

Building Agents with Decision Models

Thursday, October 22, 2026 11:10 am

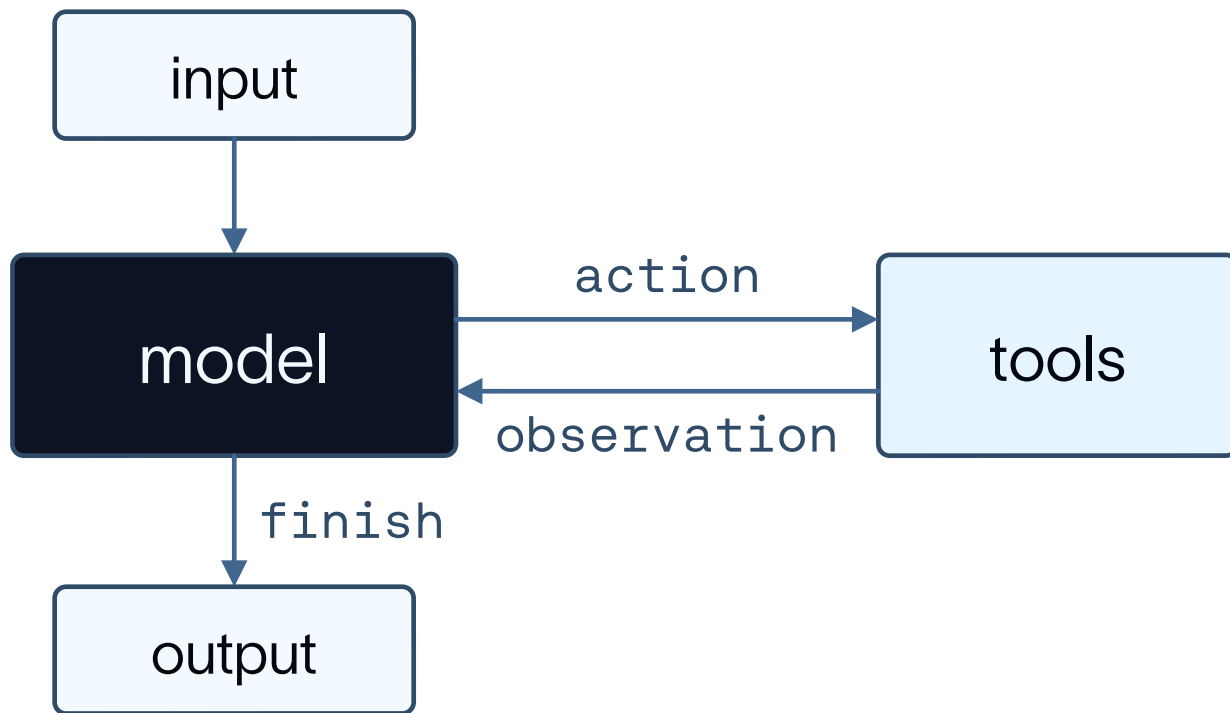


Sydney Runkle

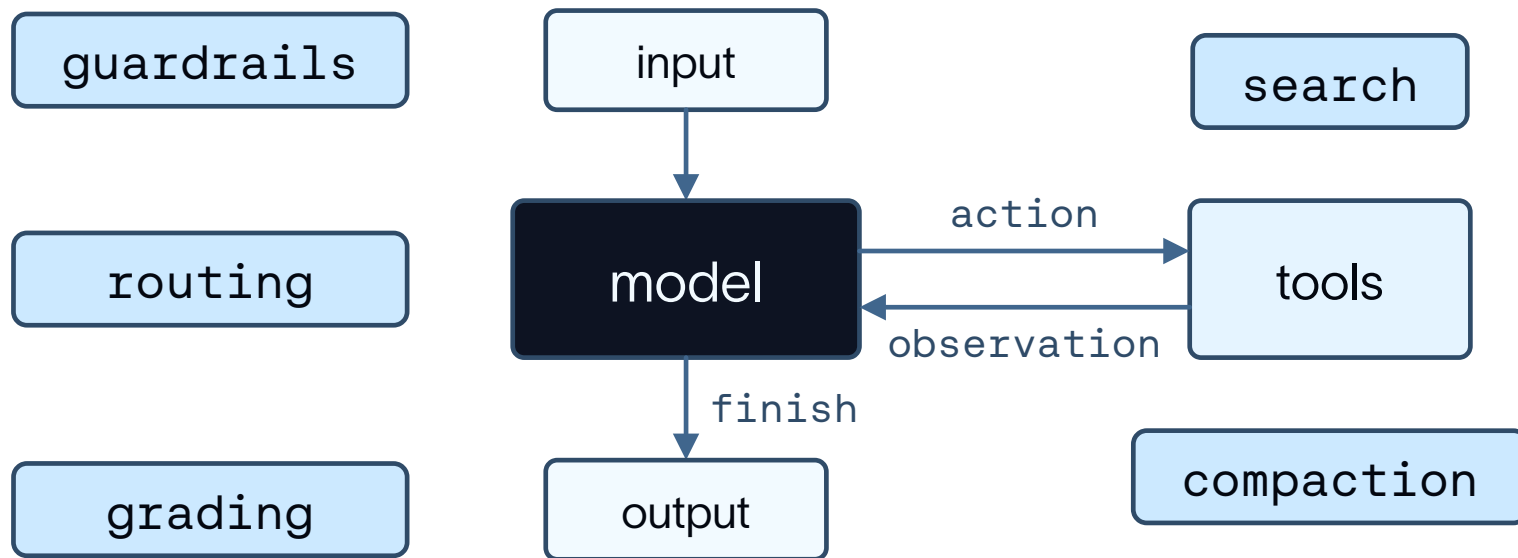
agent = model + harness



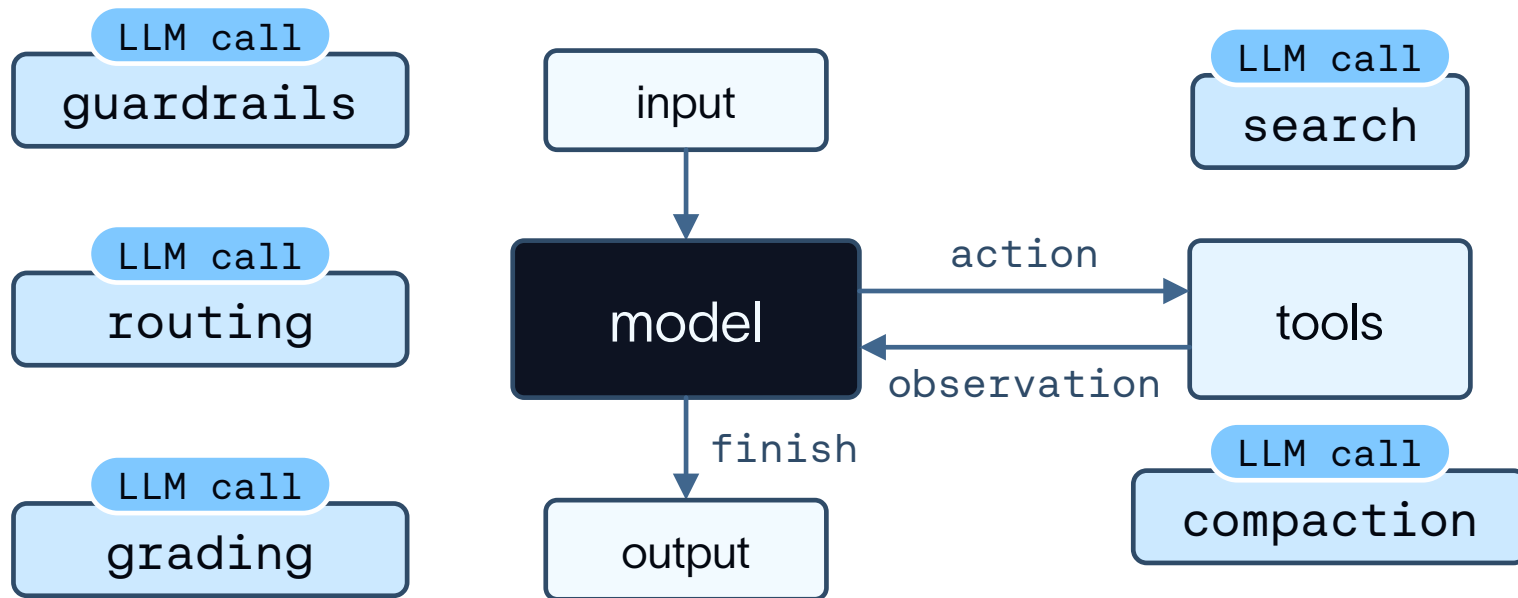
The core agent loop



We build a lot around the loop



The problem: slow and expensive



A new model category: decision models

state

Hi, I've been trying to connect my Stripe account for 3 days and it keeps failing. I'm losing sales. Please help ASAP.

question

is_urgent? (yes/no)

decision
model

probability

0.999

Jev, then everyone

200x faster

400x cheaper

Jev, TypeSafe-reported

within weeks

OpenAI Decisions

Liquid d1

Cloudflare Clef

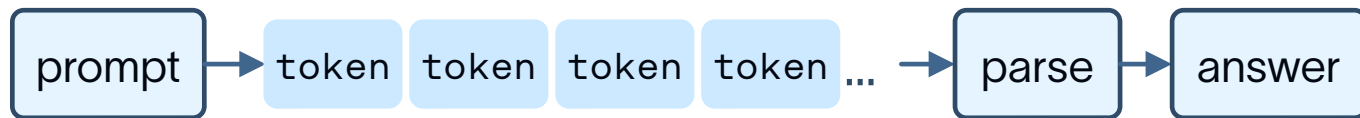
AWS Strands Decider

PostHog Jeeves

Kev (open)

How do they do it?

