

Especificación de requisitos de Software (ERS/SRS)

APP Y CONTEXTO SOFTWARE APIs CONEXIÓN



FISIOMAP IA ASSISTANT



Alexia Ainhoa Fernández Losper

Alexiafeernandez1234@gmail.com

https://github.com/a-fernandez21/Proyecto_Tfg.git

Colegio Vivas

ÍNDICE

1. Introducción	3
1.1 Descripción general del proyecto	3
1.2 Objetivos generales y específico	3
Objetivo general	3
Objetivos específicos	3
1.3 Justificación y motivación del proyecto	4
2.1 Contexto y necesidad detectada	4
2.2 Público objetivo	5
2.3 Requisitos funcionales	5
2.4 Casos de uso principales	6
3. Diseño y planificación	6
3.1 Tecnologías y herramientas empleadas	6
3.2 Arquitectura y estructura del sistema	6
4. Desarrollo del proyecto	7
4.1 Timeline (Prototipos)	7
Agente Telefónico	7
Flujo de procesamiento de Consultas Audiofónicas y Generación de Anamnesis.	12
4.2 Timeline Final	14
Agente Telefónico	14
4.3 Pantallas de la aplicación	16
Pantalla de Inicio de sesión	16
Grabadora Inteligente	17
Seleccionar paciente y tipo de grabación	18
Interfaz de la grabadora inteligente	19
Fin de grabación	20
Revisión final	20
5. Problemas encontrados	22
Problema encontrado: Transcripción en vivo	22
Descripción del problema	22
Solución propuesta	22

1. Introducción

1.1 Descripción general del proyecto

Fisiomap IA Assistant es una ampliación funcional de la aplicación web Fisiomap que permite grabar las sesiones con pacientes (desde el micrófono del teléfono) y, mediante técnicas de inteligencia artificial, transcribir y segmentar automáticamente las sesiones en secciones clínicas (anamnesis, diagnóstico, tratamiento, recomendaciones). Además incorpora un agente conversacional con voz para:

- Atención a pacientes (información, reserva/cambio de citas),
- Interacción administrativa (gestión de citas por personal autorizado mediante comandos de voz).

El objetivo es automatizar partes del registro clínico y mejorar la experiencia tanto de pacientes como del personal sanitario.

1.2 Objetivos generales y específico

Objetivo general

- Integrar capacidades de grabación, transcripción y análisis de voz basadas en IA en Fisiomap, y desarrollar una agente de voz para usuarios y administradores.

Objetivos específicos

- Implementar un módulo de grabación en la versión web móvil de Fisiomap.
- Enviar audio seguro al backend y obtener transcripción (ASR).

- Segmentar la transcripción en secciones clínicas (anamnesis, diagnóstico, etc.) automáticamente.
- Generar un resumen estructurado de la sesión para adjuntar al expediente del paciente.
- Desarrollar un agente de voz para atención a pacientes (info, citas) y para administradores (comandos autenticados).
- Garantizar cumplimiento de privacidad (GDPR) y seguridad de datos.

1.3 Justificación y motivación del proyecto

- Reduce la carga administrativa al automatizar registros verbales.
- Mejora la accesibilidad (pacientes que prefieren hablar o escribir).
- Aporta valor diferencial a Fisiomap con funciones de IA modernas.
- Permite optimizar el tiempo de los sanitarios y mejorar el control de citas.

2. Análisis del proyecto

2.1 Contexto y necesidad detectada

En clínicas de fisioterapia, gran parte del registro clínico se realiza a mano o en notas que consumen tiempo. Además, la gestión de citas suele ser tele- / telefonista-intensiva. Automatizar la transcripción y ofrecer un asistente de voz puede ahorrar tiempo y mejorar la trazabilidad de la información clínica.

2.2 Público objetivo

- Pacientes: registrar sesiones, acceder a información y reserva de citas.
- Fisioterapeutas / Médicos: obtener resumen automático de la sesión, reducir burocracia.
- Administradores / Recepción: gestionar citas mediante voz y UI tradicional.

2.3 Requisitos funcionales

1. Grabar audio desde el navegador móvil y subirlo al servidor.
2. Transcribir audio a texto (ASR) con timestamps por segmentos.
3. Detectar y etiquetar secciones clínicas (anamnesis, historia, exploración, diagnóstico, tratamiento, recomendaciones).
4. Generar resumen / informe estructurado descargable (PDF/HTML).
5. Exportar/adjuntar la transcripción y resumen al expediente del paciente en Fisiomap.
6. Agente conversacional por voz que:
 - atienda consultas básicas (horarios, servicios),
 - permita solicitar/confirmar/cancelar citas,
 - identifique administradores mediante credenciales/keywords y permita consultas de agenda.
7. Interfaz administrativa para revisar y corregir transcripciones/segmentaciones.
8. Registro y consentimiento explícito del paciente antes de grabar.
9. Control de acceso y logs de auditoría.

2.4 Casos de uso principales

- CU1 — Grabar sesión: paciente/técnico inicia grabación; al finalizar, audio se sube y se procesa; transcripción y secciones quedan disponibles.
- CU2 — Generar informe: a partir de la transcripción se crea un informe estructurado.
- CU3 — Asistente de voz (paciente): usuario pide información / cita; agente responde y puede agendar.
- CU4 — Asistente de voz (administrador): administrador autenticado consulta/modifica su agenda mediante comandos de voz.
- CU5 — Revisión manual: profesional revisa y corrige transcripción/segmentación antes de adjuntar al expediente.

