

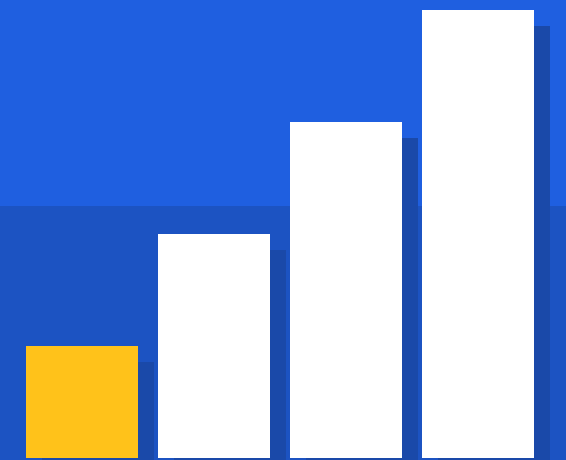
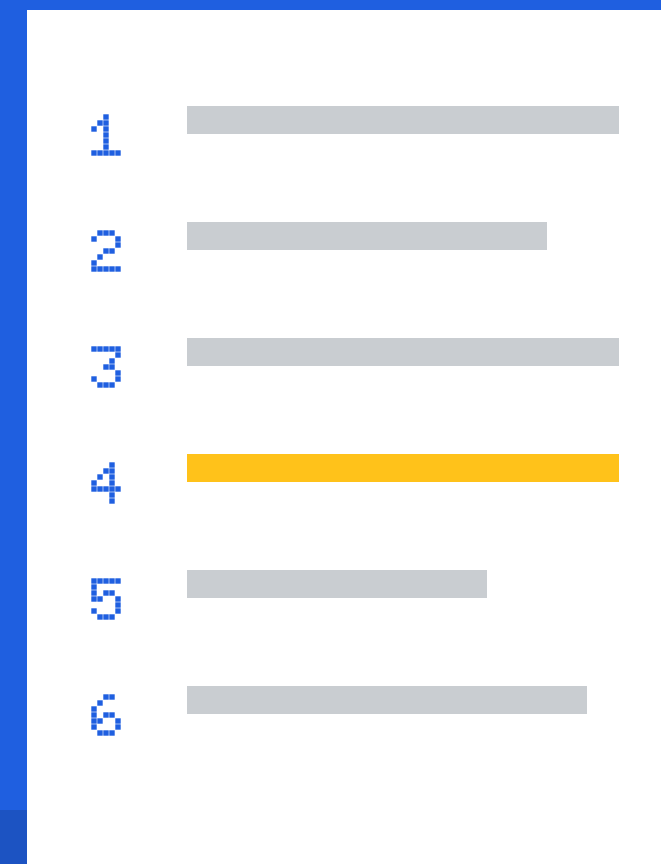
Brief an Agent

A 60 minute session on handing work to AI agents, and staying in charge.

BRIEF IT

CHECK IT

KEEP IT SAFE



WHAT CHANGED, AS OF OCT 2026

Claude, ChatGPT and Gemini now take a goal and do the work.

What you can see depends on your plan, your country and your admin.

ANTHROPIC

Claude

Since 16 Sep 2026, chat and Cowork are one Claude. Hand over a task in any conversation: there is no mode to pick.

For software: Claude Code

OPENAI

ChatGPT Work

Since 9 Jul 2026, one app holds Chat, Work and Codex. Switch the toggle to Work for a task that ends in a file or a finished job.

For software: Codex

GOOGLE

Tasks in Gemini

Click Task below the text box: personal accounts on Google AI Pro or Ultra, for now. At work, Gemini lives inside Gmail, Docs and Sheets.

For software: Antigravity

Claude first, ChatGPT second, Gemini third.

