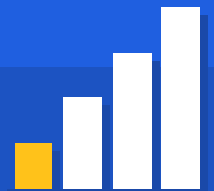
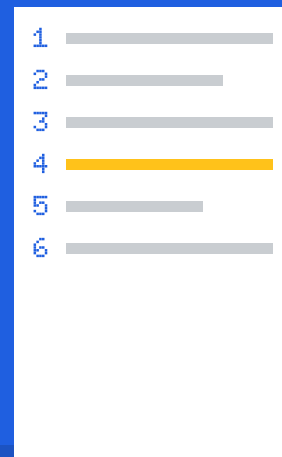


A 60 MINUTE SESSION

Brief an Agent

A session any team lead can run, with no trainer in the room. Your team learns to hand work to an AI agent with a clear brief, a level it may not go past, and a check before anything counts. You leave with your first task agreed, its owner and its check.



INSIDE THE BOX

- 01 **This facilitator guide**
The 60 minutes, what to say, both exercises with their answers, questions people ask, three emails and the sources.
- 02 **17 slides**
PowerPoint and PDF, with speaker notes that match this guide.
- 03 **A two-page handout**
The six-part brief to write on, and where everything is in Claude, ChatGPT and Gemini.
- 04 **Brief cards and task cards**
Brief cards with room to write, and ten made-up tasks to sort.
- 05 **A one-sheet on prompt injection**
How a web page can hijack an AI, for anyone whose agent reads the web or email.
- 06 **Three emails and a feedback form**
The invite, the follow-up, a two-week habit check, and a one-question form.

60 MINUTES

4 TO 16 PEOPLE

CLAUDE, CHATGPT OR GEMINI

NO TRAINER NEEDED

What this session does

AI assistants now take a goal and do the work: they research, draft, file, book and send. Hand one a vague request and you get back nothing useful, or an agent that does too much. In one hour your team agrees how to brief an agent, how far to let it go, and how to check what it did.

WHY NOW, AS OF OCT 2026

- Claude, ChatGPT and Gemini now take a goal and do the work. Since 16 Sep 2026 Claude does tasks in any conversation; ChatGPT has Work, in one app with Chat and Codex; Gemini has tasks, for personal accounts (page 5).
- Their makers ask for the outcome, not clever wording. OpenAI, on ChatGPT Work: add your files, then "**describe the outcome you want**" with any rules and how you will judge it.
- The UK's National Cyber Security Centre (20 Aug 2026) says to 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**"

WHO IT IS FOR

- Teams of 4 to 16 starting to hand work to Claude, ChatGPT Work or tasks in Gemini.
- The team lead who runs it. You don't need to be an AI expert: this guide says what to do, what to say and what to watch for.
- Any of the three. The brief, the levels and the checks are the same in all of them; the handout's back shows where each part is. My recommendation is on page 14.
- Not a course on one tool, and not a policy. Your organisation's own AI rules come first; this session works inside them.

WHAT YOUR TEAM LEAVES WITH

- 01 **The six-part brief.** Goal, Done looks like, Context, May touch, must not, Limits, Stop and ask.
- 02 **The four levels.** 1 Read, 2 Draft, 3 Ask first, 4 Alone, and when a task moves up.
- 03 **The three questions** that set how far an agent may go, and the tool chooser for when it shouldn't be an agent at all.
- 04 **Your first level 1 task**, with its owner and its check, agreed in the room.

WHAT'S IN THE BOX

FILE	WHAT IT IS	WHEN YOU USE IT
01 Facilitator guide A4 and US Letter	The plan, what to say, both exercises with answers, the emails	Read it once before. Keep it open during the session
02 Slides PowerPoint and PDF	17 slides with speaker notes	On the screen
03 Participant handout A4 and US Letter	Two pages: the brief to write on, and your tool start to end	One per person, printed on both sides
04 Brief and task cards A4 and US Letter	Page 1: two brief cards. Pages 2 and 3: ten task cards and three reference cards	Page 1 per pair; pages 2 and 3 per group, cut up
05 Prompt injection one-sheet A4 and US Letter	How a web page can hijack an AI	Hand out at the end
06 Emails, 07 form	Invite, follow-up, habit check. One question	Before, after, two weeks on

Free licence: free from Neosaipien (LICENCE.txt). Run it as often as you like, with as many teams as you like, change the files to fit, and hand them to your own people, keeping the credit and the date.

Before the session

ONE WEEK BEFORE

- ☐ Book 60 minutes and a room with a screen. Online works too: see Running it online.
- ☐ Send the invite (on the right, and in the emails file).
- ☐ Check which AI tools your organisation approves, and what may go into them.
- ☐ **Read page 14 once:** where each part of this session lives in Claude, ChatGPT and Gemini, and which one I recommend.
- ☐ **Find the approval setting** in the tool your team uses: where it asks before it acts, and whether anyone has switched that off. You'll talk about it at 00:31.
- ☐ **Ask whoever looks after your AI tools** which apps are connected with write access, and which agent tasks already run on a schedule.