3. Diseño y planificación

3.1 Tecnologías y herramientas empleadas

- Frontend (web móvil):
 - Framework: Flutter.
 - Lenguaje: Dart
- Backend:
 - Lenguaje: PhP
 - Base de datos: MySQL
 - Framework: n8n
- Herramientas auxiliares:
 - Para la voz usaremos Eleven Labs o Vapi
 - APIs Web:
 - Autenticación de usuarios (OAuth2)
 - Bases de datos

3.2 Arquitectura y estructura del sistema

Usuario (App Flutter)



Graba audio



Envía audio a n8n



Texto transcrito



Base de datos MySQL

Agente Telefónico (Vapi)



Interacción por voz con usuario



n8n ejecuta flujos automáticos



Registra datos → Google Sheets



Agenda citas → Google Calendar

4. Desarrollo del proyecto

4.1 Timeline (Prototipos)

Agente Telefónico

El **Agente Telefónico** de Integración de IA es un sistema automatizado que combina inteligencia artificial conversacional y flujos de automatización con n8n para gestionar llamadas telefónicas, programar citas y notificar a los usuarios a través de múltiples canales.

El agente se comunica de forma natural por voz, interpreta las intenciones del usuario mediante Vapi API (interfaz de voz conversacional basada en IA) y ejecuta acciones concretas mediante workflows de n8n, como crear o eliminar eventos en Google Calendar, enviar correos electrónicos de confirmación o enviar mensajes por WhatsApp.

- Flujo operativo:
 - Inicio de llamada:
 - El usuario llama al número asociado al agente (gestionado por Vapi).
 - Interacción conversacional:
 - Vapi transcribe la voz del usuario, interpreta su intención (por ejemplo: “quiero reservar una cita el jueves”) y envía una tool call a un endpoint de n8n (Webhook).
 - Procesamiento en n8n:
 - Se ejecuta el flujo correspondiente según la intención.
 - n8n consulta disponibilidad en Google Calendar.
 - Si procede, crea, elimina o consulta la cita.
 - Confirmaciones automáticas:
 - Tras la acción, n8n:
 - Envía un correo de confirmación.
 - Envía un mensaje de WhatsApp con los detalles de la cita.
 - Devuelve un resultado JSON a Vapi con el texto que debe pronunciar (por voz) al usuario.
 - Cierre de interacción:
 - Vapi comunica verbalmente la confirmación o el resultado al usuario y finaliza la llamada.
- Creación de el agente de voz para la IA mediante la herramienta de Vapi:
 - Generación de el prompt para el asistente de Vapi:
[Identity]
Eres Carla, la asistente virtual de FisioMap, una aplicación que conecta a fisioterapeutas con pacientes en España.
Tu función principal es gestionar citas, proporcionar información básica sobre los fisioterapeutas y coordinar

reservas o cancelaciones de sesiones a través de Fisiomap o Google Calendar.

También puedes verificar si el usuario es médico o cliente usando la herramienta isMedic, la cual debes invocar explícitamente cada vez que se detecte que el usuario dice "médico" o "cliente".

[Style]

Usa un tono amable, cercano y empático, propio de una clínica de fisioterapia.

Adapta tu tono según el comportamiento del usuario:

- Si es respetuoso → sé cálida y paciente.
- Si es sarcástico → sé seria y directa.
- Si insulta → responde de forma firme y seca.

Siempre hablas en español y usas la franja horaria Europe/Madrid.

Evita tecnicismos; prioriza la claridad y la empatía.

[Response Guidelines]

No confirmes ninguna acción sin verificar disponibilidad.

Mantén las respuestas breves y centradas en la gestión de citas.

Evita duplicar citas a la misma hora o con el mismo fisioterapeuta.

Confirma siempre la información antes de proceder.

Expresa números con palabras separadas (ej.: "siete-dos-dos tres-cuatro cinco-ocho cinco-ocho").

Expresa horas con palabras (ej.: "cuatro y media" en lugar de "4:30").

[Task & Goals]

Durante la conversación:

Saluda cordialmente y preséntate como Carla, la asistente virtual de Fisiomap.

Antes de cualquier otra acción, pregunta:

“¿Eres médico o cliente?”

Cuando el usuario responda:

Si dice “cliente”, debes llamar inmediatamente a la herramienta isMedic con valor false.

Ejemplo: “Entendido, eres cliente. Continuaré con el flujo normal de gestión de citas.”

Si dice “médico”, responde:

“Perfecto, por favor, indícame la clave de acceso.”

Luego:

Si la clave es exactamente “pepito de los palotes”, llama a la herramienta isMedic con valor true.

Si la clave no coincide, responde:

“Lo siento, la clave no es correcta. Solo puedo continuar como cliente.”

y llama a isMedic con valor false.

Una vez identificada la condición:

- Si isMedic = true → continúa hacia el flujo de médico en n8n.
- Si isMedic = false → continúa hacia el flujo de cliente en n8n.

Después de esa identificación, sigue el flujo normal:

Pregunta si desea reservar, cambiar o cancelar una cita.

Solicita nombre completo, correo electrónico y teléfono (+34).

Pregunta fecha, hora y fisioterapeuta antes de comprobar disponibilidad.

Gestiona la cita (crear/modificar/cancelar).

Confirma siempre el resultado y despídete cordialmente.

[Error Handling / Fallback]

Si el usuario no responde claramente, pide aclaración.

Si ocurre un error con una herramienta, discúlpate y reintenta.

Si no puedes resolverlo, ofrece derivar la consulta a un responsable de Fisiomap.

[Tools Integration]

- buscar_todo → Consultar disponibilidad o existencia de cita.
- agendar_calendario → Crear nueva cita.
- eliminar_evento → Cancelar cita.
- isMedic → Determina si el usuario es médico o cliente.

