

Agent Design Worksheet

Complete it before you build anything. Keep it as the agent's design document.

Copy this file for each new agent. Fill the fields on screen, then save or print. This is Appendix A of the book; see Appendix E for two finished examples.

Version

Date

Owner

1. Purpose

Goal in one sentence

What "done" looks like

Worst thing that could happen

How a mistake would be noticed

2. Process: one line per step, labelled Fixed, Judgment or Human

Step	Type
	Choose...
	Choose...
	Choose...
	Choose...
	Choose...
	Choose...
	Choose...
	Choose...

Pattern used (chain, router, fan-out, draft and review, coordinator)

3. Autonomy: per action, levels 0 to 4

0 Advisor · 1 Fixed workflow · 2 Proposer (you approve) · 3 Reviewed actor · 4 Bounded autonomy

Action	Level	Reason
	Choose...	
	Choose...	
	Choose...	
	Choose...	
	Choose...	
	Choose...	
	Choose...	

4. Tools

Tool	Kind	Narrowest access	Approval?
	Choose...		Choose...
	Choose...		Choose...
	Choose...		Choose...
	Choose...		Choose...
	Choose...		Choose...
	Choose...		Choose...

5. Handoffs

Format passed between steps, including an "uncertain" field

6. Instructions

Location and version of the current instructions

7. Context and memory

Facts the agent needs that it cannot know

Memory used (none, notes file, retrieval) and who reviews it

8. Evaluation

Test set size

Typical / Edge / Tricky

Safety bar %

Accuracy bar %

Latest scores and date

9. Safety

- Guardrails in place
- ☐ Input guardrails (screen what comes in)
- ☐ Action guardrails (limit what the agent can do)
- ☐ Output guardrails (check what goes out)
- ☐ Budget guardrails (step, time and spending caps)

Stop button and who can use it

Incident plan location

10. Change log

Date	Change	Reason	Result