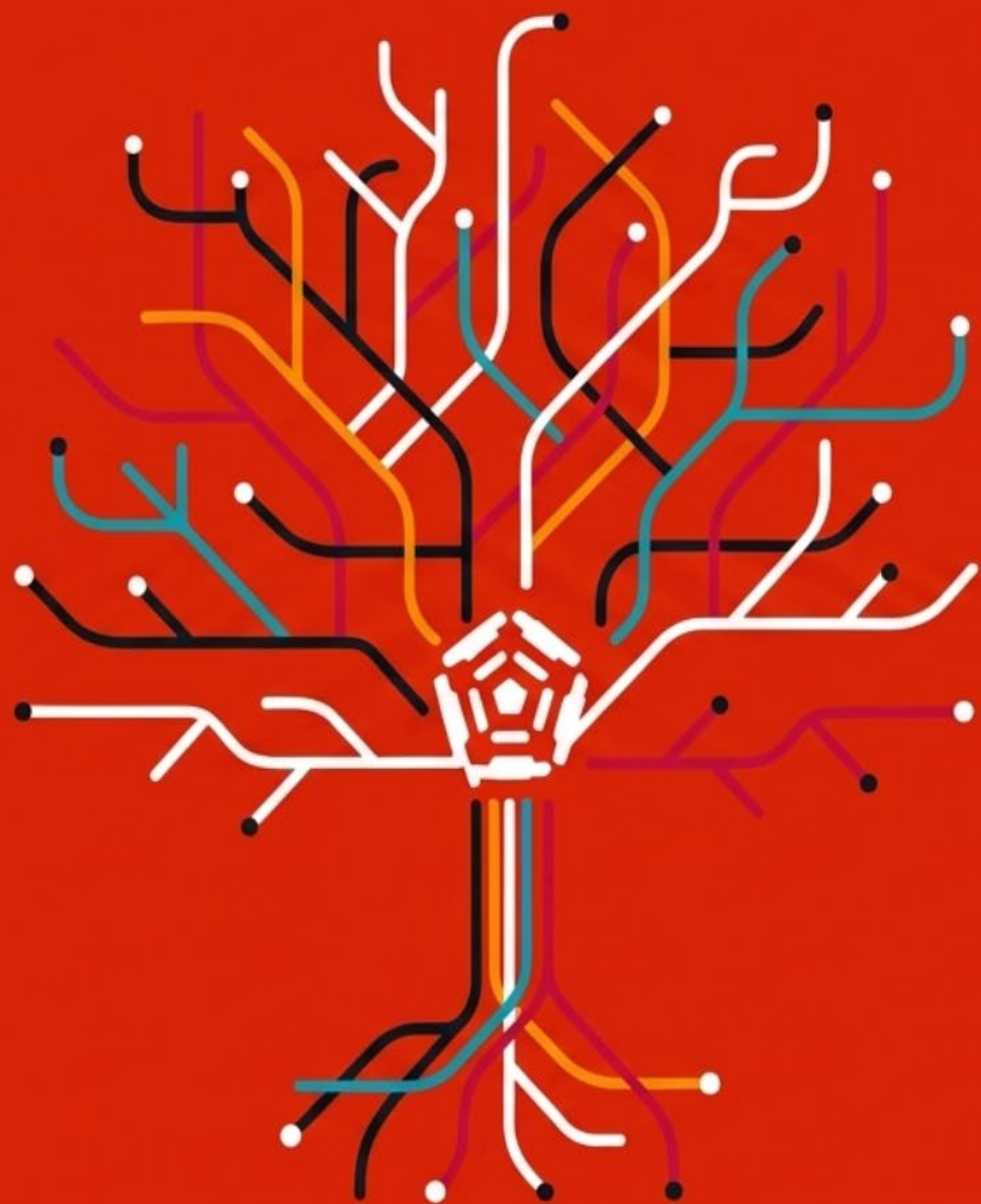




Universidad de Zaragoza | Curso 2026

¿QUÉ TIENE LA INTELIGENCIA ARTIFICIAL PARA MÍ ARCHIVO?

De la teoría a la práctica.

Universidad de Zaragoza | Curso 2026

¿QUÉ TIENE LA INTELIGENCIA ARTIFICIAL PARA MÍ ARCHIVO?

TALLER 5

Metadato no supervisado de colecciones de audio

Construye tu propio explorador de audio con IA agéntica

¿Qué tiene la inteligencia artificial para mi archivo De la teoría a la práctica · Jaca 2026

De la teoría a la práctica.

Metadato no supervisado de colecciones de audio

https://colab.research.google.com/drive/1odI1W2FBG5QXlCC_tT?usp=sharing

¿Qué vamos a hacer?

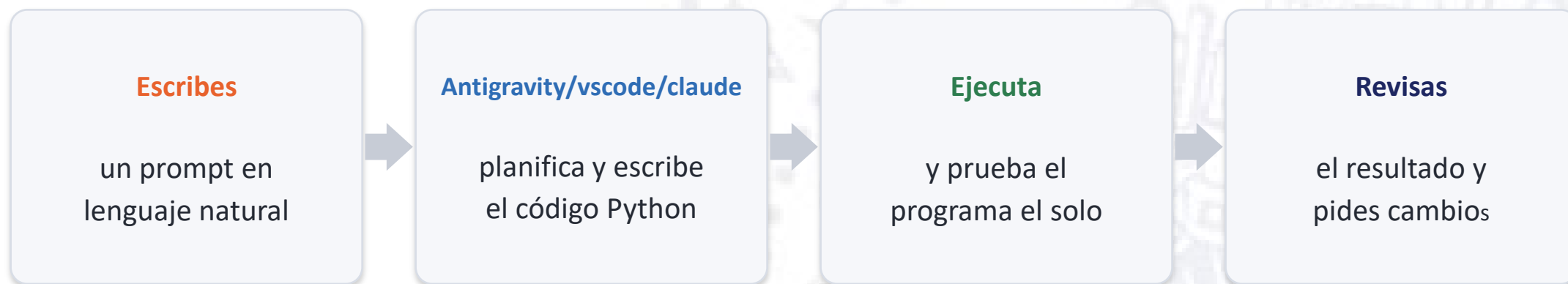
Todo con lenguaje natural, sin escribir código a mano: describiremos y el agente construye la herramienta.



Cómo trabaja un agente

Antigravity/vscode+agente/Claude: no son simplemente chats que dan consejos:

Planifican, escriben el código y lo ejecutan. Tú revisas.



Es un bucle: repites prompt → revisar → ejecutar hasta que la herramienta hace lo que quieres.

Dos formas de usar la IA

Conviene distinguirlas: cambian el coste y la confidencialidad. Hoy usaremos sobre todo la primera.

CÓDIGO GENERADO POR LLM

El LLM escribe el código y se ejecuta en tu ordenador

Coste muy bajo, no depende del volumen de datos.

No saca tus datos fuera: sin problemas de confidencialidad.

Para tareas como:

Gráficos, estadísticas,
cambios de formato, unir
tablas, visualizaciones.

EJECUCIÓN CON MODELO

Cada dato pasa por un modelo de lenguaje (API)

Coste alto, crece con los datos.

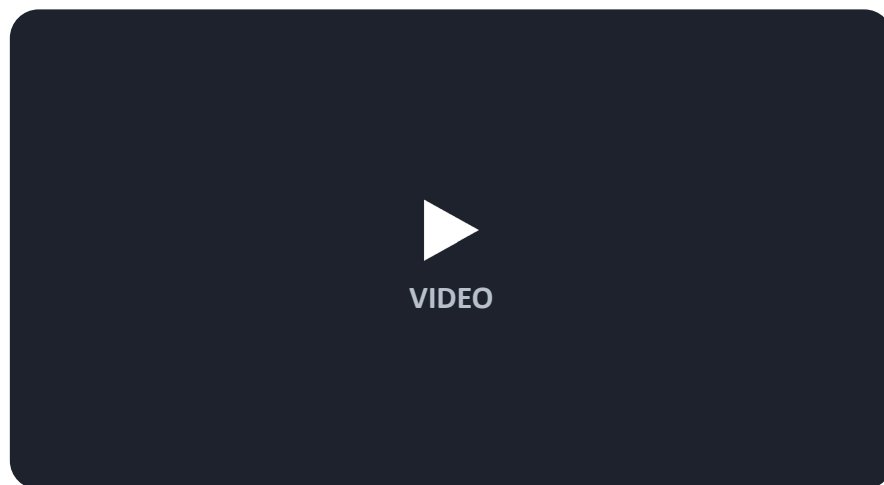
Los datos salen al modelo: ojo con la confidencialidad.

Para tareas como:

Traducir, redactar,
clasificar, resumir,
extraer información.