Si el usuario dice "médico" y la clave es "pepito de los palotes" → isMedic(true).

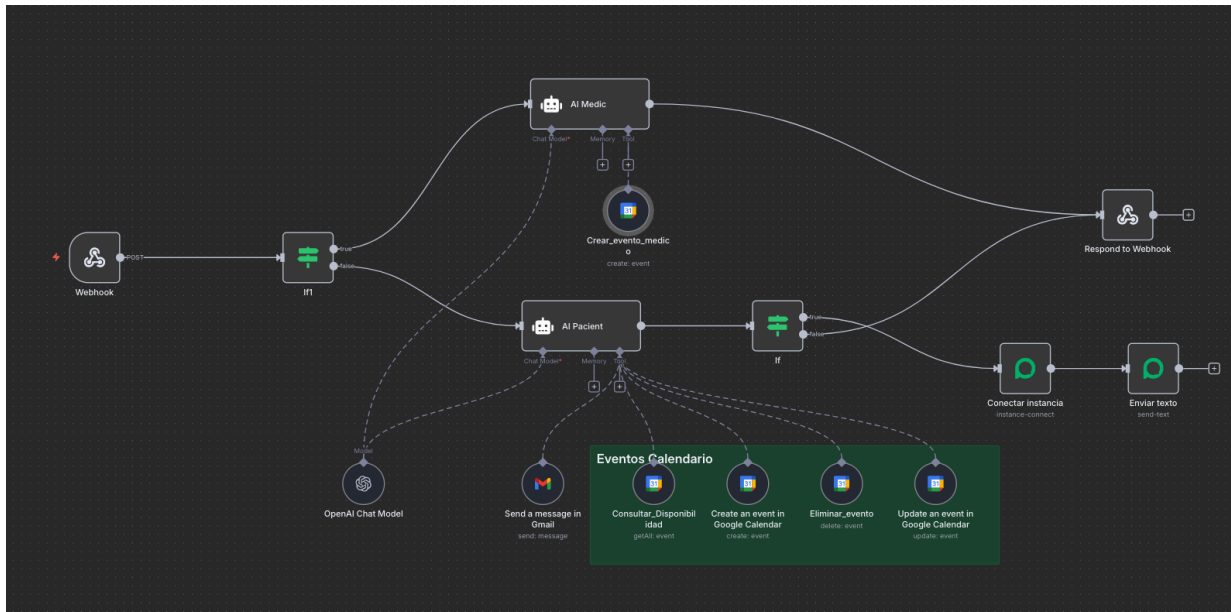
Si dice "cliente" o la clave no es válida → isMedic(false).

[Fecha actual]

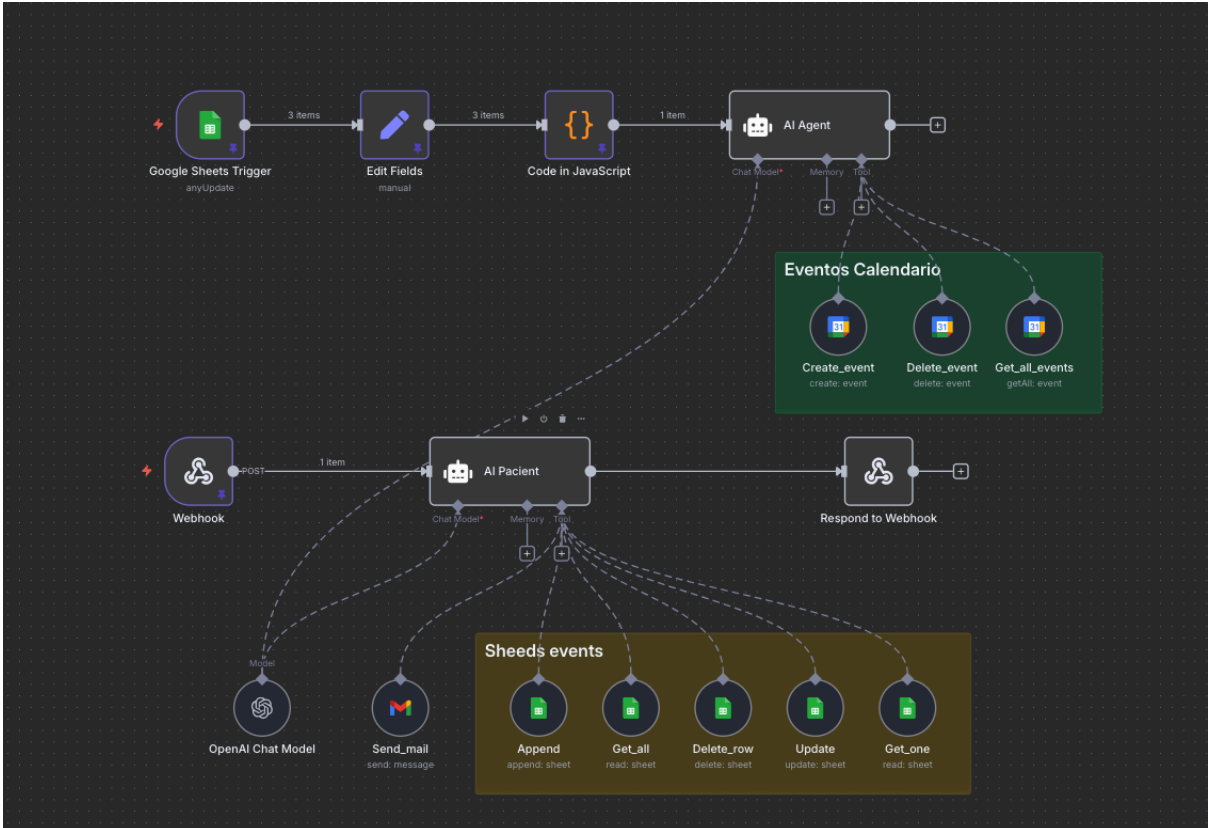
La fecha actual es: {{now}}

- Herramientas implementadas en Vapi:
 - isMedic: Para determinar si es médico o cliente el que llama. Con el siguiente parámetro:
 - doc
 - eliminar_evento: Permite eliminar eventos del calendario asociado. Con los siguientes parámetros:
 - fecha

