

Enero 2025



Sistema inteligente para apoyo a reuniones con detección de habla, generación de ideas y detección de llegadas tardías

Máster en Inteligencia Artificial Aplicada
Universidad Carlos III de Madrid

Inteligencia Ambiental

**Lucía Cordero - 100451778
Jorge Garcelán - 100442062
Laura González - 100538522**

Índice

Introducción, visión del producto y
arquitectura propuesta

01

Transcripción y visualización en tiempo real

02

Generación de resumen, palabras clave e
ideas

03

Visualización y selección interactiva de ideas

04

Detección de nuevas personas

05

Funcionalidades extra

06

Desafíos enfrentados y mejoras posibles

07



Introducción

¿Alguna vez has llegado dos minutos tarde a una reunión pero has sentido que te habías perdido la mitad? Entonces este producto es para ti.

SmartMeet presenta su nuevo software, un sistema inteligente para apoyo a reuniones con detección de habla, generación de ideas y detección de llegadas tardías. SmartMeet es el compañero inteligente que siempre has querido en tus reuniones: eficaz, útil y accesible.

¿Quieres conocer más? Te lo contamos en las siguientes páginas, ¡sigue leyendo!



Visión del producto

3

empleados

1

oficina

+600

horas de reuniones
infructíferas ahorradas

Transforma las reuniones con transcripciones en tiempo real en dos idiomas, generación de ideas innovadoras, palabras clave con estadísticas y resúmenes instantáneos que integran a todos los participantes.

SmartMeet garantiza que cada minuto cuente y cada idea brille. Con una interfaz interactiva disponible en español, inglés e italiano, se centralizan las discusiones y se fomenta la colaboración.

2025

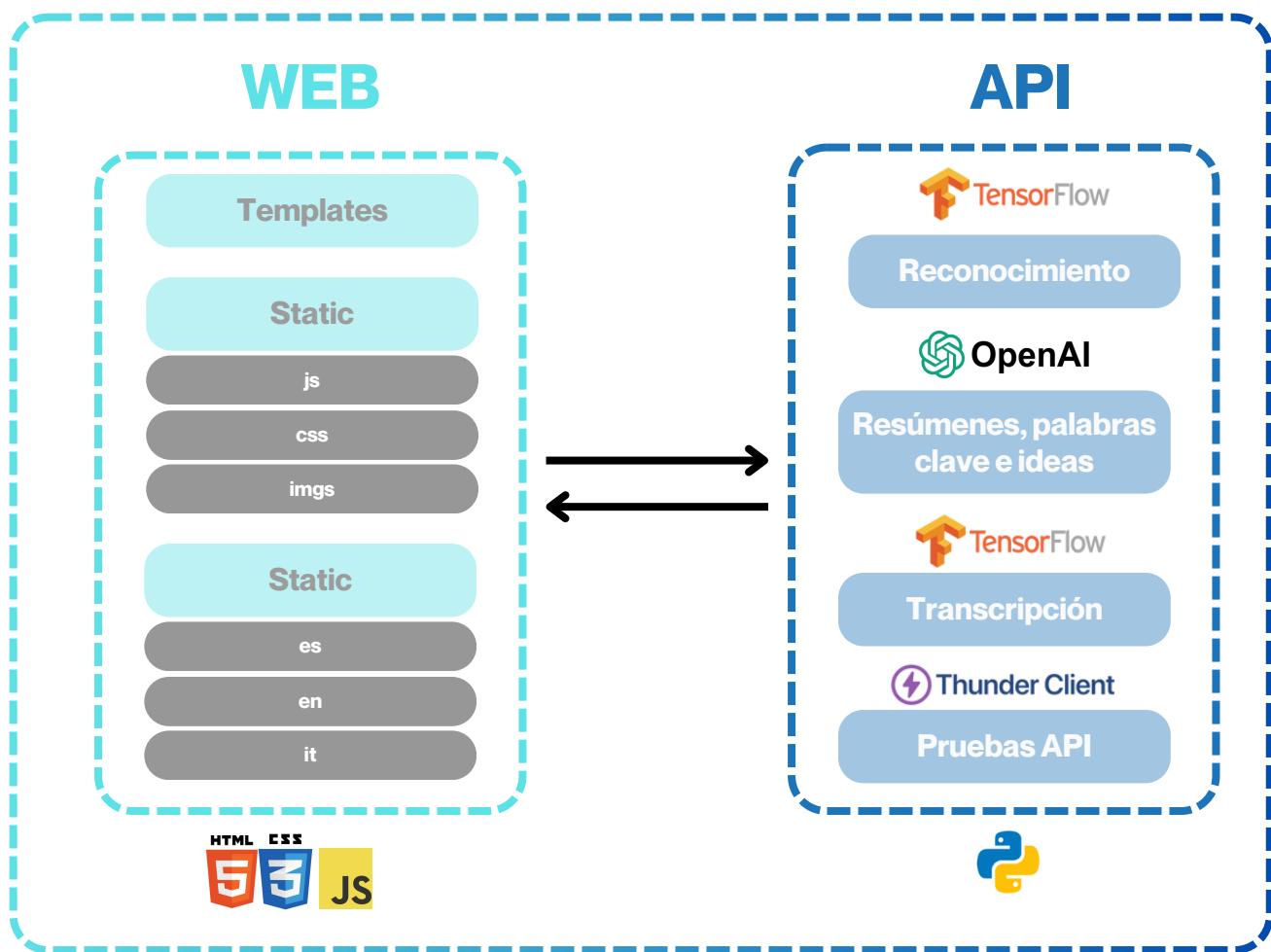
El año en el que cambia tu forma
de reunirte



