
A NO-CODE GUIDE FOR UK SMALL AND MEDIUM-SIZED BUSINESSES

Set Up Your Own AI Agent

How to run an AI agent on your own machine with Hermes. What you need, what it costs, and the one job to start with.

AUTHOR

Ted Milton
Founder, Expedient AI

EDITION

July 2026
Version 1.0

FORMAT

A4 print edition
11 pages, worksheets included

Expedient AI Ltd

INDEPENDENT · BRISTOL · UK

1 THE COLLEAGUE THAT WORKS THE NIGHT SHIFT

Not a chatbot. A colleague that works the night shift.

What Hermes actually is, and why that distinction matters for a small business.

A chatbot sits in a window and waits for you. You ask, it answers, then nothing happens until you ask again. An agent is different. It is a piece of software that runs on your own computer and can be given a recurring job, to read something, decide something, draft something, send something, on a schedule or when an event happens, without you opening a window.

The point of an agent is the bit that happens while you are not looking. A chatbot is a colleague you have to ring. An agent is a colleague who turns up at 8am with the work already done.

How it works, in plain terms

You write the job in plain English. You give it a schedule, a list of what it can read, and a list of what (if anything) it is allowed to send. At the scheduled time, the agent reads the input, decides what to do, drafts the output, and either sends it or holds it for you to review. It does this every day, every hour, or every time an event fires, until you stop it.

What it is not

It is not magic. It is a script with judgement and a schedule, and you remain accountable for what it sends. If the agent drafts a reply to your best customer, and the reply is wrong, you are the one who wrote the instruction. That is the trade. The rest of this guide is about how to manage it well.

The trade. An agent saves you time by doing the work you would otherwise have to remember to do. In exchange, you give it a job it cannot get wrong, and you check what it has done. The art is choosing the job.

2 THE HONEST CHECKLIST

What you need before you start.

Five things. If you cannot tick all five, wait.

Before you install anything, work through this list. An agent is not a piece of software you evaluate in isolation. It is a small system that needs a place to live, a way to pay, and a reason to exist. If you cannot tick all five, the install will go fine and the agent will sit doing nothing useful within a month.

The five requirements

- **A computer that can stay on.** A laptop that sleeps will miss its jobs. A desktop left on, a small home server, a mini PC on a shelf, or a cheap cloud box all work. The first option is free, the last is four to ten pounds a month.
- **An internet connection.** The agent calls the AI provider over the public internet. A standard business broadband line is fine. It does not need to be fast.
- **One AI provider account and API key.** An API key is a password that lets the agent talk to a model like Claude or GPT. You create it on the provider's website, paste it into the agent setup once, and it is stored locally on your machine. You are billed by usage, typically a few pounds a month for light use.
- **Thirty to sixty minutes of your attention, once.** The install, the setup wizard, and the first health check together take under an hour. That hour is the only one that has to be uninterrupted. The rest of the work is small and spread out.
- **One specific, repeated job you can describe in a sentence.** "Every weekday at 8am, summarise my inbox." "Every Friday afternoon, draft a sales briefing from the CRM." "Every Monday, check these five URLs and tell me what changed." If you cannot finish that sentence, you do not have a job yet. You have a wish.

STOP HERE If you cannot name the one job in a sentence, stop here and read Edition 01 first. An agent with no clear job is just an expensive way to generate text. The rest of this guide assumes you have ticked all five.

3 THREE STEPS. ONE AFTERNOON.

The install, in plain English.

No terminal required. No code to write. The whole thing is more like installing a desktop app than setting up a server.

If you have ever installed a piece of desktop software, you can install Hermes. The steps below describe what happens at a conceptual level. The exact commands change as the software evolves, but the shape has been the same for a year and is unlikely to change soon.

Install Hermes. One command in a terminal, or a desktop installer if you would rather avoid the terminal. The package is open-source, downloaded from the project's own website, and installs into a folder on your own machine. It is the same artefact whether you run it on a Mac, a Windows PC, or a Linux server.

