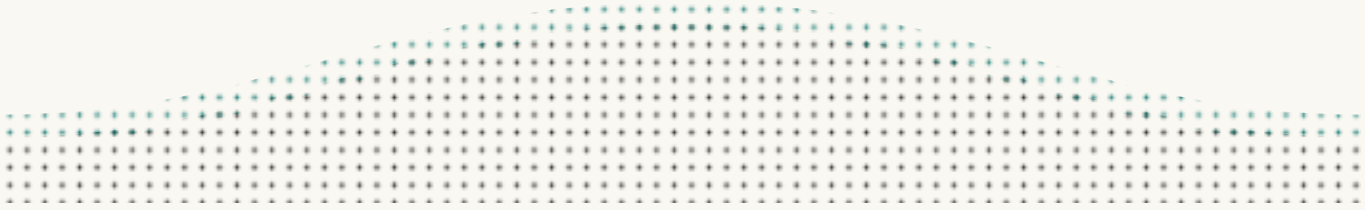
Spot the work worth automating

Most people try AI on the wrong task first, get a mediocre result, and decide the tool isn't ready. The tool was fine. The pick was wrong. Three questions fix that permanently.



README

What this is. Three copy-paste prompts that turn one week of your real work into a ranked shortlist of what to hand over — plus the one task you should keep, and why.

How to use it. Work page 3 in order with your own calendar open. Everything in a bordered box is copy-paste ready. Twenty minutes.

The running example. Throughout, you're an operations manager at a mid-size company. Your manager is Priya. A vendor onboarding rollout is in week three of six, and reconciliation lands every month.

Works with Claude, ChatGPT, Gemini, or Copilot — whichever your company already approved.
Every prompt in this guide is selectable text. Copy it straight out of the PDF.

Three questions, one verdict

Run any task through these in about thirty seconds. Three yeses and it's worth building once. A single no and you do it by hand — and that part matters as much as the first.

QUESTION 1

Does it come back?

Weekly, monthly, or every time a certain kind of request lands. One-offs never pay back the setup cost.

→

QUESTION 2

Same shape every time?

The details change completely. The structure doesn't — same inputs, same format out, same audience.

→

QUESTION 3

Bottleneck, or value?

Are you copying, reformatting and chasing? Or deciding something only you can decide?

a **Question three is the one people skip.** It's uncomfortable, because it's easier to believe all of your work requires your judgment. Most of it doesn't, and being honest about that is exactly what frees up the hours.

b **Hard is not the same as high-judgment.** Plenty of mechanical work is genuinely hard — long, fiddly, easy to get wrong. The test isn't difficulty, it's whether the hard part is *effort* or *judgment*. Effort you can hand over. Judgment you keep.

c **Start boring, not important.** Importance and repetitiveness are different axes. Pick the boring one first, because if it breaks, nothing bad happens.

Quick reference

COMMON TASK	VERDICT	WHICH QUESTION DECIDES IT
Weekly status update	Build it	All three — the textbook candidate.
Meeting notes into actions	Build it	Q3 — transcribing, not deciding.
Monthly or board reporting	Build it	Q2 — the structure never moves.
Reconciling numbers	Build it	Q3 — comparing is mechanical.
Performance review	Keep it	Q3 — your context is the product.
One-off deep analysis	Keep it	Q1 — it never pays back.

Three prompts, in order

Run these back to back in one conversation. Each feeds the next, so don't skip ahead — the inventory is what makes the scoring honest.

1 · Build the inventory

COPY-PASTE · INVENTORY

```
Here is everything I did at work last week, pulled from my calendar and my
sent folder: [paste]

List each distinct task on its own line. Group anything that is obviously the
same task repeated. Don't suggest improvements yet — I only want an honest
inventory of where the time actually went.
```

2 · Score against the three questions

COPY-PASTE · THE SCORING PROMPT

```
Score each task against three questions:

1. Does it come back? (weekly / monthly / per request / one-off)
2. Is it the same shape every time? (1-5, where 5 means the structure never
   changes and only the details do)
3. Am I the bottleneck rather than the value? Estimate what share of my time
   is copying, reformatting and chasing, versus deciding something.

Put it in a table, best candidate first. For question 3, name the mechanical
part specifically instead of giving me a percentage on its own.

Tasks: [paste the list from step 1]
```

The prompt that says no

This is the one most people never run, and it's the one that keeps you out of trouble. A shortlist without a rejected item usually means you weren't honest on question three.

3 · Find the ones to keep

COPY-PASTE · THE HONEST NEGATIVE

```
From this list, tell me which tasks I should NOT hand to AI, and why.

I'm looking for the ones where my judgment is the actual product — where the
reason it's my job is that I know things a document doesn't.

Don't be encouraging. If something scores well on paper but would be a bad
idea in practice, say so and say why.

Tasks: [paste your scored table]
```

What a good answer looks like

GREEN FLAG

It names a specific task, says which question it fails, and explains what would be lost. Vague caution ("be careful with sensitive data") is not an answer.

RED FLAG

Every task comes back approved. That means the model is agreeing with you rather than assessing. Re-run it and keep the "don't be encouraging" line — it's doing real work.

If you genuinely have nothing to reject this week, that's fine — but check that you scored question three on the hardest thing you did, not just the most annoying.

Three real tasks, scored

Same week, same person, three different verdicts. The third is the reason question three exists at all.

TASK	COMES BACK	SAME SHAPE	BOTTLENECK OR VALUE
Monday roll-up of what the team shipped	Weekly	5 / 5	Bottleneck. Five tabs open, copying lines out, reformatting into bullets, fixing tense. Verdict: build it once.
Monthly vendor reconciliation	Monthly	5 / 5	Bottleneck. Opening files and comparing numbers; judgment is only in what to do about the mismatches. Verdict: build it once.
Marcus's performance review	Twice a year	5 / 5	Value. What you've noticed over six months, what he should hear, how hard to push. Verdict: keep it. Two out of three is a no.

What the scoring prompt gave back

OUTPUT · THE PART THAT MATTERED

```
Strongest candidate: the Monday roll-up. It runs every week, the structure is
fixed, and roughly 90% of your time on it is moving text between tabs rather
than deciding anything.

Do not automate the performance review. It repeats and the form is identical,
so it passes the first two tests — but the reason the task exists is your six
months of context on this person. Automating it removes the only part that
has any value.
```

Notice what the two winners have in common. Boring, repetitive, invisible when they go right, painful when they go wrong. That's the profile you're hunting for.

Three ways people misfire

01 Starting with the most important thing

The most common failure, and it usually ends the experiment.

The big visible task feels like the obvious place to prove AI is useful. It's the worst place. Stakes are high, the shape is less standard than it looks, and one bad result becomes "we tried that." Start with the most repetitive thing instead. If it breaks, nothing happens.

02 Skipping question three

Comfortable to skip, expensive to skip.

Almost everyone scores the first two honestly and then softens the third. The result is a shortlist full of work where your judgment was the product — which produces output that's technically fine and quietly wrong.

03 Building all three at once

Ends in a folder of half-finished workflows.

A shortlist of three is a queue, not a plan for Saturday. Build the top one, run it for two real weeks, then start the second. One workflow in production beats three in progress.

RECAP

You now have

1	Three questions you can run against any task in about thirty seconds.
2	Three prompts that turn one week of real work into a ranked shortlist.
3	A worked example, including one task that deliberately fails the filter.
4	The three ways this goes wrong, and what to do instead.

briandoai.com/resources