"name" **:** "aiModel-1h4ij5h8e87"**,**

"lockScalingFlip" **:** **true,**

"fill" **:** "rgb(0,0,0)"**,**

"scaleX" **:** 1**,**

"scaleY" **:** 1**,**

"opacity" **:** 100**,**

"lockMovementX" **:** **false,**

"lockMovementY" **:** **false,**

"lockRotation" **:** **false,**

"lockScalingX" **:** **false,**

"lockScalingY" **:** **false,**

"lockSkewingX" **:** **false,**

"lockSkewingY" **:** **false,**

"lockUniScaling" **:** **false**

**}**

**],**

"thumbnailUrl" **:** null**,**

"sceneIdx" **:** 0

**}]**

**}**'

1. Características técnicas

|  |  |  |
| --- | --- | --- |
| Parámetro | Personal | Equipo |
| Precios (paquete anual, pago mensual) | 21.13 € | 158.49 € |
| Tiempo de vídeo consumible al mes | Ilimitado | Ilimitado |
| Tiempo de Streaming consumible al mes | No se especifica. | No se especifica. |
| Tiene Streaming | Sí (con SDK) | Sí (con SDK) |
| Streaming con SDK | Sí | Sí |
| Tiene traducción de video | No | Sí |
| Resolución videos de avatar basado en foto | Depende de la resolución de la imagen subida. | Depende de la resolución de la imagen subida. |
| Resolución vídeos avatar | 1080p | 1080p |
| Tiempo máximo de video | 10 minutos | 30 minutos |
| Sincronización labial | Normal | Normal |
| Número avatares | +80 | +80 |
| Avatares Personales | 3 | 5 |
| Voces | +100 | +100 |
| Expresiones del avatar | No | No |
| Clonado de voz | Sí | Sí |
| Idiomas soportados | +80 | +80 |
| Entrada alternativa | Audio | Audio |
| Licencia y uso comercial | Comercial | Comercial |
| Calidad API (Documentación) | Compleja (Baja) | Compleja (Baja) |
| Velocidad de generación de vídeo | No se especifica | No se especifica |

# **Akool**

1. Visión general

Akool se distingue por ofrecer funcionalidades como reenactment facial (lip sync de vídeos originales), face swap, traducción labial sincronizada y edición por prompt sobre imágenes y vídeos reales. Su fortaleza radica en crear contenido hiperpersonalizado con rostros reales, incluyendo traducción automática manteniendo la sincronización labial para un doblaje localizado y contenidos personalizados en múltiples idiomas. Sin embargo, no dispone de un sistema completo de generación de vídeo desde cero mediante texto, y su documentación y acceso API están más limitados que en otras plataformas más desarrolladas.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON), integrada en el proyecto.
* **Autenticación**: Bearer Token
* **API Referencia:**  <https://docs.akool.com/authentication/usage>
* **Base URL**: <https://openai.akool.com>
* **Planes:** <https://akool.com/pricing>
* **API versión**: V3
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /api/open/v3/faceswap/highquality/specifyvideo | Crea un vídeo suplantando las caras por otra imagen |
| POST /api/open/v3/content/video/createbytranslate | Traduce un vídeo |
| POST /api/open/v3/content/video/createbytalkingphoto | Crea un vídeo de un avatar mediante una foto como parámetro y un audio |
| POST /api/open/v3/content/video/lipsync | Hace sincronizar los labios con el audio. Recomendable vídeos <25fps |
| POST /api/open/v3/talkingavatar/create | Crea un vídeo de un avatar mediante un audio |
| POST /api/open/v4/liveAvatar/session/create | Crea una sesión streaming |

1. Parámetros principales de creación de video

**const** myHeaders **=** **new** Headers**();**

myHeaders**.**append**(**"Authorization"**,** "Bearer token"**);**

myHeaders**.**append**(**"Content-Type"**,** "application/json"**);**

**const** raw **=** JSON**.**stringify**({**

"width"**:** 3840**,**

"height"**:** 2160**,**

"avatar\_from"**:** 3**,**

"elements"**:** **[**

**{**

"type"**:** "image"**,**

"url"**:** "https://drz0f01yeq1cx.cloudfront.net/1729480978805-talkingAvatarbg.png"**,**

"width"**:** 780**,**

"height"**:** 438**,**

"scale\_x"**:** 1**,**

"scale\_y"**:** 1**,**

"offset\_x"**:** 1920**,**

"offset\_y"**:** 1080

**},**

**{**

"type"**:** "avatar"**,**

"url"**:** "https://drz0f01yeq1cx.cloudfront.net/1735009621724-7ce105c6-ed9a-4d13-9061-7e3df59d9798-7953.mp4"**,**