Arquitectura propuesta

La arquitectura de SmartMeet es modular y está organizada con el framework de Flask en dos componentes principales: WEB y API, conectados de manera eficiente para proporcionar una experiencia fluida e interactiva de la plataforma.

- **WEB:** Está diseñada utilizando Flask y plantillas en Jinja2, con dos páginas principales:
 - Index: La página inicial, desde donde se accede al sistema.
 - Meeting: La página principal del sistema.
- **API:** Gestiona funcionalidades como reconocimiento de audio, generación de resúmenes e ideas a través de endpoints específicos.
 - Se ha utilizado Thunder Client para probar el correcto funcionamiento de la API y de todos sus endpoints en un entorno controlado.





Transcripción y visualización en tiempo real

Utilización de Whisper

SmartMeet utiliza Whisper con transformers.js para capturar y transcribir en tiempo real lo discutido durante las reuniones. Las transcripciones se muestran dinámicamente en la interfaz interactiva, proporcionando una experiencia fluida y continua para los usuarios.

Intervalo de actualización

La frecuencia de actualización de las transcripciones en la pantalla interactiva se configura para cada 3 segundos. Esto garantiza un equilibrio entre la fluidez de las actualizaciones y el tiempo de procesamiento, evitando interrupciones en el flujo de la reunión. Se experimentó con intervalos de 2 a 5 segundos. A 2 segundos, los mensajes se acortaban en la transcripción, mientras que a 5 segundos, las actualizaciones eran demasiado lentas. La elección de 3 segundos ofrece un balance entre rapidez y precisión a la hora de completar las oraciones.

Formateo del texto

El formato del texto transcritto está optimizado mediante la adición del elemento
 y la sustitución de las etiquetas de ruido blanco, ruido de fondo y música.

En primer lugar hacemos una pseudo-dialización para garantizar que en la transcripción el texto aparezca de forma ordenada para cada persona que interviene. Esto se hace llevando una cuenta del número de periodos en los que no se haya escuchado ninguna palabra. En nuestro caso, suponemos que la conversación va a fluir de forma ordenada y sin que los interventores se interrumpan entre sí, así que se cambia de línea tras tres segundos de silencio (es decir, dos períodos donde no se genera texto, se generan o <blank_audio> o <música>).

Funcionalidades adicionales

El sistema puede transcribir en varios idiomas simultáneamente (español e inglés), ampliando su aplicabilidad en reuniones internacionales o donde los usuarios intervienen en diferentes idiomas.



