PONENCIA INAUGURAL

Por una IA más sostenible de la teoría a la práctica

Enrique Alba
Catedrático · Universidad de Málaga

XXIII Jornadas Bibliotecarias de Andalucía

Córdoba · 13 noviembre 2026



¿Cuánto pesa una consulta?

En la biblioteca consultamos libros, revistas, catálogos y webs.
Cada consulta digital consume recursos.





Libro



Revista



Catálogo



Web



CPU
Procesa la consulta



MEMORIA
Mantiene datos en uso



DATOS
Se transfieren y almacenan



ENERGÍA
Hace posible todo lo demás



¿Qué parte de ese coste vemos?

HOY, MUCHAS CONSULTAS SE HACEN **CON IA**

La IA es una disciplina informática

IA clásica

Buscar Encuentra soluciones entre opciones.

Planificar Organiza pasos para un objetivo.

Optimizar Encuentra la mejor opción posible.



IA que aprende

Aprender Detecta patrones en los datos.

Predecir Anticipa lo que puede ocurrir.

Generar Crea contenido nuevo y original.

" ChatGPT es a la IA lo que el best-seller a la biblioteca: **el más prestado, no el único libro.**"

La IA existía antes de ChatGPT

1956
IA

Conferencia de Dartmouth: nace el campo de la Inteligencia Artificial.



2012
Deep learning

Redes neuronales profundas revolucionan el reconocimiento de patrones.



2017
Transformer (atención)

La arquitectura Transformer y el mecanismo de atención cambian la IA moderna.



2022
ChatGPT

La IA generativa llega al gran público.



Historia de la IA
Ciencia
Tecnología
Sociedad

La IA tiene 70 años; ChatGPT, cuatro.
Es el nieto que sale en todas las fotos.

Nombres reales, no humo

Lo que sí nació en 2022 fue ChatGPT, sobre un avance científico de 2017: los transformers con atención.

Interfaces

ChatGPT · Claude · Gemini · Copilot · Perplexity

Modelos

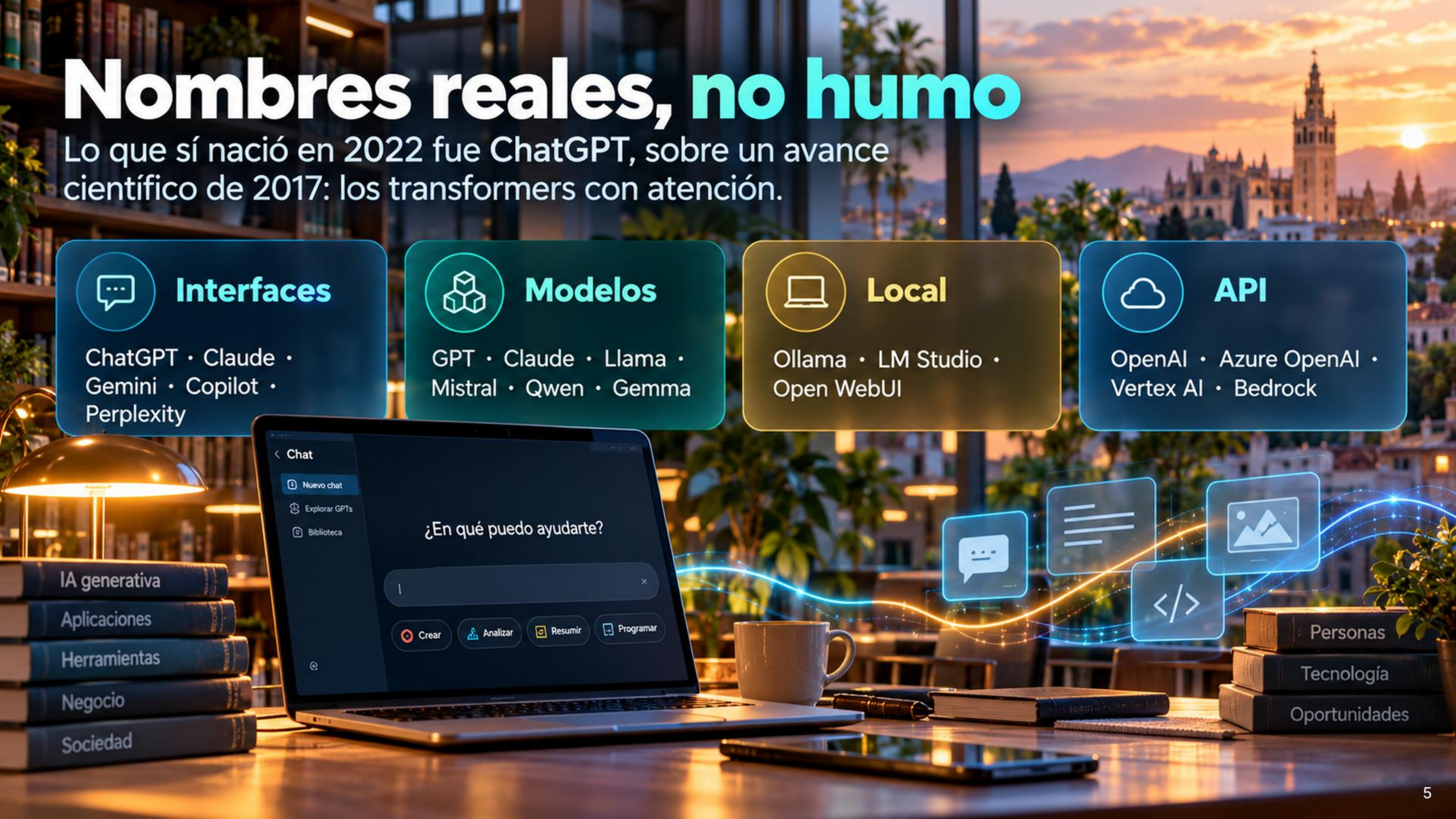
GPT · Claude · Llama · Mistral · Qwen · Gemma

