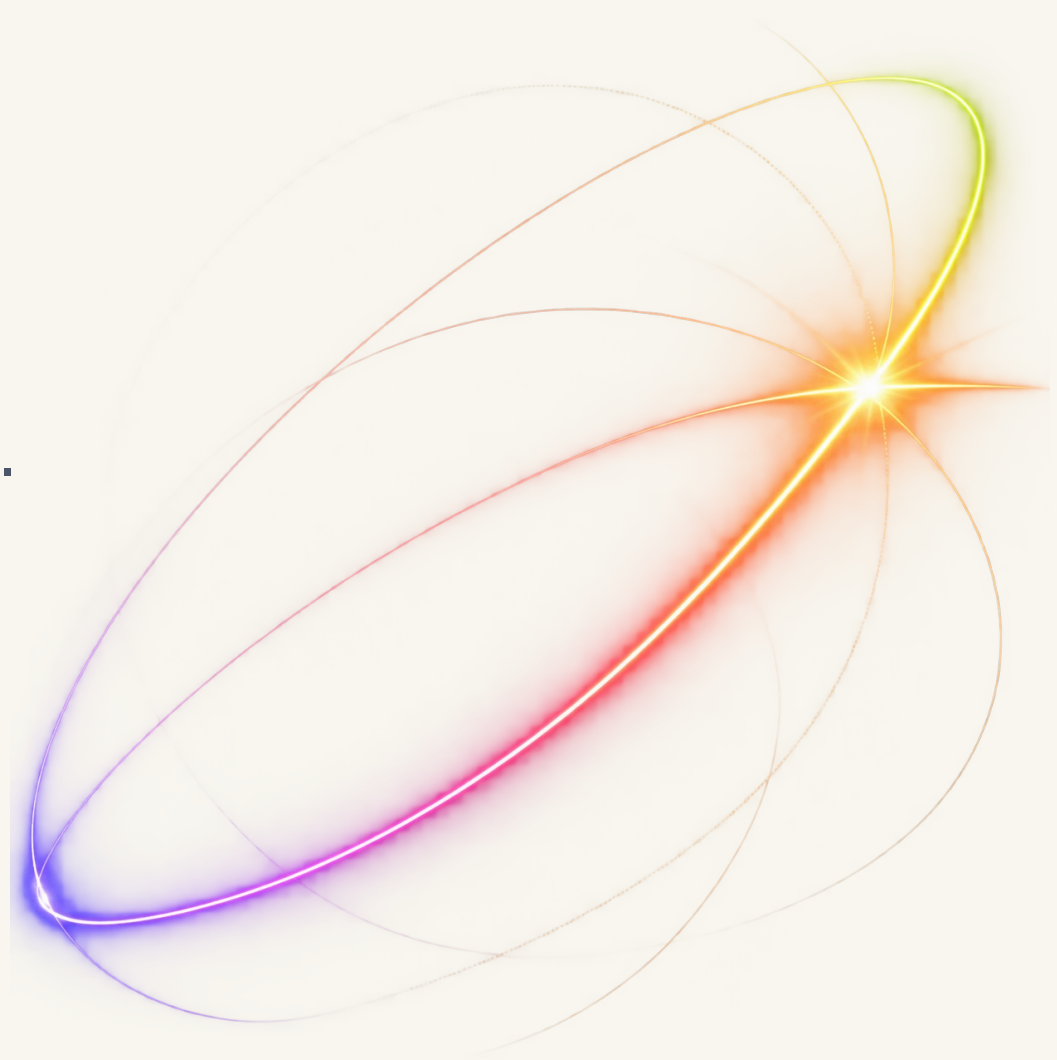


THE PRODUCTION BLUEPRINT FOR AI

Most teams do not have a
model problem.
They have a system problem.



PLANNING

BUILD

OPERATIONS

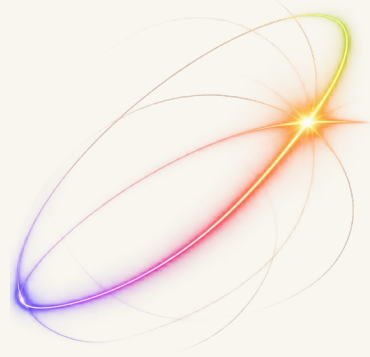
Swipe from purpose and scope to testing and operations.

SWIPE →

01

PURPOSE & SCOPE

Production starts with boundaries, not prompts.



