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](#_Toc194509716)

[1.1 Introducción 1](#_Toc194509717)

[1.2 Planificación temporal 1](#_Toc194509718)

**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

REFERENCIAS APIS:

https://www.heygen.com/blog/ai-avatar-generator

# **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 analizan bajo un conjunto común de características técnicas que sirven como parámetros de comparación.

Sin embargo, durante el proceso de búsqueda y documentación, se ha comprobado que no todos los proveedores publican información de forma homogénea. Algunas funcionalidades son promocionadas pero no documentadas en detalle, otras están limitadas a determinados planes sin especificación clara, y en muchos casos no existe un acceso público completo a sus capacidades técnicas (por ejemplo, en APIs comerciales como Synthesia o HeyGen). Esto genera una dificultad añadida para normalizar los datos y establecer criterios comparativos sólidos.

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.
* **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 facilita la toma de decisiones al comparar soluciones existentes bajo un marco técnico y funcional común.

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?gad_source=1&gbraid=0AAAAABiJW6YkZnYquF-9xswBarx7FNOPL&gclid=Cj0KCQjw2ZfABhDBARIsAHFTxGwhrU53PKKA5tt4dIFx50pVGYYiGiEc6VSPh9p4Iy5OjL46MuuMQw4aAhYUEALw_wcB>

# **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 |
| 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 |
| 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)
* **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 |
| 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) |
| 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) |

# **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).

elai (no date) *Compare Elai Vs D-ID*, *Elai.io*. Available at: https://elai.io/d-id-vs-elai/ (Accessed: 21 April 2025).