

Claude Code, Cowork o API

Qué herramienta corresponde a cada caso de uso, y cuál te va a costar dinero si la elegís mal.

La confusión es entendible: los tres se llaman Claude, los tres hacen cosas con inteligencia artificial y los tres los vende la misma empresa. Pero dos de ellos vienen dentro de la suscripción que ya pagás y el tercero se factura aparte, por volumen. Si armás un flujo de atención al cliente sobre la herramienta equivocada, o pagás de más, o se te cae cuando empieza a funcionar bien.

Este documento es lo que te prometí en el video: casos de uso concretos, qué herramienta va en cada uno y las señales de que estás por elegir mal. Está pensado para dueños de negocio y gente de operaciones, no para programadores, aunque si programás también te sirve.

Los tres, en una tabla

	Claude Code	Claude Cowork	Claude API
Qué es	Un agente que construye y modifica software. Le pasás requerimientos y trabaja sobre el código de principio a fin.	La versión de escritorio para tareas de oficina: ordenar archivos, generar reportes, cruzar datos, redactar.	La puerta para que tus sistemas usen Claude sin que haya una persona adelante.
Quién lo dispara	Vos o alguien de tu equipo, sentado frente a la computadora.	Vos o alguien de tu equipo, sentado frente a la computadora.	Un servidor, un webhook, un cliente que manda un mensaje a las 3 de la mañana.
Cómo se cobra	Incluido en la suscripción paga (Pro, Max, Team Premium). Tiene límites de uso por sesión.	Incluido en la misma suscripción. Comparte el mismo pool de uso.	Aparte. Por token consumido, sin cuota fija. Escala con tu volumen, no con tu cantidad de usuarios internos.
Escala con	Cuánto lo usás vos.	Cuánto lo usás vos.	Cuántos clientes tenés y cuánto hablan.

La regla que resuelve el 80% de los casos

Preguntate quién aprieta el botón. Si la tarea la arranca una persona de tu equipo, va por suscripción (Code o Cowork). Si la arranca un cliente, un sistema o un horario, va por API. El resto de este documento son variaciones de esa regla.

Casos de uso

Ordenados de menos técnico a más técnico. En cada uno, la herramienta correcta, el motivo, y el error típico.

01 / Ordenar y renombrar 400 facturas en PDF

Situación	Tenés una carpeta con facturas de proveedores de todo el año, con nombres tipo 'scan_0034.pdf', y necesitás que queden por proveedor y por mes.
Herramienta	Claude Cowork
Por qué	Es una tarea de oficina que corre una vez sobre archivos que ya tenés en tu máquina. Cowork lee los PDF, entiende qué es cada uno y los mueve.
Ojo con	Si esto pasa todos los meses con cientos de facturas y querés que ocurra solo, sin que nadie lo dispare, ya estás hablando de un flujo automatizado y eventualmente de API.

02 / El reporte semanal que arma tu asistente los lunes

Situación	Alguien de tu equipo abre tres planillas, copia números, arma un resumen y lo manda por mail. Le lleva dos horas.
Herramienta	Claude Cowork
Por qué	Persona adelante, archivos locales, resultado que alguien revisa antes de mandar. Es exactamente para lo que existe Cowork.
Ojo con	No intentes esto con la API 'para que sea automático' antes de haberlo hecho a mano con Cowork tres o cuatro semanas. Primero validás que el reporte sirve, después automatizás.

03 / Cruzar la exportación del CRM con la planilla de ventas

Situación	Dos fuentes que no hablan entre sí. Querés saber qué clientes compraron pero nunca fueron cargados, o al revés.
Herramienta	Claude Cowork
Por qué	Cruce de datos puntual, con criterio humano en el medio para decidir qué hacer con las diferencias.
Ojo con	Si el cruce tiene que pasar en tiempo real cada vez que entra una venta, eso es una integración entre sistemas. Ahí entra Claude Code para construirla, y quizás API si la lógica necesita IA en cada corrida.