Run the setup wizard. The wizard asks which AI provider you want to use (Claude, GPT, or a local model), then asks for the API key from the previous page. The key is stored in a configuration file on your machine, not sent to a stranger's server. You can read the file, back it up, and delete it whenever you like.

Run the first health check. The agent sends a test prompt to the model and confirms a reply comes back. If it does, the agent is live. If it does not, the wizard tells you why (usually a misspelled key, or the provider's billing page needs a top-up).

Where your data lives. The agent's memory, configuration, job instructions, and history live in a folder on your own machine, at `~/hermes` on Mac and Linux, or in a matching folder on Windows. You own it. You can back it up. You can delete it. You can open it in a text editor and read every line. That is the opposite of a SaaS tool that holds your data hostage. See Edition 03 for the data principles behind this.

What to do if you get stuck

The error message in the wizard is usually the answer. Read it. If it does not make sense, copy the last line into a search engine. Most install problems have been seen and written about by someone else. If you would rather not debug, this is also a sensible place to ask for help.

4 ONE JOB, ON A SCHEDULE

Your first agent: one job, on a schedule.

The single most important page in this guide. Read it twice if you have to.

An agent becomes useful when it has one job, on a schedule, with a clear output. The mechanism is a *cron*, a clock that wakes the agent at a time you choose. Pick the simplest useful job you can describe, set the schedule, write the instruction, and pick a delivery channel. That is the whole thing.

WORKED EXAMPLE***The morning inbox summary*****The job, in one sentence**

Every weekday at 8am, look at the unread messages in my inbox from the last 24 hours, group them by what they need (a reply, a decision, a booking, junk), and email me a six-line summary before I open my laptop.

The schedule

Weekdays 08:00 UK time. The agent wakes at 8am, does the work, delivers the summary. It does not run at weekends.

The instruction (the part to write carefully)

Read the inbox. For each unread message from the last 24 hours, decide which of four groups it belongs to: *reply* (someone is asking me a question), *decision* (someone is asking for a choice), *booking* (a calendar or meeting request), or *junk* (newsletters, marketing, notifications). For each group, list the sender and the topic in one line only. End with a single line that counts the total unread and the total worth opening. Output as plain text. Six lines per group maximum.

What the agent must NOT do

Never reply on my behalf. Never mark anything as read. Never move messages between folders. Never open attachments. Never send anything to anyone. The job is to read, group, and email me the summary.

The delivery

Email the summary to ted@example.co.uk with the subject "Morning inbox - [date]". Time the delivery so it arrives in my inbox by 8:05am.

DRAFT ONLY For the first two weeks, have the agent draft only and send the draft to you. Do not let it act on your behalf until you trust the output. This is the permission ladder from Edition 06, rung two. Earn rung three by reading what it produces, every day, for two weeks.

The instruction above is the part to write well. For the structure of a reliable instruction, see Edition 04: the six-part prompt (role, context, task, format, constraints, examples) is what makes the difference between an agent that is useful on Monday and an agent that is still useful on Friday.

5 THE CONNECTIONS THAT MATTER

Connecting it to the tools you already use.

An agent is only as useful as what it can reach.

Out of the box, an agent can read what you put in front of it and write text. That is enough for some jobs (the morning summary on the previous page reads the inbox through a connected email account). For most other jobs, you need to connect the agent to the systems you already use. The list below covers the common ones. Each row is a connection, what it unlocks, and the risk to keep in mind.

CONNECTION	WHAT IT UNLOCKS	WATCH OUT FOR
Email	Read unread messages, draft replies, flag what needs you.	Never let it send on your behalf until you have audited the drafts for a week.
Calendar	Read free/busy, propose meeting times, draft agenda reminders.	Booking meetings without confirmation is a fast way to annoy people.
A folder of documents	Summarise a long contract, compare two versions, answer questions about a folder.	Be sure the folder is the one you meant. Documents can be sensitive.
A website	Watch a page, alert when it changes, summarise new content.	Some sites change layout daily and break the watch. Pick stable pages.
CRM or spreadsheet	Read rows, group, summarise, draft commentary.	Never grant write access to the CRM without a whitelist and a daily review.
Filesystem	Move, rename, copy or create files in a nominated folder.	Restrict to one folder. Never the whole drive.

