

Deja que Claude trabaje solo. Sin firmar cada paso.

CATEGORÍA · PRIMEROS PASOS CON CLAUDE

El modo automático quita las preguntas de permiso una por una y pone en su lugar un filtro que revisa lo que Claude va a hacer antes de que lo haga. Esta guía te lo activa en cinco minutos y te deja un candado puesto para lo que sí quieres aprobar tú.

QUÉ VAS A ENCONTRAR

- 01 Por qué existe y qué problema resuelve
- 02 Como funciona el filtro por dentro
- 03 Los tres lugares donde se activa
- 04 Qué revisar si no te aparece la opción
- 05 El candado humano para dormir tranquilo
- 06 Qué hacer cuando te bloquea de más

FICHA DEL RECURSO

Qué es, para qué sirve y cómo se implementa

FUNCIÓN

El modo automático quita las preguntas de permiso una por una y pone en su lugar un filtro que revisa lo que Claude va a hacer antes de que lo haga. Esta guía te lo activa en cinco minutos y te deja un candado puesto para lo que sí quieres aprobar tú.

PARA QUÉ SIRVE

Cuando le pides a Claude una tarea de veinte pasos, te pide permiso en cada uno. Nadie revisa los veinte. Le das aceptar, aceptar, aceptar, hasta que dejas de leer lo que estás aprobando.

DÓNDE APLICARLO

Cubre: Por qué existe y qué problema resuelve; Como funciona el filtro por dentro; Los tres lugares donde se activa; Qué revisar si no te aparece la opción; El candado humano para dormir tranquilo.

CÓMO IMPLEMENTARLO

- 1 Empecemos Por Aquí: El 97% se aprueba sin leer.
- 2 Como Funciona Por Dentro: Tres filtros, en este orden.
- 3 Activación: Tres lugares.
- 4 Si No Te Sale La Opción: No es una caída.
- 5 Arranque Seguro: Ponle un candado a lo que te importa.

EMPECEMOS POR AQUÍ

El 97% se aprueba sin leer.

Cuando le pides a Claude una tarea de veinte pasos, te pide permiso en cada uno. Nadie revisa los veinte. Le das aceptar, aceptar, aceptar, hasta que dejas de leer lo que estás aprobando.

- Según Anthropic, el 97% de esas preguntas de permiso se aprueban. A eso le llaman fatiga de aprobación, y es donde se cuela el error, porque el permiso número dieciocho recibe la misma atención que los diecisiete anteriores, o sea ninguna.
- El modo automático quita esas preguntas. Pero no lo hace dejando que Claude se apruebe a sí mismo, que sería como escribir un trabajo, calificarlo tú y ponerte tu propia nota.
- Lo que hace es meter un segundo modelo, aparte, que revisa cada acción antes de que se ejecute. Compara lo que tú pediste contra lo que Claude está a punto de hacer, y frena lo que se pasa de ahí.
- Y ese segundo modelo no ve el razonamiento de Claude, a propósito. Solo ve lo que tú escribiste y la acción concreta que está por ejecutarse. Así no lo convence ningún argumento, solo los hechos.
- El resultado práctico: Claude corre tareas largas de corrido, y tú solo apareces cuando de verdad hace falta. En sus pruebas de ataques para secuestrar a Claude por medio de páginas web, con esto

activado la tasa de éxito fue cero. Lo que no es. Esto no es el modo que se salta todas las revisiones. Ese existe aparte y no tiene ninguna protección. El automático sí revisa, solo que lo revisa una máquina en vez de ti.

COMO FUNCIONA POR DENTRO

Tres filtros, en este orden.

Cada cosa que Claude quiere hacer baja por estos tres pasos. El primero que aplique decide, y ahí se acaba la revisión. Tus reglas mandan sobre el filtro. Esa jerarquía no cambia nunca.

ACTIVACIÓN

Tres lugares.

Elige el tuyo. Depende de dónde uses Claude Code. La opción se llama Auto y aparece junto a los otros modos. 1 En la app de escritorio. Busca el selector de modo que está al lado del botón de enviar, ábrelo y elige Auto. Ojo, aquí no funciona ningún atajo de teclado para esto, es con el selector. Y lo que elijas se queda guardado para esa carpeta. 2 En la terminal. Presiona Shift y Tab para ir rotando entre los modos hasta llegar a Auto. Abajo vas a ver el aviso de cuál está activo. 3 Para que arranque siempre así, sin elegirlo cada vez. Abre el archivo de configuración de tu usuario, que está en la carpeta .claude de tu computadora, y pégale esto.

