

The CV Method

Set up your first AI agent.

A plain-English guide to standing up a small, reliable AI “worker” that does one job well and never goes rogue — using tools you already have. This is how I actually set one up, minus the hype.

BY CORINA • THE CV METHOD • @THE_CV_METHOD

Current as of Aug 2026 — verify tool details, this space moves fast.

An “agent” sounds mystical. It isn’t.

It’s an AI model given one clear job, the right context, a few tools it’s allowed to touch, and guardrails — running with a human in the loop. Small and specific beats big and magical every time. Everything below rests on one rule.

RULE ZERO

It proposes. You approve.

Your first agent should never post, send, spend, or deploy on its own. It drafts and suggests; a human ships. That single rule is what makes an agent safe to actually use.

STEP 1 • PICK YOUR STACK

Four simple pieces

Here’s what I use and the general equivalent — swap in whatever you already have.

PIECE	WHAT IT DOES	I USE	SWAP FOR
The brain	An AI model that can follow instructions and call tools.	Claude	Any strong tool-use LLM
The runner	Where the agent actually runs and takes its actions.	Claude Code / an agent runner	A simple script · a workflow tool
Its tools	The connectors it’s allowed to touch (files, calendar, a doc, an API).	MCP connectors	Plugins · a few scoped API keys
Its memory	A place for its instructions + the files it reads and writes.	A notes vault (files)	Plain folders · a doc

Give it the fewest tools that get the job done. Every extra permission is a new way for it to surprise you.

STEP 2 GIVE IT ONE JOB

NARROW BEATS MAGICAL

Scope it to a single task

Don't build "an assistant that does everything." Build "the thing that drafts my weekly update," or "the thing that reads a receipt into a row." Write the job as a short brief: its role, the goal, the constraints, and what "done" looks like.

If you can't describe the job in two sentences, it's still too big. Split it.

STEP 3 GIVE IT THE RIGHT CONTEXT

CONTEXT IS THE WHOLE GAME

Point it at what it needs — and nothing more

An agent is only as good as what it can see. Hand it the specific docs, examples, and data the job needs (a style doc, last week's update, the folder of receipts) — not your entire world. Curated context makes it reliable; a firehose makes it wander.

Keep a short "how we do this" doc the agent reads every run. That doc is where most of your quality lives.

STEP 4 SET THE GUARDRAILS

WHAT IT MAY AND MAY NOT DO

Write the fence before you open the gate

Spell out the limits in its instructions: it drafts, never publishes; it asks before anything irreversible; it stays in its lane (only the tools and files it needs). Give it the least access that works, and a clear “when in doubt, stop and flag me.”

The goal isn't an agent that can do anything. It's one you can trust with a small, real thing.

STEP 5 TEST BEFORE YOU TRUST IT

EARN THE AUTONOMY

Run it by hand on real cases first

Before you automate anything, run the agent manually on a handful of real inputs. Read every output. Where it's wrong, that's not failure — that's the instruction or the context telling you what to tighten. Repeat until it's boringly consistent.

Only after it's reliable by hand should you put it on a schedule — and even then, add a fail-safe so it flags you instead of failing silently. It still never auto-posts.

The one-job brief

Fill in and hand to your agent.

You are my [role] agent. Your ONE job: [the single task – e.g. "draft my weekly update from these notes"].

Context you may use: [the specific docs/files – not everything].

Guardrails: you draft only – never send, post, spend, or publish; ask before anything irreversible; use only [these tools]; if you're unsure, stop and flag me.

"Done" looks like: [describe the finished output].

Do the task now on this input: [paste the real input].

KEEP IT HONEST & SAFE

- You approve before anything ships
- No invented results — read the outputs
- Respect terms of anything it touches
- Least access — only the tools it needs
- Keep keys & secrets out of prompts/logs
- A human owns the irreversible actions

Work down this list once

- | | |
|---|--|
| ☐ Write Rule zero: it proposes, you approve. | ☐ Pick the 4 pieces (brain, runner, tools, memory). |
| ☐ Name ONE job in two sentences. | ☐ Write the brief: role, goal, constraints, "done." |
| ☐ Give it the specific context — a "how we do this" doc. | ☐ Set guardrails + least access. |
| ☐ Run it by hand on real cases; tighten. | ☐ Only then schedule it — with a fail-safe, never auto-posting. |

Want a fleet of these, built for your business?

One reliable agent is a great start. A set of them running your back office — safely, with the guardrails — is the thing I build for people.

hello.thecvmethod.com →