Local

Ollama · LM Studio · Open WebUI

API

OpenAI · Azure OpenAI · Vertex AI · Bedrock



¿Respuesta o conocimiento?



La IA genera **texto plausible**

La biblioteca aporta **contexto y confianza**

La supervisión humana **sigue siendo clave**

La **IA** es ese usuario que **nunca dice ‘no lo sé’.**

De la biblioteca tradicional a la aumentada



Tradicional
colección · sala · personas

Digital
catálogo · web · préstamo

Conectada
repositorios · acceso remoto · datos

Aumentada por IA
RAG · asistentes · automatización

IA de cara al público

Búsqueda semántica

Encuentra lo que realmente necesitas, aunque uses tus propias palabras.

Asistente del catálogo

Te ayuda a descubrir, explorar y conocer mejor los fondos de la biblioteca.

Recomendación

Sugerencias personalizadas de libros, artículos, películas y más.

Traducción y accesibilidad

Contenido en múltiples idiomas y herramientas de apoyo para todas las personas.

Consulta de colecciones digitalizadas

Accede a manuscritos, fotografías, prensa histórica y más, desde cualquier lugar.

Multilingüe

Servicio en varios idiomas para una biblioteca más abierta, inclusiva y global.

Más servicio, no menos biblioteca.

IA que el usuario no ve



COLECCIONES Y PROCESOS

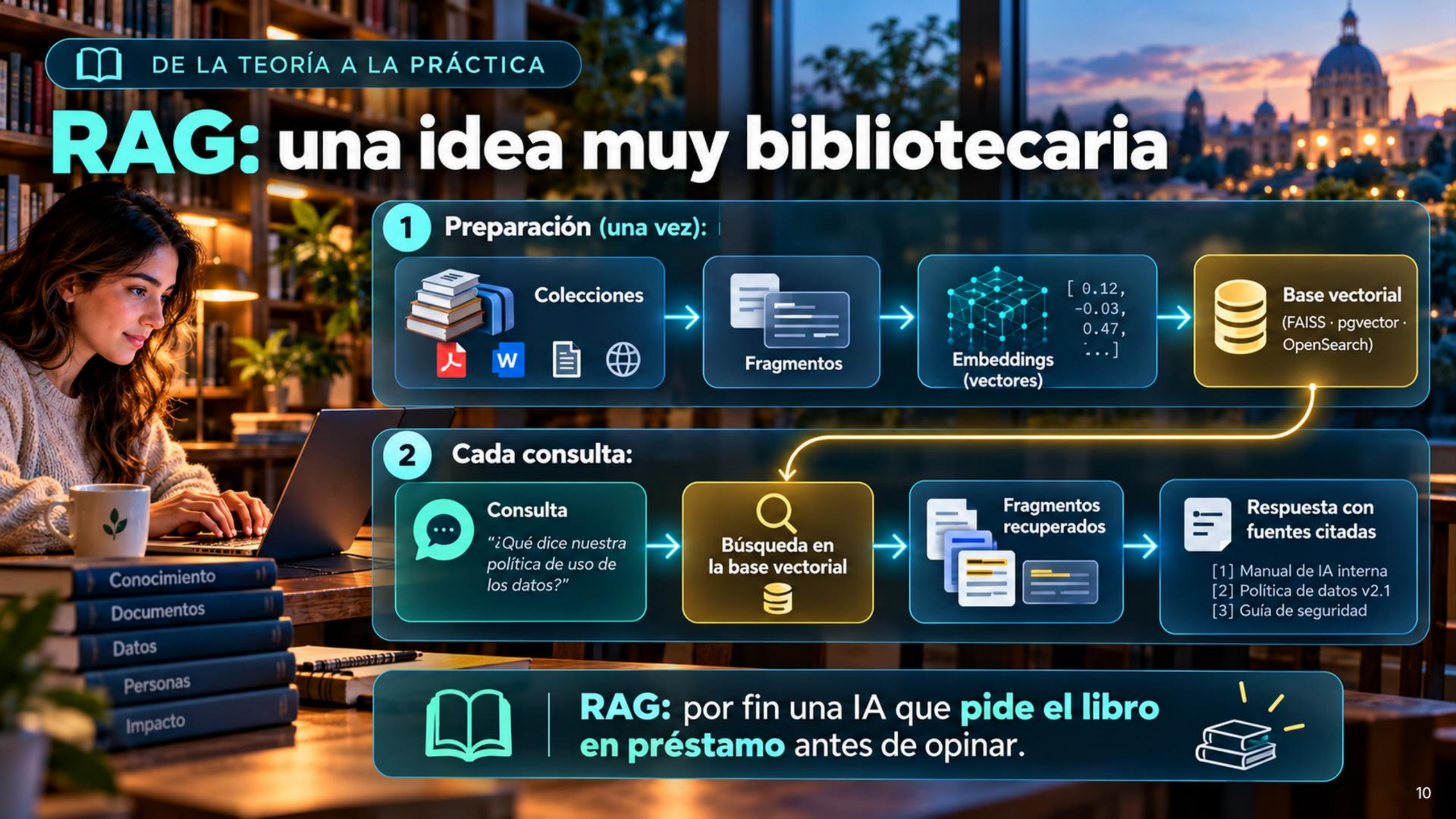
- Catalogación
- Metadatos
- OCR
- Clasificación
- Resúmenes

