THE BUILDER'S STACK · THE PRACTICE · AGENTIC SYSTEMS

The Agent Charter

What your agent may do, must ask before, and must never touch: the action surface, every action's rung on the autonomy ladder, the trifecta check, the caps, the receipts, and the eval gates that earn promotion. Fill it before you ship, sign it, and re-sign it whenever the surface changes.

AGENT / FEATURE

OWNED BY

DATE

1 The job and the line

One sentence each: the job the agent does for the user, and the line it never crosses no matter what the input says.

THE JOB

THE LINE IT NEVER CROSSES

2 The action surface

Every tool is delegated authority. Reads and writes flagged, reversibility honest, and a reason for each key on the ring.

ACTION 1

TOOL / ACTION

IT READS

IT WRITES

WHY IT EXISTS

☐ Reversible

ACTION 2

TOOL / ACTION

IT READS

IT WRITES

WHY IT EXISTS

☐ Reversible

ACTION 3

TOOL / ACTION

IT READS

IT WRITES

WHY IT EXISTS

☐ Reversible

ACTION 4

TOOL / ACTION

IT READS

IT WRITES

WHY IT EXISTS

☐ Reversible

ACTION 5

TOOL / ACTION

IT READS

IT WRITES

WHY IT EXISTS

☐ Reversible

ACTION 6

TOOL / ACTION

IT READS

IT WRITES

WHY IT EXISTS

☐ Reversible

3 The autonomy ladder

Per action, never per product: suggest, draft, act with approval, act with undo, act silently. Start one rung lower than you believe.

ACTION 1

RUNG

ACTION 2

RUNG

ACTION 3

RUNG

ACTION 4

RUNG

4 The trifecta check

Private data, untrusted content, an outbound channel: all three must never coincide without a gate.

- [ ] Holds or reaches private data
- [ ] Reads untrusted content (email, web, uploads)
- [ ] Can communicate outward (send, post, call)

THE LEG WE GATE

THE GATE

5 Caps and the kill switch

Bound what one bad turn can reach, and make stopping possible mid-action.

SPEND CAP

ACTION / RATE CAP

VOLUME CAP (ROWS, RECIPIENTS, FILES PER TURN)

THE KILL SWITCH: WHERE IT IS, WHO CAN PULL IT

- [ ] Kill switch tested mid-action, on purpose

6 Receipts and undo

If it acted, it shows. If it shows, most of it can be taken back, and what cannot is labeled.

WHAT THE USER-READABLE LEDGER SHOWS

UNDO STRATEGY (WHAT COMPENSATES WHAT)

LABELED IRREVERSIBLE

7 Eval gates

Trajectory and outcome, against sandboxed tools, with pass rates doing the promoting.

PROCESS RUBRIC (TOOLS ALLOWED, MAX CALLS, LINES)

OUTCOME BAR

- [ ] Evals run against sandboxed tools only

THE PASS RATE THAT PROMOTES A RUNG

8 Sign-off and review

The charter is re-signed when the surface changes, and amended after every incident.

SIGNED BY

LAUNCH DATE

RE-SIGN WHEN

AFTER THE FIRST INCIDENT: WHAT CHANGED