THE DAY BEFORE

- ☐ Print the handout on both sides, one per person.
- ☐ Print the cards at 100% (actual size). Page 1, one per pair. Pages 2 and 3, one set per group of three or four, cut up.
- ☐ Print the one-sheet, one per person, to hand out at the end.
- ☐ Write one brief yourself, for request A on page 7. Ten minutes, and you'll have your own example.

IN THE ROOM

- ☐ The slides on a screen, and a timer
- ☐ Pens, and scissors if the cards aren't cut yet
- ☐ A whiteboard or flip chart for the team's first task

THE INVITE / A WEEK BEFORE

Subject: 60 minutes on handing work to AI agents

Hi all,

On [day] at [time] we're spending an hour on one practical skill: handing work to an AI agent so it does the job we meant, and nothing we didn't.

AI assistants can now do the work, not just answer: they research, draft, file, book and send. So the skill that matters is the brief. What should it achieve, what may it touch, and when must it stop and ask?

Please think of one task you'd like to hand over: something that eats time every week. We'll practise on made-up tasks, so please don't bring anything confidential.

Bring a pen. If you have [our approved AI tool: Claude, ChatGPT or Gemini], bring a laptop or phone with it, but you won't need it to take part. Nobody will be tested.

[Your name]

RUNNING IT ONLINE

Share the slides. Send the handout and the cards as PDFs the day before. For Exercise 1, pairs write the brief in a shared document and swap links. For Exercise 2, groups sort the ten tasks in a shared table with the same columns as the cards.

OPTIONAL: A LIVE DEMO

Only in a tool your organisation already approves, and only if you have run the same brief yourself first. Run one read-only, level 1 brief: the venue brief on slide 8. In Claude, an ordinary conversation; in ChatGPT, switch to Work. Set the tool to ask before every action. **Never connect email, messaging or payment tools in the room.** It adds about 5 minutes (an estimate): skip the volunteer tapping to make room.

YOUR JOB IN THE ROOM

Keep the clock. The exercises are where people learn, so protect their time.

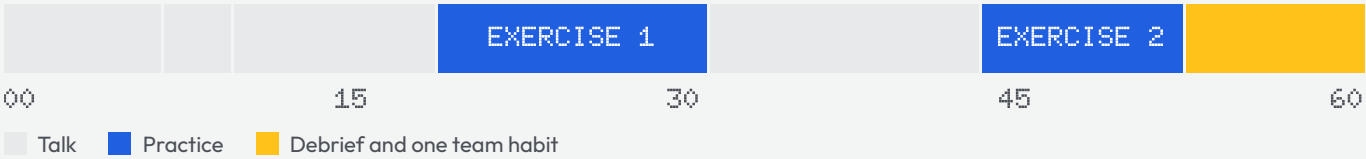
You don't need to know more than the room. If someone knows a lot about agents, ask them to help a pair.

Ask, don't answer. When a pair asks if their brief is right, ask what the agent would do with it.

Keep it made up. Real tasks are welcome as ideas, but nothing confidential goes on a card.

The 60 minutes

Talk takes 31 minutes, practice 21, and the debrief 8. The timings are estimates: a first run with a real team will tell you where yours differ.



STARTS	MIN	WHAT HAPPENS	SLIDES	PAGE
00:00	7	Welcome, what changed in Claude, ChatGPT and Gemini, and which one for what	1 to 4	5
00:07	3	Why your brief only makes sense to you: the tapped song	5 and 6	5
00:10	9	The six-part brief, grown from a vague request	7 and 8	6
00:19	12	Exercise 1: turn a vague request into a brief	9	7
00:31	12	How far to let it go: four levels, five checks, connectors, two rules	10 to 13	9, 10
00:43	9	Exercise 2: sort ten team tasks	14 and 15	11
00:52	8	Debrief, and one team habit	16 and 17	12

KEEPING TIME

- If you fall behind, cut talk, not practice.
- Start each exercise by reading its steps on the slide aloud, then start the timer.
- Give a one minute warning before each exercise ends.

ONLY 45 MINUTES?

Skip the volunteer tapping and keep slide 5 to one minute. Point at slide 3 without reading it. Take 3 minutes off each exercise and 2 off the six-part brief, and ask two debrief questions.

HAVE 90 MINUTES?

Run the live demo (page 3). Give each exercise 8 more minutes so pairs share with the whole room, and spend 8 more on the team's first task: write its brief together on the board.

HOW TO READ THE NOTES

- SAY
 - Points to make in your own words. Where a line is in bold, use those words: they are the facts, and they have been checked. Quotation marks show a source's own words.
- DO
 - What you and the group do.
- WATCH FOR
 - What might go wrong, and how to steer it.
- KEEP IT ACCURATE
 - The limits of the facts. Stay inside them, even when someone asks for more.