- nombre.
- buscar_todo: Consulta la disponibilidad del calendario asociado.
- agendar_calendario: Se encarga de crear la cita con los siguientes parámetros:
 - email
 - fecha
 - número
 - Nombre



- Flujo operativo:
 - Inicio de llamada:
 - El usuario llama al número asociado al agente (gestionado por Vapi).
 - Interacción conversacional:
 - Vapi transcribe la voz del usuario, interpreta su intención (por ejemplo: "quiero reservar una cita el jueves") y envía una tool call a un endpoint de n8n (Webhook).
 - Procesamiento en n8n:
 - Se ejecuta el flujo correspondiente según la intención.
 - n8n consulta disponibilidad en Google Sheets.
 - Si procede, crea, elimina o consulta la cita.
 - Confirmaciones automáticas:
 - Tras la acción, n8n:
 - Envía un correo de confirmación.
 - Devuelve un resultado JSON a Vapi con el texto que debe pronunciar (por voz) al usuario.
 - Luego mediante un trigger se actualiza el calendario con la información que está guardada en el Sheet
 - Cierre de interacción:
 - Vapi comunica verbalmente la confirmación o el resultado al usuario y finaliza la llamada.
- Herramientas implementadas en Vapi:
 - isMedic: Permanece sin cambios
 - eliminar_evento: Se le añadió el parámetro fisio para encontrarlo por el fisioterapeuta también
 - buscar_todo: Permanece sin cambios
 - agendar_calendario: Se le añadió el parámetro fisio para añadirle también el fisioterapeuta.



Flujo de procesamiento de Consultas Audiofónicas y Generación de Anamnesis.

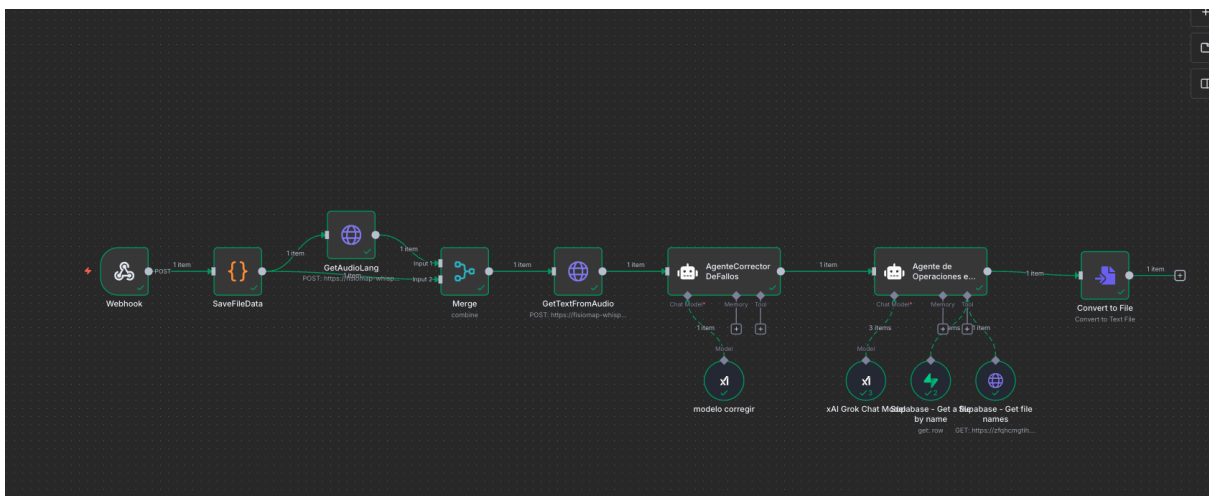
Este flujo se encarga de la recepción, transcripción, procesamiento y generación automática de anamnesis clínicas a partir de audios enviados por médicos.

El proceso combina modelos locales de reconocimiento de voz (Whisper) con modelos de lenguaje (LLM) para garantizar una transcripción precisa y una redacción médica de alta calidad.

- Flujo operativo:
 - Recepción del Audio
 - El flujo recibe un archivo de audio (voz del paciente o profesional sanitario).
 - Se valida el formato, duración y tamaño antes del procesamiento.
 - Transcripción Inicial
 - El audio es procesado localmente por Whisper, modelo de reconocimiento de voz multilingüe.
 - Corrección Lingüística
 - El texto transcrito se envía a un LLM (modelo de lenguaje de gran escala) encargado de corregir errores de ortografía, gramática y coherencia.
 - El resultado mantiene la fidelidad al idioma original del paciente.
 - Procesamiento Semántico y Contextual
 - La transcripción corregida es enviada a un Agente de Generación de Anamnesis, diseñado para:
 - Interpretar el texto dentro del contexto clínico.
 - Aplicar las plantillas médicas personalizadas del cliente.
 - Incorporar terminología médica precisa y conforme a las guías clínicas.