"scale\_x"**:** 1**,**

"scale\_y"**:** 1**,**

"width"**:** 1080**,**

"height"**:** 1080**,**

"offset\_x"**:** 1920**,**

"offset\_y"**:** 1080

**},**

**{**

"type"**:** "audio"**,**

"url"**:** "https://drz0f01yeq1cx.cloudfront.net/1729666642023-bd6ad5f1-d558-40c7-b720-ad729688f814-6403.mp3"

**}**

**]**

**});**

**const** requestOptions **=** **{**

method**:** "POST"**,**

headers**:** myHeaders**,**

body**:** raw**,**

redirect**:** "follow"

**};**

fetch**(**"https://openapi-test.akool.io/api/open/v3/talkingavatar/create"**,** requestOptions**)**

**.**then**((**response**)** **=>** response**.text())**

**.**then**((**result**)** **=>** console**.**log**(**result**))**

**.catch((**error**)** **=>** console**.**error**(**error**));**

1. Características técnicas

|  |  |  |  |
| --- | --- | --- | --- |
| Parámetro | PRO | PRO MAX | STUDIO |
| Precios (paquete anual, pago mensual) | 18.47 € | 69.48 € | 307.83 |
| Tiempo de vídeo consumible al mes | 10 mins | 40 mins | 200 mins |
| Tiempo de Streaming consumible al mes | 60 mins | 340 mins | 2000 mins |
| Tiene Streaming | Sí | Sí | Sí |
| Streaming con SDK | Sí | Sí | Sí |
| Tiene traducción de video | Sí | Sí | Sí |
| Resolución videos de avatar basado en foto | Depende de la resolución de la imagen subida. | Depende de la resolución de la imagen subida. | Depende de la resolución de la imagen subida. |
| Resolución vídeos avatar | 1080p / 4K | 1080p / 4K | 1080p / 4K |
| Tiempo máximo de video | No especificado | No especificado | No especificado |
| Sincronización labial | Alto\* | Alto\* | Alto\* |
| Número avatares | +60 (60 fotos) | +60 (100 fotos) | +60 (500 fotos) |
| Avatares Personales | 3 | 5 | 10 |
| Voces | No se especifica | No se especifica | No se especifica |
| Expresiones del avatar | No | No | No |
| Clonado de voz | Sí | Sí | Sí |
| Idiomas soportados | +155 | +155 | +155 |
| Entrada alternativa | No | No | No |
| Licencia y uso comercial | Comercial | Comercial | Comercial |
| Calidad API (Documentación) | Normal | Normal | Normal |
| Velocidad de generación de vídeo | No se especifica | No se especifica | No se especifica |

\*Con un procesado especial, pagando (POST /api/open/v3/content/video/lipsync)

# **Vidnoz**

1. Visión general

Vidnoz Su principal fortaleza es ofrecer funciones profesionales a precios asequibles, con un plan gratuito limitado, aunque su API es básica y no está orientada a grandes integraciones técnicas.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Bearer Token
* **API Referencia:**  <https://es.vidnoz.com/docs/authentication.html>
* **Base URL**: <https://devapi.vidnoz.com>
* **Planes:** <https://www.vidnoz.com/api-pricing.html>
* **API versión**: V2
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /v2/task/ai-headshots | Crea una imagen para currículum a partir de una imagen genérica |
| POST /v2/task/generate-taking-head | Crea un vídeo de un avatar a partir de una imagen |

1. Parámetros principales de creación de video

curl **--**request POST \

**--**url https**:**//devapi.vidnoz.com/v2/task/generate-talking-head \

--header "accept: application/json" \

--header "content-type: multipart/form-data" \

--header "Authorization: Bearer $API\_KEY" \

--form 'voice\_id="YmUxSTFBcVA1V1A3N2Y3L2hxMHhGQT09"' \

--form 'text="Input your text here"' \

--form 'avatar="YmUxSTFBcVA1V1A3N2Y3L2hxMHhGQT09"' \

--form 'avatar\_url="https://static.vidnoz.com/system/openapi/sample/talking-photo.jpg"' \

