Is your environment ready to open the throttle?

Michael Rishi Forrester, AI Startup Advisor, Neo4j Startup Program

A self-check from Deploying Agents at Scale: What Enterprise Adoption Actually Looks Like, Agentic Engineering Days Zurich, 20 November 2026

Is your environment ready to open the throttle on coding agents? Answer these ten questions yes or no, for one team or one system, as things stand today. Each question is about your own environment and needs no tool to answer. They are in fix-first order: if you answer no to any, the first no is the first thing to fix. Each comes with the evidence behind it and the kind of evidence it is, so you can check the work.

Answer yes only if it is true today, not planned. This page sends nothing and stores nothing: your answers stay on your screen and are gone when you close it.

The ten checks, in fix-first order

yes no **1. A baseline taken before scaling.** Do you have delivery measures recorded from before agent use grew, such as lead time, change fail rate, rework rate, recovery time and total monthly incidents, and do you still collect them the same way?
Evidence-index rows: metr-expected-24-believed-20, metr-19-percent-longer, s10-dora-2024-each-25-percent-increase, faros-q3-code-churn-71-6 **(unverified)**

yes no **2. Limits the agent cannot lift.** Are the things an agent must never do, such as deleting data, resetting infrastructure or using production credentials, blocked by permissions it does not hold, and not only by instructions in its prompt or by an approval a person clicks through?
Evidence-index rows: s9-adaptive-attacks-bypass-12-recent-defenses, s7-prompt-of-481-lines-and-8, s7-reset-script-targets-9-nodes, s9-sandboxing-cut-permission-prompts-by-84

yes no **3. Agent reports checked against reality.** Is every claim an agent makes about what it built, changed or deployed checked against the real system before anyone acts on it?
Evidence-index rows: s7-150-clusters-are-provisioning-concurrently-19 **(unverified)**, s7-0-attendee-clusters-at-workshop-time **(unverified)**

yes no **4. A written review rule by risk.** Does every change that merges get a review under a written rule that says which changes need a human, which an agent and which both, and do you know what share of last month's merges had no review?
Evidence-index rows: s10-faros-q3-prs-merged-without-review, s1-rejected-agent-prs-228-of-600

yes no **5. Automated gates before human review.** Must every agent-written change pass the same build, test, static analysis and secret checks as human code before a person reviews it?
Evidence-index rows: s5-veracode-2026-the-average-security-pass, s5-veracode-2026-syntax-correct-nearly-100, s5-gitguardian-claude-code-assisted-commits-3, s5-gitguardian-24-008-unique-secrets-in, s1-agent-prs-each-failed-ci-check

yes no **6. Tests that check the behavior.** Do agent-written changes come with tests that check the behavior that was asked for, written or reviewed by someone other than the agent that wrote the code?
Evidence-index rows: s11-metr-failure-table-runs-failing-tests, s11-mathews-and-nagappan-gpt-4-solved, s8-meta-testgen-llm-75-percent-of

yes no **7. Small changes, practiced rollback.** Do you keep each change small, and can every team undo a bad change quickly because it has practiced doing so?
Evidence-index rows: s10-faros-q3-average-pr-size-up

yes no **8. Review capacity grown with volume.** Has the capacity for review and QA grown in step with the volume of change agents produce, so that the time changes wait for review and spend in QA is not rising?
Evidence-index rows: s8-after-ai-agents-submission-to-merge, s10-faros-incidents-per-pr-up-242, s11-faros-q3-2026-22-000-developers

yes no **9. Done written down first.** Is what done means written down before an agent starts a task, and do your repositories give agents short, human-written notes on what they could not work out from the code?
Evidence-index rows: s11-lulla-et-al-10-repositories-124, s11-ambig-swe-models-struggle-to-distinguish, s11-gloaguen-et-al-context-files-do

yes no **10. Reviewers who can debug unfamiliar code.** Can the engineers who approve agent changes read and debug code they did not write, and do your junior engineers still get practice debugging code themselves?
Evidence-index rows: s11-52-mostly-junior-developers-ai-group

What to fix first

Start at your first no and work down this order. The order is argued, not measured: no source read ranks these fixes.

A baseline first. 1. A baseline taken before scaling

Cap what cannot be undone. 2. Limits the agent cannot lift; 3. Agent reports checked against reality

Keep the gate on what merges. 4. A written review rule by risk; 5. Automated gates before human review; 6. Tests that check the behavior

Make the volume hold. 7. Small changes, practiced rollback; 8. Review capacity grown with volume

Improve what goes in. 9. Done written down first; 10. Reviewers who can debug unfamiliar code

Ten yeses means ready by this check. That a yes to every check means ready is argued, not measured.

From the talk *Deploying Agents at Scale: What Enterprise Adoption Actually Looks Like*, Agentic Engineering Days Zurich, 20 November 2026. Every figure on this page is in the talk's fact check with its source and the date it was read. The evidence-index rows named under each question are in the talk's evidence index, on the talk site, agenticday.michaelrishiforrester.com.