

Write the handoff rule before the permissions

Earned autonomy. Most agent deployments specify what the agent may do and leave the escalation to the model's judgment. Do it the other way round: define when it must stop, and let it earn everything above that line.

A handoff rule is one sentence: **"The agent stops and a person decides when..."** Four things finish it.

1

The confidence floor.

Below this, the agent recommends instead of acting. Set it per decision, not per system. A markdown is not a reorder.

2

The categories where a human always answers.

New suppliers, regulated lines, anything with a named account relationship behind it. Written down, not implied.

3

The value threshold.

The cost of a wrong call above which a person signs. Money, units, or shelf-space: whichever the business actually feels.

4

What gets logged at the handoff.

Why it stopped, what it would have done, and what the person decided instead. That log is how the agent earns the next level.

Define the exceptions first.

Autonomy is what is left after the exceptions are honoured.