Read first, write last. Connect read access first, write access last, and only after the agent has drafted correctly for a week. This is least privilege, the same principle as the permission ladder from Edition 06. The agent that can read everything and write nothing is harder to misuse than the one that can do both.

6 AN HONEST BREAKDOWN

What it costs, honestly.

API fees, the machine it runs on, your time. And the math that decides whether it is worth it.

The headline cost of an agent is small. The total cost, including setup and the attention you have to keep paying it, is not. Be honest about both before you build. The table below is what you should expect to write down when you are weighing the decision.

COST LINE	MONTHLY	NOTES
AI model usage (light)	£3 to £10	A few hundred requests a month. One small workflow.
AI model usage (moderate)	£10 to £40	A few thousand requests. Two or three workflows.
AI model usage (heavy)	£40 and up	Tens of thousands of requests. Worth checking whether a cheaper model would do most of the work.
The computer it runs on	£0 to £10	Your existing machine, left on, is free. A small always-on cloud box is four to ten pounds a month.
Your setup time	One afternoon	Amortise it. An agent that runs for a year costs you a few pounds a month for the setup.
Your weekly attention	10 to 30 minutes	Reviewing what it produced, tightening the instruction, catching drift. This is the one line that is easy to underestimate.

Compare to the human task, not to zero

The right comparison is the cost of the human task it replaces, not the cost of doing nothing. An agent that saves a person two hours a week at a loaded cost of thirty pounds an hour is worth two hundred and forty pounds a month. Spending twenty pounds a month and twenty minutes a week on it is a clear win. Spending twenty pounds a month to automate a task nobody minds doing is a hobby, not a business decision. The break-even calculation from Edition 01 applies here.

Two costs people forget. The time you spend tuning the instruction in the first month, and the time you spend paying attention after. Both are real. Both are why the first agent should be small.

7 FOUR FAILURES, EACH WITH A CONTAINMENT

What can go wrong, and the guardrails.

Agents fail in predictable ways. Knowing the categories in advance lets you build the right guardrail.

None of the failure modes below are exotic. All of them are preventable. All of them have been seen in production in the last year. None of them should put you off building an agent, but each one should make you build the containing guardrail before you turn the agent on.

FAILURE MODE	CONTAINMENT
It sends the wrong thing	Nothing goes out without your approval until you opt in. Even after opt-in, only to a recipient whitelist you have written yourself. Anything outside the list pauses for human review.
It runs up a bill	Set a monthly spend cap at the provider (most have one in the billing page). Set a per-run token limit inside the agent. Both. Belt and braces.
It silently stops working	A weekly "are you alive?" check. The agent emails you on Friday if a scheduled job has not run. If you do not get the check, the agent is not running. Investigate before Monday.
It drifts and starts doing the job badly	Review a sample of its output each week for the first month. Then monthly. Five random outputs, every Friday. Ten minutes of attention catches drift long before it matters.

The monitoring is the job. An unmonitored agent is not an employee. It is a liability. The most expensive sentence in this guide is "I set it up six months ago and have not looked at it since." The five minutes a week is what makes the agent useful and safe. If you cannot pay that, the task is not a candidate.

8 A CLEAN DECISION FRAME, NOT A PITCH

When to do it yourself, and when to call us.

Both paths are sensible. The right one depends on the job, the risk, and the afternoon.

Doing it yourself is the cheaper option and the slower one. Asking someone to help is the faster option and the more expensive one. Neither is better in the abstract. The frame below picks the right one for the situation you are in.

DO IT YOURSELF

You are comfortable following written steps. The job is low-risk: drafts, summaries, watches, briefings. Nothing the agent sends goes out without your review. You have the afternoon.

Cost. Your time. Twenty to forty pounds a month, plus the weekly attention.

