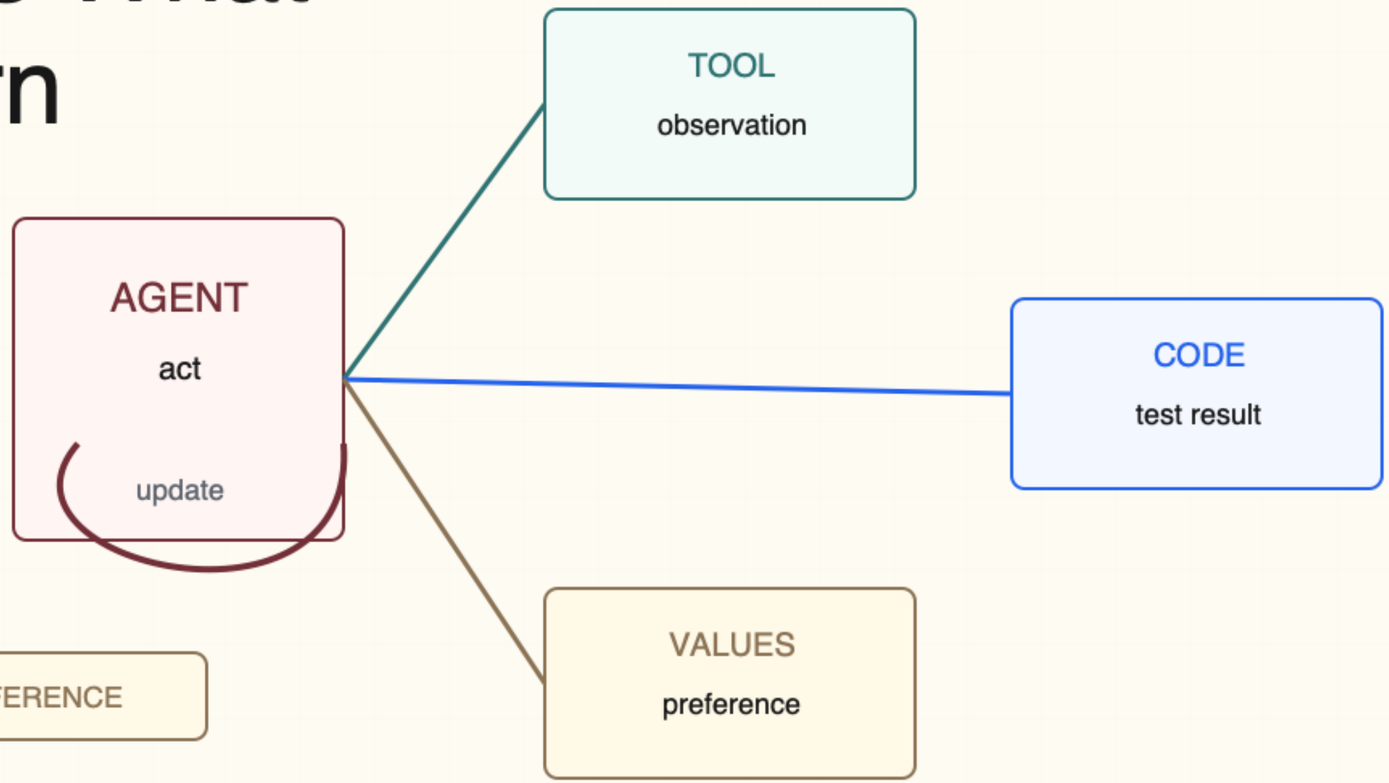
FEEDBACK CONTRACTS

PART 4 00:07-01:48 / 60:28-63:20

Feedback Changes What an Agent Can Learn

Stanford CS329A Part 4 - tools, executable tests, and constitutional preferences



OBSERVATION TEST PREFERENCE

