Production Readiness Self-Score

10 questions to classify your agent program: Experimental • Scaling • Production-Ready.

Scorecard (1 point each)

- We have clear success metrics (quality, latency, cost) for the agent.
- We maintain an evaluation suite with regression checks before deployment.
- Tools are allowlisted and permissions are least-privilege per environment.
- High-impact actions require approvals and have rollback capability.
- We log prompts/tool calls/outcomes with audit-friendly traces.
- We monitor drift and run canary tests on a schedule.

Interpretation

- 0–3: Experimental — stabilize scope and add guardrails.
- 4–7: Scaling — strengthen monitoring, governance, and release gates.
- 8–10: Production-Ready — focus on optimization and continuous improvement.