Lo que vas a construir

Una web que abres en Chrome: sueltas el video y el SRT, y exploras el debate de forma visual. Descarga el resultado de la primera parte del taller: dihana.unizar.es/~cadrete/debate.zip



Timeline coloreado por locutor · clic = saltar

Inicio Locutor Texto

- 00:00 SPK_00 ...texto
- 00:07 SPK_04 ...texto
- 00:14 SPK_03 ...texto
- 00:21 SPK_02 ...texto
- 00:28 SPK_00 ...texto
- 00:35 SPK_06 ...texto
- 00:42 SPK_03 ...texto

PANEL

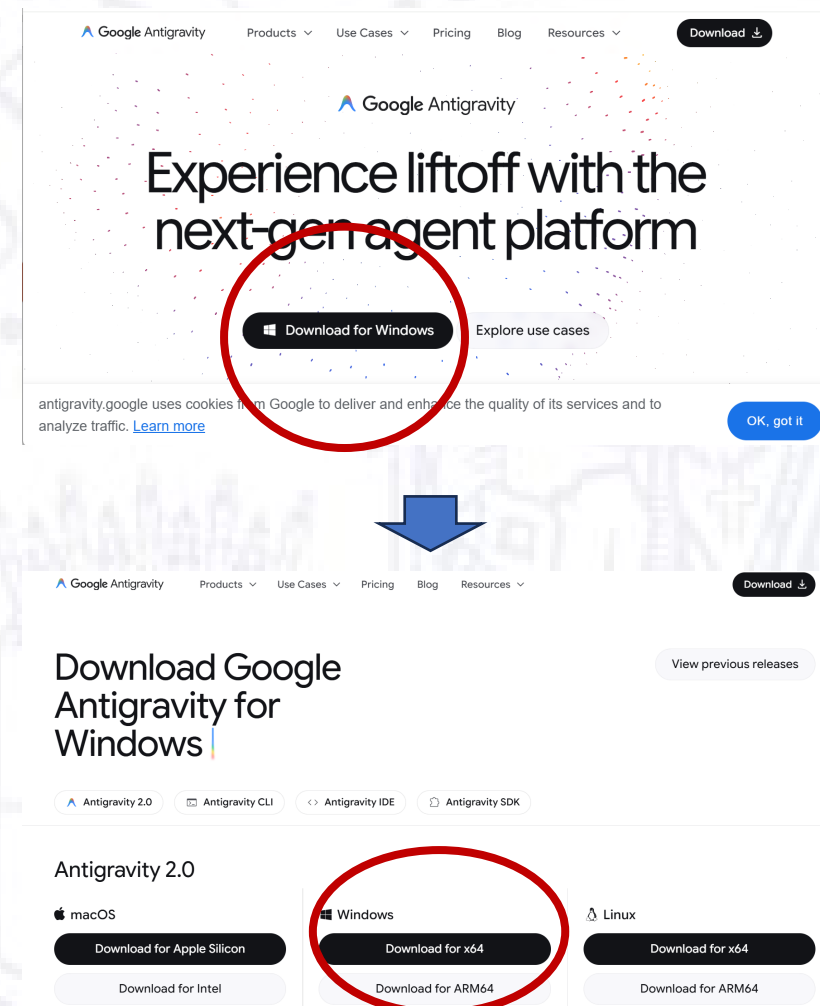
locutor
tiempos
texto

estadísticas

Descargar e instalar Antigravity

- 1 Abre Chrome y entra en <https://antigravity.google/>
- 2 Pulsa Download y elige tu sistema
(Windows, Mac o Linux)

Nota necesitas una cuenta de Gmail personal (no de organización) y Chrome instalado.
Alternativamente se puede usar vscode <https://code.visualstudio.com/>
pero conseguir tener configurado un agente lleva más pasos.



INSTALACIÓN · 2 DE 4

Descargar e instalar Antigravity

1 Abre Chrome y entra en <https://antigravity.google/>

2 Pulsa Download y elige tu sistema

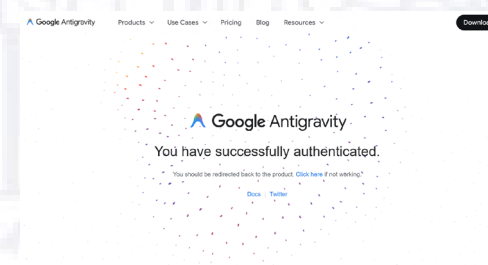
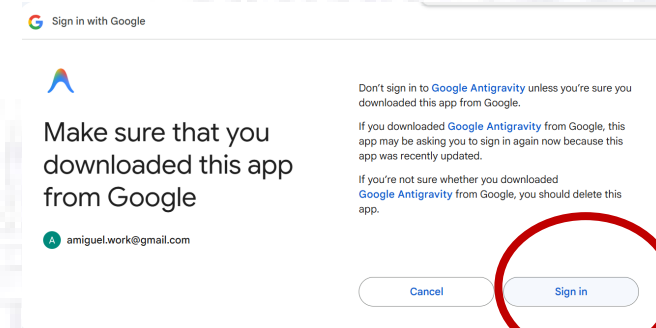
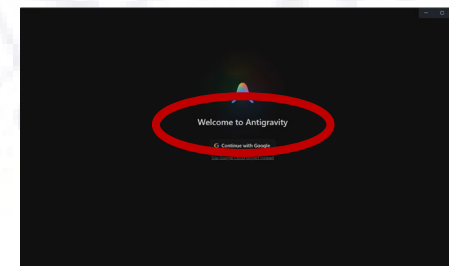
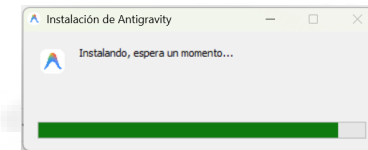
(Windows, Mac o Linux)

3 Ejecuta el instalador

“Continue with Google”

Nota necesitas una cuenta de Gmail personal (no de organización) y Chrome instalado.

Alternativamente se puede usar vscode <https://code.visualstudio.com/> pero conseguir tener configurado un agente lleva más pasos.



Descargar e instalar Antigravity

1 Abre Chrome y entra en <https://antigravity.google/>

2 Pulsa Download y elige tu sistema

(Windows, Mac o Linux)

3 Ejecuta el instalador

“Continue with Google”

4 Finaliza la instalación:

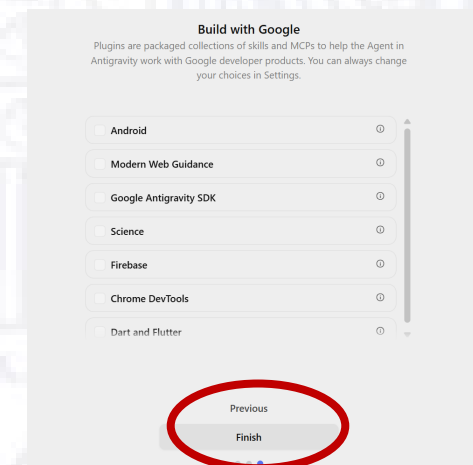
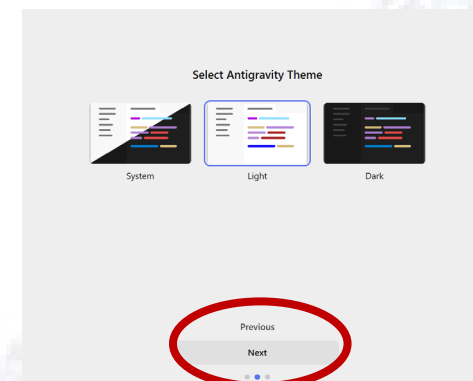
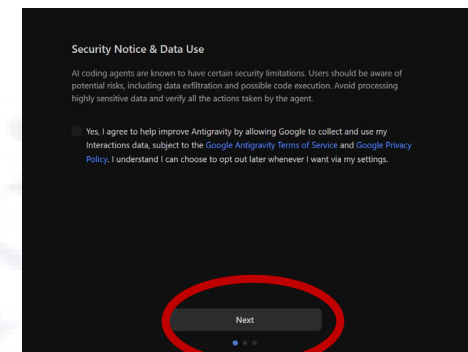
Elige si dejas acceder a datos de uso -> Siguiente

Elige tema: oscuro/claro/...

“Build with Google”, no ponemos nada y “Finalizar”

Nota necesitas una cuenta de Gmail personal (no de organización) y Chrome instalado.

Alternativamente se puede usar vscode <https://code.visualstudio.com/> pero conseguir tener configurado un agente lleva más pasos.