1. Características técnicas

|  |  |  |
| --- | --- | --- |
| Parámetro | PRO | ESCALA |
| Precios (se paga por mes) | 69.27€ (1000 créditos) | 262.16€ (6000 créditos) |
| Tiempo de vídeo consumible al mes | 250 mins | 1500 mins |
| Tiempo de Streaming consumible al mes | - | - |
| Tiene Streaming | No | No |
| Streaming con SDK | No | No |
| Tiene traducción de video | No | No |
| Resolución videos de avatar basado en foto | Depende de la resolución de la imagen subida. | Depende de la resolución de la imagen subida. |
| Resolución vídeos avatar | No especificado | No especificado |
| Tiempo máximo de video | No especificado | No especificado |
| Sincronización labial | Normal | Normal |
| Número avatares | +50 | +50 |
| Avatares Personales | No. Sólo fotos. | No. Sólo fotos |
| Voces | No se especifica | No se especifica |
| Expresiones del avatar | No | No |
| Clonado de voz | Sí | Sí |
| Idiomas soportados | +100 | +100 |
| Entrada alternativa | Audio | Audio |
| Licencia y uso comercial | Comercial | Comercial |
| Calidad API (Documentación) | Baja | Baja |
| Velocidad de generación de vídeo | No se especifica | No se especifica |

# **Pipio**

1. Visión general

Vidnoz Su principal fortaleza es ofrecer funciones profesionales a precios asequibles, con un plan gratuito limitado, aunque su API es básica y no está orientada a grandes integraciones técnicas.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Bearer Token
* **API Referencia:**  <https://docs.pipio.ai/>
* **Base URL**: <https://generate.pipio.ai>
* **Planes:**
  + <https://www.pipio.ai/pricing>
  + <https://www.g2.com/products/pipio/pricing>
* **API versión**: -
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /sigle-clip | Crea un vídeo con un avatar |
| POST /project/generate/dubbing | Traduce un vídeo |
| POST /Project/{PROJECT\_ID}/template | Modifica el template de un video |

1. Parámetros principales de creación de video

curl **-**X POST https**:**//generate.pipio.ai/single-clip \

-H 'Authorization: Key YOUR\_API\_KEY' \

-H 'Content-Type: application/json' \

-d '{

"actorId"**:** "62701420d70c840289b9ca00115bde05"**,**

"voiceId"**:** "0fdc62c9d2007363443ad76665551a6a"**,**

"script"**:** "Welcome to Pipio API. Lets create amazing avatar videos together"

**}**'

1. Características técnicas

|  |  |  |
| --- | --- | --- |
| Parámetro | PRO | TEAM |
| Precios (se paga por año, precio por mes) | 14.06€ (10 créditos) | 505.11€ (500 créditos) |
| Tiempo de vídeo consumible al mes\* | 10 mins | 500 mins |
| Tiempo de Streaming consumible al mes | - | - |
| Tiene Streaming | No | No |
| Streaming con SDK | No | No |
| Tiene traducción de video | Sí | Sí |
| Resolución videos de avatar basado en foto | No | No |
| Resolución vídeos avatar | 1080p | 4k |
| Tiempo máximo de video | No especificado | No especificado |
| Sincronización labial | Normal | Normal |
| Número avatares | +60 | +60 |
| Avatares Personales | 2 | 2 |
| Voces | +140 | +140 |
| Expresiones del avatar | No | No |
| Clonado de voz | Sí | Sí |
| Idiomas soportados | +40 | +40 |
| Entrada alternativa | Audio | Audio |
| Licencia y uso comercial | Comercial | Comercial |
| Calidad API (Documentación) | Baja | Baja |
| Velocidad de generación de vídeo | Normal | Alta |

\* Tiempo de vídeo en créditos: <https://help.pipio.ai/en/articles/9886186-what-are-credits>

# **Creatify**

1. Visión general

Creatify genera vídeos de avatares, su fortaleza radica en automatizar la creación de anuncios, no en ofrecer control detallado sobre el avatar o permitir integraciones profundas en sistemas conversacionales dinámicos.

Sólo el plan PRO tiene acceso a la API.

1. Arquitectura general

* **Tipo de API**: RESTful (HTTP sobre JSON)
* **Autenticación**: Bearer Token
* **API Referencia:**  <https://creatify.mintlify.app/api-reference/introduction>
* **Base URL**: <https://api.creatify.ai>
* **Planes:** <https://creatify.ai/pricing>
* **API versión**: -
* **Formatos admitidos**: JSON en entrada y salida.

