

BUILDER · 6 MINUTE READ

Agent-to-agent workflows: hand-offs, rules and checkpoints

When agents pass work along, clear hand-off notes, firm limits and a person's approval before any real action keep work safe and traceable.

What it is

An agent-to-agent workflow is a chain of work where one AI agent finishes its part and hands the result to the next. An agent is an AI assistant that carries out a task in steps and can use tools. The hand-off is the moment of risk: if the note passed along is vague, the next agent guesses, and guesses grow into errors.

A good workflow has four parts. A hand-off note in a shared format, so every agent knows what it received. Clear inputs and outputs for each step. Checkpoints, where a person reads the work and approves it before it moves on. And a log: a record of what each agent did and when.

Think of a hospital shift change. The nurse going off duty does not say "all fine". They use a set format: what was done, what to watch, what is due next. The next nurse can act without guessing.

Why it matters

- Many errors in a chain of agents start at a hand-off, not inside a single step.
- Checkpoints keep a person in charge of decisions that matter.
- A log lets you find out what went wrong and fix the right step.
- Limits on actions protect you if an agent misreads its instructions or is fed misleading content.

How to do it

1. Map the chain: each step, which agent does it, and what it passes on.
2. Agree one hand-off format for every step, such as: task, inputs used, output, open questions, confidence.
3. State for each agent what it may do (read, draft, suggest) and what it may never do.
4. Use the pattern draft, review, approve, act: one agent drafts, another reviews, a person approves, and only then does anything happen.
5. Put a checkpoint, a person's approval, before any action that sends, pays, deletes or changes a record.
6. Keep a log of every hand-off and every approval, with the time and who approved.
7. Set limits on rounds and time, and tell agents to stop and ask a person when unsure.
8. Test the full chain with made-up data before any real data goes through it.

Example

Imagine a finance team that handles supplier invoice queries. Before, one agent drafted replies and a second was told to "send it if it looks fine". One reply promised a refund nobody had agreed.

After, the chain looks like this, with a hand-off note between each step:

```
HAND-OFF NOTE
Task: [what was asked]
Inputs used: [documents or records read]
Output: [the draft or result]
Open questions: [anything unclear or missing]
Confidence: [high / medium / low, and why] (a hint, not a measure)
Next step: [review / person approval]
```

Agent one drafts the reply. Agent two checks it against the invoice and the payment rules, and adds its findings to the note. A person reads both, edits if needed, and presses send. The agents can draft and check, but they cannot send or

change payment records.

Common mistakes

- **Checkpoints after the action.** A review after sending is too late. Put the approval before anything leaves or changes.
- **Agents with too many powers.** A drafting agent does not need to send email. Remove that access.
- **No record.** Without a log, you cannot tell which step caused a problem. Keep one, even a shared document.
- **Trusting content from outside.** An email or web page can contain text that tries to instruct an agent. Tell agents to treat such content as information only. Instructions help but are not a reliable defence; the real protection is limited access and a person approving actions.

Try it this week

1. Pick a two-step task, such as drafting and checking a reply to a customer.
2. Copy the hand-off note above into a document.
3. Use your approved assistant to draft a reply to a made-up question, then fill in the note.
4. Open a new conversation, paste the note, and ask the assistant to review the draft and add its findings.
5. Note where the hand-off note helped and what was missing.

Going further

- MCP (Model Context Protocol) is an open standard for connecting an agent to tools and data, such as files or a calendar. It covers agent to tool.
- A2A (Agent2Agent) is an open protocol that lets agents built on different systems find each other, share tasks and exchange results. It covers agent to agent.
- Neither standard decides what an agent may do. Permissions and approvals remain your choice.
- Review the log regularly and tighten the hand-off format where questions repeat.

Key words

- **Hand-off note:** a structured message passed from one agent to the next, in the same format every time.
- **Checkpoint:** a point where a person reads the work and approves it before it moves on.
- **Log:** a record of what each agent did, what it passed on and who approved it.
- **MCP:** an open standard for connecting AI agents to tools and data.
- **A2A:** an open protocol for AI agents on different systems to communicate and share tasks.