EN ~/.CLAUDE/SETTINGS.JSON

```
{
  "permissions": {
    "defaultMode": "auto"
  }
}
```

El error más común con esto.

Si pones esa configuración en la carpeta del proyecto en vez de la de tu usuario, Claude la ignora a propósito.

Es para que ningún proyecto que te descargues se conceda permisos solo . Tiene que ir en la de tu usuario.

SI NO TE SALE LA OPCIÓN

No es una caída.

Te falta algo. Si Auto no aparece en tu selector, no es que el servicio esté fallando. Es que tu cuenta no cumple uno de los requisitos. Revísalos en este orden, porque el primero es el que le pasa a casi todos.

- El modelo, y esta es la causa número uno. Tiene que ser uno reciente. Sonnet 5, Opus 4.7 o posterior, o Fable 5. Los anteriores como Sonnet 4.5, Opus 4.5 y cualquier Haiku no lo soportan. Cambia el modelo en tu selector y vuelve a mirar.
- El plan. Aquí no hay problema, funciona en todos los planes de Claude, incluso los básicos. Si alguien te dice que necesitas el plan más caro, está mal informado.
- La organización. Esto solo aplica si usas Claude a través de la cuenta de una empresa. En los planes de equipo viene activo por defecto, pero el administrador lo puede apagar para todos. Si estás en ese caso, pregúntale a quien maneja las cuentas.
- La versión de la app. Si llevas meses sin actualizar, actualiza antes de seguir buscando. Es una función nueva y las versiones viejas no la tienen.
- Y un detalle que ahorra tiempo. Si el modo automático te aparece pero después te sale un aviso de que no puede determinar si una acción es segura, ese es otro problema distinto. Ahí la opción sí está disponible y lo que falló fue la consulta al filtro. Casi siempre se arregla solo, así que vuelve a intentar. Antes de cambiar nada, revisa el modelo. Es el requisito que más gente incumple sin darse cuenta, porque el modelo se queda pegado en el que usabas antes. Míralo primero, es un click, y te ahorra andar tocando archivos de configuración sin necesidad.

ARRANQUE SEGURO

Ponle un candado a lo que te importa.

El filtro ya bloquea solo lo grave: borrar cosas que existían antes, mandar datos tuyos afuera, publicar a producción, bajar código de internet y ejecutarlo. Pero hay acciones que igual quieres aprobar tú, y para eso existen las reglas de aviso.

- Una regla de aviso se revisa antes que el filtro y siempre te pregunta, aunque el modo automático esté encendido. Es tu palabra final. Va en el mismo archivo de configuración de la página anterior.
- El ejemplo de abajo hace que te pregunte antes de subir cambios a un repositorio. Cambia lo que va dentro del paréntesis por lo que tú quieras frenar.
- También puedes decirlo hablando. Si le escribes que no haga algo hasta que tú revises, el filtro lo respeta. La diferencia es que eso se puede perder si la conversación se hace muy larga, y una regla escrita no se pierde nunca.
- El consejo del propio equipo que lo construyó: empieza cerrado, mira qué te bloquea en la práctica, y ve abriendo de a poco. Al revés, no.

EL CANDADO, EN EL MISMO ARCHIVO

```
{  
  "permissions": {  
    "ask": [  
      "Bash(git push *)"  
    ]  
  }  
}
```

Si te bloquea de más.

Cuando el filtro frena tres acciones seguidas, el modo automático se pausa solo y vuelve a preguntarte. No te deja atascado. Y todo lo que bloqueó queda en la lista de rechazos para que lo revises y lo vuelvas a lanzar.

SIGUIENTE PASO

Actívalo hoy. Con el candado puesto.

El orden importa. Primero pon la regla de aviso para lo que no quieres que pase sin ti, y después enciende el modo automático. Así la primera tarea larga que corras ya arranca protegida, y no tienes que estar de niñera mirando la pantalla. Empieza con una tarea que ya hayas hecho antes. Vas a ver la diferencia de una.

MÁS GUÍAS GRATIS EN [377CREATIVE.COM/RECURSOS](https://377creative.com/recursos)