04 / Redactar propuestas comerciales a partir de una plantilla

Situación	Tenés una plantilla base y cada propuesta cambia según el cliente y lo que se habló en la reunión.
Herramienta	Claude Cowork (o el chat de Claude con un Proyecto)
Por qué	Redacción asistida con revisión humana. Un Proyecto con tu plantilla y tus casos anteriores hace que cada propuesta salga con tu voz.
Ojo con	Cuando el equipo comercial pide 'que las propuestas se generen solas desde el CRM', eso es API. Pero preguntá primero cuántas propuestas por semana son. Con menos de 20, la suscripción alcanza y sobra.

05 / Analizar 5.000 reseñas de clientes una sola vez

Situación	Querés saber de qué se quejan, qué elogian y si hay un patrón por sucursal.
Herramienta	Claude Cowork
Por qué	Volumen grande pero corrida única. Cowork procesa el archivo, agrupa y te devuelve el análisis. No hace falta infraestructura.
Ojo con	'Una sola vez' tiene la costumbre de volverse 'todos los meses'. Si a la tercera vez seguís haciéndolo a mano, ya amortiza construir un flujo con Claude Code que lo corra solo.

06 / Armar un panel interno o un MVP para probar una idea

Situación	Necesitás una aplicación simple: un formulario, una tabla, un dashboard que lea de tu base de datos.
Herramienta	Claude Code
Por qué	Es construcción de software. Claude Code trabaja sobre el código, crea los archivos, corre las pruebas. Si tenés alguien técnico en el equipo, lo termina en días.
Ojo con	Claude Code construye la app; no la hace funcionar con IA adentro. Si tu app necesita que Claude responda cosas cuando la usan tus clientes, esa parte se paga por API, aparte de la suscripción.

07 / Actualizar un sistema viejo que nadie quiere tocar

Situación	Tenés un sistema de hace ocho años, el programador que lo hizo ya no está, y hay que agregarle una función.
Herramienta	Claude Code
Por qué	Lee el código existente, entiende cómo está armado y hace los cambios sin reescribir todo. Es el caso de uso donde más tiempo ahorra.
Ojo con	Necesitás igual una persona técnica que revise lo que hace. No es 'sin programadores'; es 'con muchos menos programadores'.

08 / Un script que sincroniza dos sistemas todas las noches

Situación	A las 2 AM, tomar los pedidos del e-commerce y cargarlos en el sistema de facturación.
Herramienta	Claude Code para construirlo. Sin API si la lógica es fija.
Por qué	La sincronización en sí no necesita inteligencia artificial: es mover datos de A a B con reglas claras. Claude Code lo escribe; después corre solo y no consume nada de Claude.
Ojo con	Si en el medio hay una decisión que requiere entender texto libre ('¿este pedido con la nota rara es urgente?'), esa decisión sí es una llamada a la API, y se paga por cada pedido que pase.

09 / Un asistente de WhatsApp que responde a tus clientes

Situación	Llega un mensaje, hay que entenderlo y responder en segundos, a cualquier hora, para 5, 500 o 3.000 conversaciones por día.
Herramienta	Claude API
Por qué	Nadie de tu equipo está apretando un botón. Tu servidor recibe el mensaje (por Twilio, por la API de WhatsApp Business), se lo pasa a Claude y devuelve la respuesta. Escala con tus clientes.
Ojo con	Esto no viene con la suscripción. Es una cuenta separada, con tarjeta separada, que factura por token. Y hay que estimar el volumen antes de prender: el costo es mensajes por día multiplicado por tokens por mensaje.

10 / Clasificar tickets de soporte y derivarlos solos

Situación	Cada ticket que entra tiene que ir a facturación, a técnica o a ventas, y hoy lo hace una persona a mano.
Herramienta	Claude API
Por qué	Corrida por evento: cada ticket dispara una clasificación. Sin persona adelante, volumen atado a tu operación.
Ojo con	Es uno de los casos más baratos de API porque cada llamada es corta. Pero si son 40 tickets por día, una persona con Cowork lo hace en 20 minutos y no hace falta construir nada.

11 / Resumir llamadas o transcripciones del IVR

Situación	Después de cada llamada, querés un resumen en el CRM con motivo, resolución y próximos pasos.
Herramienta	Claude API
Por qué	El disparador es el fin de la llamada, no una persona. Se integra con la telefonía (Twilio, tu central) y escribe en el CRM solo.
Ojo con	Las transcripciones son largas: una llamada de 10 minutos son miles de tokens de entrada. El volumen de llamadas manda el costo, así que medilo antes.