1. Funcionalidades principales

|  |  |
| --- | --- |
| Endpoint | Descripción técnica |
| POST /api/link\_to\_videos | Crea un vídeo para un anuncio mediante un enlace a una url |
| POST /api/lipsyncs\_v2 | Convierte texto o audio reproducido por una persona hablando |
| POST /api/api\_scripts | Genera un vídeo con IA con una combinación entre la URL introducida y un texto descriptivo |
| POST /api/ai\_shorts | Crea un vídeo viral a través de un texto |
| POST /api/text\_to\_speech | Genera un vídeo con un avatar |
| POST /api/personas\_v2 | Crea un avatar. Tarda entre 2 y 5 días |

1. Parámetros principales de creación de video

**const** options **=** **{**

method**:** 'POST'**,**

headers**:** **{**

'X-API-ID'**:** '<api-key>'**,**

'X-API-KEY'**:** '<api-key>'**,**

'Content-Type'**:** 'application/json'

**},**

body**:** '{"script":"<string>","accent":"3c90c3cc-0d44-4b50-8888-8dd25736052a","webhook\_url":"<string>"}'

**};**

fetch**(**'https://api.creatify.ai/api/text\_to\_speech/'**,** options**)**

**.**then**(**response **=>** response**.**json**())**

**.**then**(**response **=>** console**.**log**(**response**))**

**.catch(**err **=>** console**.**error**(**err**));**

1. Características técnicas

|  |  |
| --- | --- |
| Parámetro | PRO |
| Precios (se paga por año, precio por mes) | 42.97€ (2400 créditos) |
| Tiempo de vídeo consumible al mes | 20 mins |
| Tiempo de Streaming consumible al mes | - |
| Tiene Streaming | No |
| Streaming con SDK | No |
| Tiene traducción de video | No |
| Resolución videos de avatar basado en foto | No |
| Resolución vídeos avatar | 4k |
| Tiempo máximo de video | 10 mins |
| Sincronización labial | Normal |
| Número avatares | +700 |
| Avatares Personales | 3 |
| Voces | +170 |
| Expresiones del avatar | No |
| Clonado de voz | NO |
| Idiomas soportados | +29 |
| Entrada alternativa | Audio |
| Licencia y uso comercial | Comercial |
| Calidad API (Documentación) | Alta |
| Velocidad de generación de vídeo | Alta |

# **Normalización de los datos**

1. Normalización de columnas para el sistema CBR

A continuación se describe el proceso de normalización aplicado a las columnas de la tabla maestra de proveedores de APIs de avatares virtuales, con el objetivo de transformar los datos a un formato uniforme, interpretable y comparable por un sistema de Razonamiento Basado en Casos (CBR).

**1. Columna: “Precio (€/mes)”**

* **Tipo**: Numérico continuo
* **Formato**: Decimal en euros (€)
* **Criterio**: Se utiliza tal cual está para comparación directa o se puede normalizar entre [0, 1] si se requiere similitud relativa. Puede aplicarse un rango máximo o escalado min-max.

**2. Columna: “Vídeo mensual (min)”**

* **Tipo**: Numérico entero
* **Formato**: Minutos
* **Criterio**:
  + Valores como “Ilimitado” se transforman en un valor alto estandarizado (por ejemplo, 10000).
  + Vacíos o no definidos se convierten en 0.
  + Posible normalización entre [0, 1] si se requiere similitud relativa.

**3. Columna: “Streaming mensual (min)”**

* **Tipo**: Numérico entero
* **Criterio**:
  + Mismo criterio que el anterior: “Ilimitado” → 10000, vacío → 0.

**4. Columna: “Tiene Streaming”**

* **Tipo**: Booleano
* **Formato**: 0 (No), 1 (Sí)

**5. Columna: “Streaming con SDK”**

* **Tipo**: Booleano
* **Formato**: 0 (No), 1 (Sí)

**6. Columna: “Traducción”**

* **Tipo**: Booleano
* **Formato**: 0 (No), 1 (Sí)

**7. Columna: “Resolución”**

* **Tipo**: Escala ordinal
* **Formato propuesto**:
  + 0 = SD (<= 720p)
  + 1 = HD (1080p)
  + 2 = 4K
* **Criterio**: Se extrae el número de la resolución (720, 1080, 4000) y se transforma según los rangos anteriores. No especificado = “0”

**8. Columna: “Duración máxima vídeo”**

* **Tipo**: Numérico entero
* **Formato**: Minutos
* **Criterio**:
  + Valores como “5 segundos” se transforman a 0.08 min
  + “Ilimitado” → 10000
  + Sin definir → 5 miuntos

**9. Columna: “Sincronización Labial”**

* **Tipo**: Escala ordinal
* **Formato**:
  + 0 = Baja
  + 1 = Normal
  + 2 = Alta

**10. Columna: "Clonado de Voz"**

* **Tipo**: Booleano
* **Formato**: 0 (No), 1 (Sí)
* **Criterio**: Se asigna 1 si el proveedor ofrece clonación de una o más voces. Se asigna 0 si no ofrece o si está indicado como "No definido".