01 USER Who uses it?

02 OUTCOME What must improve?

03 LIMITS What must never happen?

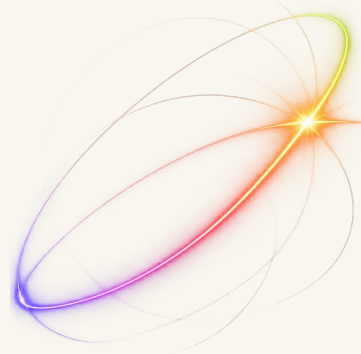
PRODUCTION CHECK

Can you define success without mentioning the model?

FAILURE MODE

The team optimizes answer quality for a workflow no user owns.

NEXT →



02

PROMPT & BEHAVIOR

Behavior must be specified before it can be trusted.

01 ROLE

What judgment should it apply?

02 INSTRUCTIONS

How should it act?

03 GUARDRAILS

Where must it stop?

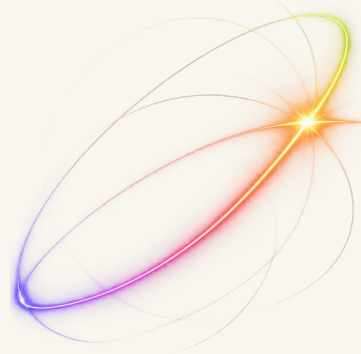
PRODUCTION CHECK

Can two reviewers predict the same intended behavior?

FAILURE MODE

A vague role produces different decisions for the same case.

NEXT →



03

MODEL & RUNTIME

Choose for the workload—not the leaderboard.

01 CAPABILITY

What must the model do well?

02 LATENCY

What response-time budget applies?

03 COST

Who owns spend by workflow?

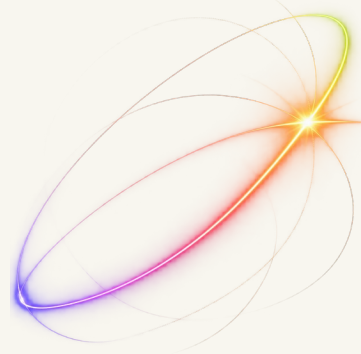
PRODUCTION CHECK

Does the workload pass quality, latency, and cost budgets together?

FAILURE MODE

A leaderboard winner misses the latency or cost budget.

NEXT →



04

DATA, TOOLS & INTEGRATIONS

Reliability depends on what the application can access.

01 SOURCES

Which information is authoritative?

02 PERMISSIONS

What is the minimum required access?

03 DEFENSES

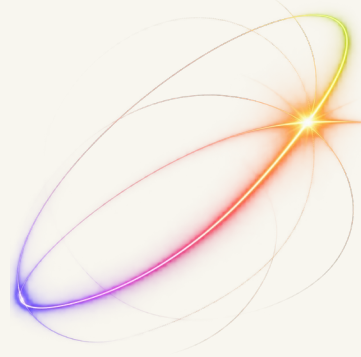
How are injection and exfiltration contained?

PRODUCTION CHECK

Can untrusted input reach data or tools it should never control?

FAILURE MODE

One injected document turns a useful tool into a data-exfiltration path.



05

STATE & MEMORY

Memory is a product decision, not a default.

01 **SESSION** What disappears when the task ends?

02 **LONG-TERM** What persists across sessions?

03 **DELETION** Can users inspect, correct, and remove it?

PRODUCTION CHECK

Can memory be justified, inspected, corrected, and deleted?

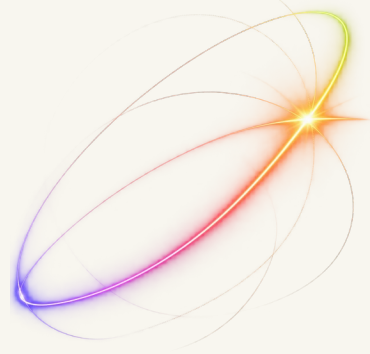
FAILURE MODE

Personalization that cannot be deleted becomes a compliance liability.

06

WORKFLOW & ORCHESTRATION

The happy path is not production.



01 **ROUTES** How does work move between steps?

02 **RECOVERY** What compensates or rolls back?

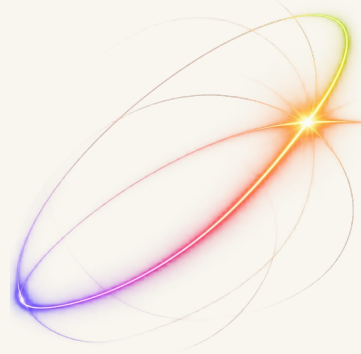
03 **HANDOFF** When does a person take control?

