Inside stilltrue

Three different questions — the package answers two, delegates the third, and **the AI never lives inside the package**.

QUESTION 1 · DRIFT

“Is this exact fact still on the page?”

String matching. You stored “Radloff” and “141 schools” when you curated; drift re-fetches the page and checks the strings still appear. Deterministic — same page, same answer, every time.

QUESTION 2 · VERIFY

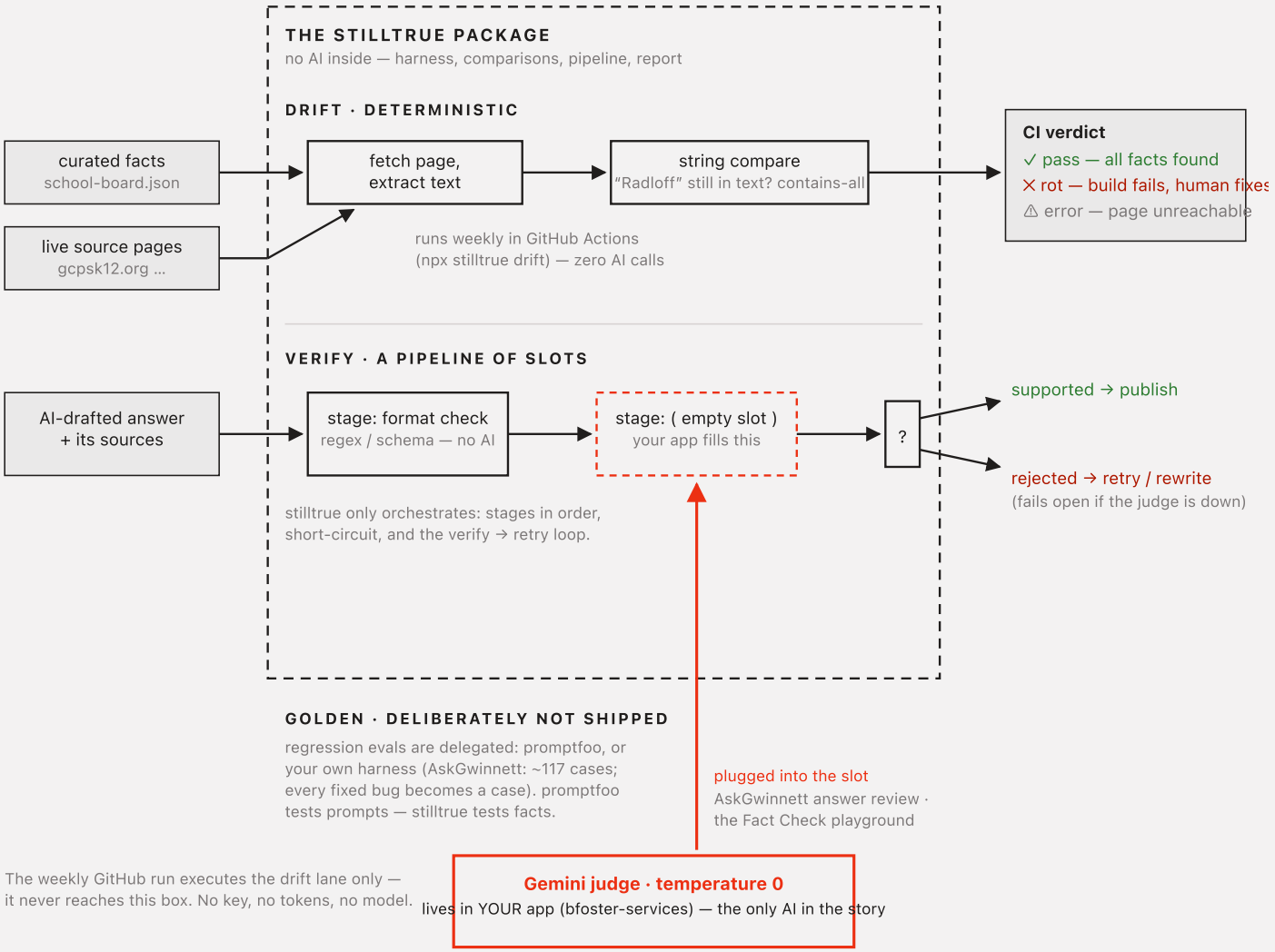
“Does this prose support that claim?”

Reading comprehension. “Meets Tuesday” vs “moved permanently to Thursday” never string-matches — so your app plugs a judge model into the pipeline. The model is a stage you supply, not part of stilltrue.

QUESTION 3 · GOLDEN — DELEGATED

“Did this change break an answer?”

Regression evals — deliberately *not* shipped. promptfoo (or your own harness) owns this lane: promptfoo tests your prompts, stilltrue tests your facts. AskGwinnett runs its own ~117-case harness; every fixed bug becomes a case.



The dashed boundary is what ships on npm: the drift engine (pure string matching) and the verify pipeline (empty stage slots) — golden is deliberately left to promptfoo or your own harness. The one red box — the Gemini judge — sits outside it, in bfoster-services, and is plugged into a slot only for jobs that need reading comprehension.

WHERE IT RUNS	WHICH LANE	AI CALL?
GitHub Actions, weekly (AskGwinnett curated facts)	drift → report	**No — string matching only**
AskGwinnett, before publishing each answer	verify, judge stage supplied by the app	**Yes — Gemini, temp 0, fails open**
rabinforest.com Fact Check playground	the verify judge as an interactive demo	**Yes — same judge pattern**
`golden` regression evals	delegated — promptfoo / your own harness, not part of the package	**No AI in the checker itself**

stilltrue — GitHub · npm. Your tests check your code — stilltrue checks your facts.