

Armstrat methodology

Cover photo: Waranont (Joe), Unsplash

The Agent Authority Ladder

Armstrat

Business Transformation Office (BTO™)

What this booklet is for

A one-page method that tells a founder or CEO what each AI agent may do, up to what limit, and who can stop it.

AI agents now arrive inside the software a business already pays for: the office suite, the accounting system, the helpdesk. Each one holds some authority, whether or not anyone decided to give it.

This booklet sets that authority on purpose. It needs no security software, only a spreadsheet, a stopwatch and a named owner for each agent.

Read it in one sitting. Run it over four weeks with the people who own each process.

Contents

The decision 3

The ladder in six parts 4

The five rungs 5

Limits and reversibility 6

Stop drill and triggers 7

Worked example 8

Where we say no 9

The agent authority sheet 10

About Armstrat 11

The decision this answers

How much authority should each AI agent in the business hold, and who can take it away?

The test: for every agent, can you name its owner and say how many minutes it takes to stop it?

Every business has a delegation of authority for people, even if it is informal: who may sign a purchase order, approve a refund or change a price, and up to what amount. AI agents rarely get the same treatment. They inherit the access of whoever connected them, often an owner or an administrator.

In November 2026 California's Attorney General sued OpenAI, alleging that its agents escaped control in testing and that one run took about two hours to stop after it was detected. Those were lab incidents, not business deployments. The lesson travels anyway: if the maker needed hours, do not assume you need minutes.

The ladder settles the decision on one page per agent: what it may read, draft or do in each system, the limits on what it does, whether its actions can be undone, and who stops it, tested against a clock.

A kill switch is a brake, not a control plan.

The ladder in six parts

Run the parts in order. Each one produces a line on the worksheet.

1	**Agent register**	List every agent, including ones bundled into software you use. Give each a named owner and a purpose.
2	**Five rungs**	For each system it can reach, set one rung: read, draft, act with approval, act within a limit, never.
3	**Limits**	Write the spend cap, the most records it may change in a day, and the credentials it holds.
4	**Reversibility test**	Can the action be undone within an hour, by whom, from what backup? If not, it needs approval.
5	**Stop drill**	Before any write access, the owner stops the agent against a stopwatch. Repeat every quarter.
6	**Re-authorization**	A new model, tool, permission or an incident sends the agent down a rung until it is reviewed.

The register and the five rungs

Every agent gets a named owner first, then one rung for each system it can reach. Each rung has a test to pass before the next.

Rung	What the agent may do	Example	Before it climbs
Read	Look at data and answer questions. Changes nothing.	Lists invoices that are over 30 days late.	A timed stop drill. Drafting is write access.
Draft	Prepare work as a draft that a person reviews and sends.	Drafts a reply to a complaint for the support lead.	A written process and a named approver.
Act with approval	Take each action only after a named person approves it.	Raises invoices that the bookkeeper approves one by one.	Caps the software enforces and a passed one-hour undo test.
Act within a limit	Act alone within written caps. Every action is logged and reversible.	Recodes expense lines under $250, up to 40 a day.	Nothing. This is the top rung, and any trigger sends it down one.
Never	Hold no access to the system at all.	Changing supplier bank details or releasing payments.	Only a written decision by the CEO moves it.

The register comes first. An agent with no named owner gets no rung at all: switch it off until someone owns it.

Limits and the one-hour test

Parts 3 and 4 decide how far an agent may go on its own, and what happens when it goes wrong.

Give each agent its own login. You can then revoke it without locking out a person.

Limits turn a rung into a rule. For every system where an agent can act, write three caps: the largest amount it may commit in one action, the most records it may change in a day, and the exact credentials and systems it may hold. If the software cannot enforce a cap, the agent belongs one rung lower.

The reversibility test decides whether an action can be delegated at all. Take one real action on a test record and try to undo it. Write down who undid it, from which backup or version history, and how long it took. If the answer is more than an hour, or nobody, the action stays at act with approval.

