BAILEY FITCHETT
Senior DevOps / Platform Engineer

AI review with engineering guardrails.

INDEPENDENT CASE STUDY / LOCAL AI REVIEW

12/12	13/13	Advisory
synthetic cases passed	validator probes rejected	policy owns the decision

The problem and my contribution

A plausible model response can omit resources or ask the wrong questions about replacement order. I built a Python review adapter around Terraform plan JSON, a local Ollama model (Qwen 2.5 3B), constrained question IDs and retained evaluation evidence.

Key decisions

Constrain the request. Send generated evidence IDs and action metadata; keep resource addresses, attribute values and arbitrary plan strings outside the model request.

Validate before acceptance. Check exact resource coverage, unchanged facts, allowed question IDs and mandatory questions. Reject malformed responses; keep the deterministic policy decision in the report.

Learn from observed failures. Incorrect replacement questions and omitted evidence led to resource-specific allowed and required IDs, then a schema requiring every evidence ID.

Recorded results

One local run passed 12 synthetic cases, including three valid empty-findings cases with policy warnings. Separately constructed validator probes were rejected in all 13 cases. A separate four-resource recording retained full response coverage and a review-required decision.

Scope and limits

The output supports human review and does not deploy infrastructure. The small synthetic evaluation measures contract compliance; broader model quality and reviewer time savings remain unmeasured. Validator probes are deliberately corrupted inputs, not model-generated failures.

Read the full story and sources

Full case study: baileyfitchett.com/ai-review/

Recorded evaluation (GitHub, source pinned in link)