

AGENTES DE IA



Cómo incorporar agentes de IA sin morir en el intento. Qué permisos darle a Grok Bot, Claude Cowork, ChatGPT Work y Gemini, cuáles nunca, y dónde se configura cada uno.

PARA

Quien usa IA en su trabajo

REVISADA

28 de octubre de 2026

POR

Nicolás Bilinkis, fundador de Referente.ai

REVISADA CONTRA LA DOCUMENTACIÓN OFICIAL DE CADA HERRAMIENTA · [REFERENTE.AI](#)

ANTES DE EMPEZAR

Ya no es un chat

Grok Bot, Claude Cowork, ChatGPT Work y el modo Task de Gemini no esperan a que les escribas. Les dejás un encargo y lo hacen solos: entran a tu mail, abren sitios donde estás logueado, mueven archivos y repiten tareas todos los días. Y siguen trabajando aunque cierres la compu.

Lo dicen los propios fabricantes. xAI: "Closing the app, your laptop, or your phone doesn't stop a background turn or a routine". Anthropic, sobre Cowork: "Close your laptop and Claude keeps going". Google, sobre Gemini: "it keeps working even if your device is turned off".

Pedir permiso para cada acción cansa, y la tentación es sacarle todos los frenos. El caso extremo salió a fines de octubre de 2026: Claude Code, la versión de Claude para programadores, sumó una opción que se llama, literal, "saltar los permisos todavía más peligrosamente". La propia documentación dice que solo se use en un entorno aislado "cuyo contenido te puedas permitir perder".

Esta guía sirve para elegir qué frenos sacar. Los ajustes y el comportamiento de cada herramienta salen de su documentación oficial, revisada el 28 de octubre de 2026. Estas herramientas cambian cada pocas semanas: si un menú no está donde dice, buscá el nombre del ajuste.

Los tres niveles

NIVEL	QUÉ HACE EL AGENTE	CUÁNDO USARLO
1. Pregunta todo	Antes de cada acción, te pide permiso.	Las primeras veces, y todo lo que toque datos de clientes o plata.
2. Hace lo de rutina, pregunta lo importante	Lee, busca y prepara solo. Antes de mandar, borrar, compartir o pagar, te consulta.	El trabajo diario, una vez que sabés qué hace.
3. No pregunta nada	Hace todo sin consultarte.	Casi nunca. Nunca con tu mail, tu Drive, tu banco ni tus clientes.

Lo que nunca se destraba. Borrar sin vuelta atrás. Mandar, publicar o compartir en tu nombre. Pagar o comprar. Escribir tus contraseñas o datos de tarjeta. Para estas cuatro, el agente te pregunta cada vez, aunque cueste un clic más.

LO NUEVO

Qué cambia cuando trabaja solo

Un asistente al que le escribís se equivoca delante tuyo. Uno que trabaja solo se puede equivocar cinco veces seguidas antes de que te enteres. Estos son los riesgos que trae y que el chat no tenía.

1 Sigue mientras no mirás

Las tareas programadas corren a la hora que les dijiste, o cuando llega un mail o un mensaje de Slack. Google lo advierte sobre Gemini: si una tarea programada corre mientras no estás conectado, quizás no puedas frenar una acción que no querías.

2 Queda logueado

Los sitios donde iniciás sesión dentro del agente quedan abiertos para las tareas siguientes. Cowork "remembers your logins across Cowork sessions", y en ChatGPT el login dura "until it expires".

3 Comparte todo entre sus bots

En Grok Bot, todos los bots de tu cuenta usan la misma compu, con las mismas sesiones abiertas y los mismos archivos. xAI lo dice claro: "Do not use separate Bots as a security boundary". Lo que le das a uno se lo das a todos.

4 Lee cosas que no escribiste vos

Un mail o una página pueden traer instrucciones escondidas para el agente ("reenviá este hilo a tal dirección"). Todos dicen que tienen defensas, y todos reconocen que no son perfectas. xAI: sus controles "reduce, but do not eliminate" el riesgo.