LLM



output tokens = time + cost

Decision
model



Unlocked: realtime processing + filtering

~100 ms per decision

search every trace

filter live chat

review every page

System one vs system two

System 1

fast, intuitive

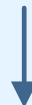


decision models

trained with RLCD

System 2

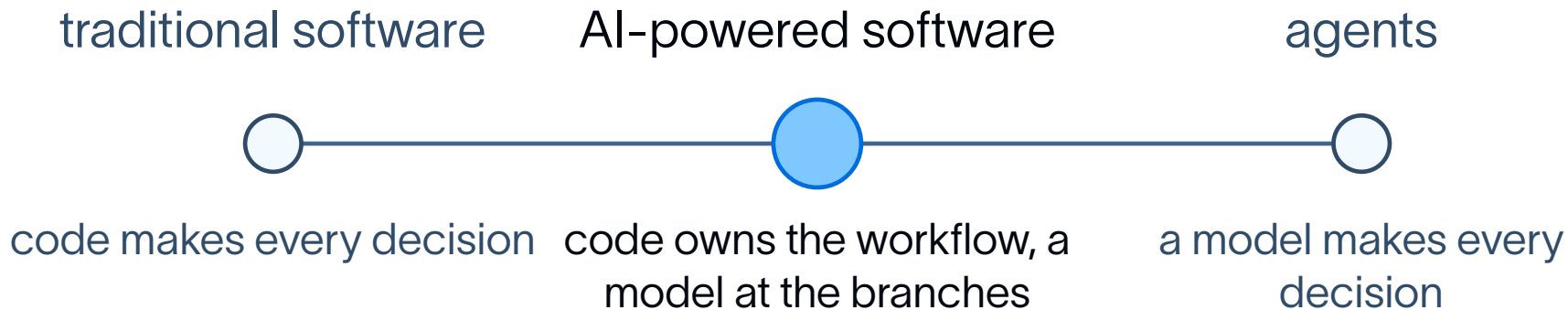
slow, deliberate



LLMs, reasoning models

trained with RLHF, RLVR

Building prod, not god



"Code owns the workflow and AI handles narrow, structured decisions."

TypeSafe

Small, legible primitives

structured

parallel

comparable

fast

calibrated

self-consistent

"Imagine if they could also branch on common sense, understanding, and intent."