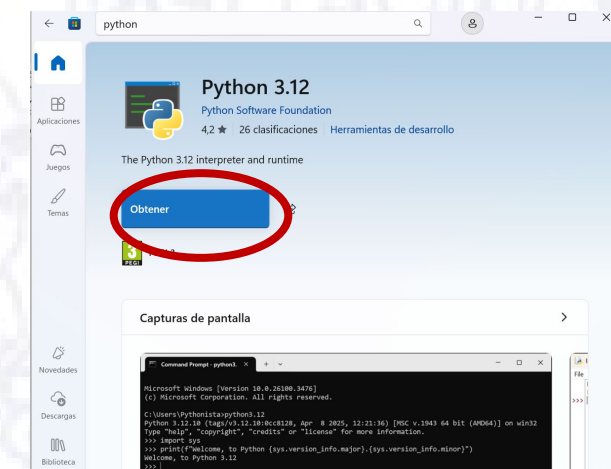
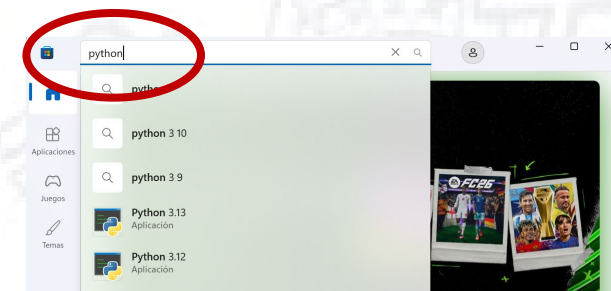
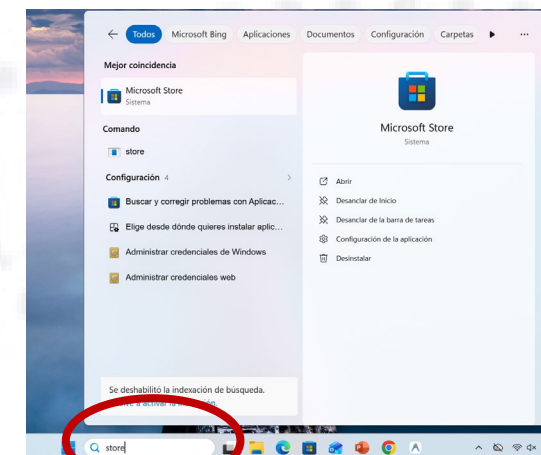


INSTALACIÓN • 4 DE 4

Instalar Python (Microsoft Store)

- 1 Abre la Microsoft Store en Windows
- 2 Busca Python 3.12 (o la versión que ofrezca)
- 3 Pulsa Obtener
- 4 Espera a que termine la instalación

Nota en Mac y Linux Python suele venir ya instalado



Cómo trabajaremos en cada paso

El mismo bucle una y otra vez. Cada transparencia tiene un prompt inicial listo para copiar. Después lo puedes mejorar.

1 En primer lugar crearemos un proyecto nuevo y seleccionamos el modelo: para este tipo de tareas sencillas Gemini 3.5 Flash (Low)

2 Copia el PROMPT de la transparencia y

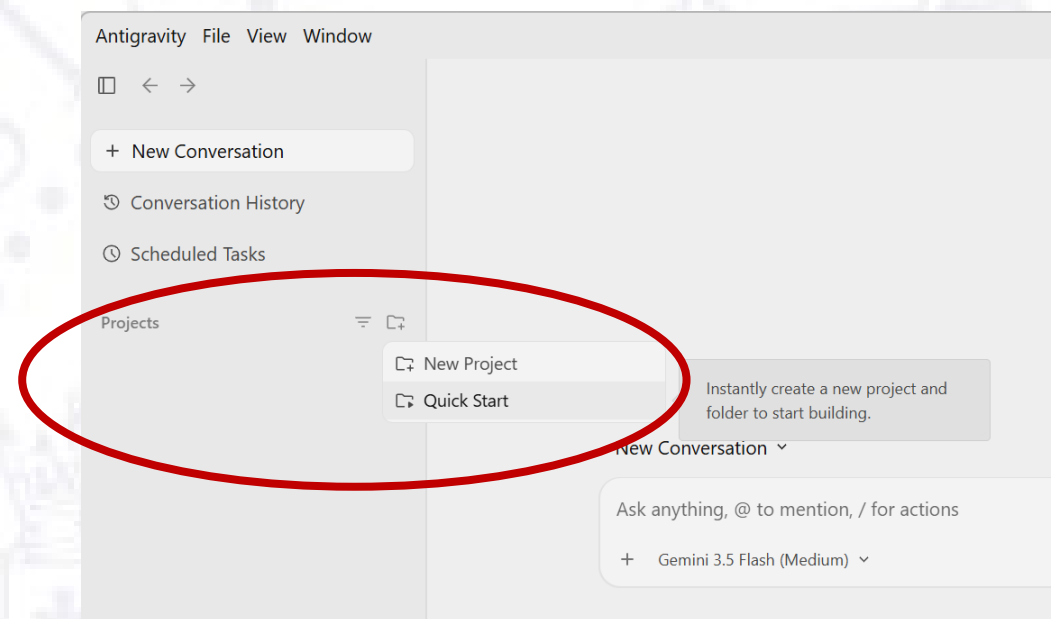
pégalo en la caja del Agente

3 El agente propone un PLAN y escribe el código:

Revisar y aceptar (Review)

4 Ejecuta el comando que te indique y mira el resultado en Chrome. Si algo falla o no te gusta, pídeselo en otro prompt

Cuota diaria limitada: trabajad en parejas y reutilizad la misma sesión del agente.



Selecting a Thinking Level

New model option: Gemini 3.5 Flash (Low). Use low thinking for simple tasks that do not require frontier intelligence or deep reasoning. In such cases, you will get similar results with less token usage.

Dismiss

1

PROYECTO 1

De una tabla a un panel

Tratamos una tabla como si fuera tu inventario: la descargamos, normalizamos, medimos y la exploramos.

Ingesta + normalización automatizar la entrada de datos y limpiar las fechas

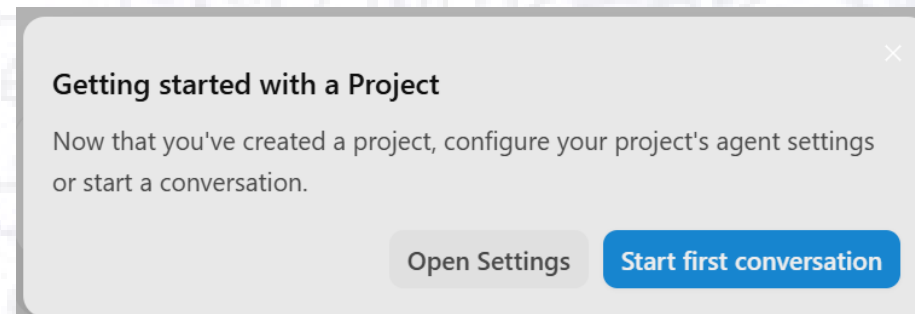
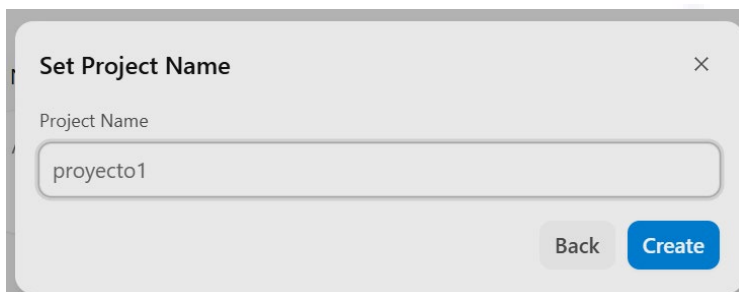
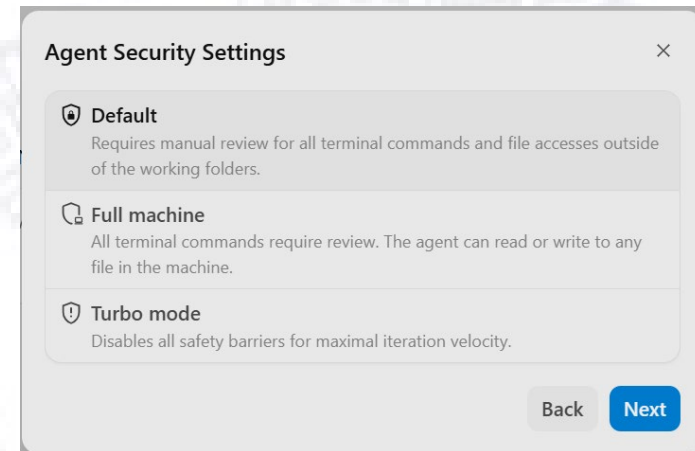
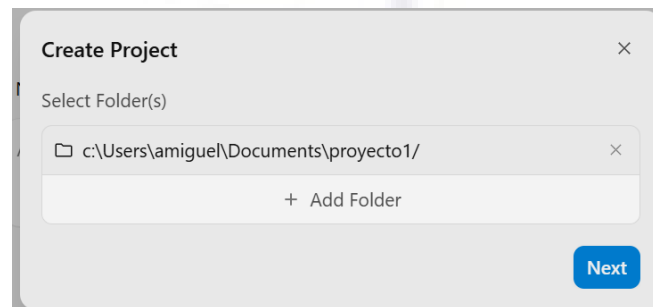
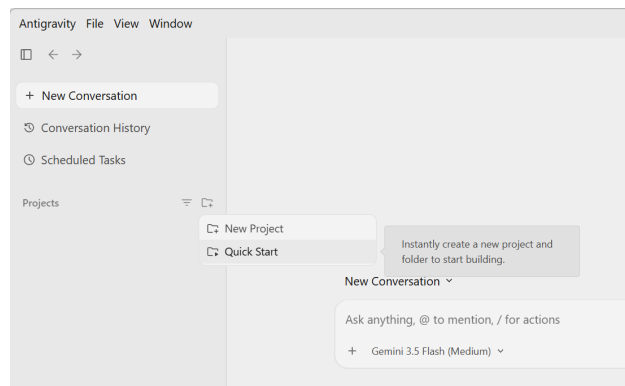
Medir el fondo estadísticas por tipo y por comunidad autónoma

