Complete one canvas per agent BEFORE building. If any block is vague, the agent will fail in production.

1. Agent Role

Job title for the agent: what outcome does it own? (e.g. 'Triage support tickets in under 60 seconds')

2. Users & Triggers

Who uses it and what starts a run? (schedule, message, event)

3. Inputs & Tools

Data sources, APIs and tools the agent may call - and what is off-limits.

4. Workflow Steps

The 3-7 steps the agent follows, in order. Mark where it must ask a human.

What the agent must NEVER do. Tone, compliance and safety rules.

6. Human Handoff

Exactly when and how work passes to a person - with what context attached.

7. Success Metrics

Task success rate, time saved, escalation rate - with targets.

8. Failure Modes

Top 3 ways this could go wrong - and the mitigation for each.