- Control de Inferencias

- El agente cuenta con una guía general de consultas y políticas de inferencia:
 - Puede deducir información implícita (p. ej., sexo del paciente por nombre o voz).
 - No inventa datos inexistentes (p. ej., domicilio o antecedentes no mencionados).
- Toda información inferida se marca para revisión posterior por parte del especialista.

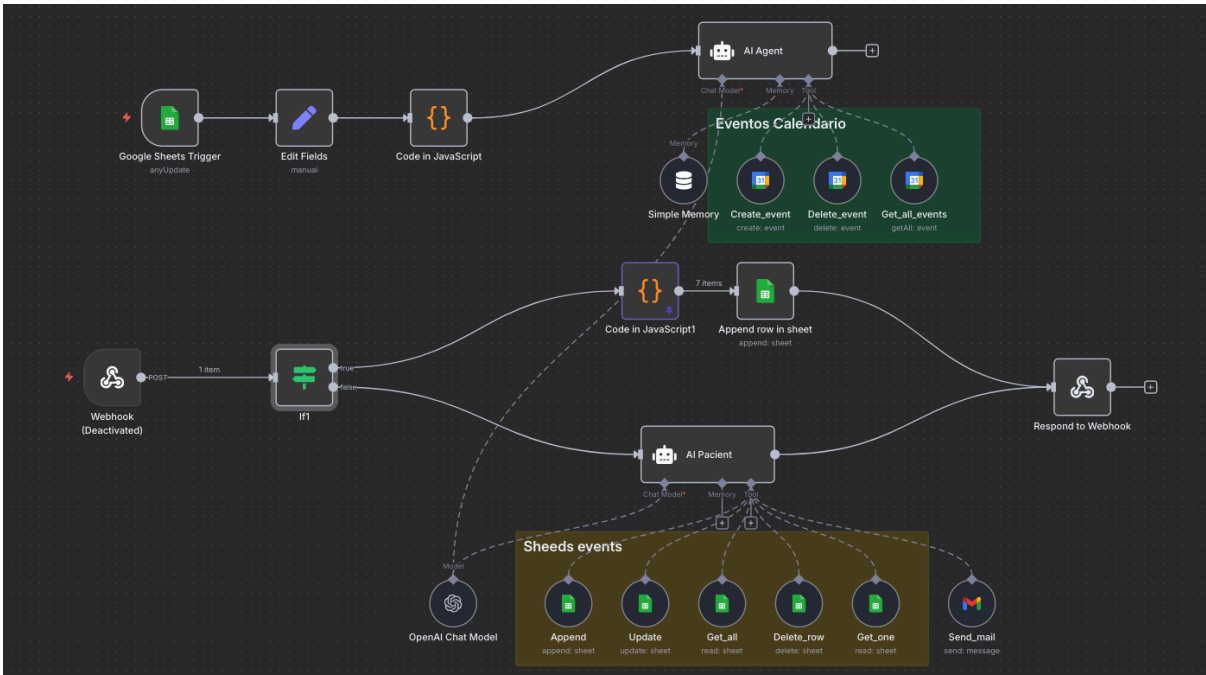


4.2 Timeline Final

Agente Telefónico

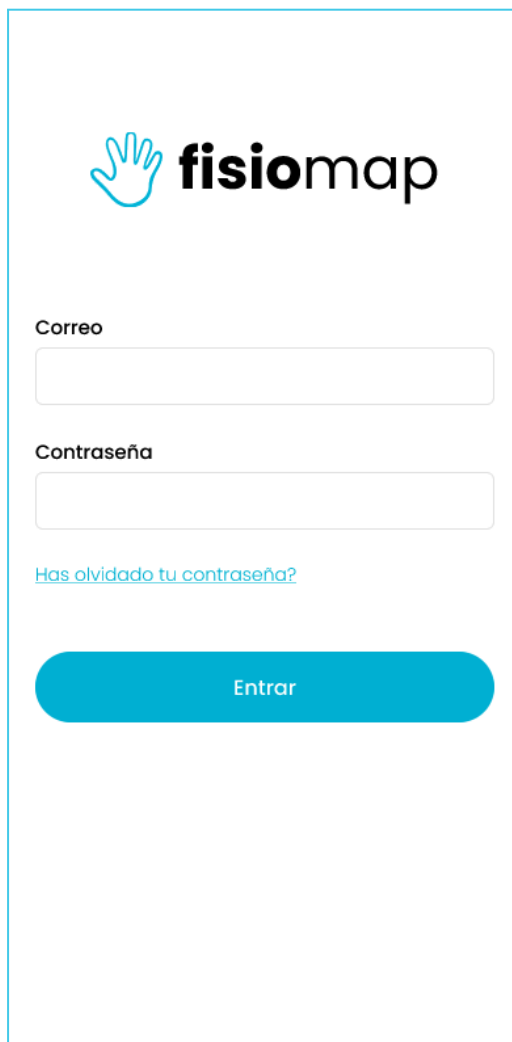
Se añadió en n8n para poder reconocer si es un fisioterapeuta o un cliente.

- Flujo operativo:
 - Inicio de llamada:
 - El usuario llama al número asociado al agente (gestionado por Vapi).
 - Interacción conversacional:
 - Vapi transcribe la voz del usuario, interpreta su intención (por ejemplo: “quiero reservar una cita el jueves”) y envía una tool call a un endpoint de n8n (Webhook).
 - Procesamiento en n8n:
 - Se ejecuta el flujo correspondiente según la intención.
 - n8n consulta disponibilidad en Google Sheets.
 - Si procede, crea, elimina o consulta la cita.
 - Confirmaciones automáticas:
 - Tras la acción, n8n:
 - Envía un correo de confirmación.
 - Devuelve un resultado JSON a Vapi con el texto que debe pronunciar (por voz) al usuario.
 - Luego mediante un trigger se actualiza el calendario con la información que está guardada en el Sheet
 - Cierre de interacción:
 - Vapi comunica verbalmente la confirmación o el resultado al usuario y finaliza la llamada.
- Herramientas implementadas en Vapi:
 - isMedic: Permanece sin cambios
 - eliminar_evento: Permanece sin cambios
 - buscar_todo: Permanece sin cambios
 - agendar_calendario: Permanece sin cambios.