GESTIÓN DE LA BIBLIOTECA

- Ofimática
- Solicitudes de financiación
- Memorias anuales
- Estadísticas de uso y préstamos

DE LA TEORÍA A LA PRÁCTICA

RAG: una idea muy bibliotecaria



RAG: por fin una IA que **pide el libro en préstamo** antes de opinar.

IA completamente local

Privacidad y control dentro de la biblioteca



Herramientas

Ollama · LM Studio
Open WebUI · llama.cpp

Ventajas

privacidad · control · coste predecible

Costes

hardware · mantenimiento · modelos más modestos

Local *no* significa automáticamente mejor; sí significa **mayor control** sobre datos y medición.

IA remota por API

No compramos el cerebro: pagamos por preguntarle



APIs y servicios

OpenAI API · Claude API · Gemini API · Azure OpenAI · Vertex AI · Bedrock

Ventajas

- potencia
- escalabilidad
- arranque rápido

Atención

- datos
- dependencia del proveedor
- opacidad energética

ENS en contextos públicos

Una API puede ser excelente, pero conviene saber qué datos salen y a dónde van.

Arquitectura híbrida

Equilibrio entre control, coste y potencia



AI Act: riesgo y transparencia





Para bibliotecas

La mayoría de usos ordinarios estarán en riesgo mínimo o en transparencia.

inventario · formación · supervisión

No se trata de frenar la IA, sino de usarla con **criterio** y **responsabilidad**.

Datos personales: GDPR + LOPDGDD



Préstamos, búsquedas y chats pueden revelar mucho sobre una persona.

- **finalidad**
- **minimización**
- **base jurídica**
- **seguridad**
- **privacidad desde el diseño**

Preguntas clave | qué datos salen · a dónde van · durante cuánto tiempo

Los datos tienen valor, pero también requieren **protección y buen gobierno.**

Del prompt al transformer

Prompt
(del usuario)

Explícame de forma sencilla qué es un agujero negro.

Texto

"Explícame de forma sencilla qué es un agujero negro."

Tokens

Explícame
de
forma
sencilla
qué
es
un
agujero
negro
.

Transformer



Probabilidad de la siguiente palabra

que	42%
y	18%
porque	12%
es	10%
se	8%
por	6%
...	4%

Respuesta
en lenguaje natural

Un agujero negro es una región del espacio donde la gravedad es tan intensa que nada, ni siquiera la luz, puede escapar. Se forma cuando una estrella muy masiva colapsa sobre sí misma y comprime toda su materia en un espacio muy pequeño.



