CHECKLIST · PRINT-AND-FORWARD ONE-PAGER

Agent Rollout Pre-Flight Checklist

Ed Schaefer · Version 1.0 · August 2026 · edward-schaefer.com/templates/agent-rollout-pre-flight-checklist · Free to share

A one-page companion to the AI-Native Ways of Working field guide.

Most teams answer these questions for the first time only after something has already shipped that shouldn't have. Every item below is cheaper to work through before an agent rollout than after one goes wrong.

What its job is

- [ ] The agent has a role, constraints, and an escalation path.

What it can reach

- [ ] Credentials cover only what the task needs. Splitting work across agents keeps least privilege per agent; widening one agent's access does the opposite.
- [ ] Tuned credentials can be issued on demand and revoked as needed.
- [ ] Someone has written down what the agent's credentials could reach, not only what it's instructed to do.

What it can spend

- [ ] A hard spend cap exists and stops the workload on its own.
- [ ] If this went wrong tonight, something would stop the spend before morning without a person having to notice.
- [ ] Anything that spends money on a model call persists what it gets back, enforced by a check that fails loudly.

What tier the work is

- [ ] A blast-radius tier is assigned before the first run, and revisited as the work changes: what breaks if this is wrong, and whether it can be undone. The five tiers come from Human-in-the-Loop Is a Coaching Problem; the Definition of Verified below summarizes them.
- [ ] Review depth for this rollout matches its tier: automated checks if low, sampled human review in the middle, full human verification if high.
- [ ] Before the first real run, the agent has done the same work somewhere safe: a test run or a replay.

Who verifies

- [ ] The team's Definition of Verified is filled in with the team's own work types, and the default example row is replaced.
- [ ] The verifier is never the author.
- [ ] No sign-off without evidence of what was checked. Never rubber-stamp.

How it stops

- [ ] A kill or pause control is named for every place the agent runs.
- [ ] Anyone on the team can stop the line the moment they spot a defect, regardless of role.
- [ ] A rollback or restore path exists and has been tested. Stopping the agent doesn't undo what it already did.

When it goes wrong

- [ ] Action-level logs exist before the first run: every step the agent takes, in order. The runbook's capture step depends on them.
- [ ] The Agent Incident Runbook is filled in before the first incident.
- [ ] The team already knows who's told, and in what order.

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.