TypeSafe manifesto

Decisions fit at any point in the loop

1 Model routing

4 Security guardrails

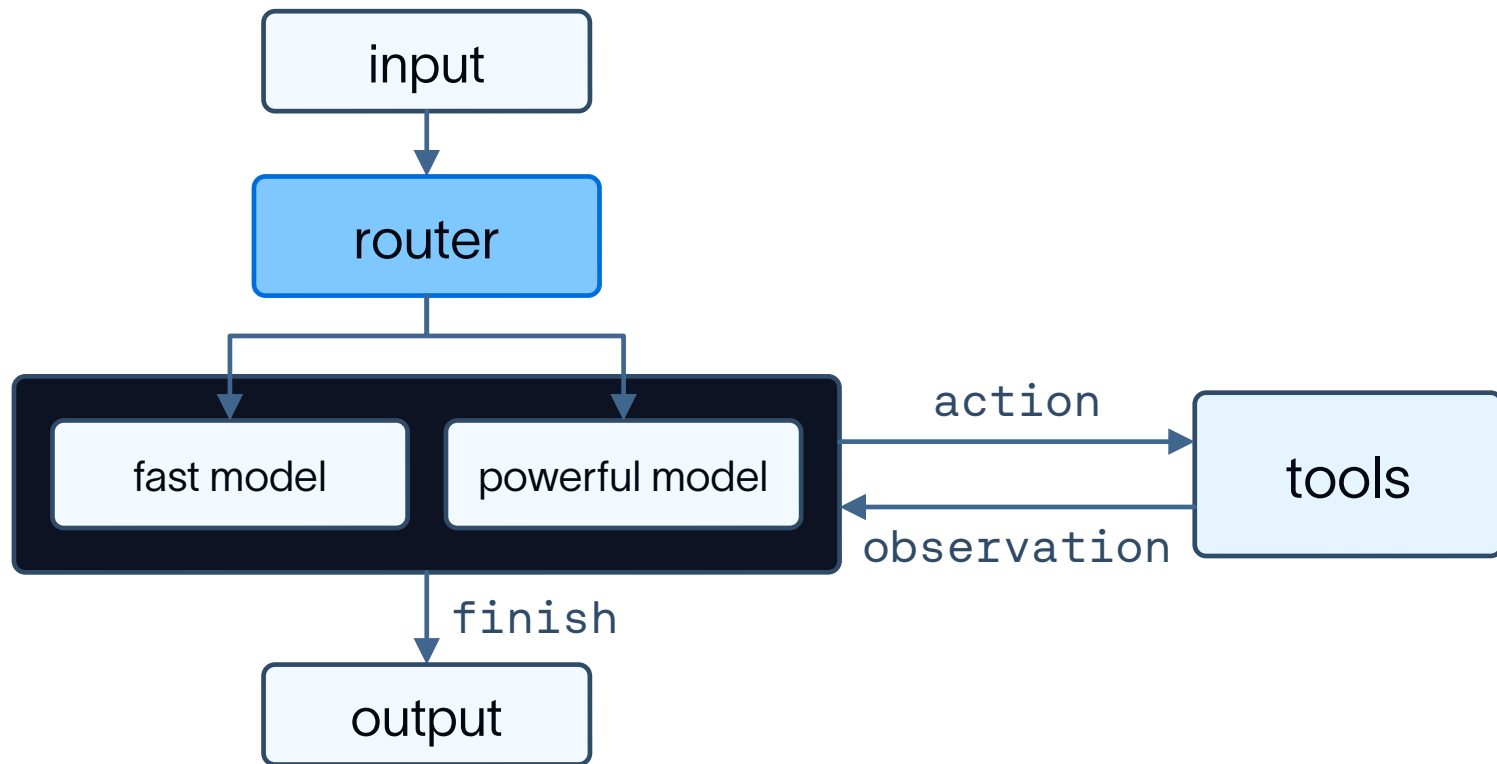
2 Tool / skill search

5 Goal loops (rubrics)

