Northlight

STRATEGY · INSIGHT · AI · APPS

AI AGENT DEPLOYMENT GO / NO-GO

BEFORE GIVING ANY AGENT AUTONOMOUS ACCESS

IDENTITY & PERMISSIONS

- Agent has its own identity, separate from any human developer ☐
- Agent permissions are explicitly defined and documented ☐
- Agent is operating at the appropriate autonomy level (1–4) for the task ☐
- High-stakes actions require explicit human approval before execution ☐

BOUNDARIES & CONTAINMENT

- Network access is restricted — agent cannot reach untrusted external domains ☐
- Agent cannot write directly to protected branches or production systems ☐
- Agent has been told its boundaries explicitly (not just blocked silently) ☐
- A "controlled failure" path exists — blocked actions are logged and surfaced, not just dropped ☐

AUDITABILITY

- Every agent action is logged with full attribution ☐
- Compliance team can audit agent actions with same rigor as human actions ☐
- Security team has been briefed and signed off ☐
- An incident response plan exists if the agent behaves unexpectedly ☐

READINESS

- Agent has been given structured context upfront (not starting cold) ☐
- A pilot has been run with a low-risk use case first ☐
- Success metrics are defined before deployment, not after ☐

Autonomy without guardrails is risk.
Guardrails without clarity are friction.
Get both right, and agents can scale with you.