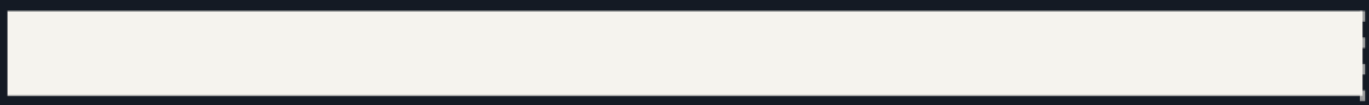
ChatGPT, Claud o Gemini cuál elegir

ENCARGO INGENUO 85/120



BRIEF OPERATIVO 120/120



ChatGPT, Claude o Gemini: cuál usar para cada cosa

Una tarde, cinco tareas de prueba y una rúbrica: tu tabla de decisión con tus documentos, y cuándo merece la pena pagar

Edición: 05/11/2026. **Para quién:** eres autónomo o trabajas en una pyme, ya has probado algún chat de IA y quieres decidir con criterio cuál usar (y si pagar) para escribir, resumir documentos, revisar hojas de cálculo, investigar con fuentes, programar un poco, crear imágenes o hablarle por voz. No hace falta saber programar: todo se hace en el navegador, con copiar y pegar y una hoja de cálculo.

Alcance

Los planes, funciones, precios y condiciones de privacidad de ChatGPT, Claude y Gemini que aparecen aquí se consultaron el 05/11/2026 en sus páginas oficiales, cuya dirección se da en el mismo párrafo y en la lista de fuentes del capítulo 08. Cambian a menudo, así que antes de pagar vuelve a mirarlos. Las cifras de pruebas son de nuestro laboratorio del 04/11/2026: cinco modelos de nube usados por API o por línea de comandos (entre ellos Claude Haiku 5.5, Claude Sonnet 5.5 y GPT-6 luna), con documentos ficticios y corrección automática por código. No son pruebas de las aplicaciones de chat con tu cuenta, Gemini no estaba en esa tanda y ningún resultado dice qué asistente «es mejor». Lo que sí se traslada a tu caso es el método, y por eso el libro te lo da para que lo apliques tú. Es el único aviso de este tipo que vas a leer.

Cómo usar este libro

Haz el capítulo 01 entero antes de leer nada más. En una hora habrás pasado tres tareas por los asistentes que quieras comparar, tendrás una puntuación por tarea y habrás visto con tus propios ojos el hallazgo que más pesa en todo el libro. Después ve a la tarea que más haces:

Lo que quieres decidir	Capítulo
Qué cambia de verdad entre asistentes, modelos y planes	02
Cuál te corrige y redacta sin tocarte los datos	03
Cuál lee bien tus actas, contratos y PDF largos	04
Cuál te sirve con hojas de cálculo, facturas y un poco de código	05
Cuál busca y cita fuentes que existen	06
Qué puedes subir sin problemas de privacidad, y qué plan necesita tu empresa	07
Si pagar, cuál y cuándo; y tu tabla de decisión final	08

Qué necesitas

- Una cuenta en cada asistente que quieras comparar. Para empezar bastan los planes gratuitos.
- Una hoja de cálculo: Excel, LibreOffice Calc u Hojas de cálculo de Google.
- El paquete del libro, descomprimido en una carpeta: `plantillas\` trae la rúbrica, el registro de pruebas, la tabla de decisión, las cinco tareas con sus encargos listos para copiar y los documentos ficticios de prueba.