Generación de resumen, palabras clave e ideas

Utilización de OpenAI API

SmartMeet analiza las transcripciones para detectar palabras clave relevantes. Estas palabras son procesadas mediante la API de OpenAI, que hemos seleccionado por su rendimiento superior en generación de texto mediante la realización de diversas pruebas y añadiendo las experiencias en la práctica 2. Asimismo, hemos ajustado parámetros como `max_tokens` (1000), `temperature` (0.7), `top_p` (None) y `presence_penalty` (1) para optimizar la generación de ideas, equilibrando la diversidad con la relevancia de las propuestas. Como forma de transparencia hemos utilizado el modelo “gpt-4o-mini” y consumido \$0.60 durante todo el desarrollo del proyecto.

Generación de resumen

El sistema utiliza las capacidades de OpenAI para generar automáticamente un resumen conciso y claro de los puntos discutidos hasta el momento, evitando interrupciones en la dinámica de la reunión.

Generación de palabras clave

SmartMeet utiliza el transcript generado automáticamente para extraer palabras clave relevantes que reflejan los temas principales discutidos en la reunión. Este proceso se realiza mediante algoritmos de procesamiento de lenguaje natural integrados en la API de OpenAI.

Generación de ideas

A partir de las palabras clave obtenidas, el sistema utiliza la API de OpenAI para generar ideas asociadas, expandiendo las temáticas y sugiriendo posibles líneas de discusión, acciones futuras o puntos importantes a considerar.

Integración en la interfaz

Con el objetivo de no retener la atención de la reunión en las nuevas ideas y palabras que se generen, se añade el siguiente aviso informativo:

 El contenido inteligente será generado en 17 segundos.



Prompts

Los siguientes prompts han sido diseñados específicamente para los cuatro endpoints de la API. Durante su desarrollo, se emplearon técnicas de prompt engineering para optimizar su efectividad, incorporando características clave como el idioma del sistema y el contenido del transcript, garantizando respuestas precisas y contextualizadas.

system

Eres un asistente virtual que ayuda a generar **resúmenes** claros y concisos de reuniones.

user

Genera un resumen en {readable_language} para esta transcripción: {transcript}

Generación de resumen

system

Eres un asistente que ayuda a identificar **palabras clave** importantes en reuniones.

user

Extrae palabras clave en {readable_language} de esta transcripción: {transcript}

Generación de palabras clave

system

Eres un asistente que ayuda a crear **ideas** en reuniones.

user

Extrae un número de ideas entre 1 y 3 en {readable_language} dadas las siguientes palabras clave: {keywords}

Generación de ideas

system

Eres un asistente experto en **análisis de texto**.

user

Analiza la siguiente transcripción en {readable_language} y calcula estadísticas para las siguientes palabras clave.

Para cada palabra clave, indica:

- El número total de apariciones.
- El porcentaje de importancia de esa palabra con respecto a la transcripción.
- El porcentaje de positividad asociado a las frases donde aparece.
- El porcentaje de neutralidad asociado a las frases donde aparece.
- El porcentaje de negatividad asociado a las frases donde aparece.

Transcripción:

{transcript}

Palabras clave: {keywords}

Generación de estadísticas de palabras clave

Visualización y selección interactiva de ideas

El sistema de SmartMeet convierte las ideas generadas en tarjetas interactivas, facilitando una experiencia visual y funcional que impulsa la colaboración. A continuación, se detalla cómo funciona esta funcionalidad.



Alta Prioridad

Torneo de Domino en la Oficina:
Organizar un torneo de Domino durante una reunión de marketing para fomentar la camaradería y el trabajo en equipo entre los trabajadores.

Palabras Clave del Domino: Utilizar palabras relacionadas con el juego del Domino para inspirar nuevas ideas de marketing, como 'estrategia', 'conexiones' y 'competencia'.

Media Prioridad

Juegos de Estrategia en Marketing:
Incorporar juegos de estrategia, como el Domino, en las reuniones de marketing para estimular la creatividad y mejorar la dinámica del grupo.

Captura de Momentos: Crear un espacio en la reunión donde se tomen fotos divertidas mientras se juega al Domino, para luego usarlas en redes sociales o como parte de una campaña de marketing.

Baja Prioridad

Marketing a Través del Juego:
Desarrollar una campaña de marketing que utilice el concepto de jugar al Domino como metáfora para trabajar juntos y construir relaciones sólidas entre los empleados.