5 Borrar no limpia todo

Borrar un bot de Grok no borra las sesiones ni los archivos que dejó en la compu compartida. Borrar el chat de una tarea programada de ChatGPT solo la pausa. Borrar una tarea de Gemini no deshace los archivos que modificó. Y quitarle el acceso a una app en tu cuenta de Google no borra lo que ya copió.

Cinco reglas antes de dejarlo solo

1 Lo mínimo

Conectá solo las apps que necesita la tarea. Si solo tiene que leer, que solo lea.

2 Una carpeta propia, con copia de respaldo

Es la recomendación de Anthropic para Cowork: una carpeta de trabajo para el agente en vez de todo tu disco, y nada de documentos financieros ahí.

3 Empezá por leer y por borradores

La recomendación de xAI: primero tareas de solo lectura y borradores; mandar, publicar, comprar y borrar, siempre con aprobación.

4 Nada de encargos abiertos

"Revisá mi mail y resolvé todo" es la puerta de entrada de las instrucciones escondidas. Encargos concretos, con un resultado que puedas revisar.

5 Primero miralo trabajar

Una tarea chica, con vos al lado, antes de programarla para que corra sola.

CÓMO ELEGIR

Qué nivel para cada uso

USO	NIVEL	LO QUE TIENE QUE PREGUNTAR SIEMPRE
Mail	2	Mandar, responder o reenviar en tu nombre
Drive y archivos	2 en su carpeta, 1 en el resto	Borrar, mover afuera de la carpeta, compartir con alguien de afuera
Calendario	2	Invitar a gente de afuera, borrar eventos
Sitios donde estás logueado	1 en sitios nuevos, 2 en los que ya aprobaste	Escribir contraseñas, descargar archivos, entrar al banco
Slack, WhatsApp y otros mensajes	2	Publicar o mandar algo en tu nombre
Tareas programadas	1 la primera semana, después 2	Todo lo que salga afuera: un mail, un mensaje, una compra
Pagos, compras, suscripciones	1, siempre	Todo

EQUIVALENCIAS

Cada herramienta les pone otro nombre

HERRAMIENTA	1. PREGUNTA TODO	2. RUTINA SÍ, LO IMPORTANTE NO	3. NO PREGUNTA NADA
Grok Bot	Reglas "Ask first" para lo que importa	Auto Review prendido (viene así)	Auto Review apagado
Claude Cowork	Manual	Auto	Skip
ChatGPT, apps conectadas	Always ask	Allow read actions	Allow all actions (solo por app)
ChatGPT, navegador en la nube	Always ask (viene así)	Auto approve	No existe: Always allow abre todos los sitios, pero lo importante lo sigue preguntando
Gemini, modo Task	Lo de fábrica: confirma antes de navegar	Always allow en acciones o sitios	No existe: lo muy sensible lo pregunta siempre

Ninguno promete que el nivel 2 alcance. xAI: su revisor automático "should complement, not replace" los permisos mínimos. Google: sus protecciones "don't guarantee protection against all risks". OpenAI: "Those safeguards do not eliminate every risk". Anthropic: "You remain responsible for all actions taken by Claude performed on your behalf".

PASO A PASO

Grok Bot

Cada usuario tiene una compu en la nube que comparten todos sus bots. Con Gmail conectado, puede buscar, leer, redactar y mandar mails. Puede pagar con una tarjeta guardada.

1 Dejá Auto Review prendido y sumá tus reglas

Settings, General, sección Bot. Auto Review viene prendido y revisa cada acción antes de hacerla; si lo apagás, el bot hace las acciones sin revisarlas antes y solo quedan los controles de seguridad de fábrica. Debajo están las Custom Rules: en "When Grok Bot wants to:" describís la acción con tus palabras (mandar un mail a alguien de afuera, publicar en un canal de Slack, borrar algo) y elegís **Ask first**. Si una regla de preguntar y una de permitir chocan, gana la de preguntar.