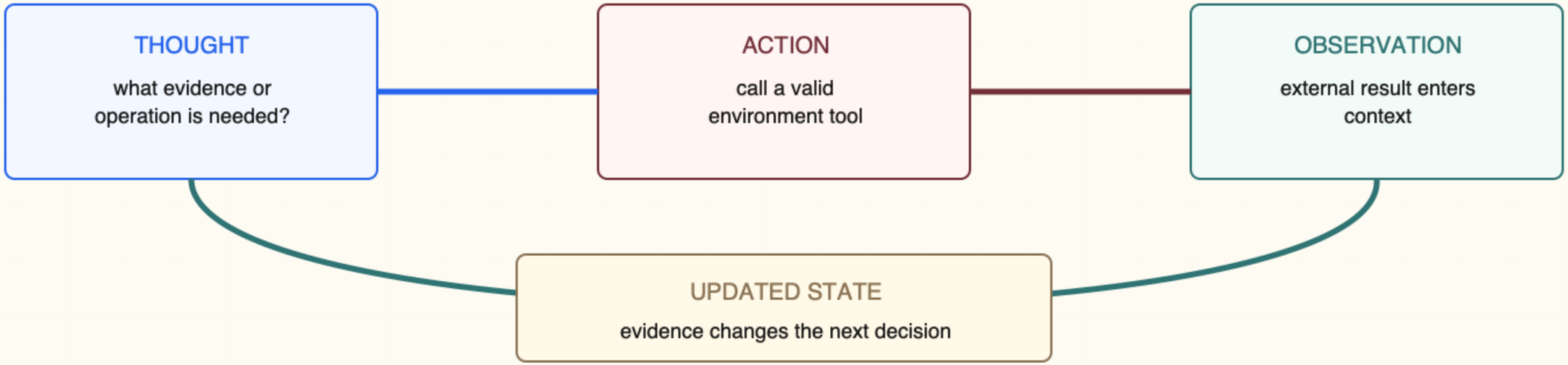
three signals != one meaning

REACT LOOP

PART 4 01:48-12:03 + arXiv:2210.03629

ReAct Lets Evidence Change the Next Decision

Thought selects an action; the environment returns an observation; the state updates.