00:00 7 MIN / SLIDES 1 TO 4

Welcome, and what changed

SAY

- This hour is about handing work to an AI agent: one that does the job, not just answers.
- **SLIDE 2 Claude, ChatGPT and Gemini now take a goal and do the work.** In Claude, any conversation; in ChatGPT, switch to Work; in Gemini, click Task. All as of Oct 2026.
- **SLIDE 3** Each one wins somewhere: point at the row your team cares about. **Samson Mathew, who made this workshop, recommends Claude first, ChatGPT second, Gemini third.** The tool your organisation approves comes first.
- **SLIDE 4** All three makers now ask for the outcome. So the skill isn't wording any more. It's the brief.

DO

- Hand out the handout. Point to the brief at the top: they'll fill it in at 00:19.
- Ask: "Who has handed a task to an agent already? What happened?" Take one answer.

KEEP IT ACCURATE

- Say "as of Oct 2026" for every product line. Claude's one conversation is rolling out, Pro and Max first. Gemini's Task needs a personal account on Google AI Pro or Ultra; at work, Gemini lives inside Gmail, Docs and Sheets.
- Features arrive in waves. If someone can't see one, it's their plan, their country or an admin setting. Claude Code, Codex and Antigravity are for software: this session doesn't need them.
- The claims in slide 3 about templates and research are the vendors' own.

00:07 3 MIN / SLIDES 5 AND 6

Why your brief only makes sense to you

DO

- Ask a volunteer to tap out a song everyone knows on the table, with one finger and no humming. Or tap one yourself: Twinkle, Twinkle, Little Star works. Give the room 20 seconds to guess.

SAY

- **SLIDE 5 In one study, people tapped out well-known songs, and expected the listener to name them half the time.** That's the tappers' average guess. **But the listeners got one in forty.** That's 3 songs out of 120.
- The tapper hears the whole tune in their head. The listener only hears taps.
- **SLIDE 6** Your brief is the tapping. You know why it matters, who it's for and what's off limits. The agent only gets what you type, and it may not stop to ask about what you left out. It acts on its best guess.

KEEP IT ACCURATE

- "Half the time" is the average of the tappers' guesses. Keep the word average with it.
- Say 1 in 40 or one in forty, and 3 of 120 songs. Never any other count.
- Don't say which songs were tapped or named: the study doesn't report that. Don't say the volunteer's song was one of them.
- The study was about songs, not AI. Slide 6 is our comparison: keep the numbers on slide 5.
- Source: Newton, Stanford, 1990, a PhD study.

00:10 9 MIN / SLIDES 7 AND 8

The six-part brief

SAY, SLIDE 7

- Brief an agent the way you'd brief a capable colleague on their first day. Six things, before it starts.
- The UK's National Cyber Security Centre asks the same questions. 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" and those are Goal, May touch, must not, and Stop and ask.
- It also says to "Instruct the agent in both what it should do and what it should not do" and so part 4 always has both halves.

THE SIX PARTS

- 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.

SAY, SLIDE 8: THE EXAMPLE

EXAMPLE BEFORE **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 from the office, room for 20, step-free access, and a link to the venue's own page. Best first.
3 CONTEXT	Last year's feedback and our travel policy (both 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.

LEVEL 1 Read CHECK The facts: open each link and find the room size on the venue's own page.

SAY, THEN

- Notice what the vague version left the agent to guess: how many, by when, what matters, what it may do, when to stop.
- It reads the web, so it gets no power to book or pay. That's a rule we'll come back to at 00:31.

KEEP IT ACCURATE

- The venue brief is made up for this workshop. It names no real venue, company or person.
- The NCSC lines are from its blog of 20 Aug 2026, which says formal guidance will one day replace it.

00:19 12 MIN / SLIDE 9 / EXERCISE 1

Turn a vague request into a brief

Goal: every pair writes a six-part brief on a brief card, then another pair tests it.

STEPS, ON THE SLIDE

- 1 **1 MIN** Pick a request, A to D on the right, or a task of your own with nothing confidential in it.
- 2 **5 MIN** Write the six parts on a brief card. Make up any details you need. Tick a level and a check.
- 3 **3 MIN** The ten second test. Swap cards with another pair. Could you start the job without coming back to ask something? Circle every part where you'd have to ask.
- 4 **2 MIN** Swap back and fix one circled part.
- 5 **1 MIN** One pair reads out its Done looks like, and its May touch, must not.

WATCH FOR

- A goal that's an activity, such as research venues. Ask: what is it for, and by when?
- Only half of part 4. Most pairs write what it may touch and forget what it must not. Ask for both.
- A vague stop rule, such as if unsure. Ask: which moment, exactly? A deposit, a clash, two files that disagree.
- Briefs that only get longer. Length isn't the point: each part should take one guess away.

FOUR MADE-UP REQUESTS

- 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.

SAY

- Before step 3: read it the way an agent would, literally and without asking.
- After step 5: what did the test catch?

KEEP IT ACCURATE

- The ten second test is our name for what Anthropic's prompting guide calls its golden rule: show your prompt to a colleague with minimal context on the task. If they'd be confused, the model will be too.
- Source: Anthropic's guide Prompting best practices, in its section Be clear and direct.