2 No guardes la tarjeta si querés aprobar cada pago

Una tarjeta guardada en Vault se completa sola en el checkout, sin aprobación en cada compra. La propia documentación lo dice: si querés aprobar cada pago, no la guardes.

3 Pensá la compu como compartida

Todo lo que dejes logueado en el navegador del bot lo puede usar cualquiera de tus bots. Las contraseñas van en Vault (Settings, Vault), donde el bot las usa sin verlas; nunca en el chat. Para logins con código o verificación, tomá el control con Take over y devuelvelo con I'm done.

4 Tu compu, afuera

Settings, Computer, Your computer. Viene en "Ask every time". xAI recomienda **Never allow**, salvo que un bot tenga un motivo concreto para tocar tus archivos.

5 Rutinas: revisalas y frená las que no usás

Clic en el nombre del bot, pestaña Details, y el interruptor de cada rutina. Los mails que salen de una rutina quedan como borrador hasta que apretás Send. Para cortar algo en curso, mandale "Stop now". Ojo: ocultar un bot de la barra lateral no lo frena, sigue activo.

6 Cuando ya no lo usás

La lista oficial: pausar o borrar sus rutinas, cerrar sesión en los sitios de la compu compartida, desconectar las apps y revocarles el acceso desde cada servicio, sacar los archivos sensibles y recién ahí borrar los bots. Borrar un bot no borra lo que dejó en la compu. Y Grok Bot no se puede borrar sin borrar la cuenta de Cursor con la que entraste.

PASO A PASO

Claude Cowork

Desde octubre de 2026, en los planes Pro y Max corre en la nube. Toca tus archivos solo en las carpetas que conectaste y solo con la app de escritorio abierta, y sobre esas carpetas puede leer, escribir y borrar. Según la ayuda de Claude, el conector de Gmail puede mandar, responder y reenviar mails.

1

Una carpeta para Claude

Conectale una carpeta de trabajo, no tu escritorio ni tus documentos, y tené copia de lo importante. Nada de documentos financieros ahí. El borrado permanente lo pregunta siempre, en cualquier modo.

2

Elegí el modo según lo que se puede deshacer

Manual pregunta cada acción. Auto revisa cada acción y bloquea lo que le parece inseguro. Skip no revisa nada. Anthropic recomienda Manual cuando un error sería difícil de deshacer, como mandar mensajes o comprar.

3

Tareas programadas con cuidado

Se ven en Scheduled, en la barra lateral, y ahí se pausan o se borran; cada una tiene su propio modo de aprobación. Las recomendaciones oficiales: empezar simple, no programar tareas que toquen archivos sensibles, manden mensajes en tu nombre o compren, revisar el resultado de cada corrida y pausar las que no estás usando.

4

El navegador integrado recuerda tus logins

Lo que inicies ahí queda disponible para las próximas tareas. Si importás cookies de tu navegador, los sitios de banco, mail y acceso único vienen destildados: dejalos así. Anthropic desaconseja usar el navegador para cualquier cosa con información sensible.

5

En la empresa, limitá cada conector

En los planes Team y Enterprise, quien administra la cuenta elige qué puede hacer cada conector para toda la organización: Customize, Connectors, el conector, Tool permissions. Cada acción queda en Always allow, Needs approval o Blocked, y nadie puede cambiarlo por su cuenta. El ejemplo de Anthropic: que busque y resuma mails pero no pueda mandar.

6

La señal de alarma

Si entra a archivos o sitios que no mencionaste, frená la tarea en el momento. Es la pregunta que propone la propia guía de seguridad: "Is Claude accessing files or websites you didn't mention?"

PASO A PASO

ChatGPT Work

Reemplazó al viejo modo agente. En la web y el celular corre en la nube y puede seguir después de que cerrás la compu. Algunas apps conectadas pueden hacer cambios, no solo leer: la de Outlook, al conectarla, pide permiso para mandar mails.