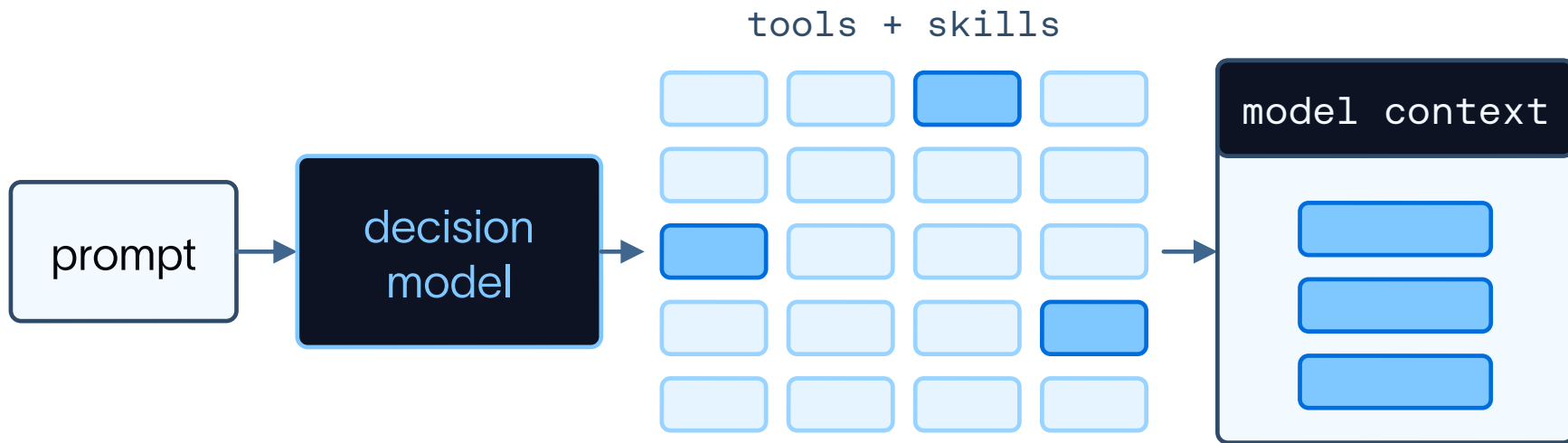
3 Tool call gating

6 Judge drift

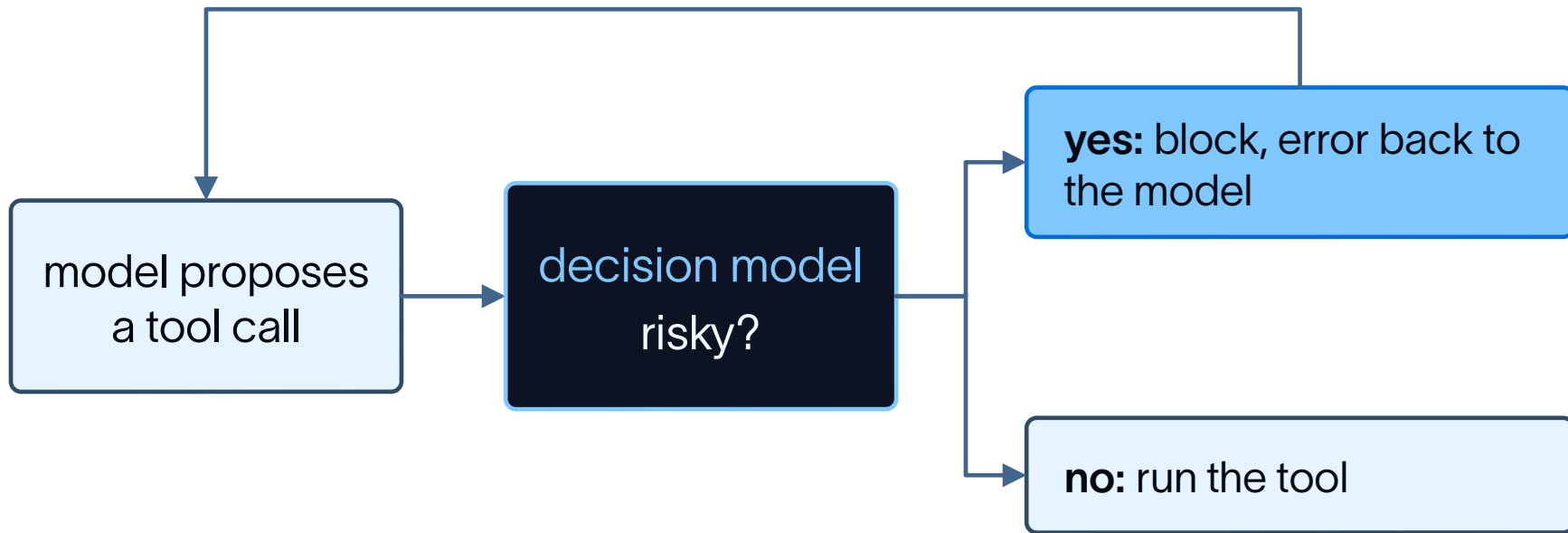
Model routing



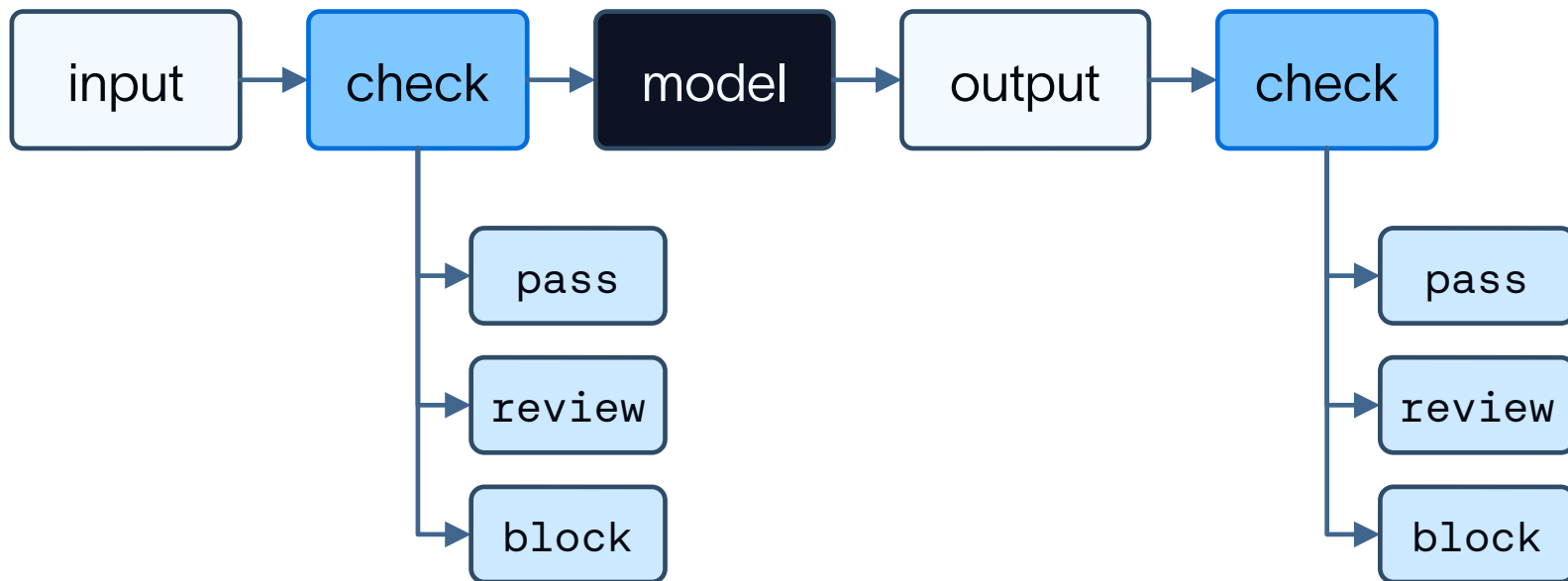