Instrumento de descripción un panel con filtros y una cronología



0 Crear un nuevo proyecto

Creamos un nuevo proyecto en antigraVity



1

De la web a tu inventario (Excel)

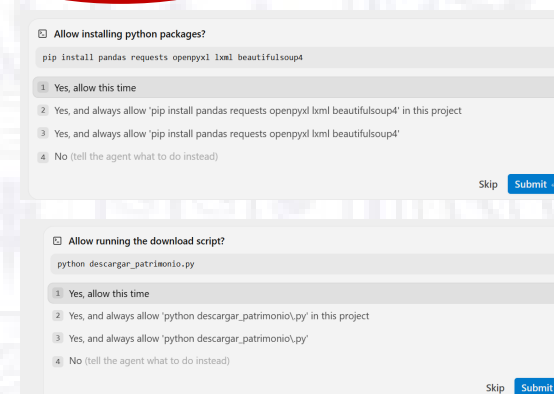
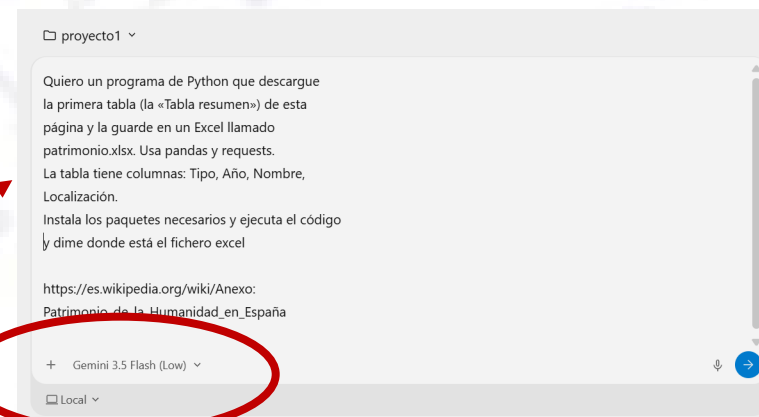
Automatizamos la entrada de datos: descargamos una tabla de Wikipedia a una Excel. Recuerda seleccionar el modelo **Gemini Flash 3.5 (Low)**

PROMPT · copiar y pegar

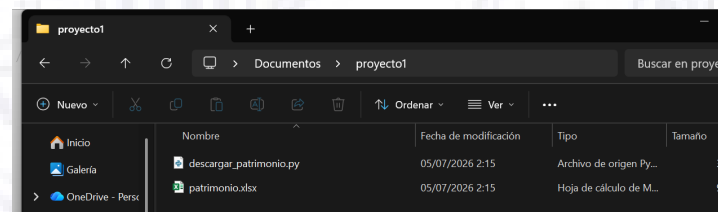
Quiero un programa de Python que descargue la primera tabla (la «Tabla resumen») de esta página y la guarde en un Excel llamado patrimonio.xlsx. Usa pandas y requests. La tabla tiene columnas: Tipo, Año, Nombre, Localización.

Instala los paquetes necesarios y ejecuta el código y dime donde está el fichero excel. Asume que python -m pip es el método correcto para instalar paquetes y úsalo directamente.

https://es.wikipedia.org/wiki/Anexo:Patrimonio_de_la_Humanidad_en_España



Si hemos elegido el modo seguro nos hará una pregunta cada vez que ejecute un comando



2

Normalizar las fechas (trabajo de archivo)

La tarea archivística por excelencia: dejar fechas y categorías limpias y coherentes.

PROMPT · copiar y pegar

Con patrimonio.xlsx: la columna «Año» a veces trae varios años (por ejemplo «1984, 1994»: inscripción y ampliación).
Crea una columna «Año_inscripcion» con solo el primer año, como número entero.
Limpia también «Tipo» para quedarnos con: Cultural, Natural, Mixto e Inmaterial.
Guarda el resultado en patrimonio_limpio.xlsx

He procesado la limpieza solicitada y he guardado el resultado.

Ubicación del nuevo fichero

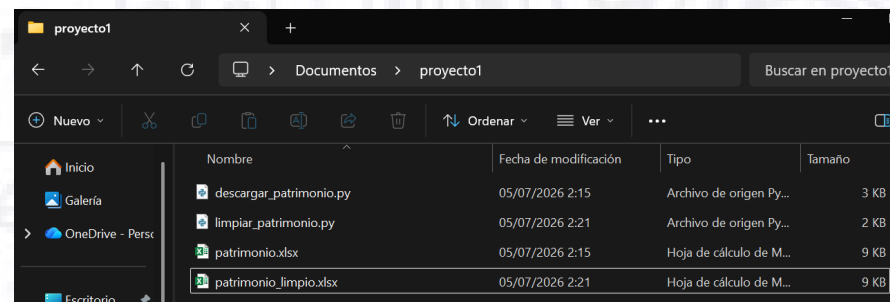
El archivo limpio se ha guardado en: `c:\Users\amiguel\Documents\proyecto1\patrimonio_limpio.xlsx`

Detalles del proceso de limpieza

1. **Creación de Año_inscripcion**: Se extrajo el primer año de la columna **Año** mediante una expresión regular que identifica la primera secuencia de dígitos y se convirtió a un entero (1984, 1994 → 1984).
2. **Normalización de Tipo**: Se removió el prefijo "Bien " de todos los valores y se aplicó formato capitalizado, resultando exactamente en: Cultural, Natural, Mixto e Inmaterial.

Aquí tienes el código del script de limpieza que se encuentra en [limpiar_patrimonio.py](#):

```
python
```



Nombre	Fecha de modificación	Tipo	Tamaño
descargar_patrimonio.py	05/07/2026 2:15	Archivo de origen Py...	3 KB
limpiar_patrimonio.py	05/07/2026 2:21	Archivo de origen Py...	2 KB
patrimonio.xlsx	05/07/2026 2:15	Hoja de cálculo de M...	9 KB
patrimonio_limpio.xlsx	05/07/2026 2:21	Hoja de cálculo de M...	9 KB

3

Estadísticas del fondo (dos gráficos)

Medimos la colección, como al describir un fondo: cuántos bienes de cada tipo y lugar.

PROMPT · copiar y pegar

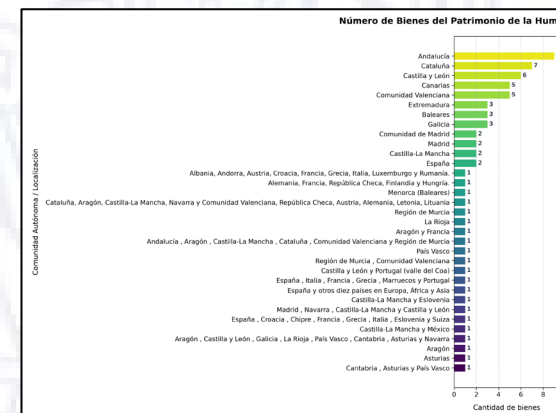
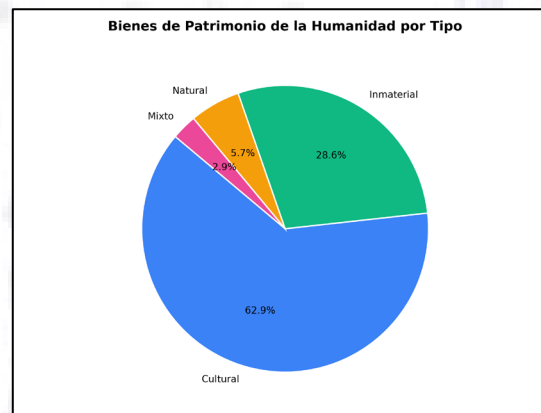
Con patrimonio_limpio.xlsx crea dos gráficos y guárdalos como imagen:

- 1) una tarta con el número de bienes por «Tipo»;
 - 2) barras horizontales con el número de bienes por «Localización» (comunidad autónoma), de mayor a menor.
- Ábrelos al terminar.

