

Set the brakes: a 15-minute check

Before your team hands work to an agent, check six settings on each person's account: approvals, connected apps, website access, scheduled tasks, memory, and the account itself. There is one card each for Claude, ChatGPT and Gemini, from the vendors' own pages as of 16 Oct 2026. Where a page doesn't say, the card says so.

HOW TO RUN IT

- 1 Take the card for your tool. Sit with each person for 15 minutes, their settings open beside the card.
- 2 Go down the six rows. Tick each one when it is set.
- 3 While the team learns, set everything for levels 1 to 3: nothing goes out, and nothing changes, unless a person approves it in the tool.
- 4 Write the date on the card. Check again whenever your tool changes its menus, and at least every [month].

THE LEVELS THE CARDS USE

- **1 Read.** It looks and reports. It changes nothing and sends nothing.
- ■ □□ **2 Draft.** It drafts. Only you send or share.
- ■ ■ □ **3 Ask first.** It acts, and asks before each send, pay, delete or change.
- ■ ■ ■ **4 Alone.** It acts inside fixed limits. You review after.

Up one level after five clean runs. Down at once after a surprise.

Red flags

Stop the task if you see any of these.

- It opens files or websites you didn't mention.
- It asks for a password, a code or card details.
- It suddenly talks about something unrelated.
- It wants to send, forward or share with someone not in the brief.
- Sent items, posts or invites you never approved.
- A scheduled task, app or connector nobody remembers adding.
- It does more than you asked, or goes round in circles.

If an agent does something odd

- 1 **Stop it.** Stop the task. Pause any schedule it belongs to.
- 2 **Look where it acted.** Sent items, drafts, edited files and their version history, the calendar, the records it could change.
- 3 **Disconnect.** The apps and folders it used. Change any password or code it may have seen.
- 4 **Tell [name].** Say what it did and when. Nobody is blamed for reporting.

Why the settings, and not only the brief

A sentence in a brief is not a lock. "Ask first" works when it is an approval step in the tool. The UK National Cyber Security Centre puts it this way: "do not rely on prompting alone" (20 Aug 2026). And the person who sets an agent off answers for what it does.

Checked on the vendors' pages on 16 Oct 2026. These products change monthly. Check the settings in your own tool. Each card lists the pages it was checked against.

Claude: set the brakes

Checked by [name] on [date]

From Anthropic’s own pages, read 16 Oct 2026. Menus move: if a setting is not where the card says, search your tool’s help for its name.

	SETTING, AND WHERE	DEFAULT	SET IT TO	
1	Approvals. In each conversation. Claude asks questions when a request is unclear.	“By default, Claude asks before taking an action.” You can let it keep working and check in only when something needs a closer look.	1 to 3: keep it asking. 4 only: let it keep working, for one proven task.	☐
2	Connectors. Customize > Connectors; in a chat, “+” > Connectors. Each tool: Always allow, Needs approval or Blocked.	On Team and Enterprise an owner adds connectors first (you may see “Request”). Custom connectors “aren’t verified”.	1 and 2: write tools Blocked. 3: Needs approval, then Allow once in the chat.	☐
3	Website access. The desktop app’s built-in browser (Pro, Max, Team); Claude in Chrome (paid plans); computer use (research preview).	The built-in browser is “separate from your own browser, logins, and tabs”. Computer use asks permission for each app.	1 to 3: only sites you trust. You type passwords into the site yourself, never into the chat.	☐
4	Scheduled tasks. Schedule repeat work, such as a report every Monday. Dispatch (Pro and Max, desktop) runs background tasks you can start from the phone.	Long tasks carry on after you close the laptop.	Only the book’s level 1 recipes, with a named owner.	☐
5	Memory. Settings > Memory. One chat without it: turn off Memory in the “+” menu before the first message, or use an incognito chat. A project: turn off “Use account memory”.	On for Free, Pro and Max. Team and Enterprise: an owner enables it and each person opts in.	Check the Topics. Memory off, or a project of its own, for sensitive work.	☐
6	Your account. Sign in with the account your company gives you.	Task work “consumes limits faster than Chat”. Heavier models, such as Fable 5.1, use limits fastest.	Company data only in the company’s account. Start on the default model.	☐

What an admin sets for everyone: On Team and Enterprise, an owner decides which connectors are available and turns memory on for the organisation. Enterprise admins get at least 30 days’ notice before the merged Claude reaches them.

NOTES: WHAT WE CHANGED TODAY

PULL THE PLUG

IF IT ACTS ODDLY

Stop the task. Switch off its connector in that chat, or disconnect it in Customize > Connectors. Take back the folders you gave it.

Stop it. Look where it acted. Disconnect. Tell [name]. The red flags are on the first page.

Checked against Anthropic’s own pages, read 16 Oct 2026: [claude.com/resources/articles/ cowork-is-now-claude](#); [claude.com/product/ cowork](#); [claude.com/docs/ cowork/guide/dispatch.md](#), [third-party/claude-desktop/unified-claude.md](#), [connectors/overview](#); [support.claude.com/en/articles/ 11817273, 11724452](#).