**11. Columna: "Expresiones del avatar"**

* **Tipo**: Booleano
* **Formato**: "Sí" = 1, "No" = 0
* **Criterio**: Se asigna "1" si existe configuración de expresiones, aunque no se detalle cuáles. "0" si el campo es "No", "No configurable" o está vacío.

**12. Columnas numéricas con símbolo "+" (ej. "+170")**

* **Tipo**: Numérico entero
* **Criterio**: Se elimina el signo y se toma el valor numérico conocido. Si el campo es "No definido" o vacío, se asigna 0.
* **Columnas afectadas**:
  + "Número avatares"
  + "Avatares Personales"
  + "Voces"
  + "Idiomas soportados"

**13. Columna: "Entrada alternativa"**

* **Tipo**: Numérico categórico
* **Formato**:
  + 0 = Solo texto
  + 1 = Audio
  + 2 = Audio + imagen/vídeo
* **Criterio**: Se evalúa el contenido textual y se asigna el valor según la combinación de entradas.

**14. Columna: "Licencia y uso comercial"**

* **Tipo**: Booleano (representado como numérico)
* **Formato**: 0 (No comercial), 1 (Comercial)
* **Criterio**: Se asigna 1 si el proveedor permite uso comercial. Caso contrario, se asigna 0.

**15. Columna: "Calidad API (Documentación)"**

* **Tipo**: Escala ordinal
* **Formato**:
  + 0 = Baja
  + 1 = Media (o "Normal")
  + 2 = Alta

**16. Columna: "Velocidad de generación de vídeo"**

* **Tipo**: Escala ordinal
* **Formato**:
  + 0 = Baja
  + 1 = Media
  + 2 = Alta
* **Criterios específicos**:
  + "25s-1min", "Normal", "Media", "1 minuto", "No especifica" → 1
  + "300ms latencia", "1 segundo latencia", "Alta" → 2
  + "5 min por 1 min procesado", "Baja" → 0





1. Cálculo de la Similitud Total en el Sistema CBR

Para evaluar la idoneidad de un proveedor respecto a un caso de entrada definido por el usuario, se utiliza una función de similitud ponderada. Cada proveedor se representa como un caso con atributos normalizados entre 0 y 1. La similitud total se calcula como:

**Fórmula General:**

***SimilitudTotal = ∑ (Similitud\_i \* Peso\_i)***

Donde:

* *Similitud\_i* es la similitud individual del atributo *i*, normalizada entre 0 y 1.
* *Peso\_i* es el peso asignado a ese atributo según su importancia relativa.

**Ponderación por Característica Técnica**

|  |  |  |
| --- | --- | --- |
| Atributo | Peso | Fórmula de Similitud Individual |
| Licencia comercial | 0.15 | 1 si igual, 0 si distinto |
| Precio | 0.15 | 1 si ≤ deseado, si >: max(0, 1 - (precio - tope)/tope) |
| Tiene streaming | 0.10 | 1 si igual, 0 si distinto |
| Calidad API | 0.10 | 1 - abs(valor\_usuario - v\_proveedor)/2 |
| Clonado de voz | 0.10 | 1 si igual, 0 si distinto |
| Resolución del video | 0.05 | 1 - abs(valor\_usuario - v\_proveedor)/2 |
| Minutos de video/mes | 0.05 | 1 si ≥ deseado, si <: valor/deseado |
| Minutos de streaming/mes | 0.05 | 1 si ≥ deseado, si <: valor/deseado |
| Traducción automática | 0.05 | 1 si igual, 0 si distinto |
| Sincronización labial | 0.05 | 1 - abs(valor\_usuario - v\_proveedor)/2 |
| Expresiones del avatar | 0.03 | 1 si igual, 0 si distinto |
| Avatares personales | 0.02 | min(valor/valor\_deseado, 1) |
| Número de voces | 0.01 | min(valor/valor\_deseado, 1) |
| Idiomas soportados | 0.01 | min(valor/valor\_deseado, 1) |
| Duración máxima de video | 0.01 | 1 si ≥ deseado, si <: valor/deseado |
| Entrada alternativa | 0.01 | 1 - abs(valor\_usuario - v\_prov)/2 |
| SDK de streaming | 0.01 | 1 si igual, 0 si distinto |
| Velocidad de generación | 0.01 | 1 - abs(valor\_usuario - v\_proveedor)/2 |

1. Elección de proveedor para el trabajo

Se pretende seleccionar el proveedor más adecuado de servicios de generación de avatares virtuales con base en las necesidades técnicas y limitaciones presupuestarias propias para un TFM.