The tool your organisation approves comes first, for company data. Each one wins somewhere:

	CLAUDE	CHATGPT	GEMINI
Hand over a task	One conversation. It asks before acting, by default.	Work, with a cloud browser for sites you sign in to.	Task, on personal accounts only, for now.
Documents, slides, sheets	Docs and Slides on every plan. Claude for Excel.	Following company templates (OpenAI's own claim).	Inside Docs, Sheets and Slides at Google companies.
Research	The web plus your connected work tools.	Most control: edit the plan, name the sites.	Free Deep Research. Gemini Notebook for your own files.
Repeat routines	Scheduled tasks, and Dispatch on the desktop app.	Tasks that start when an email or Slack message arrives.	Schedules in tasks, on personal accounts.
Meetings, pictures	Give it the transcript or the notes.	Meetings plugin (beta, Mac). ChatGPT Images 2.5.	Take notes for me in Google Meet. Nano Banana images.

ALL THREE ASK FOR THE OUTCOME

ANTHROPIC, ON CLAUDE

Describe the result you want, attach files, and choose the folders it may use.

Anthropic's pages on Claude, 16 Sep 2026, in plain words

OPENAI, ON CHATGPT WORK

Add your files, then "describe the outcome you want" with any rules, and how you will judge it.

OpenAI, ChatGPT Work and Codex help page, read 16 Oct 2026

GOOGLE, ON TASKS IN GEMINI

Describe the task, and when it should run.

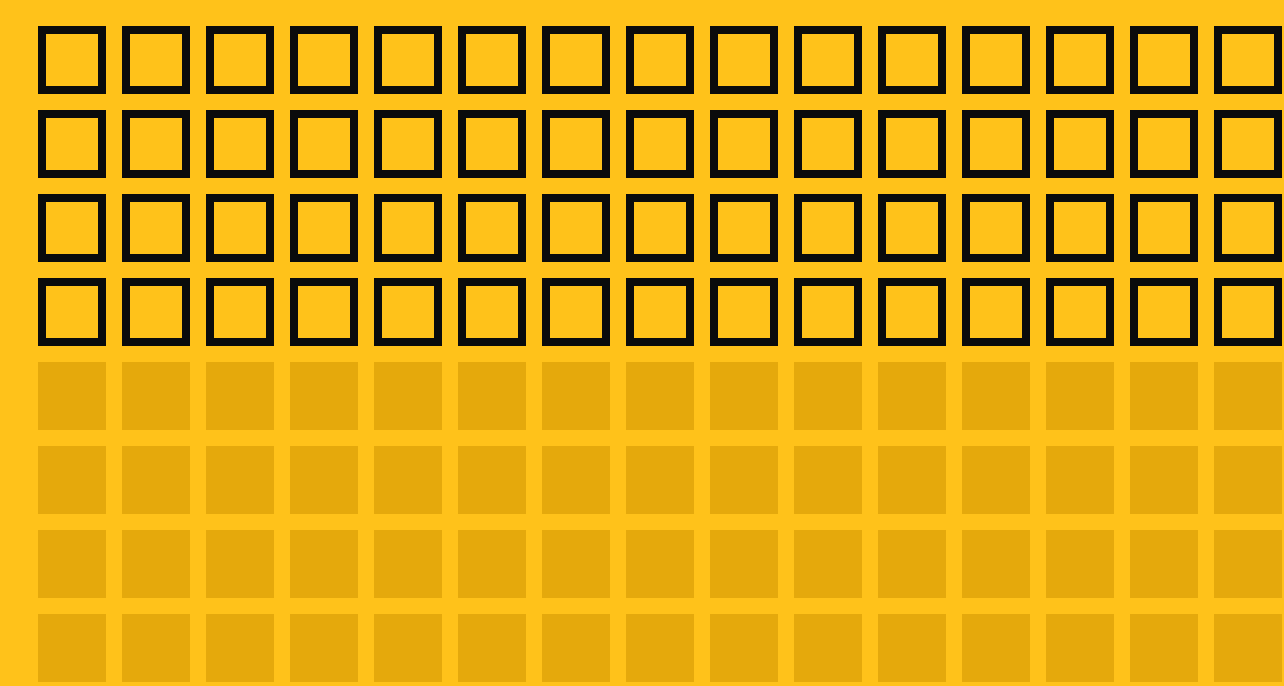
Google, Gemini Apps Help on tasks, read 16 Oct 2026, in plain words

So the skill isn't clever wording any more. It's the brief: the outcome, the limits and when to stop and ask.

