

# What to agree before the agent

# ACTS

## Here is what I would do

- 1 Cancel your 10:05 flight
- 2 Rebook the 14:30 direct
- 3 Email the new plan to you and Jane

**Go ahead**

Change it

I'll do it myself

# Where the research came from

We went through what Google, Anthropic, Microsoft, LangChain, Asana and Notion have shipped in their agent products, alongside the studies published in 2025 and 2026.

Twenty patterns came out of it. **Five sit in the same place: the handover**, where a person and an agent settle what is about to happen. Nothing has run yet.

A minute spent agreeing the plan beats an hour spent undoing the wrong one. **Every serious agent product now does this up front.**

# 1: The plan preview

The agent states what it is about to do. The person accepts, changes it, or takes it back.

## Plain words

"Cancel your 10:05 flight", not system jargon.

## Three ways out

Go ahead, change it, or I'll do it myself.

### Your flight was cancelled

- 1 Cancel the 10:05, refund to your card
- 2 Rebook the 14:30 direct
- 3 Tell the hotel you arrive late

Go ahead

Change it

I'll do it myself

## Always do this when

It costs money, cannot be undone, or someone else sees it.

## Handing it back

Is a normal option, not a failure.

The test: could the person read this and know exactly what is about to happen to them?

## 2: The editable plan

The agent drafts the steps. The person edits them before anything runs.

### Reorder

Move a step earlier.

### Adjust

Change a limit in place.

#### The plan, not running yet

- 1 Find flights to San Francisco under \$400 · direct only · edit
- 2 Use the company discount added by the agent · remove
- 3 Hold it, do not buy  
changed by you

+ Add a step

Start

### Remove

Delete a step it added.

### Why it matters

Otherwise your only option is rejecting all of it.

Microsoft's research agent ships this. Rejecting six steps to fix one is what it replaces.

# 3: The clarifying question

The agent asks instead of guessing.  
The person answers with a tap.

## Two or three

More than that  
and people give  
up.

## Tap to answer

Not a blank box.

Before I start

You said "next Tuesday".

30 September

7 October

How thorough should this  
be?

Quick look, 2 min

Thorough, 15 min

## Ask about

Vague dates,  
unclear scope,  
how good it needs  
to be.

## Do not ask

Anything it could  
look up itself.

Research assistants already do this. It is the cheapest habit on this list to copy.

# 4: The agreed limits

The person sets time and money. The agent says what it will do at the ceiling.

## Show the estimate

Next to the limit.

## Cap the help

How many helpers it may bring in.

| Limits for this job |                       |
|---------------------|-----------------------|
| Time                | 20 minutes            |
| Budget              | \$4.00, likely \$1.80 |
| Helpers             | up to 5               |
| At the limit        | stop and show me      |

## Say what happens at the limit

Stop, or carry on and tell you.

## Anthropic's problem

Its own agent pulled in dozens of helpers for simple jobs.

A thorough research job can cost fifteen times a normal question. That belongs on screen.

# 5: The autonomy setting

The person sets how far it may go alone, job by job. The agent stays inside it.

## Watch

Points things out.  
Plans nothing.

## Suggest

Plans it. Runs  
nothing.

| How far it goes alone |             |
|-----------------------|-------------|
| Book a meeting room   | Go ahead    |
| Draft a reply         | Check first |
| Email a customer      | Suggest     |
| Spend money           | Watch       |

## Check first

Prepares it. You  
say go.

## Go ahead

Acts inside limits  
you set, and  
explains  
afterwards.

One switch for everything is why people turn a whole feature off after a single bad experience.

# Five things to settle first

- 1 The plan preview.** The agent says what it will do, in plain words.
- 2 The editable plan.** The person changes a step without rejecting all of them.
- 3 The clarifying question.** The agent asks rather than guesses.
- 4 The agreed limits.** The person sets time and money, and what happens at the ceiling.
- 5 The autonomy setting.** The person sets how far it goes alone, job by job.

All five sit in the handover, before any work starts. That is why they are cheap to add now and expensive to add later.

Which one is missing from the AI tool you use most?

**Name it in the comments.**

If you change one thing

# Start with the plan preview

## Why this one

The agent shows three lines in plain words and three buttons: go ahead, change it, I'll do it myself. One screen, and the other four sit on top of it.

Victor Yocco argues this should be mandatory for anything that **costs money, cannot be undone, or is visible to someone else.**

Go deeper

# The whole pattern library

## On the site

Seven families, forty-one patterns.  
What each one prevents, and the  
product or the paper it came from.  
Free, and no sign-up.

[thinkbigleaders.com/resources/agent-patterns](https://thinkbigleaders.com/resources/agent-patterns)

## Or talk it through

Curious, and nowhere near a decision?  
Good. Have a use case in mind? Also  
good. Already building one, or want  
agents inside your own product? Bring  
what you have.



**Thirty minutes with Think Big Leaders.**

Scan the code to pick a time.