

Human-in-the-loop & earned autonomy

Ed Schaefer · Version 1.0 · August 2026 · edward-schaefer.com/cards/human-in-the-loop · Free to share

Where the human sits as agents earn trust.

- To me, human-in-the-loop is a coaching relationship: an agent earns autonomy the same way a person does, and an approver who checks every single move just becomes the bottleneck everyone waits on.
- Autonomy granted equals earned trust times blast radius: how much I trust this specific agent, built up over time and revocable, multiplied by how bad it is if this specific action goes wrong. Either axis alone gives you the wrong answer.
- Blast radius resolves to five tiers: Tier 1 is just do it, no ask needed; Tier 5 is fully human, we walk through it together. This scale is the one piece of this practice area that keeps showing up elsewhere: the Pre-Flight Checklist, Definition of Verified, and the AI Operating Model all cite it by name.
- As trust gets earned, my own job moves up the loop: doing the work myself, then reviewing what comes back, then coaching the setup, then setting intent and letting the agent run without me watching every step — borrowed from David Marquet's intent-based leadership.
- I've named the ways this goes wrong because I've watched each one happen:
- **The Rubber Stamp:** rubber-stamping approvals once checking gets tiring.
- **The Bottleneck:** turning into the approval chokepoint everything queues behind.
- **Automation Complacency:** going automation-complacent and quietly dropping the checking.
- **Context Collapse:** losing the context that made oversight useful in the first place.
- **Escalation Fatigue:** burning out on the escalations themselves.
- **The Static Ladder:** setting a trust tier once and never coming back to it.

What's behind this card

Each named anti-pattern gets worked through in full in [Human-in-the-Loop Is a Coaching Problem, Not a Gate](#).

This is one of six cards today. The field guide maps seventeen practice areas, and I'm growing this set as each one earns its own depth.

This is a fit question, not a recipe. Where your team actually sits against this list, and what that's costing, is what the [Engineering Ways of Working Diagnostic](#) is built to find.