Las condiciones requeridas para el caso son:

* El plan debe costar **menos de 50 € al mes**.
* Al menos **50 minutos de vídeo generables al mes**, para cubrir pruebas, prototipos y posibles entregas funcionales.
* Al menos **100 minutos de streaming al mes**, para demostraciones en vivo o contenido dinámico.
* El proveedor debe ofrecer **streaming en tiempo real**.
* Debe contar con **licencia comercial**, para permitir la publicación de resultados o demostraciones en plataformas externas.
* Debe incluir **clonado de voz**, con el fin de personalizar los avatares utilizados.
* Debe permitir al menos **100 avatares distintos**, para evaluar diferentes estilos o identidades virtuales durante el desarrollo.



1. Proveedor obtenido mediante CBR

El proveedor seleccionado mediante el sistema CBR ha sido D-ID, en su modalidad Launch, al obtener una **similitud total de 0.59** respecto a los requisitos establecidos en el caso de entrada.

# **Proveedores fuera de estudio**

Este apartado identifica a los proveedores que no entran dentro del estudio del arte; los motivos son dos: no tienen API o es necesaria invitación para poder utilizar la API.

* Fliki.ai: <https://fliki.ai/enterprise/text-to-speech-api>
  + Requiere autenticación privada para usar la API.
* Zebracat.ai: <https://www.zebracat.ai/>
  + No tiene API
* InVideo.io: <https://invideo.io/ai/ai-image-generator/>
  + No tiene API.
* Picsart.com: <https://picsart.com/blog/api-program/>
  + Requiere invitación para usar la API
* SoulsMachines: <https://soulsmachines.com>
  + No tiene API.

# **Especificación de Requisitos**

1. Introducción

Este apartado recoge los requisitos funcionales y particulares que definen el comportamiento esperado del sistema a desarrollar. Estos requisitos sirven como base para el diseño, implementación y validación de la aplicación.

* 1. Objetivos generales

El objetivo principal del proyecto es desarrollar una aplicación que permita la creación de vídeos tanto en modo offline como online, utilizando avatares realistas basados en inteligencia artificial, excluyendo avatares animados o modelos 3D.

La aplicación permitirá difundir estos vídeos en redes sociales si el administrador lo considera oportuno.

La aplicación ofrecerá la posibilidad de generar hasta tres avatares personalizados y una voz personalizada que podrá ser utilizada en los vídeos, permitiendo que el avatar se exprese con dicha voz mediante tecnologías de IA.

* 1. Catálogo de requisitos

**REQUISITOS FUNCIONALES**

* **R.F.-1 Inicio de sesión (Loguin):** El administrador debe poder autenticarse para acceder a la aplicación.
* **R.F.-2 Gestión de parámetros**: La aplicación debe mostrar las opciones de “Crear Avatar”, “Crear Voz” y “Configuración de Avatar en Streaming”. La configuración del avatar en streaming deberá adaptarse al contexto de uso, ya sea comercial o divulgativo.

La API de D-ID, utilizada para la creación de agentes virtuales en tiempo real, permite establecer un *prompt* contextual (a través de OpenAI) que condiciona el comportamiento del avatar. Para ello, la aplicación deberá contar con una interfaz que permita configurar estos parámetros.

* + **R.F.-3 Creación de Avatar:** El administrador podrá crear o modificar hasta tres avatares personalizados.
  + **R.F.-4 Creación de voz**: Se podrá crear o editar una única voz personalizada para su uso en vídeos generados con avatar.
* **R.F.-5 Conversación con Avatar en Tiempo Real**: El usuario externo podrá mantener sesiones de conversación en tiempo real con el avatar, con una duración máxima de cinco minutos por sesión.
* **R.F.-6 Gestión de Vídeos**: El administrador podrá visualizar, organizar y gestionar todos los vídeos generados dentro de la aplicación.
* **R.F.-7 Crear / Traducir Vídeo:** La aplicación ofrecerá una interfaz para crear vídeos mediante avatares (propios o del proveedor), con posibilidad de traducirlos a otros idiomas.
* **R.F.-8 Publicar en redes sociales**: Se mostrarán las opciones necesarias para publicar vídeos ya procesados en redes sociales.

