

WORKSHEET ONE OF TWO

Define the tasks.

If you cannot write this page for a pilot, the pilot is not ready to scale. That is the whole test. Fill it out for *one* process: the one actually carrying weight, not the easiest one.

01 What tasks does it perform?

One process. Name the tasks the way the person who owns the work would describe them, not the way a demo would.

02 What may it do on its own?

Be specific about the limits. "Use good judgment" is not a limit.

03 What must it hand back to a human?

The escalation rule. A system that cannot refuse cannot be governed.

04 What check runs on it, and when?

A check that runs automatically costs the team nothing. A check that depends on someone remembering is the one that fails.

05 Who owns the work, and what needs their sign-off before it leaves the company?

A named person, and a written list of what they sign off. Not a team, not a role, not "whoever is around."

WORKSHEET TWO OF TWO

The leverage test.

Run this on every AI pilot you have underway right now. Answer honestly, because the purpose is to find the one worth keeping, not to justify the ones you have already paid for.

01 What did this pilot change?

Hours, defects, cycle time, cost per run. If the answer is nothing, it is a cost wearing a productivity costume.

02 What was the baseline before it?

Without a baseline you cannot tell improvement from enthusiasm.

03 Does it touch a system the business already runs?

Or did someone stand up a second permissions model nobody audits?

04 If it is wrong, is the damage contained?

Start where the money is not. Non-financial workflows first.

05 Who is accountable when it is confidently wrong?

Functional testing asks whether it ran. The operating question is what happens when it is confidently wrong.