PRODUCTION CHECK

Can every partial failure recover, compensate, or escalate?

FAILURE MODE

A workflow with no rollback collapses on the first partial failure.



07

EXPERIENCE & HANDOFF

Trust depends on what happens when AI is uncertain.

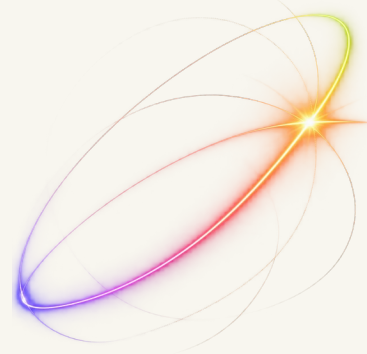
- | | | |
|----|-----------|--------------------------------|
| 01 | INTERFACE | What does the user see first? |
| 02 | EXPLAIN | How is uncertainty made clear? |
| 03 | ESCALATE | When does a person take over? |

PRODUCTION CHECK

Does the user understand what happened and what to do next?

FAILURE MODE

Silent uncertainty makes confident errors look authoritative.



TESTING & OPERATIONS

08

Production readiness is a continuous loop.

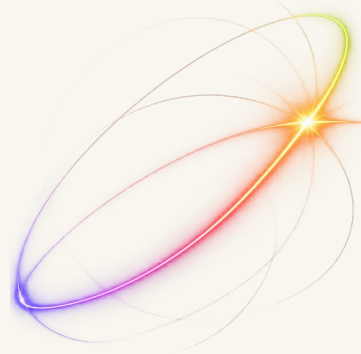
-
- | | | |
|-------|-------------|---|
| 01 | GOLDEN SETS | Which known cases must never regress? |
| <hr/> | | |
| 02 | TRACING | Can you replay prompt, tool, and data lineage? |
| <hr/> | | |
| 03 | BUDGETS | Can you attribute cost and latency by workflow? |
-

PRODUCTION CHECK

Can each release prove quality, lineage, cost, and latency?

FAILURE MODE

A prompt change ships without regression tests and breaks known cases.



Three controls are non-negotiable at every tier.

They must span the full system—not live in a launch checklist.

01 EVALUATION

Golden sets, regression suites, and human review.

02 OBSERVABILITY

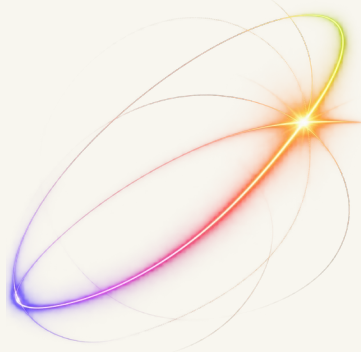
Prompt, tool, data, version, cost, and latency lineage.

03 SECURITY

Least privilege, PII handling, residency, and injection resistance.

IF YOU CANNOT TEST IT, TRACE IT, OR CONTAIN IT—

it is not production-ready.



Choose the level of control your application actually needs.

Illustrative examples—not a ranking. Graduate when control demands it.

TIER 1

READY TO USE

Start here
Prove the workflow

OPTIONS

ChatGPT
Claude
Perplexity

TIER 2

WORKFLOW BUILDERS

Graduate for triggers, integrations + approvals

OPTIONS

n8n
Make
Relevance AI

TIER 3

DEVELOPMENT FRAMEWORKS

Graduate for custom state, recovery + eval systems

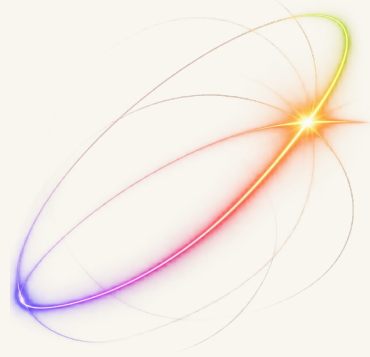
OPTIONS

LangGraph
CrewAI
LlamaIndex

NON-NEGOTIABLE AT EVERY TIER

Guardrails • Evaluation • Observability • Safe failure

Start with 01 → 06 → 08 this week.



Three checks turn the blueprint into an operating habit.

01

SCOPE

Write the success metric and prohibited outcomes.

06

RECOVERY

Choose one timeout; define rollback or human takeover.

08

EVALUATION

Build ten golden cases; block release on regression.

Where is your system weakest?

Comment: 01 Scope • 05 Memory • 06 Recovery • 08 Evals