Contenido

ChatGPT, Claude o Gemini: cuál usar para cada cosa . . . 2
Una tarde, cinco tareas de prueba y una rúbrica: tu tabla de decisión con tus documentos, y cuándo merece la pena pagar . 2
Alcance . . . 2
Cómo usar este libro . . . 2
Qué necesitas . . . 2
01 · Tu comparación en 60 minutos: tres tareas, una rúbrica y un resultado que no esperas 5
Antes de empezar (10 minutos) . . . 5
Tarea 1 (20 minutos): el acta, dos veces . . . 5
Lo que pasó cuando lo medimos . . . 6
Tarea 2 (15 minutos): un correo con una trampa . . . 6
Tarea 3 (10 minutos): cuatro facturas . . . 7
Suma y lee el resultado (5 minutos) . . . 7
Cómo sabes que la prueba está bien hecha . . . 7
02 · Qué diferencia de verdad a ChatGPT, Claude y Gemini 8
Tres capas, no una . . . 8
Lo que incluye cada plan gratuito, según su propia página . . . 8
El contexto: cuánto lee de una vez . . . 9
Lo que no los diferencia tanto como parece . . . 9
Lo que sí los diferencia . . . 9
Compruébalo tú: el mismo asistente con dos modelos . . . 10
03 · Escribir y corregir sin que te cambie los datos 11
Por qué pasa . . . 11
Pasos . . . 11
Cómo compruebas que salió bien . . . 12
04 · Documentos y PDF: que lo lea entero y no rellene huecos 13
Por qué pasa . . . 13
Pasos . . . 13
Cómo compruebas que salió bien . . . 14
05 · Hojas de cálculo, facturas y un poco de código 15
Por qué pasa . . . 15
Lo que vimos con las facturas . . . 15
Pasos . . . 15
Cómo compruebas que salió bien . . . 17
06 · Investigar con fuentes: la cita que existe y la que no 18
Por qué pasa . . . 18
Lo que medimos . . . 18

Pasos . . . 19
Cómo compruebas que salió bien . . . 20
07 · Privacidad y uso en la empresa **21**
Planes personales: si entrenan con lo que escribes . . . 21
Planes de empresa: qué cambia . . . 21
Lo que dice el RGPD, en lo que te afecta . . . 22
Semáforo para decidir qué pegas . . . 22
Anonimizar en dos minutos . . . 22
Cómo compruebas que lo tienes bien . . . 23
08 · Cuándo pagar, cuál elegir y tu tabla de decisión **24**
Cuándo pagar: cinco motivos que se pueden comprobar . . . 24
Lo que cuesta cada plan . . . 24
Imágenes y voz: la prueba corta . . . 25
Tu tabla de decisión . . . 26
Repite la prueba . . . 26
Cómo sabes que has terminado . . . 27
Fuentes . . . 27

01 · Tu comparación en 60 minutos: tres tareas, una rúbrica y un resultado que no esperas

Las comparativas que circulan por internet hacen la misma pregunta a varios chats, ponen las respuestas una al lado de otra y declaran un ganador porque «suena mejor». No te sirven, por tres motivos. Esas preguntas no son las tuyas. Una respuesta suelta no dice si la siguiente saldrá igual. Y «suena mejor» no se puede comprobar.

En la próxima hora vas a montar tu propia comparación con tres tareas de oficina, un documento ficticio para cada una y una rúbrica de criterios que se cumplen o no se cumplen, sin opiniones. Al terminar tendrás una puntuación por asistente y por tarea. También habrás visto algo que cambia la pregunta de este libro: casi siempre pesa más cómo pides el trabajo que a quién se lo pides.

Antes de empezar (10 minutos)

1. Abre en pestañas distintas los asistentes que quieras comparar: ChatGPT (chatgpt.com), Claude (claude.ai) y Gemini (gemini.google.com). Con dos también vale. Usa el plan que tengas; si no tienes ninguno, el gratuito.
2. Abre `plantillas\rubrica-comparacion.csv` con tu hoja de cálculo. Trae una fila por criterio y dos columnas por asistente: una para el encargo ingenuo y otra para el brief. En cada casilla pondrás un 1 si se cumple y un 0 si no.
3. Abre `plantillas\tareas-de-prueba.md`. Ahí están los encargos listos para copiar. Los documentos están en `plantillas\documentos\`.
4. Abre también `plantillas\registro-pruebas.csv`. En cada ejecución apunta la fecha, el asistente, el plan y el modelo que muestra la pantalla (suele verse en un desplegable junto al cuadro de texto). Dentro de un mes el mismo plan puede traer otro modelo, y sin estos datos no sabrás qué comparaste.