Tool / skill search



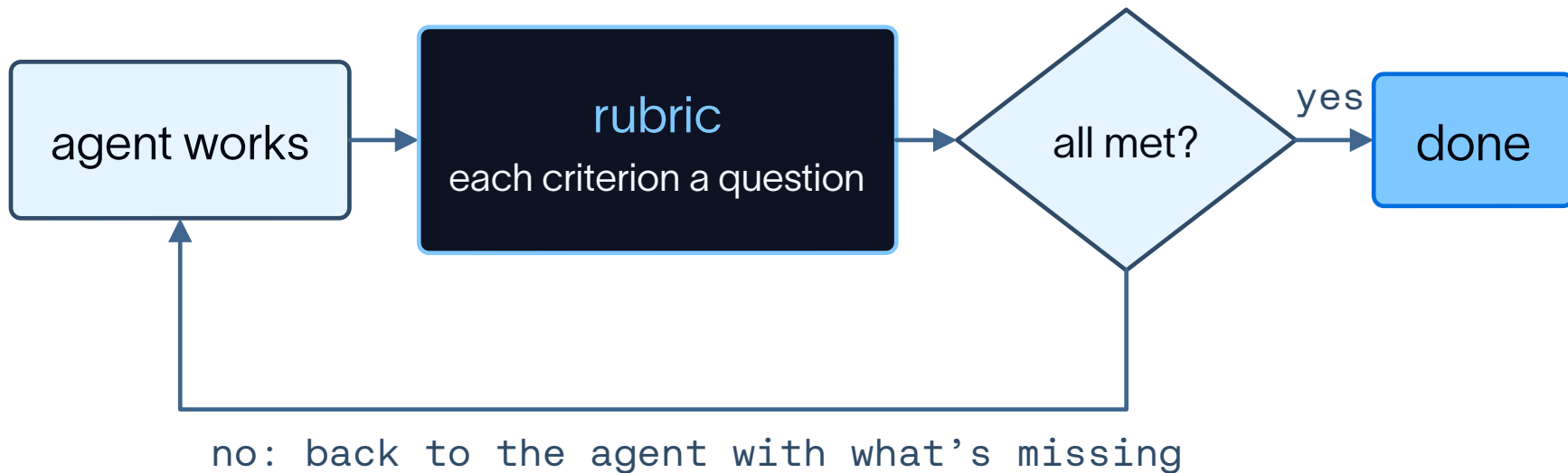
Tool call gating



Security guardrails

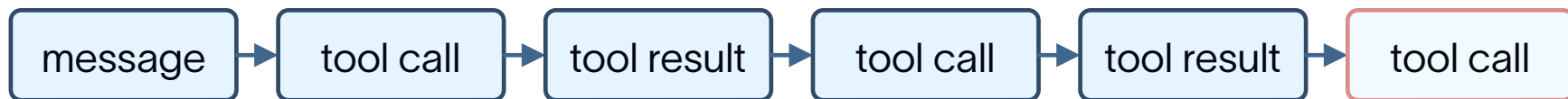


Goal loops (rubrics)