He generado y abierto los dos gráficos a partir del archivo `patrimonio_limpio.xlsx`. Las imágenes se han abierto de forma automática en tu visor de imágenes predeterminado.

Las imágenes se han guardado con los siguientes nombres en el directorio del proyecto:

1. **Gráfico de Tarta (bienes por Tipo):**  bienes_por_tipo.png
2. **Gráfico de Barras Horizontales (bienes por Localización):**  bienes_por_localizacion.png



4

Un instrumento de descripción interactivo

Cerramos con un panel con filtros: en la práctica, un instrumento de descripción.

PROMPT · copiar y pegar

Quiero un dashboard con Streamlit al que le vuelque patrimonio_limpio.xlsx.

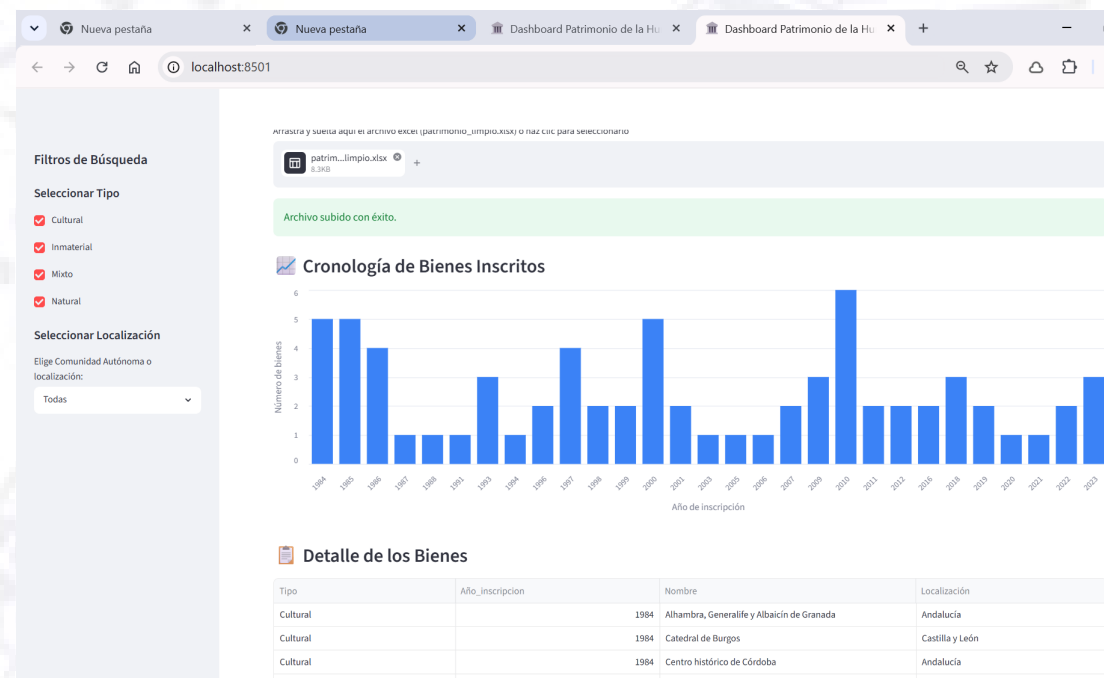
Arriba, una caja para arrastrar y soltar la Excel; si hago click en la caja abre un diálogo para seleccionar el fichero.

A la izquierda, casillas para elegir «Tipo» y un desplegable para «Localización».

A la derecha, un gráfico de bienes por año (cronología) que se filtre, y debajo la tabla.

Instala lo necesario y ejecútalo al acabar.

Asume que python -m pip es el método correcto para instalar paquetes y úsalo directamente.



2

PROYECTO 2

Explorador de Debate

El proyecto grande: una web para navegar un video ya transcrito y diarizado.

Entrada sueltas el video (.mp4) y los subtítulos (.srt)

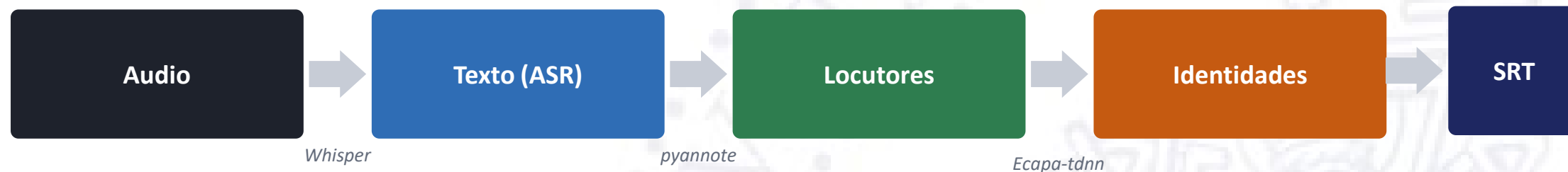
Navegas timeline con colores por locutor y clic para saltar

Metadatos locutores, tiempos, texto y estadísticas del debate



Los datos: del audio al SRT

El trabajo pesado ya está hecho antes del taller. Hoy partimos del SRT ya preparado.



Transcripción (ASR) + diarización + identidades

Whisper pasa la voz a texto con tiempos y **pyannote** separa quién habla en cada momento. Con **Ecapa-TDNN** hemos asignado las identidades. El resultado es un SRT con número, tiempos, texto y una etiqueta de locutor.

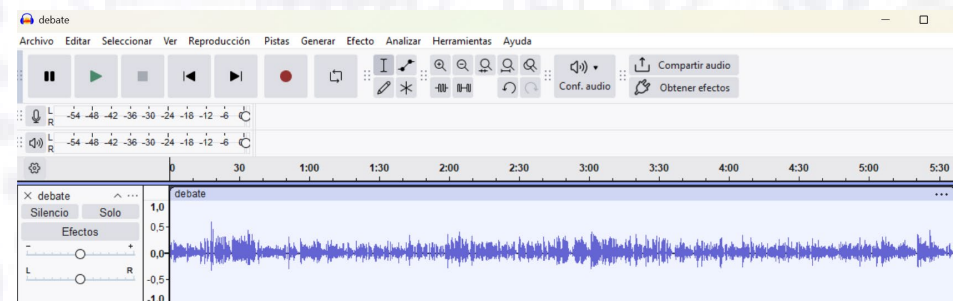
Metadato no supervisado: Mediante herramientas abiertas

debate.srt

9

00:00:33,000 --> 00:00:34,060

<Yann LeCun> Feifei, what do you think?



0

Crear el proyecto y el servidor

Arrancamos una web mínima con Python que iremos ampliando paso a paso.

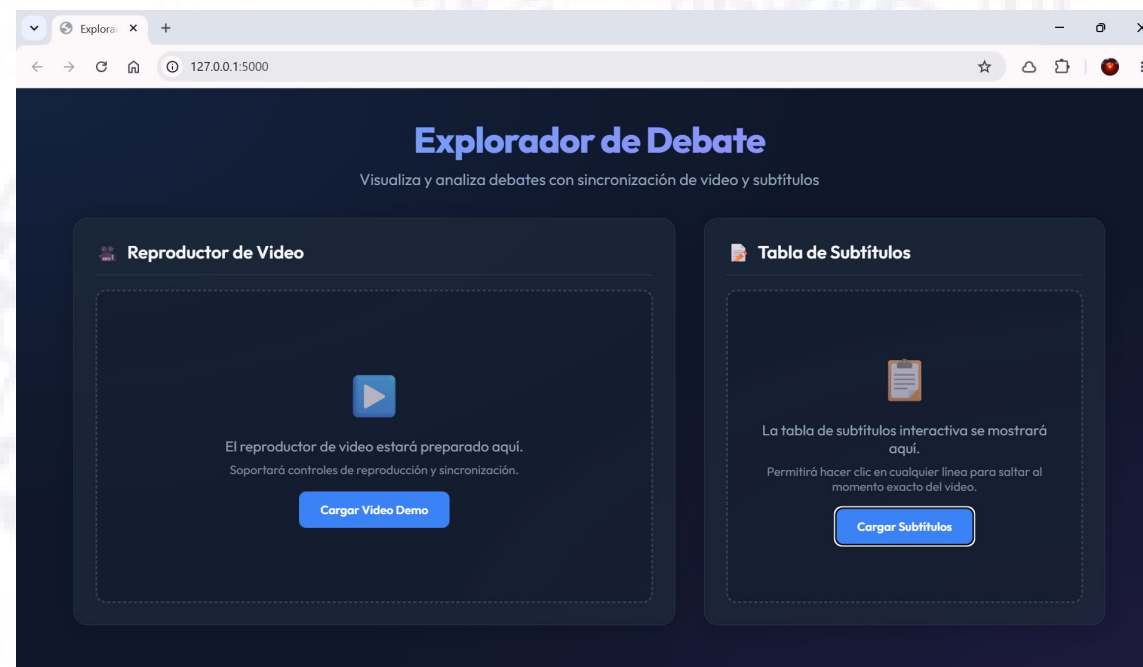
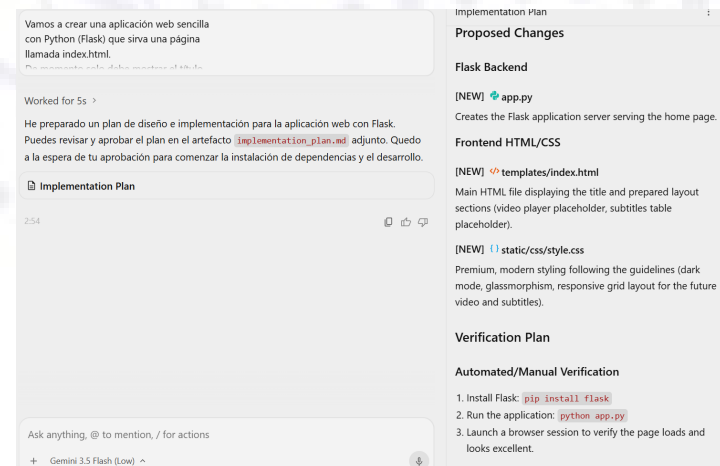
PROMPT · copiar y pegar