ChatGPT: set the brakes

Checked by [name] on [date]

From OpenAI's own pages, read 16 Oct 2026. Menus move: if a setting is not where the card says, search your tool's help for its name.

	SETTING, AND WHERE	DEFAULT	SET IT TO	
1	Approvals in Work. Choose Work (the toggle at the top on desktop). Follow its progress in the chat.	You answer its questions, change direction and "approve important actions".	1 to 3: approve each important action yourself.	☐
2	App permissions. Settings > Plugins, then the app. Always ask, Allow read actions, Allow low-risk actions, Allow all actions.	Read the current setting on screen: it can differ by plan and account.	1 and 2: Always ask. 3: Allow read actions. 4 only: Allow low-risk actions, on one app, for one proven task.	☐
3	Website access. Work's cloud browser (paid plans, not Free or Go); the desktop app's built-in browser; a browser extension.	It pauses for your sign-in or confirmation; sign-ins go through a secure form the model cannot see.	1 to 3: confirm each site and step. Never type a password into the chat.	☐
4	Scheduled Tasks. Run once, on a schedule, or when something happens: a new Gmail message, a new Slack message, a GitHub pull request.	Plus, Pro, Business and Enterprise; not Free or Go. Event-started tasks run while you are away.	Only the book's level 1 recipes, with a named owner.	☐
5	Memory. Settings > Personalization > Memory: a summary you can correct. "Sources" under an answer show what was used.	Temporary chats can use or skip memory, chosen when you start them. Shared projects are always project-only memory.	Check the summary. A temporary chat that skips memory for sensitive work.	☐
6	Projects. New project in the sidebar; instructions under Project settings. Share with chat or edit access.	Project instructions override your general ones. Work is not available in project-only memory projects.	Edit access only for people who run the brief.	☐

What an admin sets for everyone: On Business and Enterprise, ask your admin which apps and plugins the workspace allows, and connect your work account, never a personal one.

NOTES: WHAT WE CHANGED TODAY

PULL THE PLUG

Stop the task in its chat. Stop any scheduled task it belongs to. Disconnect the app in Settings > Plugins.

IF IT ACTS ODDLY

Stop it. Look where it acted. Disconnect. Tell [name]. The red flags are on the first page.

Checked against OpenAI's own pages, read 16 Oct 2026: help.openai.com/en/articles/6825453, 20001275, 10169521, 8590148, 11487775, 20001280.

Gemini: set the brakes

Checked by [name] on [date]

From Google’s own pages, read 16 Oct 2026. Menus move: if a setting is not where the card says, search your tool’s help for its name. **Tasks, skills and custom apps are for personal accounts for now. On a work account, start with Gemini inside your Google apps, and keep work out of personal-account tasks.**

	SETTING, AND WHERE	DEFAULT	SET IT TO	
1	Confirmations. In the task, and on each connected app’s action.	Tasks: some actions need your confirmation, and for sign-ins it asks you to take over. Apps: write actions always need your confirmation.	Read each confirmation before you say yes.	☐
2	Connected apps. Left menu > Customize, or Settings > Connected Apps. Mention an app with @ to use it.	On a work account your admin decides which apps are allowed. Google “does not control, monitor, or secure third-party MCP servers”.	Connect only what the task needs. Remove apps you no longer use.	☐
3	Website access. Tasks use your local Chrome or a remote browser in the cloud. Gemini in Chrome has auto browse (Pro and Ultra, US first).	Google advises not to type passwords or payment details into the task.	1 to 3: confirm each step, and do sign-ins yourself.	☐
4	Tasks and schedules. Click Task below the text box; describe the task and when it should run.	18 or over, a personal account, Google AI Pro or Ultra, Keep Activity and Personal Intelligence on; not in the EEA, Nigeria, Switzerland or the UK.	Only the book’s level 1 recipes, with a named owner.	☐
5	Memory. Customize sets how Gemini responds; past chats are now called memories. Personal Intelligence links Gmail, Photos, Search and other Google apps.	Personal Intelligence is off until you turn it on.	Leave it off for work unless your company allows it.	☐
6	Work account. Sign in with your company’s Google Workspace account.	The admin controls these features; chat history is kept 18 months by default.	Company data only in the company’s account.	☐

What an admin sets for everyone: On a work account, your Workspace admin decides which apps are allowed, and controls memory and chat history.

NOTES: WHAT WE CHANGED TODAY

PULL THE PLUG
IF IT ACTS ODDLY

Stop the task. Remove the app from Connected Apps. Do any sign-in yourself.
Stop it. Look where it acted. Disconnect. Tell [name]. The red flags are on the first page.

Checked against Google’s own pages, read 16 Oct 2026: gemini.google/release-notes/; support.google.com/gemini/answer/17094507,17209137,14959807,14620100; support.google.com/mail/answer/13952129.