Para mucha gente es una caja negra; **para un ingeniero es una gran máquina estadística.**

Lo importante aquí: **aprende patrones y genera respuestas.**

Primero entendemos qué es; **después podremos hablar de su coste.**

Entrenar no es inferir



Entrenamiento

- muchos datos
- muchas GPU
- coste muy alto
- ocurre pocas veces

Inferencia

- cada consulta
- coste menor
- latencia visible
- se repite millones de veces



Impacto total ≈ coste de entrenamiento + N consultas × coste por consulta

Una consulta puede ser **pequeña**; **millones de consultas, ya no.**

¿Por qué consume energía la IA?



cómputo
(CPU/GPU)

memoria

almacenamiento

red

refrigeración

La nube no es una nube:
son ordenadores, redes y edificios

Y además está el coste de comprar, renovar y desechar máquinas

Medimos para decidir

PTME

Precisión	Tiempo	Memoria	Energía
% aciertos	latencia	GB	J/consulta





No buscamos el mínimo de una sola métrica;
buscamos el mejor equilibrio

MLOps: la IA no se despliega y se olvida



datos → modelo → pruebas → despliegue → monitorización → mejora

versiones • evaluación • trazabilidad • seguridad • sesgos

Operar IA de forma continua y responsable

En producción, la IA también necesita mantenimiento, métricas y supervisión.

La escala del consumo

Primero medimos un servicio. Luego entendemos el sistema.



Fuente: AIE y Red Eléctrica · cifras 2025–2026

De una consulta a un año

De un gesto digital a una magnitud anual.



Google, 2025 · biblioteca: 500 consultas/día, aproximadamente

Agua y CO_2 por consulta

La energía digital también tiene huella material.



1 consulta
≈ 0,26 mL

500 consultas/día
≈ 130 mL/día

1 año
≈ 47 L

10.000 L
≈ 78 personas · 1 día



1 consulta
≈ 0,03 g CO_2e

500 consultas/día
≈ 15 g/día

1 año
≈ 5,5 kg

1 t CO_2e
≈ 33 millones
de consultas

CO_2e = CO_2 equivalente · Google, 2025 · equivalencias aproximadas

¿Cuánto cuesta usar una API?

100.000 consultas/mes · 2.000 tokens de entrada + 500 de salida



GPT-5.4 mini

- **100.000** consultas/mes
- **2.000** tokens de entrada
- **500** tokens de salida

API

≈ **345** €/mes

≈ 4.140 € / año

GPT-5.4

- **100.000** consultas/mes
- **2.000** tokens de entrada
- **500** tokens de salida

API

≈ **1.150** €/mes

≈ 13.800 € / año

No incluye caché, herramientas, alojamiento ni impuestos.

Precios oficiales · conversión a euros aproximada

Elegir bien el modelo ahorra

La mejor IA no siempre es la más grande.



Green AI

Investigación para que la propia IA consuma menos.



Cuantización

Menos bits,
misma utilidad.



Modelos pequeños

Modelos más ligeros
y eficientes para
tareas específicas.



Caché y reutilización

Aprovechar respuestas
y cálculos ya hechos.



Elegir el tamaño justo

Usar el modelo
adecuado para
cada necesidad.





La investigación y la tecnología también pueden **reducir el impacto de otras actividades.**

Adoptar IA en bibliotecas

Beneficios y desafíos, siempre con revisión humana.



Ahorro de tiempo

Mejor acceso

Más personalización

Errores y seguridad

revisión humana · fuentes citadas

Costes y dependencias

piloto pequeño · modelos abiertos · medir €/consulta

Brecha digital y sesgos

formación · acceso asistido · revisar sesgos

La IA amplía el servicio cuando se **gobierna bien**

El efecto rebote

La eficiencia es necesaria, pero no suficiente.



Cada consulta gasta

1/2

La nueva generación de modelos consume aprox. la mitad por consulta.

→

pero hacemos

×10

Ahora preguntamos mucho más: probamos, iteramos y exploramos.

→

consumo total

×5

El resultado neto puede ser un consumo muy superior, a pesar de la mayor eficiencia por consulta.

Tres respuestas no siempre son necesarias.

Medir el total es tan importante como mejorar la eficiencia.

Efecto rebote · paradoja de Jevons

Primeros 90 días

Una hoja de ruta pequeña, medible y útil.

Días 1–30

Elegir una tarea concreta y un sistema seguro de datos

Días 31–60

Pilotar con pocas consultas y revisión humana

Días 61–90

Medir calidad, € y kWh por 1.000 consultas

Empezar pequeño permite aprender rápido.



Medir antes de **escalar.**

Medir. Elegir. Optimizar. Compartir.

La IA puede ampliar el valor público de las bibliotecas si se usa con criterio.



Enrique Alba
Catedrático · Universidad de Málaga
eat@lcc.uma.es

Bibliotecas andaluzas: **tradición, servicio e innovación responsable.**