Vamos a crear una aplicación web sencilla con Python (Flask) que sirva una página llamada `index.html`.

De momento solo debe mostrar el título 'Explorador de Debate' y quedar preparada para añadir después un reproductor de video y una tabla de subtítulos.

Instala lo que sea necesario y arranca el servidor para verla en Chrome. Asume que `python -m pip` es el método correcto para instalar paquetes y úsalo directamente.

Escribe la dirección que tengo que abrir en Chrome o lánzalo directamente.



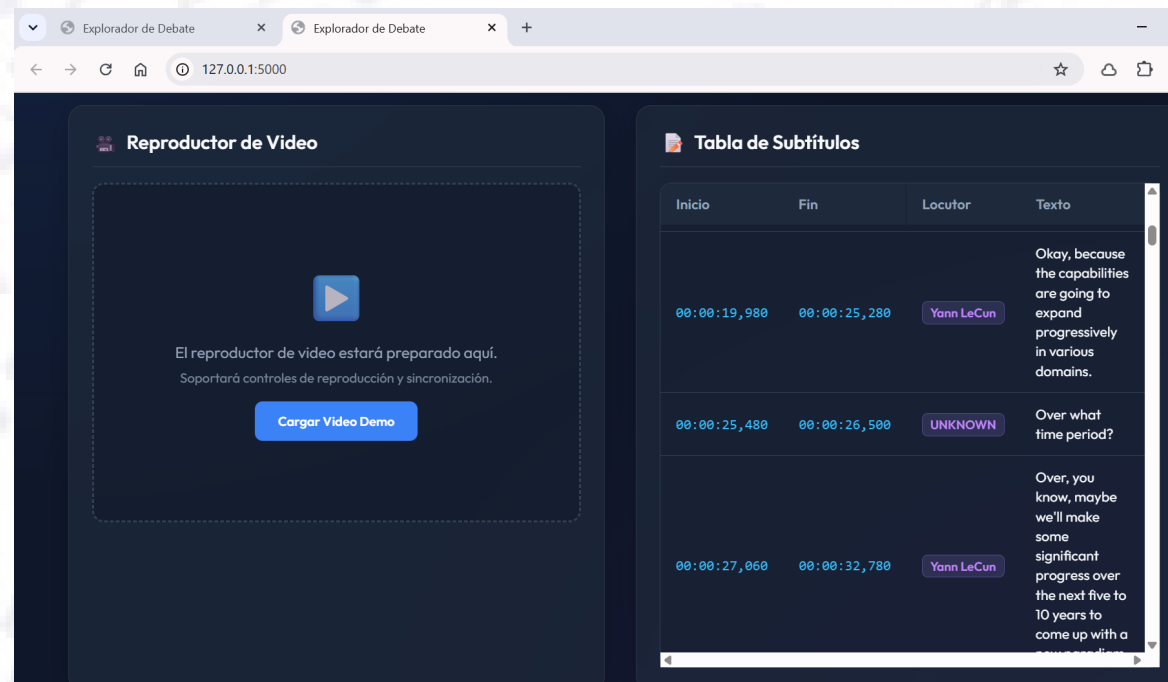
1

Cargar el SRT y verlo como tabla

Primera pieza: leer los subtítulos e interpretarlos, incluida la etiqueta de locutor.

PROMPT · copiar y pegar

Añade una caja para arrastrar y soltar un archivo .srt. Si se hace click sobre la caja se abre un diálogo para localizar el archivo .srt. Al soltarlo, léelo con JavaScript e interpreta el formato SRT: número, inicio, fin y texto. Al principio de cada texto hay una etiqueta como <Yann LeCun> que indica quién habla; sepárala como 'Locutor'. Muestra los segmentos a la derecha en una tabla con columnas: Inicio, Fin, Locutor, Texto.



The screenshot shows a web browser with two tabs labeled 'Explorador de Debate'. The address bar shows '127.0.0.1:5000'. The main content area is divided into two panels. The left panel, titled 'Reproductor de Video', contains a large play button icon and the text: 'El reproductor de video estará preparado aquí. Soportará controles de reproducción y sincronización.' Below this is a blue button labeled 'Cargar Video Demo'. The right panel, titled 'Tabla de Subtítulos', displays a table with four columns: 'Inicio', 'Fin', 'Locutor', and 'Texto'. The table contains three rows of subtitle data.

Inicio	Fin	Locutor	Texto
00:00:19,980	00:00:25,280	Yann LeCun	Okay, because the capabilities are going to expand progressively in various domains.
00:00:25,480	00:00:26,500	UNKNOWN	Over what time period?
00:00:27,060	00:00:32,780	Yann LeCun	Over, you know, maybe we'll make some significant progress over the next five to 10 years to come up with a

2

Añadir el video y el reproductor

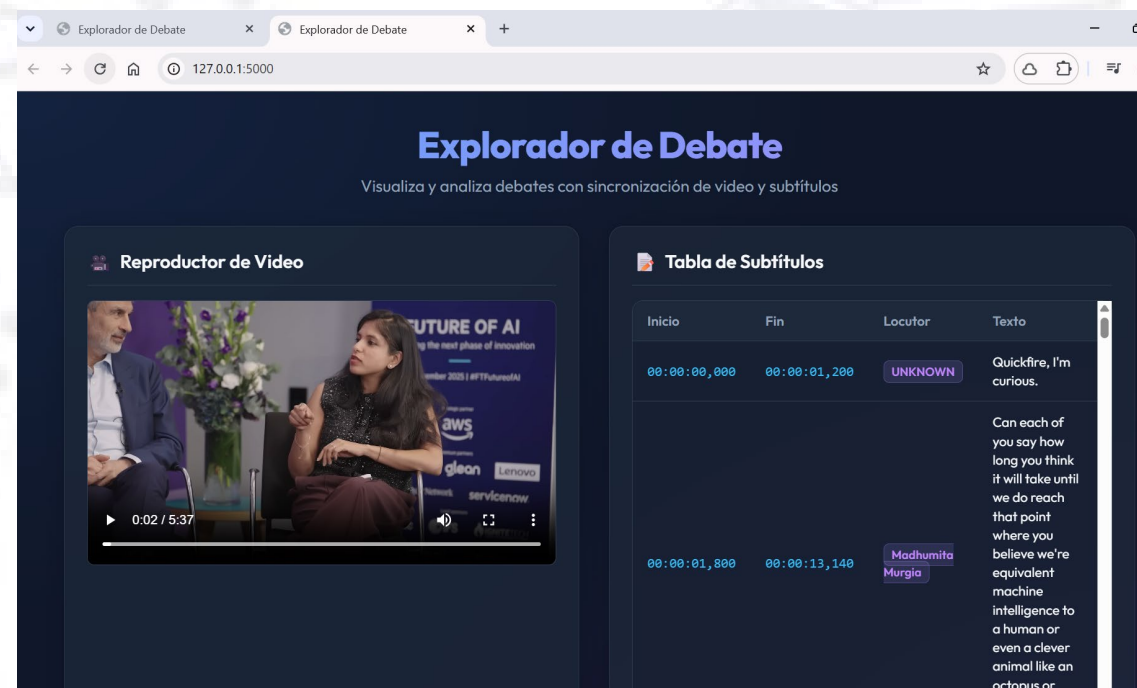
Segunda caja de arrastrar y soltar, ahora para el video del debate.