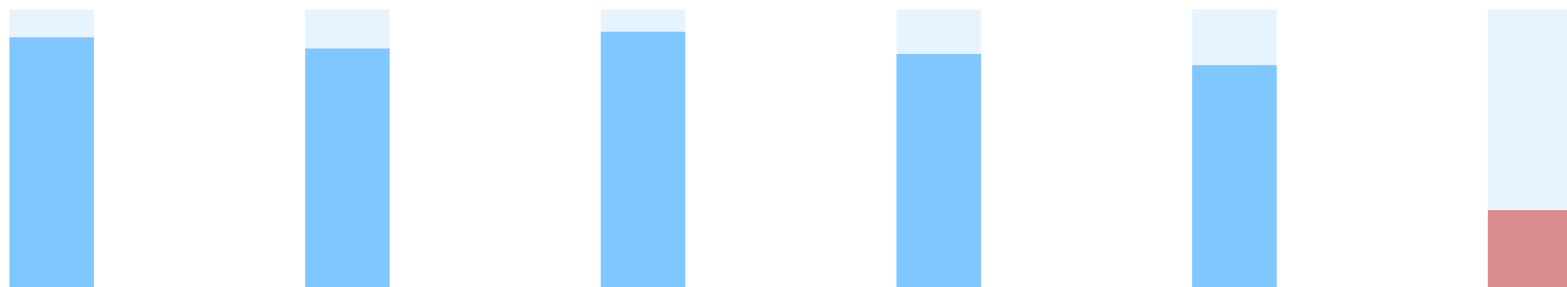


Judge drift

trajectory

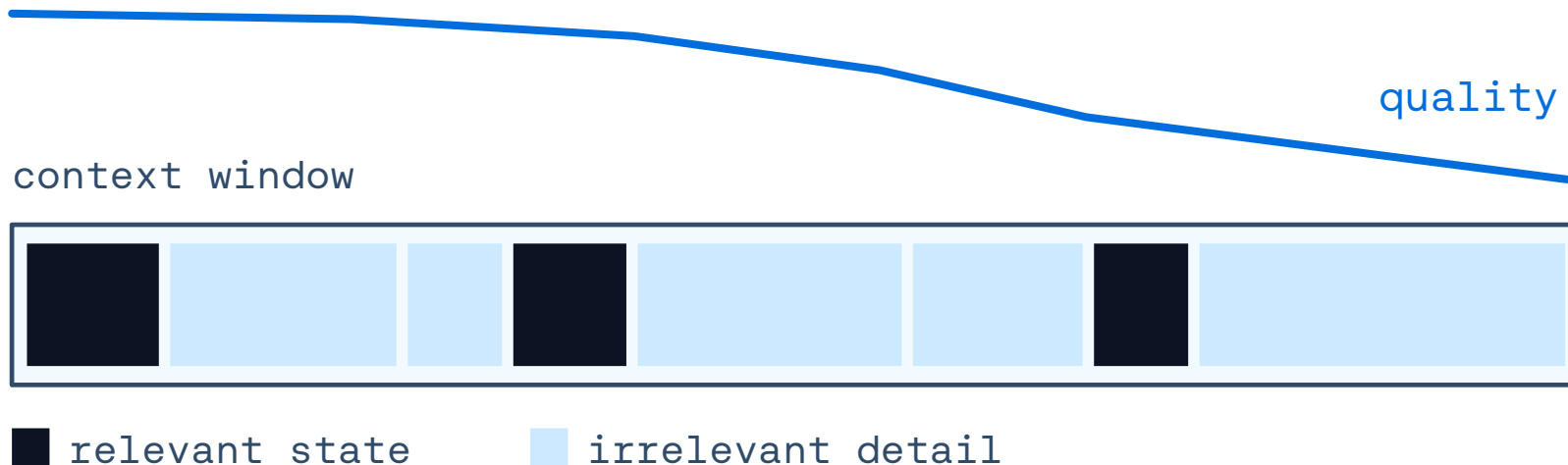


score



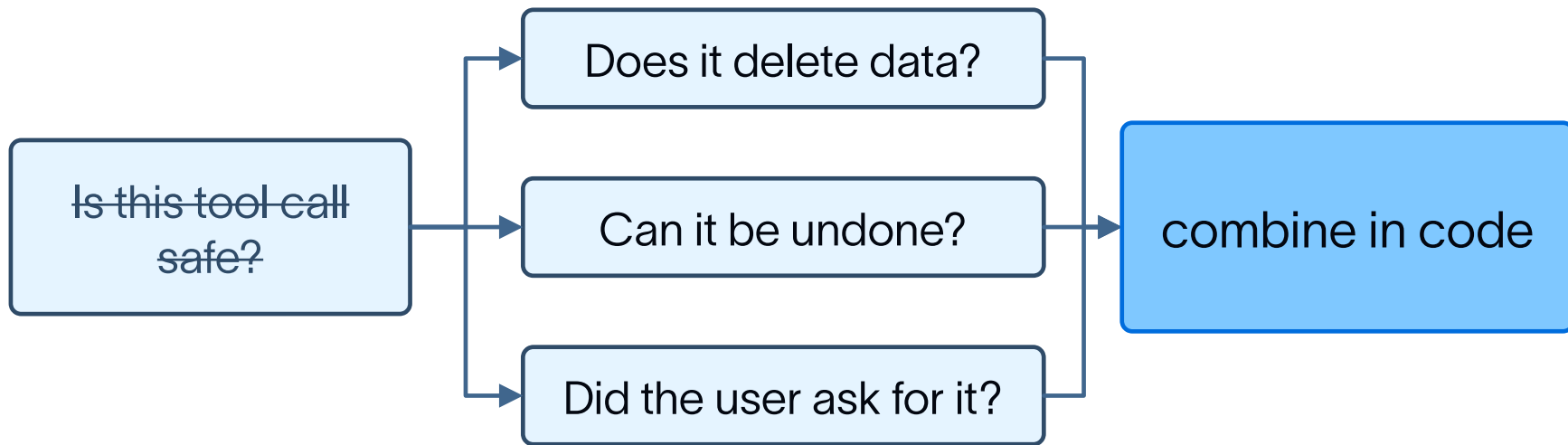
drifting

It's still context engineering

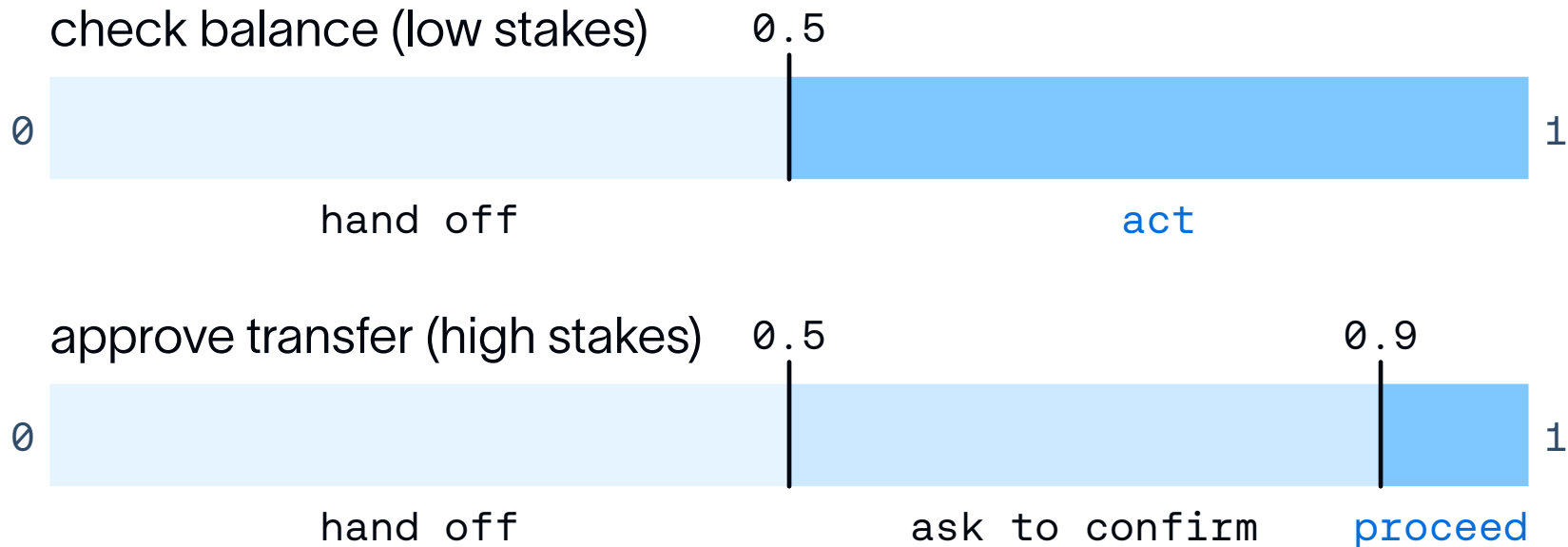


- Finite context window
- Context rot

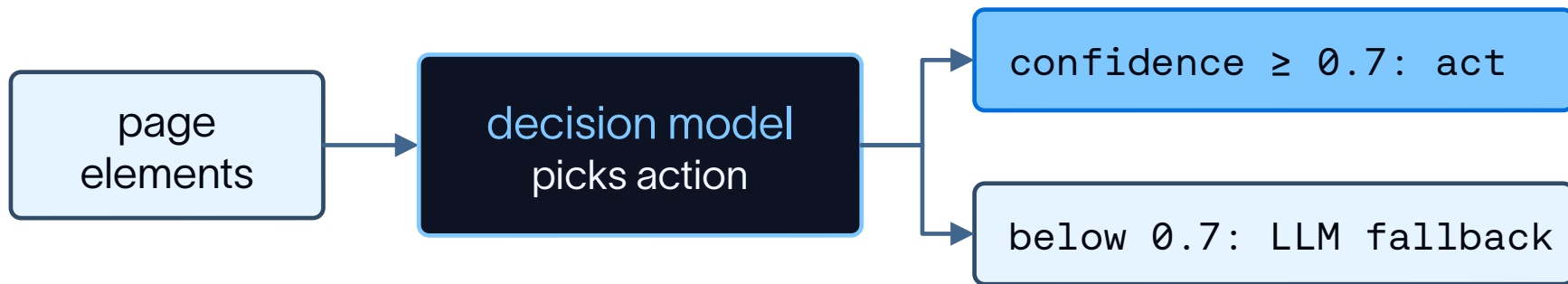
Break fuzzy problems into questions



Calibrate to scores



Cheap by default, frontier on exception



1.97s → 0.46s median act()

Stagehand from Browserbase, early testing

TL;DR

- Obsess over context quality
- Break big problems down into smaller decisions
- Calibrate to scores + test
- Progressive disclosure of intelligence (start simple)



Thank you

[@sydneyrunkle](#)