THE ANSWERS

One good brief for each request is on the next page. Keep it for yourself: any brief that passes the ten second test is a good answer.

FOR YOU, NOT THE ROOM

Exercise 1: one good brief for each request

Made up for this workshop. The details are invented: a pair's own details are just as good. What matters is that each part takes a guess away.

A / Sort out the supplier stuff before Friday.

1 GOAL	List which supplier contracts end before March, so I can decide on Friday which to renew.
2 DONE LOOKS LIKE	A table in this chat: supplier, what they supply, end date, notice period, and the clause it comes from. Soonest first.
3 CONTEXT	The contracts folder and the supplier list. Trust the signed contracts over the list.
4 MAY TOUCH, MUST NOT	Read the folder and the list. Don't email suppliers, edit files or sign anything.
5 LIMITS	About 20 minutes. Our 30 biggest suppliers only.
6 STOP AND ASK	If a contract has no end date, or the two disagree, list it and ask me.
LEVEL, CHECK	1 Read. The facts: open three contracts and check the end dates yourself.

B / Get the new starter ready for Monday.

1 GOAL	Get everything ready for the new starter who joins on Monday, so their first day runs without gaps.
2 DONE LOOKS LIKE	A checklist for me, a first-day timetable as a draft document, and draft invites for three welcome meetings. Nothing sent.
3 CONTEXT	Our onboarding checklist, the team list and the role description.
4 MAY TOUCH, MUST NOT	Read those and the team's calendars. Create one document and draft invites. Don't send invites, order equipment, create accounts or share the starter's personal details.
5 LIMITS	The first day only. About 20 minutes.
6 STOP AND ASK	If someone on the welcome list is away, suggest another person and ask me.
LEVEL, CHECK	2 Draft. The changes: open each draft invite. Right person, right time, right room.

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

1 GOAL	Find the three biggest changes in sales between the last two quarters, so I can explain them at Tuesday's meeting.
2 DONE LOOKS LIKE	One page: each change with its numbers, how sure you are (high, medium or low), and what to check next.
3 CONTEXT	The two quarterly sales exports. Treat anything else as background only.
4 MAY TOUCH, MUST NOT	Read the two exports. Don't edit them or any record, and don't email anyone.
5 LIMITS	About 30 minutes.
6 STOP AND ASK	If the exports disagree on a total, or a change rests on one customer, stop and ask me.
LEVEL, CHECK	1 Read. The facts: add up one total yourself, and open the rows behind each change.

D / Clean up the team calendar for next week.

1 GOAL	Find the clashes and double bookings in the team calendar for next week, so we can fix them at Monday's stand-up.
2 DONE LOOKS LIKE	A list in this chat: each clash, who it affects, and two ways to fix it.
3 CONTEXT	The shared team calendar, next week only.
4 MAY TOUCH, MUST NOT	Read the team calendar. Don't move, decline or delete any meeting, and don't message anyone.
5 LIMITS	Next week only. About 10 minutes.
6 STOP AND ASK	If a clash involves someone outside the team, list it and ask me.
LEVEL, CHECK	1 Read. The changes: open the calendar and confirm nothing moved.

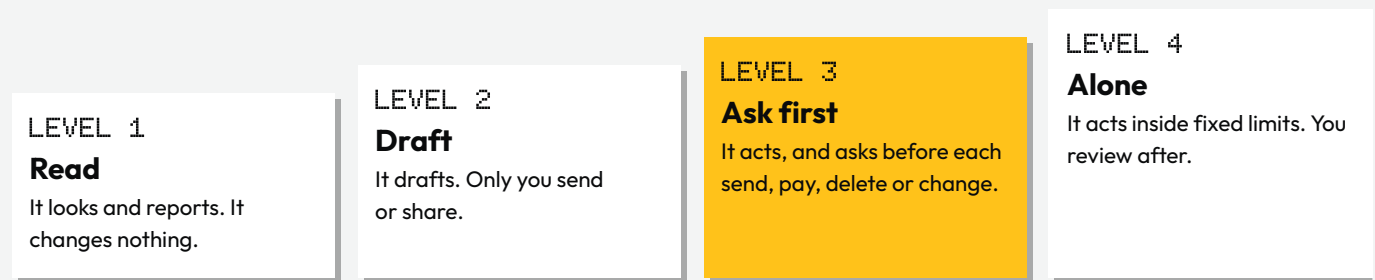
WHAT MOST PAIRS MISS

In B, letting it send the invites is level 3, Ask first: fine later, after five clean runs at level 2. In D, moving other people's meetings is an action, so start by having it list the clashes.

00:31 7 MIN / SLIDES 10 AND 11

How far to let it go

SAY, SLIDE 10: FOUR LEVELS



- Start low. **Up one level after five clean runs. Down at once after a surprise.**

SAY, SLIDE 10: THE FIVE CHECKS

Read each one out. Pairs tick one on their brief card, and a clean run needs all five.