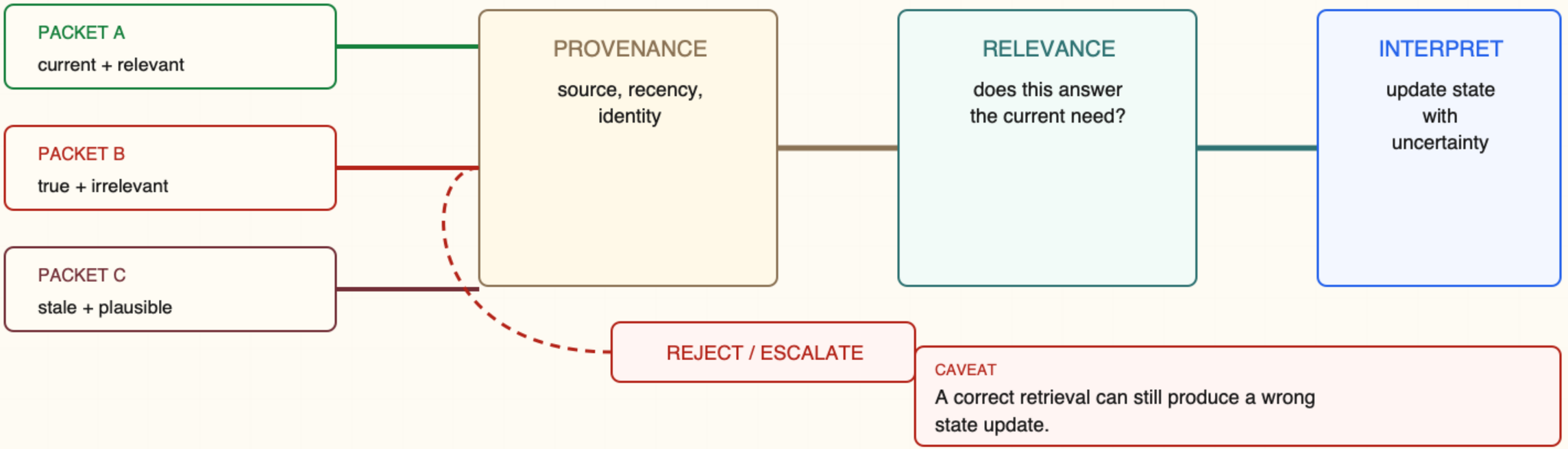


EVIDENCE BOUNDARY

PART 4 09:07-12:54 / 18:37-20:04

A Tool Observation Is Evidence, Not Truth

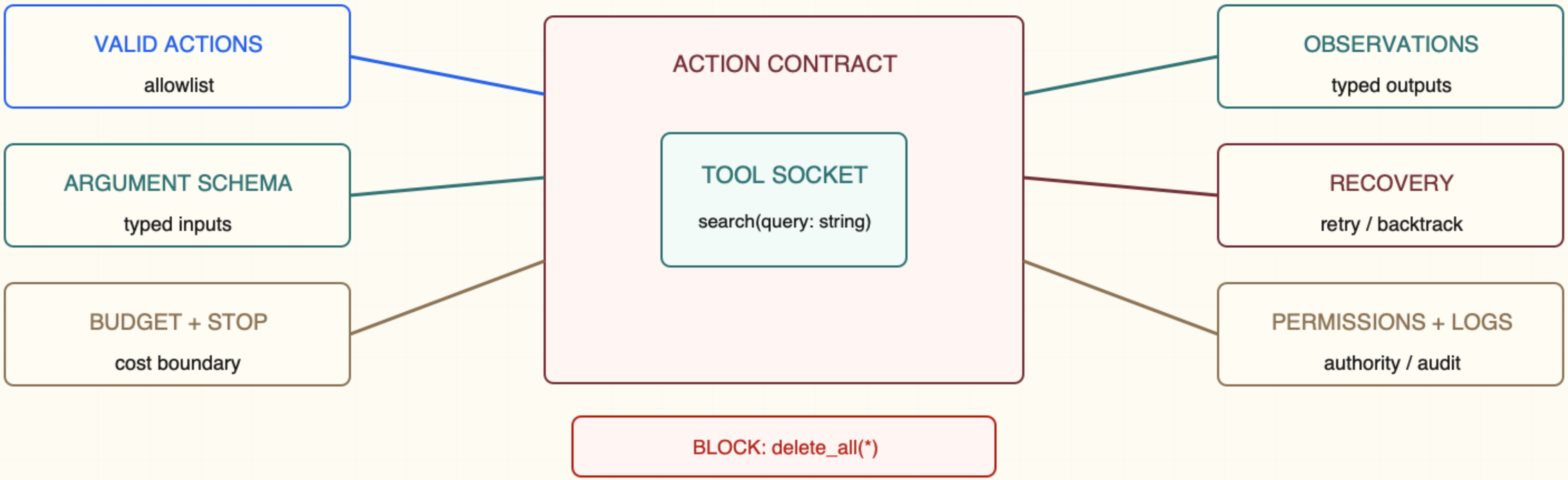
Retrieval can be relevant or misleading; interpretation still determines the state update.



TOOL BOUNDARY

Tool Use Needs an Explicit Action Contract

The operating boundary must define schemas, observations, budgets, recovery, and permissions.