In one study, people tapped out well-known songs, and expected the listener to name them half the time.

THE TAPPERS' AVERAGE GUESS

1 IN 2

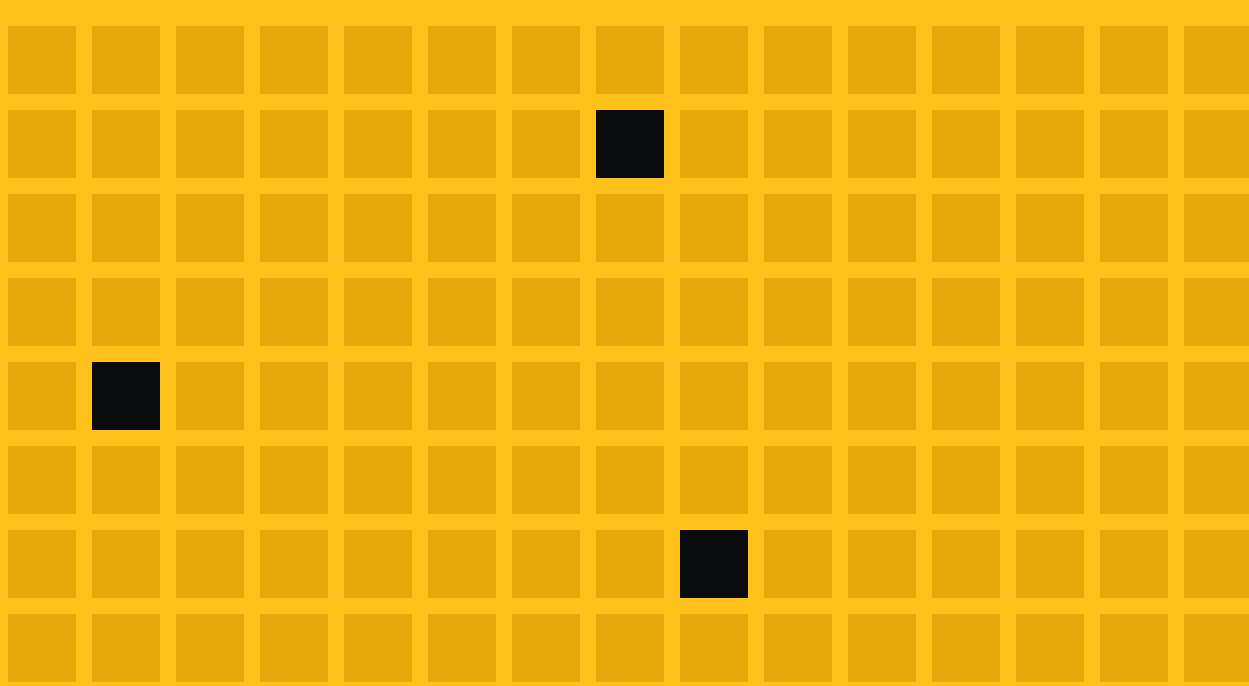


OUTLINED: THE GUESS, ON AVERAGE

SOURCE: NEWTON, STANFORD, 1990

THE LISTENERS NAMED

1 IN 40



3 OF 120 SONGS

A study of songs, not AI.

Your brief is the tapping.

You hear the whole tune: why it matters, who it's for, what's off limits. The agent only gets what you type, and it acts on its best guess.



Our comparison. The study was about songs, not AI.

Give it six things before it starts.

1

Goal

What it is to achieve, why, and by when.

2

Done looks like

The finished thing: format, length, where it lands.

3

Context

The files, links and examples to use, and which to trust.

4

May touch, must not

What it may read, use and change, and what is off limits.

5

Limits

Time, steps, pages or spend: when to stop anyway.