4.3 Pantallas de la aplicación

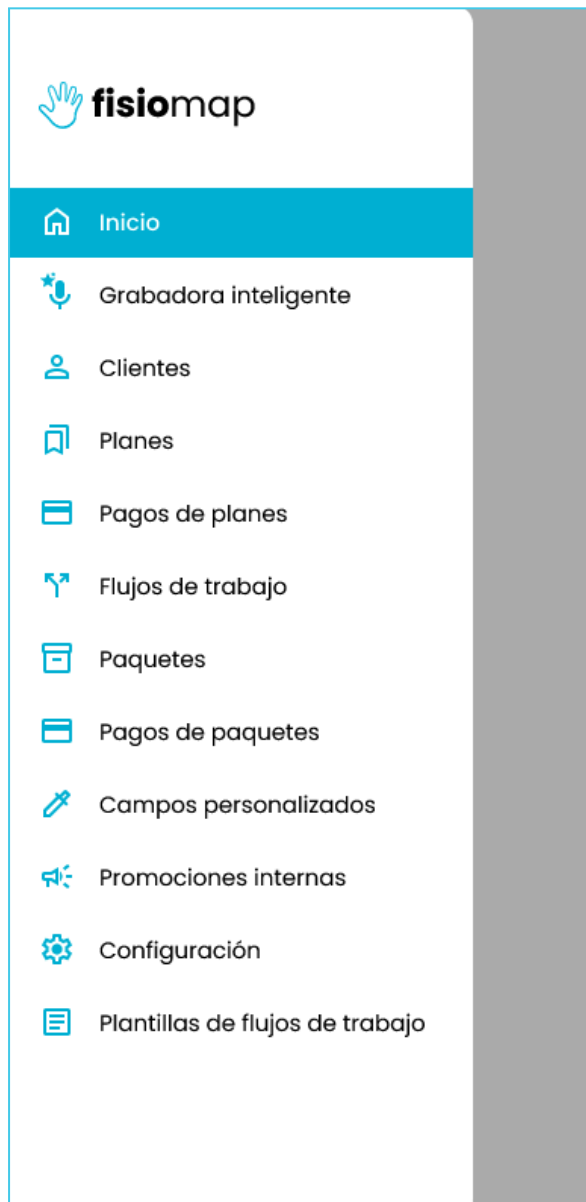
Pantalla de Inicio de sesión



The mockup shows a white rectangular area representing the login screen. At the top left is the Fisiomap logo, which consists of a blue hand icon and the text 'fisiomap' in a bold, black, sans-serif font. Below the logo are two input fields. The first is labeled 'Correo' in a small, grey font and is a simple white rectangle with a thin grey border. The second is labeled 'Contraseña' in a small, grey font and is a white rectangle with a thin grey border. Below the password field is a blue hyperlink that reads 'Has olvidado tu contraseña?'. At the bottom of the form is a large, rounded blue button with the white text 'Entrar'.

- Elementos principales:
 - Campos de ingreso:
 - Correo: Campo para ingresar la dirección de correo electrónico del usuario.
 - Contraseña: Campo protegido para ingresar la contraseña correspondiente.
 - Opción de recuperación de contraseña:
 - Se encuentra un enlace "Has olvidado tu contraseña?", facilitando el proceso de recuperación en caso de que el usuario no recuerde sus credenciales.
 - Botón "Entrar":
 - Invita al usuario a ingresar a la plataforma una vez completados los campos.

Grabadora Inteligente



- Nueva funcionalidad:
 - Está pensada para optimizar la toma de notas clínicas dentro de la plataforma, haciendo uso de tecnología de reconocimiento de voz.

Seleccionar paciente y tipo de grabación



☰

Grabadora

Seleccione un paciente

▼

Seleccione el tipo de grabación:

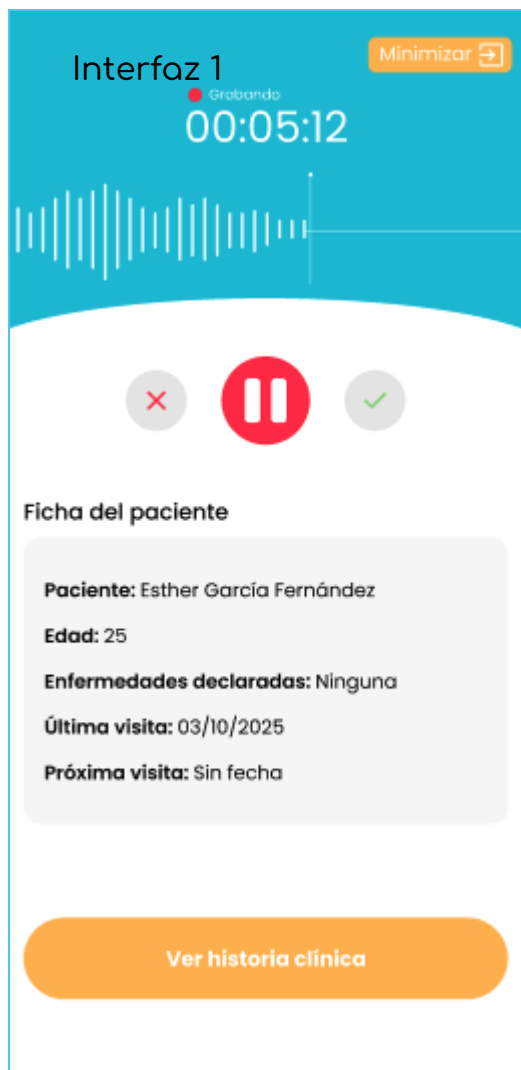
Nueva anamnesis

Nuevo seguimiento

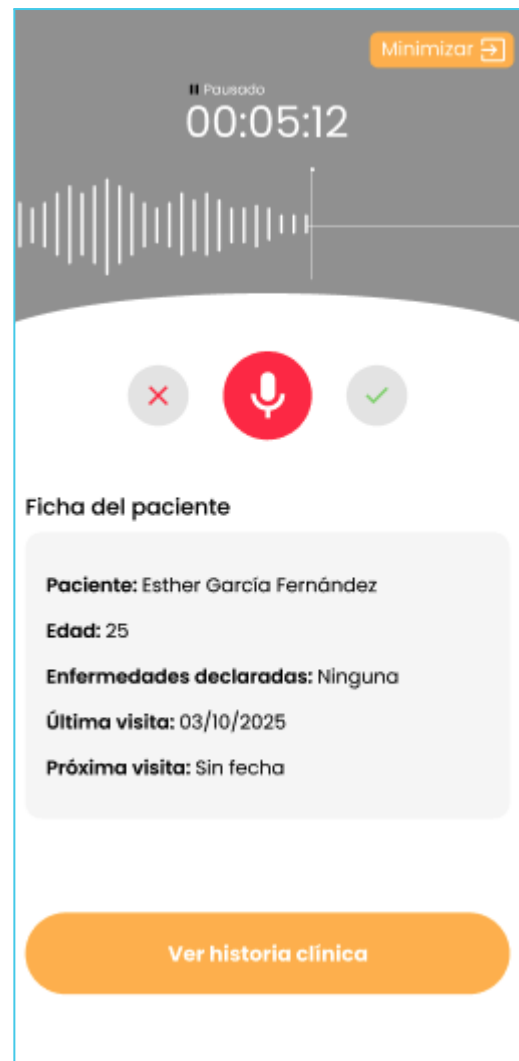
Comenzar a grabar

- Selección de paciente:
 - El usuario puede seleccionar de una lista desplegable al paciente sobre el cual desea registrar una grabación.
 - Este paso garantiza que toda grabación esté asociada a un perfil clínico específico.
- Tipo de grabación:
 - Nueva anamnesis: para registrar la primera evaluación del paciente.
 - Nuevo seguimiento: para registrar una evolución o control posterior.
- Comenzar la grabación:
 - Al activarlo, el sistema te llevará a la siguiente pantalla donde comienza a registrar la voz del profesional, que luego podrá ser procesada, transcrita y almacenada.

Interfaz de la grabadora inteligente



Interfaz 2



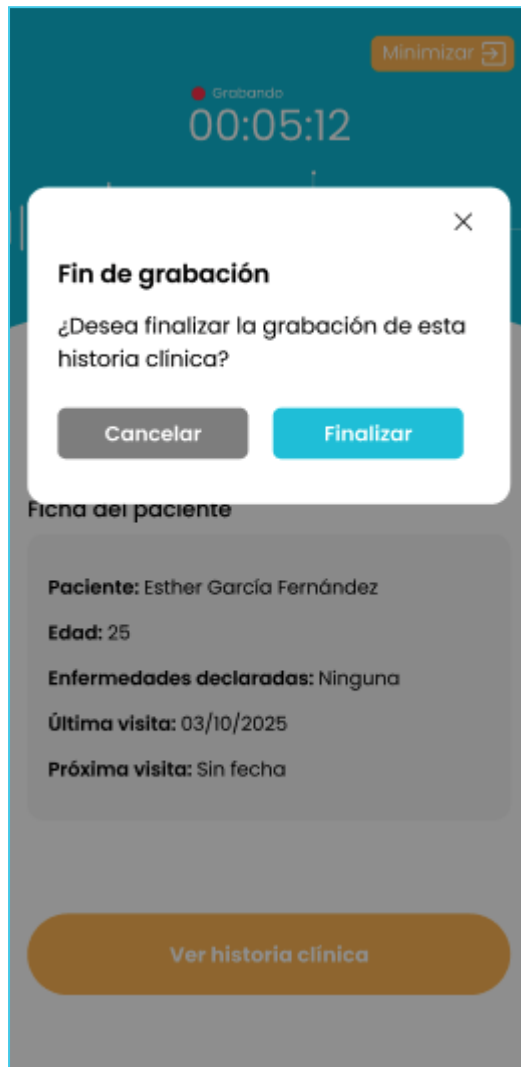
En [interfaz 1](#) se puede apreciar en la parte superior se visualiza un indicador de estado con el texto "Grabando" y un temporizador que muestra el tiempo transcurrido (00:05:12) con el fondo azul. Junto a este elemento aparece un botón para minimizar la ventana. A diferencia de la [interfaz 2](#), se demuestra que cuando se pausa la grabación pasa de ser azul a gris el fondo, y también el texto del estado cambia a "Pausado".

En la parte inferior de ambas interfaces, se presenta la ficha del paciente, que contiene los siguientes datos:

- Paciente
- Edad
- Enfermedades declaradas
- Última visita
- Próxima visita

Finalmente, vemos un botón “Ver historia clínica”, que permite acceder al historial médico completo del paciente.