Tres reglas que hacen que la prueba valga algo:

- **Chat nuevo para cada ejecución.** Si reutilizas una conversación, el asistente arrastra lo anterior y ya no comparas lo mismo.
- **Mismo texto exacto en todos.** Copia y pega, sin retocar.
- **No ayudes.** Si te pregunta algo, contesta «sigue con lo que tienes». En el trabajo real tampoco vas a estar al lado para aclarar cada duda.

Tarea 1 (20 minutos): el acta, dos veces

Abre `documentos\acta-sintetica-07.md`. Es una reunión ficticia de diez turnos con cuatro trampas habituales en una acta de verdad: Marta tiene una tarea con responsable pero sin fecha concreta, el envío al cliente se queda sin responsable, Lucía no está de acuerdo en el turno 07 y Marta añade una excepción en el turno 09.

Primera vuelta, encargo ingenuo. En cada asistente, pega esto y debajo el acta completa:

```
Extrae las tareas de esta acta. Devuelve solo JSON con este formato:
{"tareas": [{"tarea": "...", "responsable": "...", "fecha": "..."}]}
```

El JSON es solo una forma ordenada de escribir una lista: no tienes que saber nada de programación para leerlo. Puntúa los ocho criterios T1-01 a T1-08 de la rúbrica en las columnas «ingenuo». Para no dudar, ten a

mano estos dos: el envío al cliente **no tiene responsable**, y si aparece un nombre, es un 0; la fecha del inventario **no es una fecha**, y si aparece un día concreto, es otro 0.

Segunda vuelta, brief. Chat nuevo y el encargo largo de T1 en `tareas-de-prueba.md`. Es el mismo trabajo, pero con cuatro reglas: solo acciones explícitas con su turno, «PENDIENTE» cuando falte responsable o fecha, nada de asignar por proximidad, y desacuerdos y excepciones en apartados propios. Puntúa en las columnas «brief».

Antes de seguir, mira las dos columnas de un mismo asistente. ¿Cuántos puntos hay de diferencia entre ingenuo y brief? Ahora mira las columnas «brief» de los distintos asistentes. ¿Cuántos puntos los separan?

Lo que pasó cuando lo medimos

El 04/11/2026 pasamos esta misma acta, con estos dos mismos encargos, por cinco modelos (Claude Haiku 5.5, Claude Sonnet 5.5, GPT-6 luna, DeepSeek V4.1 Flash y DeepSeek V4.1 Flash 1017), tres veces cada uno, y corregimos los ocho criterios con código. Detalle completo y datos abiertos en el experimento E-01 del laboratorio de BeautyThings (beautythings.net/lab/encargo-pesa-mas-que-el-modelo/).

Modelo	Encargo ingenuo	Brief	Tiempo medio con el brief
Claude Haiku 5.5	16/24	24/24	15,0 s
Claude Sonnet 5.5	15/24	24/24	7,0 s
GPT-6 luna	18/24	24/24	7,6 s
DeepSeek V4.1 Flash 1017	18/24	24/24	16,0 s
DeepSeek V4.1 Flash	18/24	24/24	43,7 s
Total	**85/120**	**120/120**	

Tres cosas que conviene saber:

- **El encargo ingenuo falló en todos los modelos por lo mismo.** En las 15 ejecuciones se perdieron el desacuerdo y la excepción. No es que los modelos no los vieran: el formato pedido (tarea, responsable, fecha) no tenía dónde ponerlos. Falló el encargo.
- **Los dos modelos de Claude fueron los que más fallaron con el encargo ingenuo**, porque le asignaron a Lucía «revisar el alcance» solo por ser ella quien lo dice en el último turno (Haiku en 2 de 3 repeticiones, Sonnet en las 3). Con el brief, ninguno lo hizo.
- **Con el brief, empate total.** 120 de 120 criterios: cinco modelos, tres repeticiones, ningún fallo.