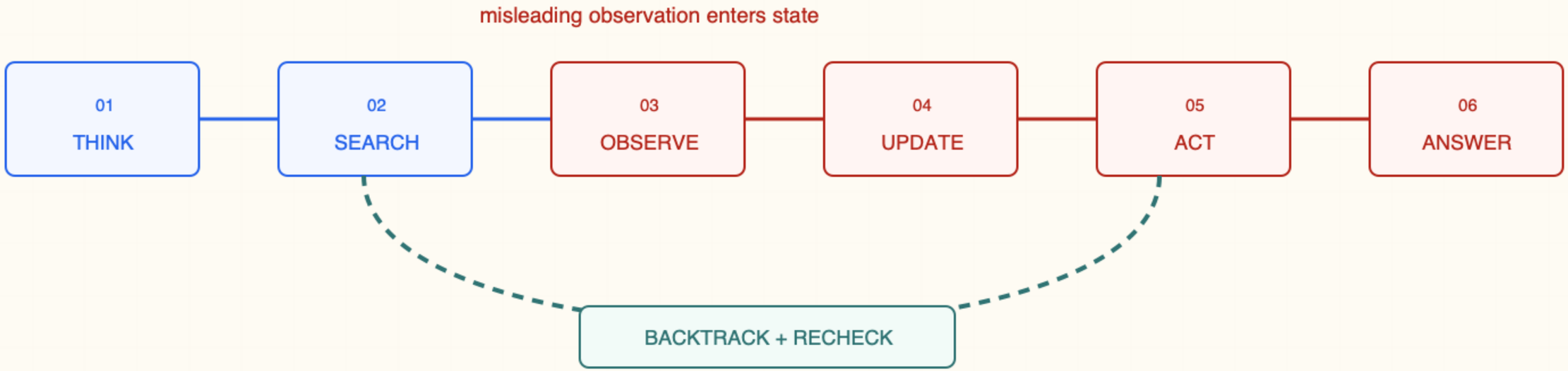


TRAJECTORY RISK

PART 4 18:37-27:32 + arXiv:2210.03629

One Early Error Can Contaminate the Whole Trajectory

Longer loops create more chances to recover and more chances to compound a mistake.



CAVEAT

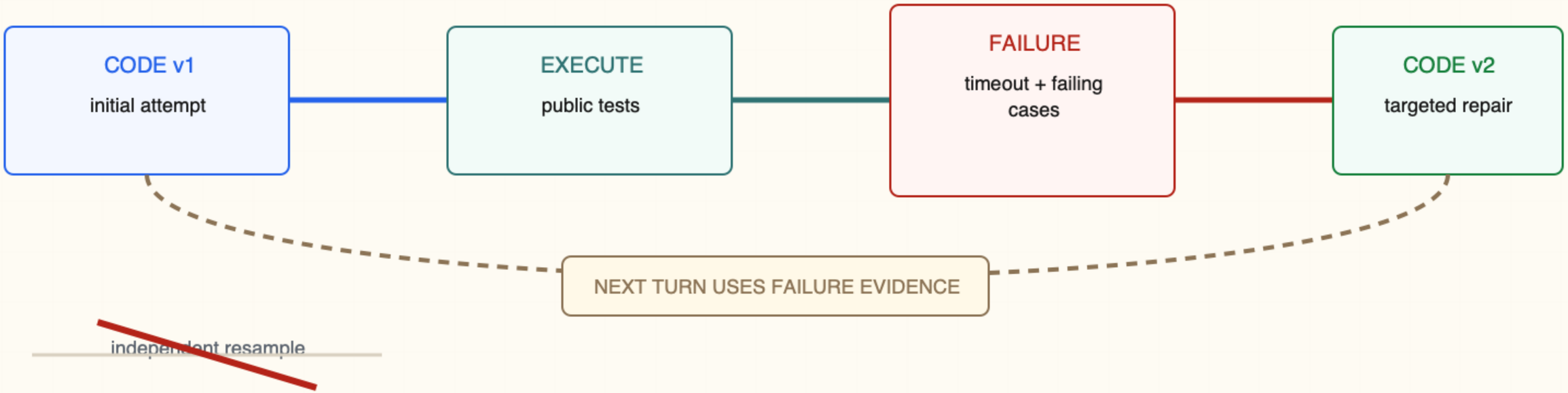
Longer trajectories add both recovery opportunities and failure surface.