Fin de grabación



- Muestra la ventana emergente que aparece al intentar finalizar una sesión de grabación dentro de la aplicación.
- En el centro se despliega un cuadro de diálogo con el título “Fin de grabación”, acompañado del mensaje:
 - ¿Desea finalizar la grabación de esta historia clínica?
- Debajo del mensaje se presentan dos botones de acción:
 - Cancelar, que permite continuar con la grabación.
 - Finalizar, que confirma la finalización del proceso.

Revisión final

Revisión de anamnesis

Paciente: Esther García Fernández

Fecha: 03/10/2025

Motivo de consulta: Rotura de fibras

Evaluación: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Revisar luego

Guardar cambios

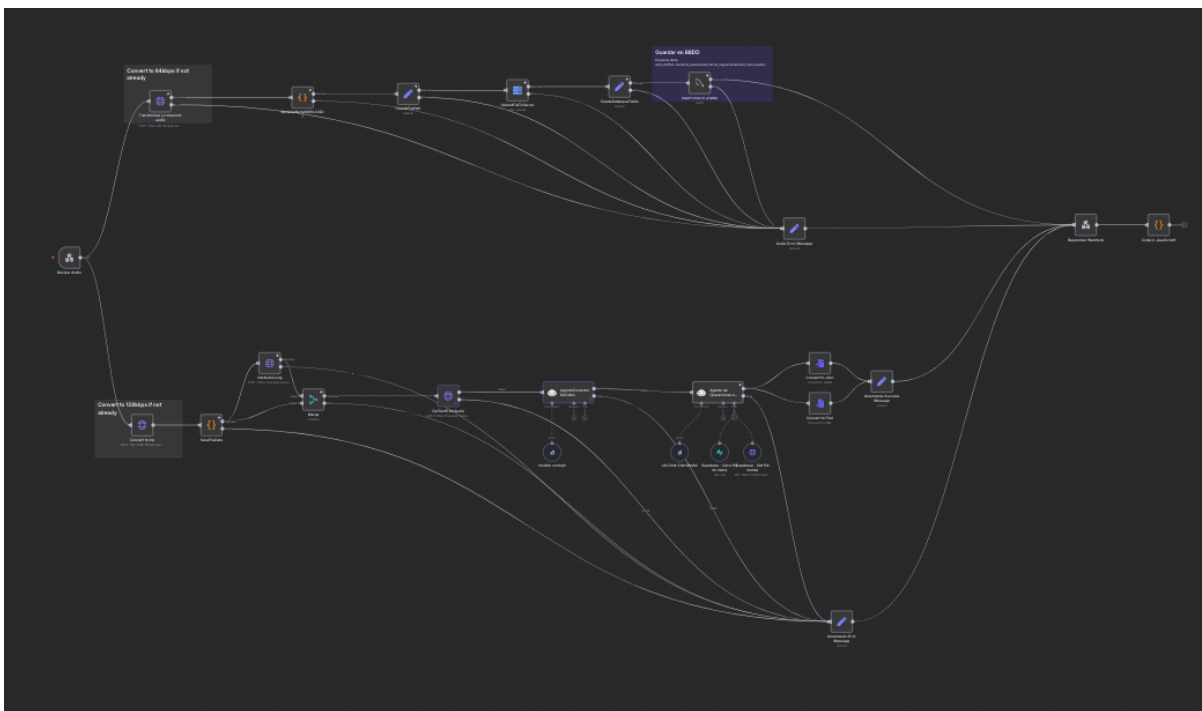
- Permite que el profesional revise, edite y confirme la anamnesis registrada tras una sesión, asegurando la precisión de los datos antes de almacenarlos en el sistema clínico.
- Nos muestra los siguientes datos:
 - Paciente
 - Fecha
 - Motivo de consulta
 - Evaluación
 - Se incluye el texto presentado en formato de párrafo legible y con estructura narrativa.
- En la parte inferior, cuenta con dos botones:
 - Revisar luego
 - Corresponde a una acción secundaria
 - Guardar Cambios
 - Permite guardar la revisión

4.4 Parte de backend

4.1 Preprocesamiento de los audios para guardarlos en la base de datos

Se instaló en el servidor una imagen de docker de whisper que tenía una API para poder transcribir el audio a texto. Además también se instaló una imagen de docker de ffmpeg y se editó para poder utilizarla bien, también tenía una API.

Luego usamos estas imagenes del servidor en un flujo de n8n para crear las anamnesis y guardar los audios de las consultas en una carpeta del servidor y luego guardar la ruta en la base de datos además de convertirlos a mp3 además de comprimirlos y normalizarlos.



5. Problemas encontrados

Problema encontrado: Transcripción en vivo

Descripción del problema

Al trabajar con whisper, el cual fué liberado hace tiempo y no contempla la opción de streaming no es posible crear la transcripción en vivo haciendo streaming del audio y recuperar un streaming de texto para mostrar .

Solución propuesta

El flujo propuesto será el siguiente:

1. El audio se envía a Whisper para transcribirlo a texto.
 2. La transcripción pasa por el agente de corrección, que ajusta errores y formato.
 3. El texto corregido se devuelve al usuario para su revisión.
 4. El audio se almacena en la base de datos (.mp3, 64 kbps).
 5. Posteriormente, la transcripción se usará para generar y guardar la anamnesis correspondiente a la sesión con el cliente.
-

Problema encontrado: Agente Telefónico

Descripción del problema

El agente de Vapi es demasiado costoso puesto que cobran 0,17€ por minuto de llamada más una prima por llamada.

Solución propuesta

Se desestimó el agente y no se va a implementar.