AgentOps KPI Quick Reference

Core metrics to operate AI agents safely, reliably, and cost-effectively in production.

Reliability & Quality

- Success rate: % of tasks completed correctly end-to-end.
- Escalation rate: % of runs requiring human intervention.
- Regression score: delta vs previous model/prompt/tool version.

Observability

- MTTD: mean time to detect incidents (quality/cost/safety).
- MTTR: mean time to resolve incidents (rollback/fix/redeploy).
- Drift frequency: how often performance shifts outside tolerance.

Safety & Compliance

- Policy pass rate: % requests/actions passing checks.
- Audit completeness: % runs with full logs + tool traces.
- Permission violations: denied tool calls and boundary breaches.

Cost & Performance

- Cost per interaction: blended model + tool + retrieval costs.
- Latency p95: user-perceived response time at the 95th percentile.
- Token efficiency: tokens used per successful outcome.