**REQUISITOS PARTICULARES**

* **R.P.-1**: El número máximo de avatares personalizados permitidos será de tres, y solo se podrá crear una voz personalizada, debido a las limitaciones impuestas por el proveedor.
* **R.P.-2**: Los vídeos generados deberán cumplir con los requisitos de calidad y formato para ser utilizados con fines comerciales.
* **R.P.-3**: La aplicación deberá informar al administrador del tiempo restante disponible para generar vídeos, en función de los minutos contratados según el plan de suscripción con el proveedor.
* **R.P.-4**: Para una mayor capacidad de control, se priorizará el uso del SDK del agente de streaming (avatar en tiempo real) en lugar de realizar llamadas HTTP (POST).
* **R.P.-5**: La aplicación debe permitir la publicación de vídeos generados en al menos una red social.
  1. Casos de Uso

|  |  |  |
| --- | --- | --- |
| CU-01 | Login de Usuario | |
| Requisitos relacionados | R.F.1 | |
| Descripción | El administrador debe poder acceder a la aplicación | |
| Precondiciones | El administrador debe conocer las credenciales y estar dado de alta en base de datos. | |
| Acciones | **Paso** | **Acción** |
| **1** | El administrador introduce su identificador y su contraseña |
| **2** | Pulsa el botón “Acceder” |
| Postcondiciones | Si las credenciales son correctas, el administrador accede a la aplicación | |
| Excepciones | Si las credenciales no son correctas, se le informa por pantalla que no tiene acceso a la aplicación | |
| Importancia | Alta | |

|  |  |  |
| --- | --- | --- |
| CU-02 | Gestión de parámetros | |
| Requisitos relacionados | R.F.2 | |
| Descripción | Permite al administrador ver los botones de “Crear Avatar”, “Crear Voz”. Puede configurar el comportamiento del avatar que visualizará el usuario en streaming. | |
| Precondiciones | El administrador debe haber iniciado sesión correctamente y haber accedido mediante el botón “Gestión de parámetros” | |
| Acciones | **Paso** | **Acción** |
| **1** | Configura el agente Streaming (para sesiones online) |
| **2** | Pulsa el botón “Guardar” |
| Postcondiciones | Si la configuración es correcta, saldrá un mensaje emergente indicando el éxito de la operación. Los datos quedarán guardados en el proveedor. | |
| Excepciones | Si la configuración no se ha guardado correctamente, aparecerá un mensaje de error. | |
| Importancia | Alta | |

|  |  |  |
| --- | --- | --- |
| CU-03 | Crear Avatar | |
| Requisitos relacionados | R.F.3 | |
| Descripción | Permite al administrador crear o modificar un avatar personalizado que podrá ser utilizado posteriormente en la generación de vídeos. | |
| Precondiciones | El usuario debe haber iniciado sesión y haber accedido desde “Gestión de Parámetros”. | |
| Acciones | **Paso** | **Acción** |
| **1** | Introduce los datos requeridos para el avatar |
| **2** | Pulsa el botón “Guardar” |
| Postcondiciones | Si la configuración es correcta, saldrá un mensaje emergente indicando el éxito de la operación y los datos quedarán guardados en el proveedor. | |
| Excepciones | Si la configuración no se ha guardado correctamente, aparecerá un mensaje de error. Si ya existen 3 avatares personalizados, mostrará un mensaje informativo. | |
| Importancia | Media | |

|  |  |  |
| --- | --- | --- |
| CU-04 | Crear Voz | |
| Requisitos relacionados | R.F.4 | |
| Descripción | Permite al administrador crear o modificar una voz personalizada que podrá ser utilizada posteriormente en la generación de vídeos. | |
| Precondiciones | El usuario debe haber iniciado sesión y haber accedido desde “Gestión de Parámetros”. | |
| Acciones | **Paso** | **Acción** |
| **1** | Introduce los datos requeridos para la voz |
| **2** | Pulsa el botón “Guardar” |
| Postcondiciones | Si la configuración es correcta, saldrá un mensaje emergente indicando el éxito de la operación y los datos quedarán guardados en el proveedor. | |
| Excepciones | Si la configuración no se ha guardado correctamente, aparecerá un mensaje de error. Si ya existen 1 voz personalizada, mostrará un mensaje informativo. | |
| Importancia | Media | |