Some actions fail by definition: a sent email, a released payment, a message to a customer, a record deleted with no backup. Those never move above act with approval, however good the agent is.

If nobody can undo it within an hour, a person approves it first.

The stop drill and re-authorization

Parts 5 and 6 test the brake before you need it, and take authority back when something changes.

Our ceiling: if the stop takes longer than 30 minutes, the agent goes back to draft.

Before an agent gets any write access, its owner stops it during business hours with a stopwatch running. Revoke the credentials or disconnect the app. Record the minutes to a safe state, what was still in flight, and who else had to help. Repeat the drill every quarter.

OpenAI now expects responders to pause the activity on its most severe alerts unless, within 30 minutes, they can show the alert is a false positive. A business with simpler agents should be at least as fast.

Authority is not permanent. Four events send an agent down one rung until the owner reviews it and reruns the drill: the model changes, the agent gains a new tool, it is given a new permission, or it is involved in an incident, however small.

Whoever must stop the agent on a Friday afternoon should have done it once already.

A worked example

A hypothetical business, not a client. It shows how the ladder reads once it is filled in.

Situation

A distributor with 80 staff connects an AI assistant to its accounting system to draft invoices and reconcile bank lines. The connection uses the finance director's login, which can also pay suppliers.

How the ladder is set

The finance director is the named owner. Bank feeds: read. Journal entries: draft. Invoices: act with approval. Expense coding under $250: act within a limit. Payments and supplier bank details: never. The agent gets its own login.

The stop drill

The owner revokes the agent's login on a weekday afternoon with a stopwatch running, notes the minutes and the drafts still in flight, and books the next drill for the following quarter.

Where it says not yet

The software also offers to pay small supplier bills. Not yet: payments stay at never until the payment run is written down and two stop drills have been timed.

Illustrative only. The business, its size and its limits are invented to show the method. Nothing on this page is a client case or a result.

Where the answer is no, or not yet

Five cases where we hold an agent back, however good the demo looks.

Nobody will own the agent

A rung without an owner is a guess. Switch it off until someone takes it.

The process is not written down

If nobody can say how the work runs, nobody can say what the agent may do.

The stop has never been timed

No access to money, customers or production records until the drill has run against a clock.

You want it to run unattended

Not yet, for most businesses. If you cannot trace what it did yesterday, it does not act alone.

A kill switch is the whole plan

A switch stops the next action. It cannot take back the last one.

The agent authority sheet

One sheet per agent. Fill it in by hand with the owner and keep it with the register.

Agent, purpose and owner

What it is, what it is for, and the one person accountable.

Systems and rungs

For each system: read, draft, act with approval, act within a limit or never.

Limits

Spend cap, records per day, credentials and systems held.

Reversibility

Undone within an hour? By whom, from what backup?

Stop drill

Date, who stopped it, minutes to a safe state, what was in flight.

Next review

Next drill date, or sooner on a trigger.

Triggers that send the agent down a rung: a model change, a new tool, a new permission, an incident.

About Armstrat

Armstrat is a Business Transformation Office (BTO™): a specialized, modern firm that embeds forward-deployed engineers (FDEs) inside your business and focuses solely on delivering outcomes.

Our methodologies are the working methods our FDEs use inside client businesses. We publish them free, with no email wall, so your own team can run them with or without us.

How to get help

Read the booklets at armstrat.com/downloads.

Take the free AI Readiness Assessment at armstrat.com/ai-readiness-assessment.

Write to hello@armstrat.com if you want our FDEs to run a method with your team.

armstrat.com/downloads

Armstrat
Business Transformation Office (BTO™)
hello@armstrat.com · armstrat.com

You may use, print and share this booklet inside your business. Keep the Armstrat name and copyright notice on every copy. To republish, teach from or resell it, ask Armstrat first.