CHECK 1 The job Did it do what the brief asked?	CHECK 2 The facts Is every name, number, date and quote real?	CHECK 3 The steps How did it get there?	CHECK 4 The changes What is different now?	CHECK 5 The sign-off Would you put your name to it?
---	---	---	--	---

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

SAY, SLIDE 11: ASK FIRST

- **"Ask first" is set in the tool, not typed in the brief.** A sentence in a brief is not a lock. Set it as an approval step in the tool.
- As of Oct 2026, Claude: **"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.
- As of Oct 2026, ChatGPT Work: 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.
- As of Oct 2026, Gemini: tasks stop for your sign-ins and confirm some actions. Write actions in connected apps always need your confirmation.
- So: find the setting in your own tool, set it to ask, and check nobody has switched it off. Settings change monthly: the Agent Playbook's dated insert lists where each one is, as of its edition.

WATCH FOR

- **So we'll keep everything at level 1.** Fine for a start. The point is that each step up is earned, with five clean runs.
- **Our tool asks anyway.** Ask who set it, and whether anyone can switch it off, or set an app to allow everything.

KEEP IT ACCURATE

- The levels, the five clean runs and the rules are ours: a way of working, not a vendor's feature.
- The three tool lines are as of Oct 2026, read on Anthropic's, OpenAI's and Google's own pages on 16 Oct 2026. Names of settings change, so say as of Oct 2026 with each one.

00:38 5 MIN / SLIDES 12 AND 13

What it can touch, and what it reads

SAY, SLIDE 12: CONNECTORS IN PLAIN WORDS

- A connector is a signed-in link from your AI 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.
- Each maker has its own word for it. Claude: a connector. ChatGPT: an app, installed inside a plugin. Gemini: a connected app.
- Those links use one shared standard, MCP, the Model Context Protocol. Its own site says: "**Think of MCP like a USB-C port for AI applications**" and the connector is one plug built on it.
- A skill is different: it teaches the AI how to do a task. A connector gives it access to a tool or data.
- This is part 4 of the brief, May touch, must not, set in the tool: switch on only what the task needs, and start on ask or read-only.

SAY, SLIDE 13: 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.
- Why: words planted in a web page or an email can be taken as an instruction. Anthropic's safety page says that for this to work "**two things must be true at the same time**" and they are that Claude "**can read information outside your trusted boundary, and can perform actions that could compromise the user**" and if you take one away, the attack gets harder.
- **These steps cut the risk. They don't end it.**

BEFORE ANYONE CONNECTS ANYTHING

- 1 **Is it allowed?** Use what your organisation approves. On team plans an owner adds connectors first.
- 2 **Who made it?** Prefer the official or verified one.
- 3 **What can it read and change?** Read the access screen before you approve.
- 4 **Start careful.** Needs approval, Always ask or read-only. Never Allow all actions.
- 5 **Right account.** Your work account for work data, never a personal one.
- 6 **Only what the task needs.** Switch the rest off in that chat, and disconnect when done.

KEEP IT ACCURATE

- Anthropic says custom connectors "**aren't verified**" and Google says it "**does not control, monitor, or secure third-party MCP servers**" so prefer the official or verified one.
- In Claude on Team or Enterprise, a person may see "Request" until an owner adds a connector. On a Google work account, the admin decides which apps are allowed.

KEEP IT ACCURATE

- Anthropic's two conditions are written about Claude. Our rule applies the same idea to any agent.
- Anthropic's own words are that attacks "**become more difficult**" if one of the two is not true, which is not the same as impossible.
- In December 2025 the UK's National Cyber Security Centre warned that prompt injection may never be fully fixed. Spell it Centre.

00:43 9 MIN / SLIDES 14 AND 15 / EXERCISE 2

Sort ten team tasks

Goal: every group picks the right tool for ten made-up tasks, and sets a level for each agent.

SAY, SLIDE 14

Pick the tool first: chat for one answer you use yourself, deep research for many sources, an agent when the next step depends on what it finds, an automation when every step can be written down. **Set each agent's level: the lowest level that does the job, never above what the three questions allow.**

STEPS, SLIDE 15

- 4 MIN** Sort the ten task cards by tool: Chat, Deep research, Agent, Automation. One card fits none.
- 3 MIN** For each Agent card, ask the three questions and write its level.
- 2 MIN** Compare with the next group. Which card did you disagree about?

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.

FOR YOU, NOT THE ROOM The answers. Q1, Q2 and Q3 are the three questions above. Tool cards get no level.