6

Stop and ask

The moments it must come back to you first.

Before running an agent, think carefully about "what it is to achieve" and "what actions it should be allowed to take" and "when it should stop and seek human approval"

UK National Cyber Security Centre, 20 Aug 2026

Sort out a venue for the team day.

1 GOAL	Find three venues for our team day in November, so I can choose one by Thursday.
2 DONE LOOKS LIKE	A table in this chat: venue, travel time, room for 20, step-free access, a link to its own page. Best first.
3 CONTEXT	Last year's feedback and our travel policy (attached). Treat review sites as background only.
4 MAY TOUCH, MUST NOT	You may search the web and read venue websites. Don't book, pay, sign up, fill in any form or email anyone.
5 LIMITS	About 20 minutes. No more than 20 web pages.
6 STOP AND ASK	If a venue asks for a deposit, a sign-in or anyone's details, stop and tell me.

Check: The facts. Open each link and find the room size on the venue's own page.

Turn a vague request into a brief

- 1** **1 MIN** Pick a request, A to D, or a task of your own with nothing confidential in it.
- 2** **5 MIN** Write the six parts on a brief card. Tick a level and a check.
- 3** **3 MIN** Swap cards. Ten second test: could you start without asking? Circle where you'd ask.
- 4** **2 MIN** Swap back. Fix one circled part.
- 5** **1 MIN** One pair reads out its Done looks like, and its May touch, must not.

MADE UP, PICK ONE

A

Sort out the supplier stuff before Friday.

B

Get the new starter ready for Monday.

C

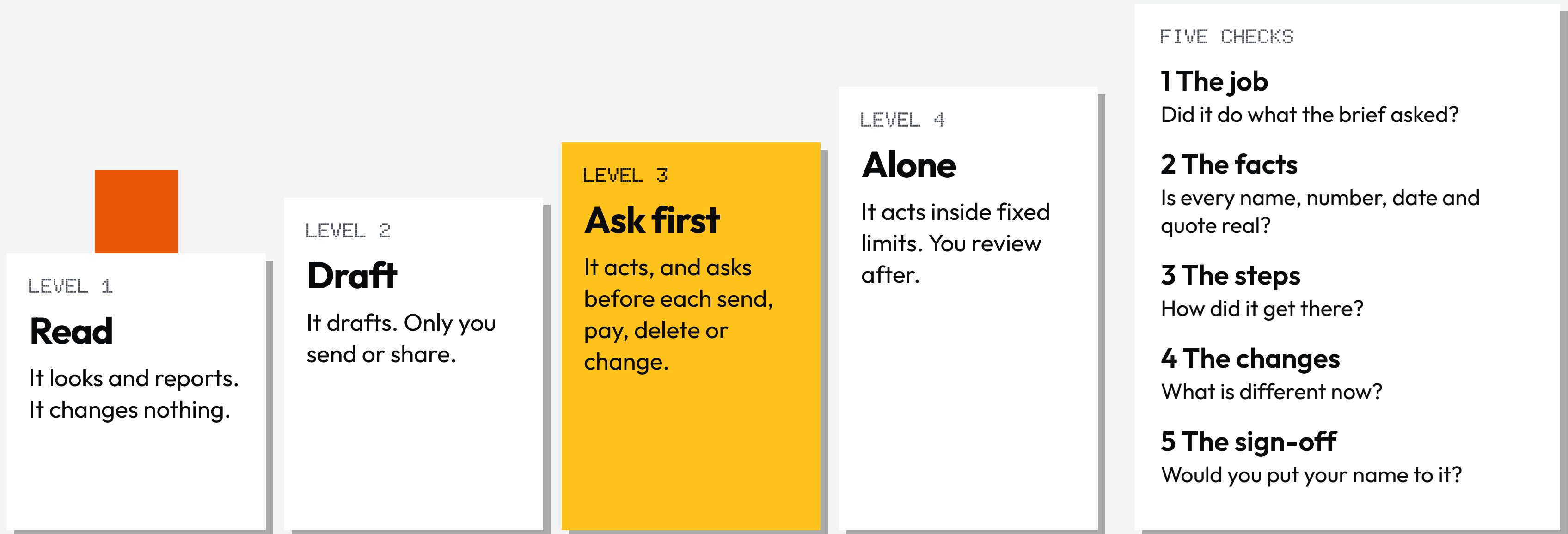
Look at last quarter's numbers and tell me what's going on.