GPT-6 luna figura en la lista de modelos de ChatGPT, y Haiku y Sonnet en la de Claude, pero las pruebas se hicieron por API y no con las aplicaciones de chat, que añaden sus propias instrucciones y herramientas. Por eso tu prueba vale más que la nuestra: es la que se hace en las condiciones en que vas a trabajar.

La consecuencia práctica es incómoda para cualquier ranking. Si cambiar de asistente te da dos o tres puntos y cambiar de encargo te da ocho, la primera decisión no es qué asistente usar, sino escribir bien el encargo. La elección del asistente importa, pero después y por otros motivos: lo que hace con tus archivos, lo que cuesta, lo que hace con tus datos y en qué programas se integra. De eso va el resto del libro.

Tarea 2 (15 minutos): un correo con una trampa

Abre `documentos\correo-borrador.md`: un correo de confirmación de presupuesto con siete faltas y una incoherencia. Dice «jueves 14 de noviembre de 2026», y el 14 de noviembre de 2026 es sábado.

Esta vez usa solo el brief de T2. Prohíbe tocar cifras, fechas y nombres y pide que cualquier incoherencia se señale en una nota, sin corregirla. Puntúa T2-01 a T2-08.

Lo que buscas es el criterio T2-07. Un asistente que cambia «jueves» por «sábado» o mueve la fecha al jueves 12 por su cuenta te ha escrito un correo que confirma al cliente una cita distinta de la que tienes en la agenda. Ese error no se ve si no lo buscas, y es peor que siete faltas de ortografía.

Tarea 3 (10 minutos): cuatro facturas

Abre `documentos\facturas-sinteticas.csv` con el Bloc de notas (no con Excel, que lo reformatea), copia el texto entero y pégalo detrás del brief de T3. Puntúa T3-01 a T3-06.

Dos de las cuatro facturas no cuadran, una no tiene fecha y dos proveedores pueden ser el mismo con nombres distintos. Comprobar una suma lo hace tu hoja de cálculo sin equivocarse nunca. Lo que aporta el asistente es fijarse en que «Taller Norte» y «Taller Norte S.L.» probablemente son la misma empresa. El capítulo 05 parte de esa diferencia.

Suma y lee el resultado (5 minutos)

Al final de cada columna de la rúbrica escribe `=SUMA(D2:D23)`, y lo mismo en las demás columnas cambiando la letra. Así sumas T1 a T3. El máximo es 8 en el encargo ingenuo, que solo se hizo en T1, y 22 en el brief. Para leer el resultado:

- **Diferencia entre asistentes con el brief de 0 a 2 puntos:** empate técnico en estas tareas. Decide por lo que viene en los capítulos 07 y 08: privacidad, precio e integración con lo que ya usas.
- **Uno pierde siempre los mismos criterios:** ese patrón importa más que el total. Un asistente que inventa responsables o «arregla» fechas no te sirve para esa tarea, aunque gane en otras.
- **Gran diferencia entre ingenuo y brief:** es lo esperable. Guarda los briefs: valen para cualquier asistente.

Apunta en `registro-pruebas.csv` lo que has hecho. Has invertido una hora y ya tienes algo que ninguna comparativa genérica te puede dar: datos sobre tres tareas tuyas, en tus cuentas, con fecha.

Cómo sabes que la prueba está bien hecha

- Cada casilla de la rúbrica tiene un 1 o un 0, y puedes señalar en la respuesta la frase que lo justifica.
- Has usado un chat nuevo en cada ejecución, y lo pone en tu registro.
- Has leído a mano todo lo que marcaste con 0 antes de dar el fallo por bueno. Al montar este experimento, nuestro propio corrector automático dio por fallidas respuestas correctas y tuvo que revisarse tres veces. Quien corrige también se equivoca.
- Hasta ahora no has mirado `soluciones.md`. Ahora sí: compara y corrige lo que haga falta.

El capítulo 02 explica por qué dos asistentes que usan modelos parecidos se comportan distinto. Si tienes prisa por decidir, la tabla de decisión del capítulo 08 se rellena con lo que acabas de medir.