#	TASK	TOOL	Q1	Q2	Q3	LEVEL	WHY
1	Rewrite one paragraph of the team newsletter so it's easier to read.	Chat					One draft that you use yourself. You do any sharing.
2	Find what has been published this year about a new rule in our sector, with a link for every claim.	Deep research					Many sources, read and cited. It only reads and reports: check The facts by opening the links.
3	Every Friday at 4 pm, post the same timesheet reminder in the team channel.	Automation					The same words at the same time, written down in advance. No agent needed.
4	Draft replies to the routine questions in the shared customer inbox, for a person to check and send.	Agent	Y	Y	Y	2 Draft	The job is drafting. Its replies would leave the team and it reads outside email, so a person sends each one.
5	Book one of our own meeting rooms and invite the team to next month's planning day.	Agent	Y	N	N	3 Ask first	It has to act, so it needs level 3. No question caps it, so it could move to 4 after five clean runs.
6	Tidy the team's shared folder: move old files into the archive and delete duplicates.	Agent	N	N	N	3 Ask first	It has to act, and deleting can't be undone easily, so it asks before each delete. It stays at 3.
7	Pay the supplier invoices that arrive by email this week.	Agent	N	Y	Y	3 Ask first	No easy undo, money leaves the team, and the invoices come in by outside email. A person approves every payment in the tool, and checks the bank details against your supplier record, never the email. Never level 4.
8	Every Monday at 8, read the project tracker and list what's overdue, for me to read.	Agent	Y	N	N	1 Read	It only reads and reports: leave the list in the task's results, with no email or message tool. Automation, or a saved view in the tracker, is a fair answer too.
9	Score the 30 applications for the new role and turn down the weakest ten.	None	N	Y	Y	None	Not a job to hand over: decisions about people stay with people. A chat may help draft the scoring criteria.
10	Copy each new entry from the event sign-up form into the team spreadsheet.	Automation					Fixed steps. Outsiders write the entries, but it only copies them: it can't send, pay or delete.

00:52 4 MIN / SLIDE 16

Debrief

Ask these four, in order. Go round the room, or take answers from anyone.

- 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?

THEN

- Read out the answers to cards 7 and 9. Card 7 is the one to remember: money, outside email, no undo. A person approves every payment.

WATCH FOR

- **A rule for everything.** You're not writing the policy today. You're picking one task to start with.

00:56 4 MIN / SLIDE 17

One team habit: your first level 1 task

DO

- Agree three things and write them in the white boxes on slide 17, or on the board: **the task**, a level 1 task that only reads and reports; **its owner**, one person, not the whole team; and **its check**, one of the five checks and how.
- Everyone copies them onto their handout.
- If the room is stuck, offer one of these:

Every Monday, list what's overdue in the project tracker. Owner: the project lead. Check: The facts, open three items.

Before each first call, a cited one-page note on the client's sector. Owner: whoever runs the call. Check: The facts, open every link.

Each month, list the contracts that end in the next 90 days. Owner: the office manager. Check: The facts, open two contracts.

SAY

- Run it at level 1 and check every run. After five clean runs, decide together whether it moves up.
- Close with the three habits: **Brief it. Check it. Keep it safe.**
- You'll send the task, the handout and the one-sheet tomorrow, and check in two weeks.

THE FEEDBACK FORM

One question: "What will you do differently the next time you hand work to an AI agent?" Hand it out as people leave, or put it in the follow-up email.

AFTER THE SESSION

- ☐ Send the follow-up email (next page) the next working day, with the handout and the one-sheet attached.
- ☐ Put the team's first task, owner and check where people will see them.
- ☐ Send the habit check two weeks later. Count the clean runs, and decide together whether it moves up.

Questions people ask

Which assistant should we use?

The one your organisation approves. If you can choose, my recommendation is Claude first, ChatGPT second, Gemini third (page 14).

Isn't this just prompt writing with extra steps?

Wording still helps, but it's now the smallest part. Anthropic's engineers wrote in Sep 2025 that **"the exact formatting of prompts is likely becoming less important as models become more capable"** and with agents, what it may touch matters more.

Can't I just write "ask me before you send" in the brief?

Write it, and set it in the tool too. A sentence in a brief is not a lock, and words planted in a page can claim the opposite.

Our tool has no approval setting.

Then keep that tool at level 1 or 2: it reads or drafts, and a person does every action.

What if it does something odd?

Stop the task. Look where it acted. Disconnect what it used. Tell whoever looks after your AI tools. Nobody is blamed for reporting.

We're on Google Workspace. Does this work?

Yes, in Gmail, Docs, Sheets and Meet. As of Oct 2026 Gemini's Task agent is for personal accounts only.

Claude or Claude Code? Chat, Work or Codex?

Code and Codex are for building software. Everything in this session happens in ordinary Claude, or in ChatGPT's Work (page 14).

Who answers for what it does?

The person who briefed it. Anthropic's own safety page tells Claude users **"You remain responsible for all actions taken by Claude performed on your behalf"** and our rule is the same for any agent.

Should an agent ever decide about people?

No. Decisions about hiring, pay, performance or discipline stay with people (task card 9). Near those decisions, ask whoever advises you on employment law. This guide is not legal advice.

THE FOLLOW-UP EMAILS, ALSO IN THE EMAILS FILE

THE FOLLOW-UP / THE NEXT WORKING DAY

Subject: Briefing agents: our first task, and the one-pager

Hi all,

Thanks for yesterday's session. The handout (its back shows each step in your tool) and the one-sheet on prompt injection are attached.

