

Who Signs Off When the AI Acts

A one-page brief for the companion gate map — deciding what your AI may do on its own, and what waits for a person.

FACTORYOS BY COMPTRIO · YOUR DATA STAYS ON YOUR MACHINE. · [COMPTRIO.COM](#)

An AI agent doesn't just answer questions — it does the legwork, and it can act in the world: send the message, post the entry, file the record. The fear that comes with that is reasonable: *what if it does something dumb, and it can't be taken back?* The answer isn't to trust it more or less. It's to decide, in advance, exactly where a human signs off — and to keep that dial in your hands, not the vendor's.

The criterion: the gate goes wherever the AI is about to change something real — and the bigger the consequence, the less optional that gate should be. A human signature where the stakes are high; everything else free to run. That line is a decision you make, not a setting the vendor makes for you.

WHAT THE MAP DOES

List the real things your AI would do. For each one, set three dropdowns — what it changes, how reversible it is, who it touches — and the sheet returns a verdict automatically: **GATE** (a person signs off first) or **Auto-run** (safe to let go). A tally shows how much of your AI's work needs a signature and how much doesn't. Flip the **Regulated-work** toggle and the rule tightens: in HIPAA, legal, or financial work, anything past an internal note waits for a person by default.

THE RULE IT USES

- **Gate it** when it touches a customer, real money, or a regulator — or it's hard to undo — or it acts in the outside world.
- **Let it run** when it only drafts or files internally, is easily undone, and stays in-house.
- **Don't gate everything.** That just rebuilds the bottleneck the AI was meant to remove. Put the human exactly where the consequences are, and nowhere else.

THE VENDOR TEST

Two questions for any AI vendor: **can your AI hand a person all four — intent, justification, provenance, action detail — before it acts? And can I set the gates myself, on the actions I choose?** If the answer is no, you're being asked to rubber-stamp a black box. That's not a human in the loop; it's a human standing next to it. Done right, the drudgery is the machine's and the judgment stays yours — the AI a supporting character that prepares the work, never the one that decides it.

The companion worksheet builds your own gate map: list your actions, set three dropdowns, read the verdict and the tally. Nothing leaves your computer — the point of the local, owned AI it's built for.

Sources: Comp trio, *How AI Agents Prepare Work for Human Approval & Designing Human-in-the-Loop Controls* (comptrio.com).

— Shawn Snarski · FactoryOS by Comp trio · your data stays on your machine. · [comptrio.com](#)

READING THE VERDICT

A FEW EXAMPLE ACTIONS	
Draft a reply for me to review	AUTO-RUN
Summarize an internal document	AUTO-RUN
Send an email to a client	GATE
Post a journal entry to the books	GATE
Issue a refund / move money	GATE
File a regulatory or insurance claim	GATE

The AI still does all of it — pulls the data, drafts, scores, builds the report. It just stops at the gate and waits, on the actions where stopping matters.

WHAT A HUMAN SHOULD SEE FIRST

A gate is only worth having if the AI hands you enough to judge by — not a yes/no box. On a GATE action it should compile all four: **Intent** (what it's about to do), **Justification** (why), **Provenance** (the exact source data, traceable), and **Action detail** (what runs the instant you approve).