|  |  |  |
| --- | --- | --- |
| CU-05 | Conversación con Avatar en Tiempo Real | |
| Requisitos relacionados | R.F.5 | |
| Descripción | Permite al usuario externo mantener una conversación interactiva en tiempo real con un avatar mediante streaming, con una duración máxima de 5 minutos por sesión. | |
| Precondiciones | El avatar en streaming debe estar correctamente configurado previamente y disponible. El usuario debe acceder a la sesión desde un enlace o interfaz habilitada. | |
| Acciones | **Paso** | **Acción** |
| **1** | El usuario accede al entorno de conversación en tiempo real con el avatar |
| **2** | El usuario inicia la conversación hablando o escribiendo. |
| **3** | Transcurrido el tiempo configurado en “Gestión de Parámetros”, la sesión finaliza automáticamente o puede finalizarse manualmente. |
| Postcondiciones |  | |
| Excepciones | Si el tiempo de sesión ha expirado, se notifica al usuario con un mensaje adecuado. | |
| Importancia | Alta | |

|  |  |  |
| --- | --- | --- |
| CU-06 | Gestión de vídeos | |
| Requisitos relacionados | R.F.6 | |
| Descripción | Permite al administrador visualizar, organizar y gestionar los vídeos que han sido generados dentro de la aplicación, ya sea para revisión, descargar o eliminación | |
| Precondiciones | El usuario debe haber iniciado sesión y accedido desde el menú principal | |
| Acciones | **Paso** | **Acción** |
| **1** | Se muestra una lista con los vídeos generados, junto con sus detalles |
| **2** | El usuario puede seleccionar un vídeo y realizar acciones como visualizarlo, descargarlo o eliminarlo. |
| Postcondiciones | Se actualiza el estado de los vídeos según las acciones realizadas. Los vídeos eliminados se retiran del proveedor | |
| Excepciones | Si no existen vídeos aparece la lista vacía. Si hay un fallo al cargar o eliminar un vídeo, se notifica el error. | |
| Importancia | Media | |

|  |  |  |
| --- | --- | --- |
| CU-07 | Crear / Traducir Vídeo | |
| Requisitos relacionados | R.F.7 | |
| Descripción | Permite al administrador crear un vídeo con un avatar (personal o del proveedor). También se da la opción de traducirlo a otros idiomas. | |
| Precondiciones | El usuario debe haber accedido a través de “Gestión de Vídeos” | |
| Acciones | **Paso** | **Acción** |
| **1** | Se introducen los parámetros y el texto o el audio para generar el vídeo |
| **2** | Se confirma la generación del vídeo. El sistema realiza las llamadas necesarias a la API. |
| Postcondiciones | Si todo ha ido bien, se vuelve a la gestión de vídeos, indicando que el proceso ha sido un éxito. | |
| Excepciones | Si ha habido algún error con la API, se notificará con un mensaje y se mantendrá en esa pantalla | |
| Importancia | Alta | |

|  |  |  |
| --- | --- | --- |
| CU-08 | Publicar en Redes Sociales | |
| Requisitos relacionados | R.F.8 | |
| Descripción | Permite al usuario seleccionar un vídeo previamente generado y publicarlo directamente en una o varias redes sociales a través de integraciones disponibles. | |
| Precondiciones | El usuario debe haber accedido a través de “Gestión de vídeos”. Además, debe tener configuradas las credenciales de acceso a las redes sociales | |
| Acciones | **Paso** | **Acción** |
| **1** | Se elige una o más redes sociales disponibles |
| **2** | Se añade la información en texto de la noticia o publicación |
| **3** | Pulsa el botón “Publicar” y el sistema realiza la publicación mediante la API correspondiente |
| Postcondiciones | El vídeo queda publicado en las redes sociales y se actualiza su estado en base de datos | |
| Excepciones | Si ocurre un error, se mostrará por pantalla | |
| Importancia | Media | |

Diagrama

El contenido generado por IA puede ser incorrecto.

# **Bibliografía**

*average time for a video to be created* (no date) *HeyGen API Documentation*. Available at: https://docs.heygen.com/discuss/64e7cee370767300739b4b55 (Accessed: 11 April 2025).

D-ID Technical Support (2024) ‘Especificar Resolución de salida’, *Specify output resolution*. Available at: https://docs.d-id.com/discuss/65cd0e2ad51b9a0017a6b182 (Accessed: 3 April 2025).

Moja (2024) ‘Razonamiento Basado en Casos: DESCUBRE Aplicaciones REALES’, 10 June. Available at: https://lanuevaia.com/razonamiento-basado-en-casos/ (Accessed: 2 May 2025).