Our first level 1 task: [the task]. Owner: [name]. Check: [which of the five checks, and how].

The six-part brief: Goal / Done looks like / Context / May touch, must not / Limits / Stop and ask.

The four levels: 1 Read, 2 Draft, 3 Ask first, 4 Alone. Up one level after five clean runs. Down at once after a surprise.

The three questions: Can it be undone easily? Does it leave the team? Does it read outside content?

Two rules: "Ask first" is set in the tool, not typed in the brief. And outside words, no power: an agent that reads the web, outside email or files others sent may not also send, pay or delete without asking.

The five checks: The job, The facts, The steps, The changes, The sign-off.

One question, a line is plenty: what will you do differently the next time you hand work to an AI agent?

[Your name]

THE HABIT CHECK / TWO WEEKS LATER

Subject: Two weeks on: our first agent task

Hi all,

Two weeks ago we agreed our first level 1 task: [the task], owned by [name], checked with [the check].

[Name], three quick answers, a line each:

1. How many runs so far, and how many were clean (all five checks passed, nothing to fix)?
2. What did the checks catch?
3. Should it stay at level 1, move up one level, or go back to a person?

Everyone else: reply with one more task we could brief at level 1 or 2. I'll pick one for next month.

A reminder of the rule: up one level after five clean runs, down at once after a surprise.

[Your name]

Your three tools

For you, before the session. Every line is as of Oct 2026, read on each vendor's own pages on 16 Oct 2026. Features arrive in waves: a person's plan, country and admin settings decide what they see.

MY RECOMMENDATION

Claude first, ChatGPT second, Gemini third.

The tool your organisation approves comes first, for company data. Here is why, next to where another tool wins.

WHY CLAUDE FIRST

- The simplest start to finish: one conversation for quick questions and handed-over tasks, with no mode to choose.
- It asks before acting by default, and Anthropic reports Opus 5.5 is harder to steer with hidden instructions.
- Finished Docs, Slides and Design on every plan, and Claude for Excel cites cells and keeps formulas intact.

WHY CHATGPT SECOND

- ChatGPT Work is a strong agent: a cloud browser for signed-in sites, and tasks that start when an email or message arrives.
- A wide set of everyday features: voice, images, Sites, and ChatGPT for Word even on Free.

WHY GEMINI THIRD

- The best fit when a company already runs on Google Workspace: Gemini in Gmail, Docs, Sheets and Meet.
- But as of Oct 2026 its Task agent, custom apps and skills are for personal accounts, and many features start in the US and in English.

WHERE ANOTHER TOOL WINS, SO SAY SO

- Pictures and video: ChatGPT or Gemini.
- Routines that start when an email or Slack message arrives: ChatGPT Work.
- A company that has standardised on one tool: use that tool.

- Meeting notes in Google Meet, and writing inside Gmail: Gemini.
- Research where you must control the sources: ChatGPT deep research.

THE JOB	CLAUDE	CHATGPT	GEMINI
Quick question, short rewrite, sort a list	Haiku 5.5 or Sonnet 5.5	GPT-6 at Instant	Auto or Flash-Lite
Everyday writing, a summary, a first deck	Sonnet 5.5 at its Default effort	GPT-6 at Medium	Flash
A careful report, a messy spreadsheet	Opus 5.5	GPT-6 at High	Pro, thinking High
Hand over a multi-step task	Opus 5.5, in an ordinary conversation	Work with Default, or GPT-6.1 Sol	Task with Flash or Argon
The hardest, high-stakes work	Fable 5.1 (it uses limits fastest)	GPT-6 at Pro, or GPT-6 Astra in Work	Argon, thinking High

Start with the default. Move up only when the answer misses something that matters; move down for speed, or to save your limit. Not every model is on every plan: Opus 5.5 needs a paid Claude plan, Fable 5.1 needs Max, Team Premium or Enterprise, GPT-6 at Pro needs Pro, Business or Enterprise, and Argon needs Google AI Pro or Ultra.

IN CLAUDE

Since 16 Sep 2026 the app has two places: Claude, for chat and tasks together, and Code, which stays separate. Open Code only to build or change software, a website or a script.

IN CHATGPT

Chat for quick help. Work for anything that ends in a file, a report or a done task. Codex only for software. Chat and Work sit behind one toggle.

IN GEMINI

At work, start with Gemini inside Gmail, Docs, Sheets and Meet. The Task agent is for personal accounts on Pro or Ultra, for now. Antigravity is for software.

The steps for each tool (hand over a task, pick the model, set ask first, connect an app, keep a job in a project, save a skill) are on the back of the handout. The claims about templates, computer use and hidden instructions are the vendors' own.

Checked against

Every product line in this box was read on the vendor's own page on 16 Oct 2026, and is true as of Oct 2026. These products change monthly: check the settings in your own tool. The facts below were checked against the primary sources named, on the dates shown. Made-up examples throughout; the timings are estimates; nothing here is legal advice. Neosaipien is not affiliated with Anthropic, OpenAI or Google.