Situación	'¿Cómo se hace una nota de crédito?' preguntado a un bot en vez de al gerente.
Herramienta	Depende de quién lo usa
Por qué	Si son tus empleados, un Proyecto en Claude con los manuales cargados resuelve el 90% sin construir nada. Si tiene que vivir dentro de tu intranet o de Slack y usarlo gente sin cuenta de Claude, entonces es API.
Ojo con	El error común es construir el bot por API para 8 empleados. Ocho cuentas Team salen menos que el desarrollo, y no hay nada que mantener.

Los tres errores que cuestan dinero

1. Usar la suscripción personal como si fuera producción

Alguien conecta su cuenta Pro a un flujo de n8n o a un script que responde clientes. Funciona una semana. Después choca contra el límite de uso por sesión en el peor momento (un viernes a las 6 de la tarde, con clientes esperando) y además viola los términos de uso. La suscripción es para personas; los sistemas van por API.

2. Construir con API algo que una persona usaría tres veces por semana

El caso opuesto. Se arma una integración, con servidor, con mantenimiento, con un programador que la cuida, para una tarea que Cowork hace en 15 minutos con alguien del equipo adelante. Antes de construir, hacé la tarea a mano con Cowork durante un mes. Si el volumen justifica automatizar, lo vas a ver en los números.

3. Prender la API sin haber estimado el volumen

La API no tiene cuota fija, y eso es bueno y malo. Bueno porque pagás lo que usás. Malo porque si no sabés cuánto vas a usar, no sabés cuánto vas a pagar. Con un agente de WhatsApp, un cliente que manda 20 mensajes cortos cuesta más que uno que manda 3 largos, porque cada mensaje arrastra todo el historial de la conversación.

Cómo estimar el costo de la API en cinco minutos

No hace falta ser técnico. Necesitás tres números:

Mensajes por día	Cuántas interacciones entran hoy, aunque las atienda una persona. Mirá una semana real, no la que te gustaría tener.
Tokens por mensaje	Un token es más o menos 3 o 4 caracteres en español. Un mensaje de WhatsApp promedio con su respuesta anda entre 300 y 800 tokens, pero el contexto (instrucciones del bot, historial) se suma en cada llamada y suele pesar más que el mensaje en sí.
Precio por millón de tokens	Está publicado en la página de precios de Anthropic y cambia según el modelo. Los modelos chicos salen una fracción de los grandes y para clasificar o responder preguntas frecuentes suelen alcanzar.

La cuenta

Costo mensual aproximado = mensajes por día × tokens por mensaje × 30 × precio por millón ÷ 1.000.000. Hacé la cuenta con el modelo más barato y con el más caro. Si los dos números te cierran, avanzá. Si solo cierra el barato, el diseño del bot tiene que respetar eso desde el día uno.

Checklist antes de decidir

- ¿Quién dispara la tarea: una persona de mi equipo o un evento externo (cliente, horario, sistema)?
- ¿Cuántas veces por semana pasa hoy? ¿Y cuántas dentro de seis meses si el negocio crece como quiero?
- ¿Si se cae un viernes a la tarde, quién lo nota primero: yo o un cliente?
- ¿Ya lo hice a mano con Cowork o con el chat al menos un mes? ¿Sé que el resultado sirve?
- ¿Tengo el número de volumen y la cuenta de costo hecha, o estoy prendiendo a ciegas?

Si las respuestas apuntan a 'persona, pocas veces, lo noto yo': suscripción. Si apuntan a 'evento externo, muchas veces, lo nota un cliente': API, con la cuenta hecha antes.

Sobre mí

Soy Gonzalo Gomez, desarrollador full-stack senior en Buenos Aires. Hace varios años que armo sistemas de comunicación con inteligencia artificial: asistentes de WhatsApp, pipelines de voz, integraciones con Twilio. Este documento sale de las mismas preguntas que me hacen los clientes antes de empezar un proyecto, y de un par de proyectos que arrancaron con la herramienta equivocada.

Más contenido técnico y de negocio en **ggomez.dev** y en el canal de YouTube. Si tenés un caso que no entra en ninguno de los doce de acá arriba, escribime: seguro me sirve para el próximo video.

-Gonza