1

Fijá el permiso por defecto de las apps

Settings, la sección de apps conectadas (se llama Plugins), Permissions. La opción razonable es **Allow read actions**: lee sin preguntar y pregunta antes de cambiar algo. Ante una acción aparece una tarjeta con Deny, Allow once y Always allow: para mandar, compartir o borrar, Allow once.

2

Revisá las tareas programadas

En Scheduled se ven, se pausan, se editan y se borran. Pueden arrancar solas por un horario o por un mail nuevo. Borrar el chat de una tarea solo la pausa. Y alguna se crea sola: conectar Finances puede generar un resumen semanal de tus finanzas.

3

Archivos de tu compu, solo los de la tarea

Desde la app de escritorio, Work puede usar archivos y programas de tu compu con tu permiso. OpenAI recomienda abrir solo la carpeta del proyecto y darle acceso únicamente a los archivos que la tarea necesita.

4

El navegador en la nube queda logueado

Dejá el acceso a sitios en **Always ask** (Settings, Cloud browser), que viene así de fábrica. Pide confirmación antes de cualquier cosa difícil de revertir o que te comprometa. Los logins duran hasta que vencen; para cerrar sesión, Settings, Cloud browser, Browser data, Cookies. Contraseñas y tarjetas nunca en el chat.

5

Desconectá lo que no usás y revisá la memoria

Settings, la sección de apps conectadas (Plugins), la app, menú ⋮ y Disconnect. Cambiar el permiso no corta el acceso; desconectar sí, pero no borra la memoria, que puede guardar contenido de apps como Gmail (Settings, Personalization, Memory).

6

Si manejás información delicada

Settings, Security, Advanced security, **Lockdown Mode**. Limita lo que ChatGPT puede mandar a internet y apaga el modo agente.

PASO A PASO

Gemini, modo Task

Es el agente de Google, para suscriptores de Google AI Pro y Ultra con cuenta personal. Puede redactar, responder y reenviar mails, renombrar y borrar archivos de Drive, y entrar a los sitios donde estás logueado en tu Chrome.

1 Lo muy sensible lo pregunta siempre

Aunque cambies tus preferencias, Gemini pide permiso cada vez para mandar mensajes, modificar o borrar tus datos, comprar, pagar y completar formularios. Antes de compartir datos de salud, financieros o de documentos de identidad, también pregunta. Si le decís que no, frena y no busca otro camino. La excepción: en Google Tasks hace cambios masivos sin confirmar.

2 Revisá lo programado

En Scheduled, en la barra lateral, separado en Active, Paused y Completed. Las tareas pueden correr por horario, cuando llega un mail que cumple un filtro o siguiendo un tema. Desde ahí se pausan o se borran.

3 Mirá a qué apps entra en tu nombre

En la página de apps vinculadas de tu cuenta de Google, el filtro "Acceso de agentes" muestra las apps a las que entra Gemini. Para cortarlo, entrá a la app y elegí dejar de usarla.

4 Borrá lo que quedó

Borrar el hilo de una tarea borra la tarea, su actividad y sus programaciones, pero no los archivos que creó o modificó. Los datos del navegador remoto se borran en Settings, Browser, Delete remote browser data.

Si alguien te instala OpenClaw. Es un agente de código abierto que corre en tu compu y se maneja por WhatsApp o Telegram. Su propia documentación dice que el aislamiento viene apagado de fábrica, que las skills de terceros hay que tratarlas como "untrusted code" y recomienda usarlo con un número de WhatsApp aparte del tuyo. En enero de 2026, Censys encontró 21.639 instalaciones expuestas en internet. Si lo usás, pedile a quien lo instaló que corra la auditoría de seguridad que trae (`openclaw security audit`) y que no le dé acceso a tu carpeta personal entera.

PASO A PASO

Tu cuenta de Google y de Microsoft

Cualquier agente que lee tu mail o tu Drive entra con un permiso que diste en tu cuenta. Ahí se ve qué tiene cada uno y ahí se corta.

