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

# **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 /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 /animations | Permite aplicar expresiones o animaciones sobre imágenes |
| GET /voices / GET /presets | Consulta voces TTS y modelos de avatares disponibles |

Tabla A1: Métodos D-ID Vídeos

1. Parámetros del modelo talk

require**(**'dotenv'**).**config**();**

**const** axios **=** require**(**'axios'**);**

// Iniciar temporizador

console**.**time**(**"⏱️ Tiempo de respuesta"**);**

// URL del endpoint de D-ID para generación de vídeo

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

**const** data **=** **{**

script**:** **{**

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

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

provider**:** **{**

type**:** 'microsoft'**,**

voice\_id**:** 'Sara'**},**

**},**

config**:** **{**

fluent**:** **true,**

pad\_audio**:** 0.5**,**

output\_resolution**:** 1280

**},**

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

**};**

**const** headers **=** **{**

Authorization**:** `Basic ${process.env.DID\_API\_KEY}`**,**

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

**};**

axios**.**post**(**url**,** data**,** **{** headers **})**

**.**then**(**res **=>** **{**

console**.**log**(**'✅ Video generado correctamente'**);**

console**.**log**(**'➡️ Id:'**,** res**.**data**.**id**);**

console**.**timeEnd**(**"⏱️ Tiempo de respuesta"**);** // Mostrar duración en consola

**})**

**.catch(**err **=>** **{**

console**.**error**(**'❌ Error en la solicitud:'**);**

console**.**error**(**err**.**response**.**data**.**message**);**

console**.**timeEnd**(**"⏱️ Tiempo de respuesta"**);** // Mostrar duración en consola

**});**

Respuesta:

Texto

El contenido generado por IA puede ser incorrecto.

**Tiempo de respuesta**: 1.659 segundos.

1. Características técnicas

|  |  |
| --- | --- |
| Parámetro | Valor / Detalle técnico |
| Tiempo medio de generación | 1–20 segundos por video, variable según duración y carga del sistema. (PRUEBA: 1.659 seg). |
| Tiene Streaming | Sí |
| Renderizado en streaming | Hasta 100 FPS (modo stream), 4x velocidad real |
| Resolución estándar | Hasta 1280x1280 (D-ID Technical Support, 2024) px para presentadores normales |
| Resolución premium | Hasta 1920x1080 (1080p) en clips con HQ presenters |
| Formato de imagen de entrada | JPG / PNG ≥512x512px (idealmente cuadrado y frontal) |
| Calidad Avatar | Alta (rostro animedo + HQ Presenters con torso/manos) |
| Número avatares | Depende del servicio contratado. |
| Voces (Proveedores) | Microsoft Azure, ElevenLabs, Amazon Polly. |
| Contenido animado | Principalmente rostro; con HQ Presenters también incluye torso y manos |
| Animaciones del avatar | Neutral, Feliz, Grave, Sorpresa |
| Duración del vídeo | Definida por el texto/audio proporcionado |
| Idiomas soportados | +100 idiomas con múltiples voces TTS (Microsoft / ElevenLabs) |
| Entrada alternativa | Permite usar audio real en lugar de texto |
| Licencia y uso comercial | Permitido según plan contratado y términos de servicio |
| Integración técnica | Accesible desde Node.js, Python, Postman, curl o SDKs personalizados |

Tabla A2: Características D-ID

1. Precios

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Plan | Precio mensual | Minutos de vídeo | Minutos de streaming | Agentes | Sesiones | Marca de agua | Controles avanzados |
| Build | 13.05 € | 16 min | 32 min | 36 | 106 | Marca de agua D-ID | Expresión, tono, estilo de voz |
| Launch | 31.71 € | 45 min | 90 min | 119 | 294 | Marca de agua AI | Expresión, tono, estilo de voz |
| Scale | 125.56 € | 200 min | 400 min | 535 | 1,165 | Marca personalizada | Expresión, tono, estilo de voz |
| Enterprise | Personalizado | Escalable según necesidad | Escalable | + | + |  |  |

Tabla A3: Precios 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) |
| GET /v2/templates | Consulta las plantillas de video disponibles |
| GET /v2/videos/{id}/download | Devuelve el archivo de video una vez generado (MP4) |
| 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 | Valor / Detalle técnico |
| Tiempo medio de generación | 1–20 segundos por video, variable según duración y carga del sistema. (PRUEBA: 1.659 seg). |
| Tiene Streaming | No |
| Renderizado en streaming | - |
| Resolución estándar | 1920x1080 (1080p) [Diferentes formatos] |
| Resolución premium | - |
| Formato de imagen de entrada | - |
| Calidad Avatar | Muy Alta (visión frontal completa, sincronización labial superior) |
| Número Avatares | +150 |
| Voces (Proveedores) | +1000. No se especifican proveedores. |
| Contenido animado | Rostro o cuerpo. |
| Animaciones del avatar | No. Implícitas en el modelo elegido. |
| Duración del vídeo | Definida por el texto/audio proporcionado |
| Idiomas soportados | +140 |
| Entrada alternativa | No tiene. Se puede utilizar un audio subido previamente. |
| Licencia y uso comercial | Permitido según plan contratado y términos de servicio |
| Integración técnica | Accesible desde Node.js, Python, Postman, curl o SDKs personalizados |

1. Precios

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Plan | Precio mensual | Minutos de vídeo | Minutos de streaming | Agentes | Sesiones | Marca de agua | Controles avanzados |
| Starter | 26.27 € | 10 min (720p) | - | 1 estándar | - | - | Voz Limitada |
| Creator | 60.69 € | 20 min | - | 70 estándar | - | - | Sesión de voz TTS |
| Pro | 226.45 € | Personalizado | - | 70 estándar + Premium | - | - | + Control narrativo |
| Enterprise | Personalizado | Escalable según necesidad | - | 150 + personalizados | - | - | API SDK |

# **Bibliografía**

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).