

Agent Guardrails and Permissions Review

Use this checklist in your next design review. Tick each statement that is true for your agent today. Thresholds and intervals in the guide are recommended starting points; calibrate to your environment.

IDENTITY AND PERMISSIONS

- ☐ Every agent runs under its own or a delegated identity, with no shared or administrator credentials.
- ☐ Each task has a written permission scope listing allowed actions and data.
- ☐ Elevated access is time-bound and expires by default.
- ☐ Delegated identity is used where the agent must not exceed the requesting user's rights.

POLICY ENFORCEMENT

- ☐ Every hard rule exists as a check in the harness, not only in the prompt.
- ☐ Each action is classified by risk class, with minimum controls applied.
- ☐ Irreversible writes and access changes route to a human checkpoint.
- ☐ Output checks cover personal data and unapproved commitments.

INJECTION AND MISUSE

- ☐ All content from documents, tickets, emails and tool results is treated as untrusted data.
- ☐ An injection test set exists and was run before pilot, with zero unauthorised actions.
- ☐ Guardrails are tested with varied phrasing, not keywords alone.

AUDIT

- ☐ All allow, deny and flag decisions are recorded in the trace.
- ☐ Denials and false positives are reviewed on a schedule.
- ☐ Permission scopes are reviewed on a fixed schedule (quarterly as a starting point).

Score: ____ of 14. Any unticked box is a gap to close before the agent handles real work.

Full guide: readyforrole.com/blog/guardrails-and-permissions.html | Series hub:
readyforrole.com/blog/enterprise-ai-agent-engineering.html | support@readyforrole.com