EXECUTION FEEDBACK

PART 4 27:32-37:47 + arXiv:2410.02089

RLEF Teaches Code Models to Repair, Not Just Resample

A new attempt is conditioned on execution evidence from the previous attempt.

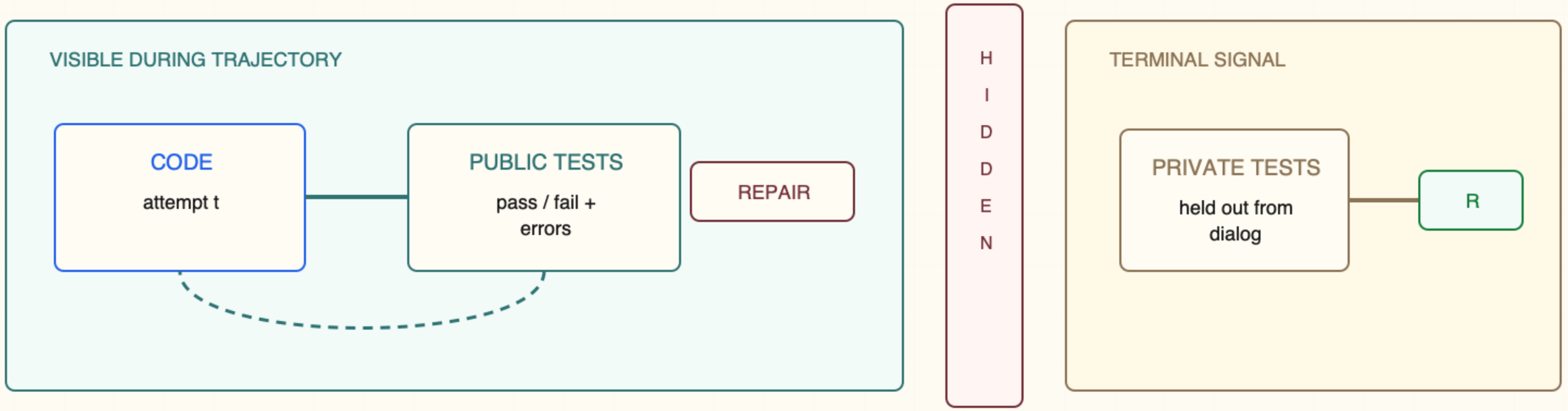


TEST VISIBILITY

PART 4 30:02-34:36 + RLEF Sections 2.1-2.2

Public Tests Teach the Repair; Private Tests Guard the Reward

RLEF separates in-trajectory feedback from the terminal correctness signal.



CAVEAT

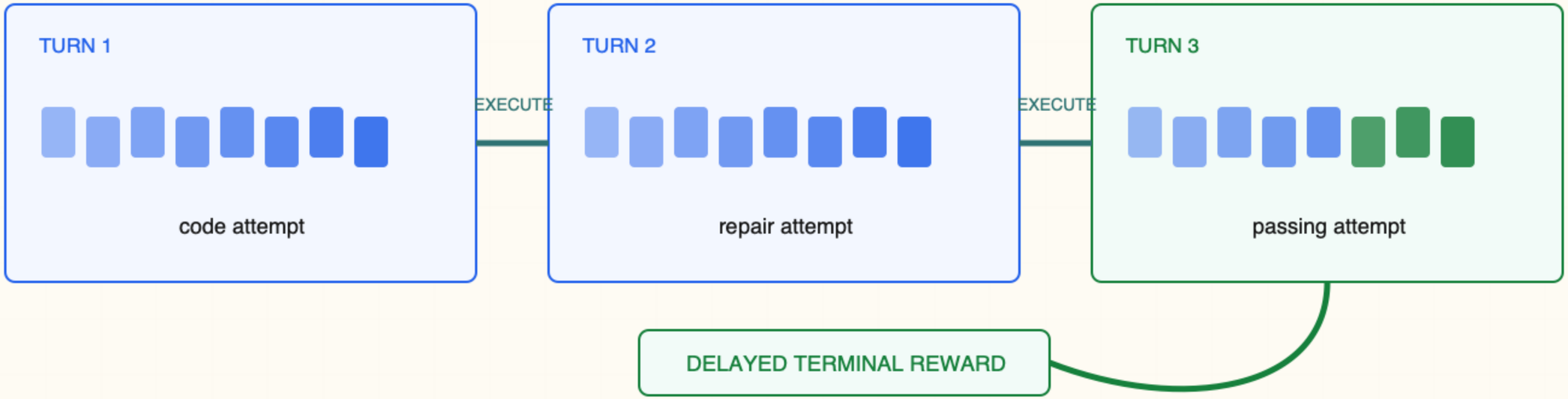
Hidden tests reduce direct exposure; they do not make the specification complete.