What you own. The agent, the configuration, the maintenance, the audit.

GET HELP

The job touches customers or money. You need the agent connected to three or more systems. You want it maintained by a named person who knows your business. Or you simply do not have the afternoon, and the job is worth doing this month rather than next quarter.

Cost. Setup and a monthly retainer. Quoted after the scoping call.

What you own. The decision. The instructions. The right to walk away.

What Expedient AI does, in plain words

If you ask, we set the first agent up with you against a real workflow, put the guardrails from page 8 in place, and maintain it with a named person and a monthly review. We will tell you on the scoping call if this guide is enough and you do not need us. We will tell you the price after the scoping call, not before. Both are policy, not sales technique.

You can read more, and start a conversation, at expedientai.co.uk.

9 THE FIRST SEVEN DAYS

The first-week checklist.

A printable list. Tick it as you go.

Day one is the install and the first run. Days two to seven are the boring part: read what the agent produced, notice where it was wrong, tighten the instruction. By the end of the week you will have a clear view of whether the agent is doing the job or whether the job was wrong.

DAY 1

- Install Hermes on the machine that will stay on.
- Run the setup wizard. Pick the AI provider. Paste the API key.
- Run the first health check. Confirm the agent can reach the model.
- Write the one-job instruction. Use the structure from Edition 04.
- Set the schedule. The one from the morning inbox example is a sensible starting point.
- Set the spend cap at the provider. Lower than you think is necessary.
- Keep the agent on draft-only. Nothing sends today.

END OF WEEK 1

- Is the output good enough to keep? If yes, draft a second iteration of the instruction and tighten the categories.
- If the output is good enough, consider whether one action can move from draft to send, behind a recipient whitelist.
- If the output is not good enough, the job was wrong, not the tool. Pick a smaller job and start again.
- Set the Friday weekly review on the calendar. This is the audit that keeps the agent safe.

WEEK 1, DAYS 2 TO 7

- Every day, read what the agent produced. Even if it is short.
- Note where it was wrong. One sentence per day is enough.
- Tighten the instruction to remove the wrong case. Do not add new cases yet.
- Keep the agent on draft-only. No outbound actions this week.
- Check the spend. Confirm it is on track.
- Check the schedule. Confirm the agent ran at the time you set.
- Sit with the discomfort of not having sent anything. This is the part most people skip.

ONE JOB Resist the urge to add a second job until the first has run cleanly for a month. One good agent beats five neglected ones. The first agent is rarely the last. Treat it as a working prototype, not a finished product.

10 COLOPHON

The series and how to reach us.

Seven field guides for UK SMEs. One Bristol-based company behind them.

About this guide

Written by Ted Milton, founder of Expedient AI. Part of a series of seven field guides for UK small and medium-sized businesses. Independent, plain-English, written by a practitioner who has built the agents covered in the other guides.

The series

01 · Where AI Actually Helps

Choose one worthwhile use case, scope a 30-day pilot and estimate the return before buying software.

02 · The 30-Day AI Pilot

A day-by-day guide to running the pilot you scoped in edition 01.

03 · Your Data, Your Rules

AI and privacy in plain English for UK businesses handling personal data.

04 · Writing AI Instructions That Actually Work

The six-part structure that turns vague prompts into reliable outputs.

05 · The AI Vendor Interrogation Kit

Eight questions to ask before signing a contract with any AI vendor.

06 · Your First AI Employee

A non-technical guide to autonomous AI agents for small businesses.

07 · Set Up Your Own AI Agent (this guide)

A no-code guide to running an AI agent on your own machine with Hermes.

Colophon

PUBLICATION

Set Up Your Own AI Agent

Edition 07 · July 2026 · Version 1.0

Author: Ted Milton, Expedient AI Ltd

CONTACT

Expedient AI Ltd

expedientai.co.uk

You can set all of this up without contacting us. If you would rather have a named person build the first agent and keep it running, that is what Expedient AI does. We will tell you on the first call if the guide is all you need.