PROMPT · copiar y pegar

Implementa la caja de arrastrar y soltar un archivo de video .mp4. Con click se abre diálogo de fichero.

Cuando lo suelte, muéstralo en un reproductor de video con controles (play, pausa, barra de tiempo) arriba a la izquierda de la página.

Deja la tabla de subtítulos a la derecha del video.



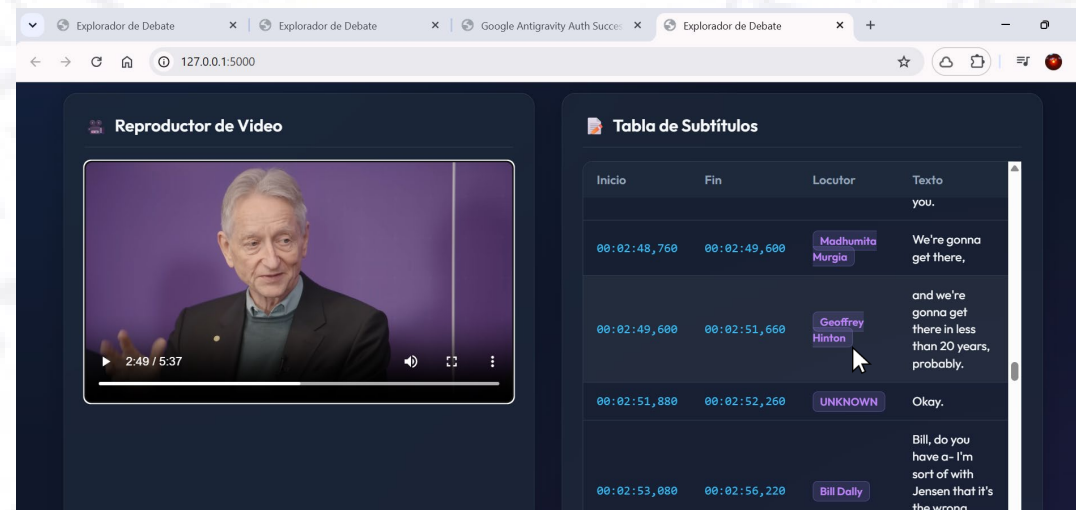
3

Clic en una fila para reproducir

Conectamos tabla y video: la primera interacción útil.

PROMPT · copiar y pegar

Ahora quiero que, al hacer clic en una fila de la tabla, el reproductor salte al tiempo de inicio de ese segmento y empiece a reproducir automáticamente.



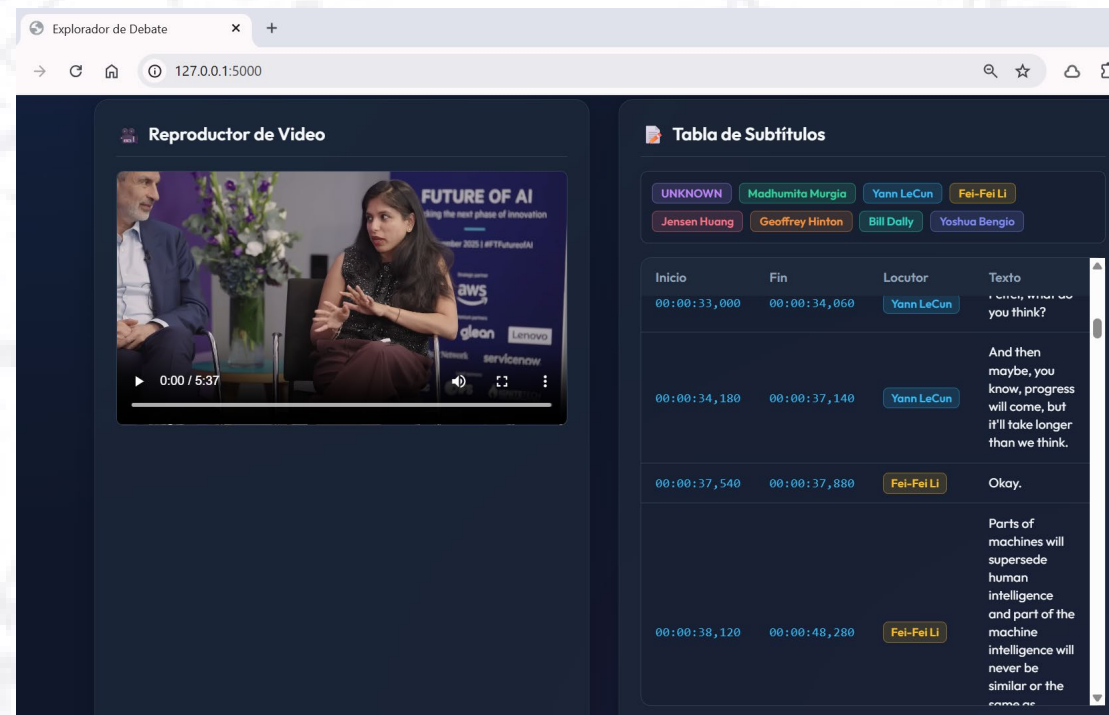
4

Un color por locutor

Damos identidad visual a cada voz. Empieza a leerse como un debate.

PROMPT · copiar y pegar

Asigna un color distinto a cada locutor.
Pinta con ese color el fondo de la etiqueta
del locutor en cada fila de la tabla.
Añade arriba una pequeña leyenda con el
color de cada locutor.



The screenshot shows a web application interface with two main panels. The left panel, titled "Reproductor de Video", displays a video player with a video of two people in a discussion. The right panel, titled "Tabla de Subtítulos", shows a table of subtitles with columns for "Inicio", "Fin", "Locutor", and "Texto". The table lists several subtitles, each associated with a specific speaker and a color-coded background. A legend at the top of the table identifies the speakers and their corresponding colors: UNKNOWN (purple), Madhumita Murgia (green), Yann LeCun (blue), Fei-Fei Li (yellow), Jensen Huang (red), Geoffrey Hinton (orange), Bill Dally (light blue), and Yoshua Bengio (dark blue).

Inicio	Fin	Locutor	Texto
00:00:33,000	00:00:34,060	Yann LeCun	... think, what do you think?
00:00:34,180	00:00:37,140	Yann LeCun	And then maybe, you know, progress will come, but it'll take longer than we think.
00:00:37,540	00:00:37,880	Fei-Fei Li	Okay.
00:00:38,120	00:00:48,280	Fei-Fei Li	Parts of machines will supersede human intelligence and part of the machine intelligence will never be similar or the same as

5

La línea de tiempo (timeline)

La pieza estrella: ver todo el debate de un vistazo y navegarlo.

PROMPT · copiar y pegar

Debajo del video quiero una línea de tiempo horizontal que represente toda la duración del debate.

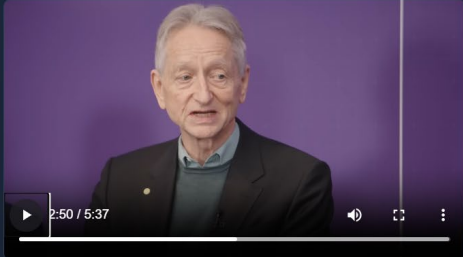
Dibuja cada segmento como un bloque de color según su locutor, en su posición temporal y con el ancho proporcional a su duración.

Al hacer clic en un bloque, el video salta a ese segmento.

Explorador de Debate

Visualiza y analiza debates con sincronización de video y subtítulos

Reproductor de Video



2:50 / 5:37

Línea de Tiempo del Debate



Tabla de Subtítulos

UNKNOWN Madhumita Murgia Yann LeCun Fei-Fei Li
Jensen Huang Geoffrey Hinton Bill Daily Yoshua Bengio

Inicio	Fin	Locutor	Texto
			animal like an octopus or whatever?
00:00:13,220	00:00:17,100	Madhumita Murgia	How far away are we, just the years?
00:00:17,680	00:00:19,040	UNKNOWN	It's not going to be an event.
00:00:19,500	00:00:19,820	UNKNOWN	Okay.

6

Resaltado sincronizado

Que la interfaz siga al video mientras se reproduce.

PROMPT · copiar y pegar

Mientras el video se reproduce, resalta en cada momento el segmento que se está escuchando: tanto la fila de la tabla como el bloque del timeline.

Además, haz que la tabla se desplace (scroll) sola para mantener a la vista la fila activa que se está reproduciendo.

Reproductor de Video



0:34 / 5:37

Línea de Tiempo del Debate


Tabla de Subtítulos

UNKNOWN
Madhumita Murgia
Yann LeCun
Fei-Fei Li
Jensen Huang
Geoffrey Hinton
Bill Dally
Yoshua Bengio

Inicio	Fin	Locutor	Texto
00:00:27,060	00:00:32,780	Yann LeCun	Over, you know, maybe we'll make some significant progress over the next five to 10 years to come up with a new paradigm.
00:00:33,000	00:00:34,060	Yann LeCun	Feifei, what do you think?
00:00:34,180	00:00:37,140	Yann LeCun	And then maybe, you know, progress will come, but it'll take longer than we think.