CREDIT ASSIGNMENT

PART 4 31:57-34:36 + arXiv:2410.02089

Credit Assignment Spans Tokens and Turns

Code is emitted token by token; execution feedback arrives after a complete attempt.



CAVEAT

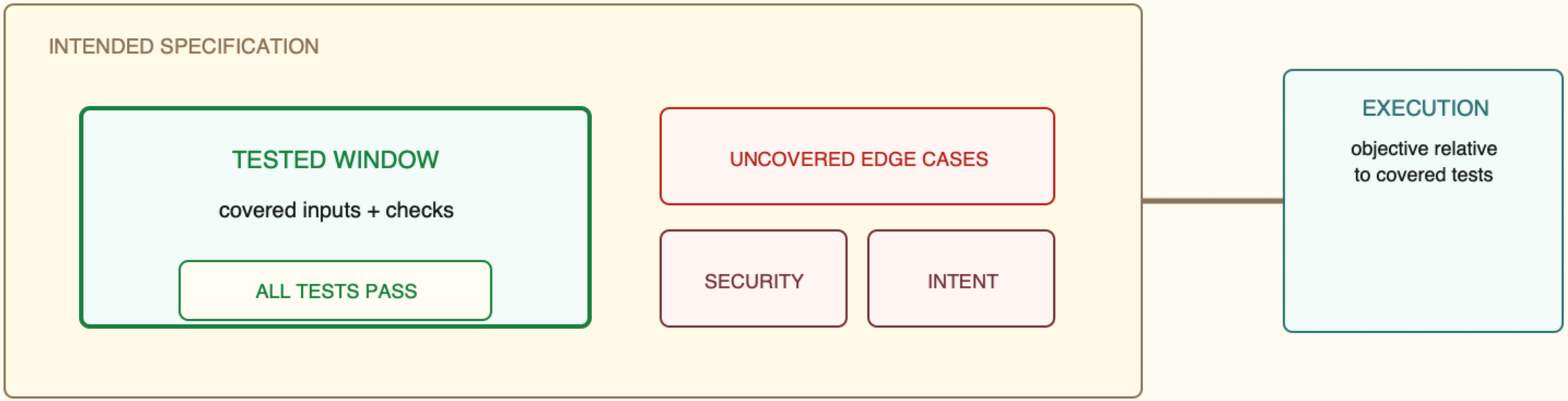
One terminal result must assign credit across many token decisions.

SPECIFICATION GAP

PART 4 33:40-46:30 + arXiv:2410.02089

Executable Feedback Is Objective, but Narrow

Passing tests proves covered behavior, not complete correctness, security, or intent.



