Guardrail Checklist

Scope

- ☐ One clear purpose.
- ☐ Out-of-scope handling defined.

Tools

- ☐ Least privilege.
- ☐ Read/write separated.
- ☐ Schemas validated.
- ☐ Timeouts/errors defined.

Actions

- ☐ Human approval before risky write.
- ☐ Idempotency designed.
- ☐ Max steps/tool calls set.

Data

- ☐ Allowed sources listed.
- ☐ Sensitive data minimized.
- ☐ Secrets stay outside prompt/log.

Reliability

- ☐ Safe fallback.
- ☐ Logs/correlation ID.
- ☐ Retry policy.
- ☐ Incident owner.