Juego de Domino en la Reunión: Integrar una sesión de juego de Domino durante la reunión de marketing para fomentar el trabajo en equipo y la creatividad entre los trabajadores.

Visualización y Organización de Ideas

Las ideas generadas se presentan en forma de tarjetas móviles en la interfaz. Estas tarjetas contienen una descripción breve y concisa de la idea y palabras clave destacadas relacionadas con la transcripción.

Los usuarios pueden reorganizar las tarjetas en tres grupos (alta prioridad, media prioridad y baja prioridad) mediante una función de arrastrar y soltar. Esto permite priorizar ideas relevantes o agruparlas según el contexto o tema discutido. Por otro lado, en cada *batch* generativo se descartan las ideas que no se hayan clasificado en un nivel de prioridad para dar paso a otras tres ideas nuevas, evitando la acumulación de propuestas innecesarias.



Visualización de palabras clave y estadísticas

Las palabras clave se muestran en burbujas debajo de la transcripción.

Palabras Clave



Al hacer clic en una palabra clave, el sistema muestra un panel emergente con estadísticas útiles:

- **Frecuencia de aparición:** Cuántas veces se mencionó la palabra en la transcripción.
- **Importancia:** Dada la transcripción y un prompt (que se detalla a continuación), el modelo generativo de OpenAI genera una valoración de la importancia de la palabra con respecto al resto de la conversación.

alerta

Apariciones: 1

Importancia:

0.14%

- **Positividad, Negatividad y Neutralidad:** Se muestran en barras porcentuales (de 0 a 100%) la valoración del contexto en el que ha aparecido la palabra clave, es decir, si se trata de una palabra utilizada mayormente para describir situaciones positivas, negativas o neutras.

Positividad:



Neutralidad:



Negatividad:





Detección de nuevas personas

Utilización de YOLOv10

SmartMeet utiliza YOLOv10 con Transformers.js para la detección de personas en tiempo real durante las reuniones. Este modelo permite identificar a los participantes y mostrar las bounding boxes asociadas a cada uno de ellos, con un número que los identifica. Las detecciones se actualizan constantemente, de modo que si una persona se mueve, su identificación la seguirá automáticamente, proporcionando una experiencia fluida y dinámica.

Utilización de COCO - SSD

El software también incorpora COCO-SSD, un modelo más rápido pero menos preciso en comparación con YOLOv10. Aunque su precisión es inferior, COCO-SSD compensa esta limitación con una velocidad significativamente mayor, lo que mejora la experiencia del usuario.

Para ofrecer la mayor flexibilidad, se permite al usuario seleccionar el modelo de detección más adecuado según la capacidad computacional de su equipo: YOLOv10, para una mayor precisión, o COCO-SSD, para una experiencia más rápida y ligera.

Tiempo de gracia

Además, se ha implementado un sistema de tiempo de gracia. Este consiste en un temporizador que comienza al inicio de la reunión. Si durante este tiempo no se detecta al menos una vez el número esperado de participantes, se genera un aviso indicando cuántos participantes faltan en la reunión.



Funcionalidades extra

Interfaz adaptable

Se ha implementado una interfaz personalizable que incluye las siguientes opciones:

- Tema claro/oscuro: El usuario puede alternar entre un tema claro u oscuro, adaptándose a sus preferencias visuales o condiciones de iluminación.
- Tamaño de fuente: Es posible ajustar el tamaño de la letra a pequeño, mediano o grande, para garantizar una mejor legibilidad y accesibilidad.

Configuración de Accesibilidad

Seleccionar Tema:

Claro

Tamaño de Letra:

Mediano

Cerrar

Botón crear/salirse reunión

Al iniciar una reunión, se presenta una pantalla donde el usuario puede configurar los siguientes parámetros: nombre de la reunión, número de participantes esperado, duración de la reunión, así como, idioma de la interfaz, tema de la aplicación.

Selección de Idioma y Transcripción

Para garantizar la versatilidad de la aplicación:

- Idioma de la interfaz: El usuario puede seleccionar el idioma entre inglés, español e italiano, facilitando su uso en diferentes regiones.
- Idioma de la transcripción: También es posible elegir el idioma en el que se generarán las transcripciones durante las reuniones, proporcionando flexibilidad en la comunicación.