- 01 **Anthropic.** Claude Cowork and chat are now one Claude, 16 Sep 2026: one Claude for chat and tasks, asking before acting by default, scheduled tasks. claude.com/resources/articles/cowork-is-now-claude. Read 16 Oct 2026.
- 02 **Anthropic.** Connectors, their per-tool approval settings and Allow once: claude.com/docs/connectors/overview. Custom connectors aren't verified: support.claude.com/en/articles/11724452. Connectors and skills: claude.com/docs/extend/overview. Read 16 Oct 2026.
- 03 **Anthropic.** Models and effort: support.claude.com/en/articles/8664678. Claude Opus 5.5, 22 Sep 2026: anthropic.com/news/claude-opus-5-5. Claude Code: claude.com/product/claude-code. Projects: support.claude.com/en/articles/9517075. Skills: claude.com/docs/skills/overview.md. Read 16 Oct 2026.
- 04 **Anthropic.** Use Claude Cowork safely: the two conditions for prompt injection, and who is responsible. support.claude.com/en/articles/13364135-use-claude-cowork-safely. Read 16 Oct 2026.
- 05 **Anthropic.** Effective context engineering for AI agents, 29 Sep 2025. anthropic.com/engineering/effective-context-engineering-for-ai-agents. Read 16 Oct 2026.
- 06 **OpenAI.** ChatGPT Work and Codex: Chat, Work and Codex in one app, describing the outcome, approving important actions. help.openai.com/en/articles/20001275-chatgpt-work-and-codex. Release notes, 9 Jul 2026: help.openai.com/en/articles/6825453. Read 16 Oct 2026.
- 07 **OpenAI.** Connected apps and their permissions: help.openai.com/en/articles/11487775. Models: help.openai.com/en/articles/20001354. Projects: help.openai.com/en/articles/10169521. Skills: learn.chatgpt.com/docs/skills-and-plugins.md. GPT-6 Astra, 3 Sep 2026: openai.com/index/gpt-6-astra/. Read 16 Oct 2026.
- 08 **Google.** Tasks in Gemini, who can use them, sign-ins and confirmations: support.google.com/gemini/answer/17094507. Gemini at work: support.google.com/mail/answer/13952129. Models: support.google.com/gemini/answer/18648722. Read 16 Oct 2026.
- 09 **Google.** Connected apps, write actions and third-party MCP servers: support.google.com/gemini/answer/17209137. Skills: support.google.com/gemini/answer/17094296. Gemini Notebook, 16 Jul 2026, and the release notes: gemini.google/release-notes/. Read 16 Oct 2026.
- 10 **MCP.** What is the Model Context Protocol: the USB-C comparison. modelcontextprotocol.io/docs/getting-started/intro. Read 16 Oct 2026.
- 11 **UK NCSC.** Managing the cyber risk of agentic AI, 20 Aug 2026. ncsc.gov.uk/blogs/managing-the-cyber-risk-of-agentic-ai. Read 16 Oct 2026.
- 12 **UK NCSC.** Mistaking AI vulnerability could lead to large-scale breaches, NCSC warns, 10 Dec 2025. ncsc.gov.uk/news/mistaking-ai-vulnerability-could-lead-to-large-scale-breaches. Reread 16 Oct 2026 and unchanged.
- 13 **Newton, 1990.** The tapped songs: Elizabeth Newton's PhD dissertation at Stanford University, The rocky road from actions to intentions. Checked against it on 14 Oct 2026.
- 14 **Anthropic.** Prompting best practices, its section Be clear and direct: the golden rule behind the ten-second test. Checked on 14 Oct 2026.
- 15 **One-sheet.** Prompt injection: checked against its four sources (UK NCSC, OWASP, Brave and Anthropic, listed on the one-sheet) on 3 Oct 2026, and reread 16 Oct 2026.

FROM ME

Want it run live, or the first agent set up with you?

In the Team workshop I run this live with your team, built around your own tools and tasks. And in Workflow automation I can set up your team's first agent with you: the brief, the level, the setting in the tool and the check.

NEXT, FOR EACH MANAGER

The Agent Playbook. The 20 ready briefs by role, and a "set the brakes" card for each assistant.

FOR THE TEAM'S RULES

The Team AI Policy Kit. The agent rules in writing: what may be connected, who approves, what never goes in.

FOR THE WHOLE ORGANISATION

The Team Pack. Every kit and both workshops in one free download, run as a 90-day programme.

FREE

The Field Guide. Free lessons on handing a task to an agent, what it may touch and reviewing its work, at neosaipien.com/learn/

SERVICES

Team workshop: I run it live. **Workflow automation:** I set up the first agent with you. Also the Leadership briefing, Executive coaching and AI implementation: neosaipien.com/services/

BOOK A CALL

A 20 minute call about your team: neosaipien.com or samson@neosaipien.com