@demotoprod · SHIP · also LEVERS

Orchestration checklist

A September 2026 security study of tool-calling agents saw one session's cumulative input reach 14,293x its first call under attack; benign runs peaked at 1.0x to 208.7x across six model families. Context sent again is billed again. Four afternoon checks show where a chain, an agent or a hand-off earns its cost.

1 Run the task as a fixed chain before adding an agent

Prompt chaining suits a task that splits cleanly into fixed subtasks, and one guide recommends the simplest solution first. A study of 1,600+ annotated traces from 7 multi-agent frameworks found 14 failure modes; its authors say gains on popular benchmarks are often minimal.

- Write the task as numbered steps. Where inputs and outputs are known, run direct calls in a fixed order, with code choosing the next step.
- Score the chain on a fixed set of real tasks; its pass rate and cost are the bar to beat.

2 Pass a compressed summary and the decisions between steps

In reruns of 12 tasks on two providers, keeping raw history raised mean session cost about 21 to 36% over compressing or dropping it; compression kept 10 or more tasks working, dropping kept 2. A post proposes compressing history into key details, events and decisions.

- Find each point where one call's output feeds another; log input tokens there across real runs.
- Pass the next step a summary plus the decisions so far; re-run your tasks and confirm they pass.

3 Give an agent only the unknown step, and cap it

One guide suggests agents for open-ended problems where the number of steps is hard or impossible to predict, with stopping conditions such as a maximum number of iterations. A second names maximum turns as an exit condition and escalates to a human past its retry limits.

- Mark each chain step known or unknown; give an agent only an unknown one, keep the rest as code.
- Set the cap from your longest good run; when hit, stop, log the task and send it to a person.

```
for step in range(MAX_STEPS):   # cap from your longest good run
    action = decide(state)      # the one open-ended step
    if action.done:
        break
    state = run_tool(action, state)
else:
    raise StepCapReached(task_id)   # log it, do not retry
```

4 Score each task, and cut hand-offs that do not earn their cost

A multi-agent research system used about 15x more tokens than a chat, and its authors say the task's value must be high enough to pay for the increased performance. A 260-configuration study put multi-agent results against one agent at +80.8% on one benchmark and -70.0% on another.

- Log one row per task run, not per agent: path, hand-offs, tokens, and whether a checker passed it.
- Run the chain and the agent on the same tasks; drop a hand-off if pass rate holds and cost falls.

```
row = {"task": task_id, "path": "chain",       # or "agent"
       "handoffs": n_handoffs, "tokens_in": t_in, "tokens_out": t_out,
       "passed": checker(output)}             # judged per task
log.append(row)
```


Nitish Gautam · indica-tech.com · hello@indica-tech.com

INDICA TECH · @DEMOTOPROD

Nitish Gautam

Founder-led AI engineering. Demo to production, under NDA, in regulated industries, on fixed scope.

Paste into your project

A POLICY FILE FOR ONE ORCHESTRATED TASK

```
# orchestration policy; set every value from your own runs
default_path: chain             # fixed steps, code picks the next one
agent_steps: [<unknown step>]   # only the steps you cannot predict
max_steps: MAX_STEPS            # from your longest good run
on_cap: send_to_person          # log the task, do not retry
handoff_payload: [summary, decisions]   # not raw history
log_per_task: [path, handoffs, tokens_in, tokens_out, passed]
keep_handoff_if: pass_rate_holds_and_cost_falls
```

Where this list comes from

Ten-plus years taking products from zero to one across seven sectors, including defence, fintech, medical and telecom, with AI shipped into regulated and air-gapped environments. The @demotoprod channel takes one production AI failure apart each day. These checklists are the same work written down: every item is a dated failure or a documented default, with the control that stops it and the primary source.

If this list found a gap

1 **Your demo works and production is stuck.**

Production Readiness Audit. Fixed price, two weeks. A five-dimension scorecard, cost and failure analysis, and a 90-day execution roadmap.

2 **You need senior AI judgement without a full-time hire.**

Fractional AI Lead. Monthly: 90-day minimum, then 30 days' notice either side. Weekly 1:1 with the founder or CTO, architecture and code review, technical hiring support, investor-ready due diligence.

3 **An AI feature has to work, by a date.**

AI Build Sprint. Fixed price, four to eight weeks. Scope signed off up front, weekly demos, CI/CD and observability included, 30-day production warranty.

4 **A regulator, an auditor or a customer's security team is asking.**

AI Governance & Compliance. Fixed price, four weeks. ISO 27001 / SOC 2-aligned controls, a model risk framework, human-in-the-loop and audit logging, regulator-ready documentation.

BOOK A CALL
indica-tech.com/#contact

EMAIL
hello@indica-tech.com

WEBSITE
indica-tech.com

ALL CHECKLISTS
indica-tech.com/checklists/

YOUTUBE
youtube.com/@demotoprod

LINKEDIN
linkedin.com/in/startup-founderx

INSTAGRAM
instagram.com/demotoprod

TIKTOK
tiktok.com/@demotoprod

Free to share. The live list, with every source link: indica-tech.com/checklists/#multi-agent