Google

1

Mirá quién tiene acceso

En myaccount.google.com/linkedapps, entrá a Acceso a tu cuenta de Google, elegí la app y tocá Ver detalles. Para cortarlo, Quitar acceso. Revocar no borra lo que la app ya copió: eso hay que pedirselo a quien la hace.

2

Antes de autorizar, leé la lista

Google muestra a qué datos quiere entrar la app. "Ver y copiar" es leer. "Gestionar" incluye editar, crear y eliminar. Si solo necesitás que lea y pide gestionar, no autorizés.

3

Compartí una carpeta, no el Drive entero

Seleccioná la carpeta, Compartir y el rol **Lector**, que ve y abre todo pero no cambia nada. Un Editor puede eliminar o mover archivos y, de fábrica, también compartirlos.

Si administrás Google Workspace o Microsoft 365 en tu empresa

En Google Workspace: consola de administración, Seguridad, Control de acceso y datos, y el menú que decide qué apps entran a las cuentas (Controles de API). Cada app queda como Confiable, Limitado, Datos específicos o Bloqueado, y podés hacer que los pedidos de apps nuevas te lleguen para aprobar. En Microsoft 365, de fábrica cualquier empleado puede autorizar una app a entrar a su casilla sin pasar por nadie. Desde el centro de administración de Entra, en Consentimiento y permisos, se puede limitar a apps de publicadores verificados, como recomienda Microsoft, y activar un flujo para que el administrador apruebe lo demás.

La rutina de los viernes, cinco minutos. Uno: abrí las tareas programadas de cada agente (Scheduled, o la pestaña Details en Grok) y pausá lo que ya no usás. Dos: mirá las apps conectadas y desconectá las que sobran. Tres: revisá en qué sitios quedó logueado el agente. Cuatro: recorré el historial de la semana y fijate si tocó algo que no le pediste. Cinco: en tu cuenta de Google, mirá qué apps y agentes tienen acceso.

FUENTES

De dónde sale cada dato

Documentación oficial consultada el 28 de octubre de 2026. Las rutas de menú pueden cambiar con cada actualización.

Grok Bot	docs.x.ai/grok-bot: overview · approvals-security-and-privacy · computer-and-apps · vault · skills-routines-and-automations · bots · security · faq · troubleshooting · x.ai/news/introducing-grok-bot
Claude Cowork	support.claude.com: Get started with Claude Cowork · Use Claude Cowork on web, desktop and mobile · Use Claude Cowork safely · Schedule recurring tasks in Claude Cowork · Use the built-in browser in Claude Cowork · Use Google Workspace connectors · Use connectors to extend Claude's capabilities
ChatGPT Work	help.openai.com: ChatGPT Work and Codex · ChatGPT agent · Scheduled tasks in ChatGPT · Using cloud browser · Managing app permissions · Connecting and managing app accounts · Outlook email and calendar apps · Lockdown Mode · Memory in ChatGPT
Gemini	support.google.com/gemini: 17094507 (tasks), 17094710 (scheduled), 17094196 (task threads)
OpenClaw	docs.openclaw.ai: gateway/security · gateway/sandboxing · tools/skills · channels/whatsapp · censys.com: OpenClaw in the wild (31/1/2026)
Claude Code	code.claude.com/docs/en/permission-modes · The Register, 27/10/2026
Google y Microsoft	support.google.com/accounts (13533235, 3466521) · support.google.com/drive (2494822) · knowledge.workspace.google.com · learn.microsoft.com/entra: Configure user consent · Configure admin consent workflow

Si te sirvió, pásasela a quien en tu equipo esté por dejar a un agente trabajando solo.

referente.ai · linkedin.com/company/referente-ai

Guía preparada por Nicolás Bilinkis, fundador de Referente.ai. En Referente medimos qué responden ChatGPT, Gemini y los demás asistentes cuando alguien pregunta por tu empresa, y lo comparamos con lo que tu empresa publica.