D

Clean up the team calendar for next week.

Read it the way an agent would: literally, and without asking.

Four levels. Start low.



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

A clean run: it passed all five checks, with nothing to fix.

RULE 1

"Ask first" is set in the tool, not typed in the brief.

A sentence in a brief is not a lock. Find the approval setting, set it to ask, and check nobody has switched it off.

CLAUDE, AS OF OCT 2026

"By default, Claude asks before taking an action" so leave it that way. Set connector tools to Needs approval, and choose Allow once until you trust it.

CHATGPT WORK, AS OF OCT 2026

It pauses so you can "approve important actions" as it works. Set each app to Always ask while you learn, and never to Allow all actions.

GEMINI, AS OF OCT 2026

Tasks stop for your sign-ins and confirm some actions. Write actions in connected apps always need your confirmation.

WHAT IT CAN TOUCH

A connector is a signed-in link to another app.

Your files, your calendar, your email. It can only reach what your own account can reach, and only what the connector allows.

MCP

The shared standard those links use. Its own site says:

"Think of MCP like a USB-C port for AI applications"

CLAUDE CALLS IT

Connector

Add it in Customize, Connectors. In a chat: + then Connectors.

CHATGPT CALLS IT

App, inside a plugin

Add it in Settings, Plugins. In a chat: type @ or +.

GEMINI CALLS IT

Connected app

Add it in Customize. In a chat: mention it with @.

Switch on only what the task needs. Start on ask or read-only.

RULE 2

OUTSIDE WORDS, NO POWER

An agent that reads outside content (the web, outside email, files others sent) may not also send, pay or delete without asking.

"For prompt injection attacks to be successful, two things must be true at the same time" and Anthropic names them: Claude "can read information outside your trusted boundary, and can perform actions that could compromise the user"

Anthropic, Use Claude Cowork safely, read 16 Oct 2026

These steps cut the risk. They don't end it.

Pick the tool. Then set the level.

CHAT

One answer, draft or idea that you use yourself.

DEEP RESEARCH

A question that needs many sources, read and cited.

AGENT

The next step depends on what it finds, in your files and apps.

AUTOMATION

You can write every step down in advance, the same each time.

QUESTION 1

Can it be undone easily?

If not: level 3 at most, and you approve each step.

QUESTION 2

Does it leave the team?

A customer, a supplier, the public, money. If so: level 3 at most, and a person presses send.

QUESTION 3

Does it read outside content?

Web pages, outside email, files others sent. If so: it must not also send, pay or delete without asking.

Set each agent's level: the lowest level that does the job, never above what the three questions allow.

Sort ten team tasks

- 1 4 MIN Sort the ten task cards by tool: Chat, Deep research, Agent, Automation.
- 2 3 MIN For each Agent card, ask the three questions and write its level: the lowest level that does the job, never above what the three questions allow.
- 3 2 MIN Compare with the next group. Which card did you disagree about?

The tasks are made up. One card fits none of the four tools.

- 1 Which part of a brief did the ten second test catch most often?**
- 2 Which task card did your group disagree about, and what settled it?**
- 3 Where is the approval setting in our own tools, and who checks it?**
- 4 Which of our tasks should stay with a person, at any level?**

OUR FIRST LEVEL 1 TASK

THE TASK

Type it here: one task that only reads and reports.

ITS OWNER

One person, not the whole team.

ITS CHECK

One of the five checks, and how.

BRIEF IT
CHECK IT
KEEP IT SAFE

neosaipien

Five clean runs, then decide together whether it moves up.