

No existe la mejor IA

La pregunta "¿cuál es la mejor IA?" está mal formulada, y ese es el aprendizaje. La correcta es: **¿cuál resuelve mejor esta tarea, hoy?** Cambia por tarea, cambia por semana y cambia según qué dato tuyo ve cada herramienta.

QUÉ VAS A PODER HACER

- 1 Explicar la diferencia entre herramienta, modelo y modo sin dudar.
- 2 Decidir entre modo rápido y modo razonamiento según la tarea.
- 3 Correr una comparación honesta, con criterios definidos antes de leer nada.

LAS TRES CAPAS QUE TODO EL MUNDO CONFUNDE

CAPA	QUÉ ES	EJEMPLO	QUÉ DECIDE
Herramienta	La app que usas	ChatGPT, Claude, Gemini, Copilot	Qué datos tuyos ve y qué puede hacer
Modelo	El motor que corre adentro	GPT, Opus, Sonnet, 2.5 Pro, Flash	Capacidad bruta y costo
Modo	Cuánto se le deja pensar	rápido / thinking / pro	Tiempo, costo y profundidad

El caso que aclara todo: **Copilot no tiene modelo propio**. Corre GPT con una capa que lo conecta a tus correos, Teams y OneDrive. Su ventaja no es el motor, es el acceso a tu contexto. Esa distinción, motor contra acceso, es la que hay que dejar clavada.

QUIÉN ES QUIÉN

HERRAMIENTA	EMPRESA	MODELOS QUE CORRE	SU VENTAJA REAL
ChatGPT	OpenAI	Familia GPT, con modos rápido, thinking y pro	Ecosistema, análisis de archivos, imágenes
Claude	Anthropic	Opus (profundo) y Sonnet (ágil)	Redacción larga, matices, código
Gemini	Google	Pro (razonamiento) y Flash (rápido)	Ventana de contexto enorme, Workspace
Copilot	Microsoft	No tiene modelo propio: GPT + capa Prometheus	Ve tus correos, Teams y OneDrive

RÁPIDO O RAZONAMIENTO

RÁPIDO

redactar · reformatear · traducir · resumir · reescribir
si la respuesta es "buscar y ordenar", usa rápido

RAZONAMIENTO

cálculos · código · debugging · análisis de varios pasos
· decisiones con trade-offs
si la respuesta empieza con "depende de...", usa razonamiento

Trade-off real: más profundidad = más tiempo y más costo.
En tareas simples, el modo pro puede darte una respuesta PEOR por exceso de elaboración.

ANTES Y DESPUÉS: COMPARAR DE VERDAD

× ANTES

¿Cuál es mejor para mi trabajo, ChatGPT o Claude?

Le estás pidiendo que se autoevalúe, sin decirle qué trabajo haces ni qué significa "mejor" para ti. La respuesta va a ser una lista de características que ya conocías.

✓ DESPUÉS

Tarea real: resumir el reporte semanal de incidentes para el cliente.

Formato obligatorio: exactamente 5 ítems, sin emojis, cada uno con una acción concreta y su responsable. Máximo 15 palabras por ítem.

[reporte pegado abajo]

El mismo prompt corre en las 4 herramientas, el mismo día. Como el formato está fijado, lo único que puedes comparar es el contenido, que es lo que importa.

EL PROTOCOLO DE COMPARACIÓN

1. Elige una tarea REAL tuya, de las que haces seguido.

2. Escribe UN prompt con R-E-L-C, con el formato fijado.

3. Anota 3 criterios y qué esperas ver ANTES de leer ninguna respuesta.

4. Córrelo en las 4 herramientas EL MISMO DÍA, en pestañas lado a lado.

5. Puntúa y elige la que resolvió ESA tarea. Repite con otra tarea distinta.

Criterios que sirven:
claridad · accionabilidad (¿lo ejecuto mañana?) · uso del contexto que di · profundidad vs brevedad según lo pedido · consistencia · latencia y costo

Las dos trampas. Una: el formato engaña. Seis bullets con emojis parecen mejores que cuatro párrafos sobrios y eso no dice nada del contenido; por eso fijas el formato en el prompt. Dos: probar hoy una y en dos semanas otra no compara nada, porque los modelos se actualizan sin avisar.

REPASO

1. ¿Qué modelo usa Copilot y cuál es su ventaja real?

No tiene modelo propio: usa GPT con la capa Prometheus. Su ventaja no es el motor sino el acceso a tus correos, Teams y OneDrive.

2. ¿Por qué hay que definir los criterios antes de leer las respuestas?

Porque si los defines después, los vas a acomodar a la respuesta que más te gustó estéticamente. Es sesgo de confirmación puro.

3. ¿Por qué hay que probar el mismo día?

Porque los modelos se actualizan constantemente. Una diferencia de una semana puede explicar toda la diferencia de resultado.

4. ¿Cuándo NO conviene un modelo de razonamiento?

En tareas de redacción o formato simples: pagas más, esperas más y a veces el resultado es peor por sobreelaboración.

5. ¿Por qué dos personas eligen ganadores distintos con el mismo resultado?

Porque los criterios son personales y dependen de la tarea. Eso no invalida el método: es exactamente el punto, no hay una mejor IA universal.

LOS 3 ERRORES MÁS COMUNES

❗ Casarte con una herramienta por costumbre y no volver a probar nunca.

❗ Comparar sin fijar el formato: terminas puntuando el diseño de la respuesta, no su calidad.

❗ Usar el modo más caro "por si acaso" para todo.