7

Barra lateral e estadísticas

Añadimos información y cifras del debate.

PROMPT · copiar y pegar

Debajo muestra estadísticas del debate:
número de locutores y tiempo total de
intervención de cada uno, en segundos en una
tabla y también en un gráfico de barras.



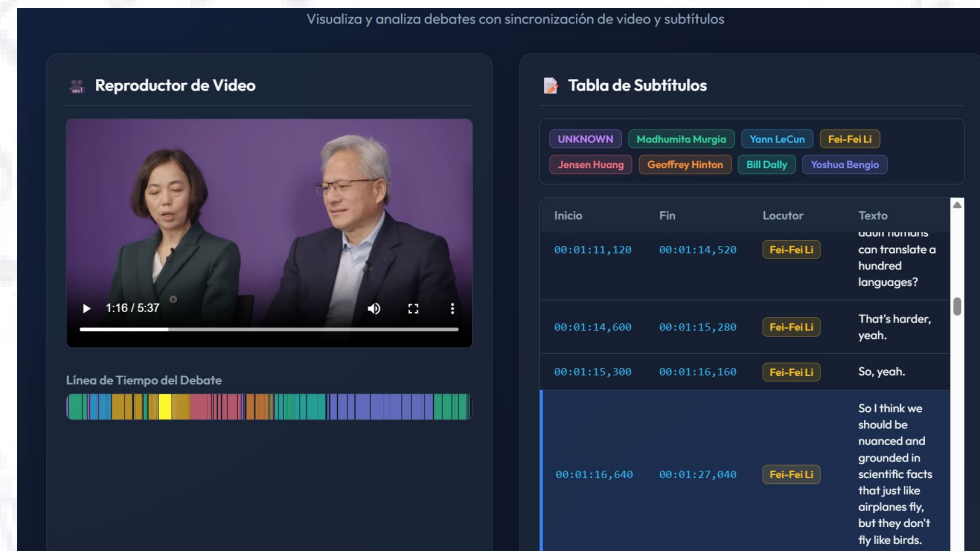
8

Controles con teclado

Añadimos unos controles de teclado.

PROMPT · copiar y pegar

Puedes añadir controles desde el teclado para el reproductor: espacio para pausa/play, flecha izquierda para retroceder 10s, flecha derecha para avanzar 10s y la coma para pausar y avanzar un solo frame.



9

Mini RAG (avanzado)

Añadimos un mini RAG para buscar, prueba a buscar: está en contra

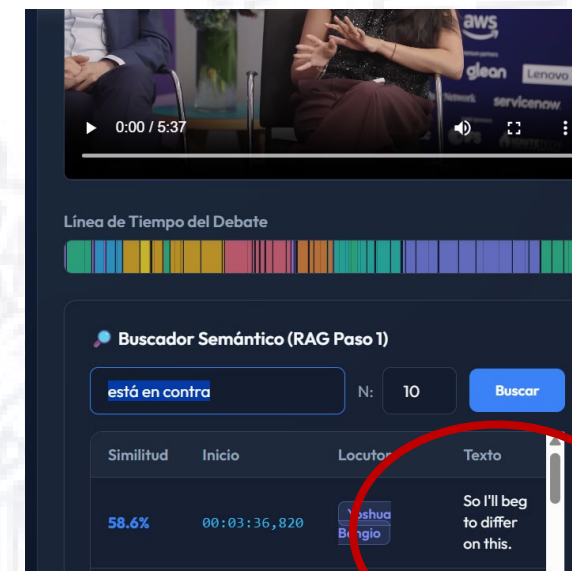
PROMPT · copiar y pegar

Quiero añadir un buscador semántico como primer paso de una demo RAG. Para evitar calcular los embeddings de todos los subtítulos en tiempo real, usaremos una matriz precalculada en formato NumPy (.npy) de dimensiones (N, 1024) que se subirá mediante un nuevo "drop-zone" en la interfaz. El backend (app.py) debe leer este archivo de forma segura usando 'file.read()', cargarlo en memoria con 'np.load(..., allow_pickle=True)' y servir la app con 'debug=False' para evitar bloqueos del puerto 5000 en Windows.

Al realizar una búsqueda por texto, el backend codificará la consulta usando el modelo local "Qwen/Qwen3-Embedding-0.6B" (mostrando un estado de "Cargando modelo..." en la UI mientras arranca). Para calcular la similitud de coseno con 'sentence_transformers.util.cos_sim', es indispensable castear explícitamente el tensor de la matriz de embeddings cargada al mismo tipo de datos ('dtype', usualmente BFloat16) del vector devuelto por el modelo Qwen. En el frontend (index.html), asegura detener la propagación del evento clic en los inputs de archivos ocultos ('e.stopPropagation()') para evitar congelar el navegador por bucles infinitos. Tras la búsqueda, se mostrarán en una tabla interactiva los N mejores resultados (N configurable, por defecto 10), permitiendo hacer clic sobre cada fila para saltar al segundo exacto de reproducción del video.

prueba a buscar:

está en contra



So I'll beg to differ on this.

Búsqueda por embedding semántica no son las palabras exactas (ni el mismo idioma!)

Mini RAG + LLM (avanzado)

Añadimos al mini RAG un LLM para preguntar sobre las búsquedas

```
sk-proj-QbqD9g3tcu76OGd1kckWkOrkDY1AGL4DGw0QLGOrWw7s85hcbFwqq_d-  
F90ERBvR1UzJnKABQtT3BlbkFJWlFU_FCsJXeiJVRPXqiY2wFhphqAvnlc6GTyQkkeEgFUvywW  
60Tc7FYbc1aPerH7aqJgZ8CsUA
```

PROMPT · copiar y pegar

ahora por último añade un cuadro de texto adicional debajo del módulo de RAG, que va a servir para hacer una consulta mediante la api de chatgpt. Para ello tendremos un campo para introducir la API key y otro para hacer la consulta. Lo que haremos es primero hacer una pregunta al RAG que hemos preparado, la pregunta del RAG la va a diseñar chatgpt mediante un prompt que le diga dada esta pregunta del usuario escribe una pregunta corta con la que vamos a buscar subtitulos adecuados en un RAG, con los 10 más parecidos, la información del locutor de cada segmento y los 4 segmentos anteriores a cada uno de los localizados para entender el contexto se lanzará la consulta final a la API advirtiéndole de que se añade información buscada en un RAG concatenada a la query y que solo conteste sobre ese contexto y si la información que se pide no está que conteste no tengo información suficiente. La respuesta final de la api de chatgpt se me mostrará en un cuadro de texto debajo del de la query y la apikey.

prueba a buscar:

qué opinan los invitados sobre la capacidad para planificar de la IA

Generación RAG (Paso 2: ChatGPT)

qué opinan los invitados sobre la capacidad

Preguntar

Respuesta de ChatGPT:

Yoshua Bengio menciona que hay datos interesantes que muestran la capacidad de las IA para planificar a lo largo de diferentes horizontes, y que han crecido exponencialmente en los últimos seis años. Él sugiere que, si esta tendencia continúa, la IA podría alcanzar el nivel de un empleado en su trabajo dentro de aproximadamente cinco años. Además, destaca que muchas empresas están enfocándose en la capacidad de la IA para realizar investigación en IA, lo que podría desbloquear muchas otras capacidades. Sin embargo, también advierte que hay mucha incertidumbre y que no se deben hacer afirmaciones contundentes sobre el futuro.

Retos para seguir explorando

Si quieres seguir, puedes trabajar sobre más mejoras:

- **Buscar por palabra** escribe un término y salta a los momentos donde se dice
- **Filtrar por locutor** muestra solo las intervenciones de una persona
- **Siguiente intervencion** un botón para saltar a la próxima vez que habla ese locutor
- **Exportar metadatos** guarda los segmentos en CSV o JSON (campos tipo EBUCore / Dublin Core)
- **Búsqueda semántica** mejora el modelo de RAG utilizando una base de datos externa como qdrant
- **Conecta con LLM** conecta con un LLM y haz preguntas directamente sobre el contenido. Necesitarás una clave de API

De la voz al metadato

La finalidad de esta demostración no la aplicación concreta, es aprender una forma nueva de trabajar el archivo.



Hemos generado metadatos

tiempos, locutores, texto y estadísticas, sin etiquetar nada a mano: eso es metadatado no supervisado.



El agente construye herramientas

no solo analiza el archivo; fabrica el visor a medida hablando en tu idioma.



Piensa en tu caso

ten claros los formatos que se van procesar y dónde se procesan y almacenan los datos: nube o local.