Configuración de Idioma

Seleccionar Idioma:

ES

Cerrar

Configuración de Idioma

Seleccionar Idioma:

✓ ES

EN

IT



Límite de tiempo

Al crear una reunión, el usuario puede establecer un límite de tiempo. Una vez alcanzado este límite, el sistema redirigirá automáticamente al usuario a la pantalla principal.

Tiempo restante: 9 minutos 41 segundos

Timeline / Línea de tiempo

Además de mostrar las ideas en el formato anteriormente mencionado, también hemos incluido una línea del tiempo que muestra las ideas con su respectiva etiqueta temporal (del momento en el que fueron mencionadas). Esto facilita que alguien que llega tarde a la reunión pueda seguir el flujo de la conversación sin problema.

- 22:22:45

Optimización de Procedimientos del Equipo de Marketing: Desarrollar un manual que detalle los procedimientos clave del equipo de marketing, para mejorar la eficiencia y coordinar mejor el trabajo en equipo.

- 22:22:45

Equilibrio Vida-Trabajo en Marketing: Implementar políticas que fomenten un equilibrio saludable entre la vida personal y laboral dentro del equipo de marketing, como horarios flexibles o días de trabajo remoto.

Botón de Informe Final

Se ha añadido un botón de informe final que permite al usuario generar un resumen de la reunión en cualquier momento, ya sea al finalizarla o cuando se añada un nuevo participante. Este informe incluye información relevante como: los datos de la reunión, la transcripción completa, las ideas generadas y el timeline de estas.

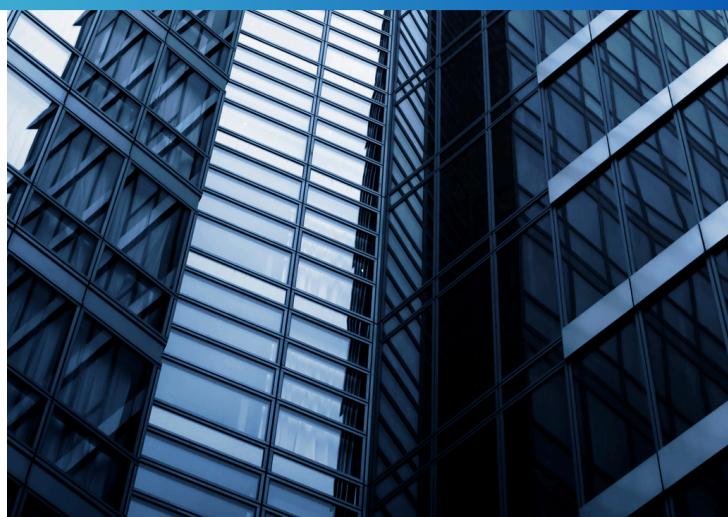
Desafíos

1. Integración de Whisper

Durante el desarrollo, uno de los principales desafíos fue la integración de Whisper para la transcripción en tiempo real. Inicialmente, surgieron problemas al coordinar a los workers, ya que el master enviaba múltiples peticiones al worker encargado de realizar la transcripción antes de que este estuviera listo, lo que generaba conflictos relacionados con promesas y asincronía. Identificar la raíz del problema no fue sencillo y requirió una cantidad significativa de tiempo.

2. Bounding boxes

El segundo desafío principal fue adaptar las bounding boxes de los modelos de COCO y YOLO al tamaño dinámico de la webcam, ya que al ser una web responsive las bounding boxes aparecían en lugares incorrecto.



Mejoras posibles

1. Despliegue del sistema

Una de las áreas clave para futuras mejoras es la traducción automática del transcripción, que actualmente está diseñada para detectar mensajes en otro idioma que no sea el de la interfaz del sistema y traducirlos automáticamente. Este enfoque establece una base prometedora para reuniones multilingües, facilitando la comunicación entre participantes que hablan diferentes idiomas

2. Reuniones multilingües

Además, a largo plazo, se podría considerar un despliegue completo del sistema, ya que actualmente solo opera de forma local. Esto permitiría desplegar la solución en servidores en la nube.

Enero 2025



¡Gracias!

www.smartmeet.com