CAVEAT

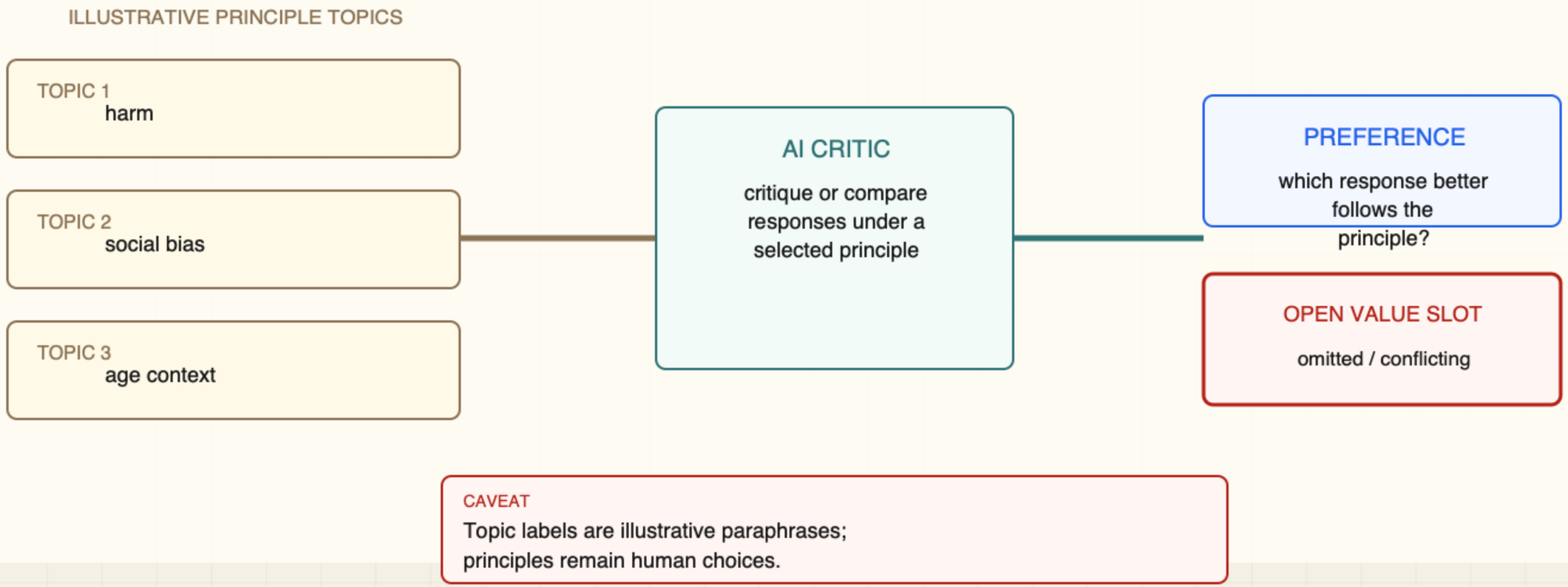
Passing tests is evidence of covered behavior, not proof of complete correctness.

NORMATIVE FEEDBACK

PART 4 46:30-50:42 / 53:20-57:14

A Constitution Makes Normative Feedback Inspectable

Human-written principles guide an AI critic, but they do not remove human judgment.

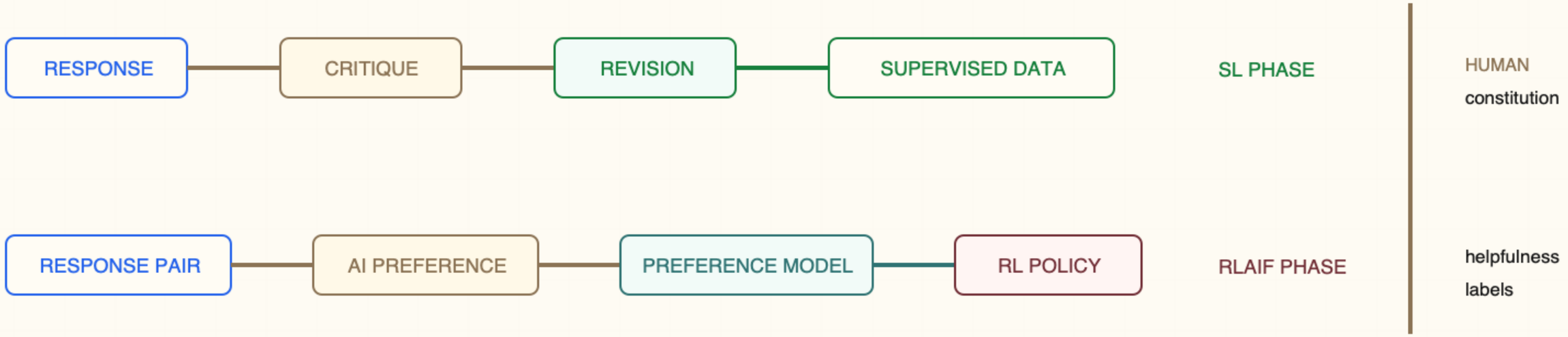


CONSTITUTIONAL TRAINING

PART 4 47:49-60:28 + arXiv:2212.08073

Constitutional AI Trains Through Revision and Preference

Supervised critique revises responses; RLAIF turns AI comparisons into a learned reward.



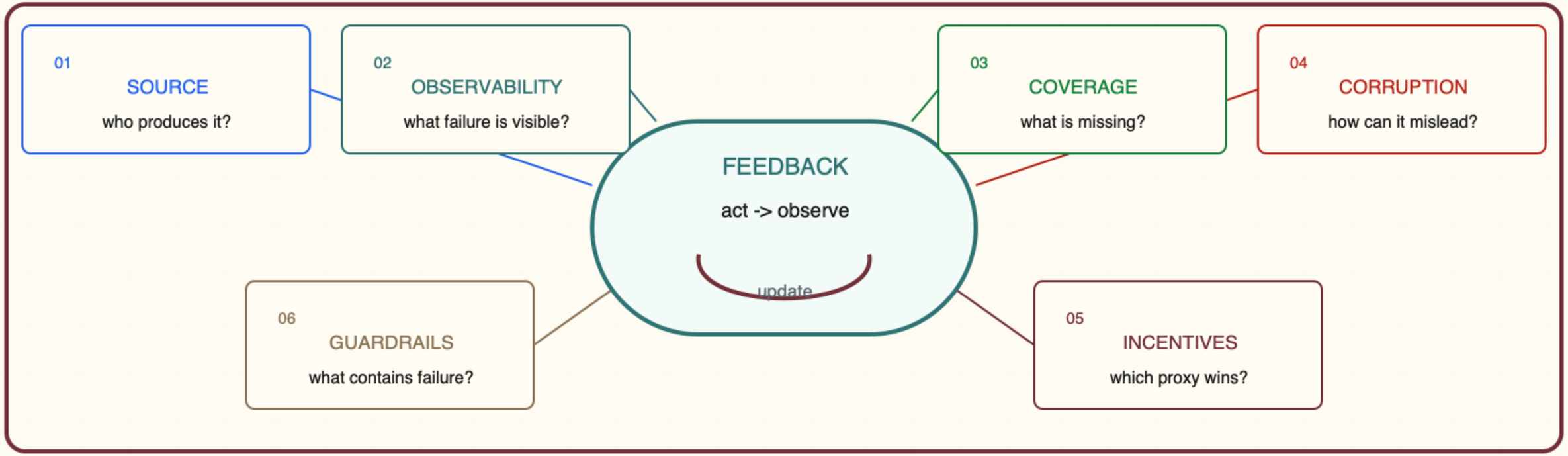
CAVEAT

AI feedback scales labeling; it does not remove humans or evaluator blind spots.

FEEDBACK AUDIT

Audit the Feedback Contract, Not Just the Model

Source, observability, coverage, corruption, incentives, and guardrails bound self-improvement.



SELF-IMPROVEMENT IS BOUNDED BY THE FEEDBACK CONTRACT