

ANCHOR PROJECT · ACCEPTED · ONE-PAGE SUMMARY

SustainAI: Predictive Sustainment & Supply-Exception Triage

Built for the problem I lived as a Marine logistician (0431): too many exceptions, too little time to find the ones that threaten the mission. An RUL model predicts turbofan degradation on NASA C-MAPSS FD001, deterministic thresholds band exceptions by severity, and a single Gemini-backed agent triages each one behind hard guardrails. Every number below carries a receipt: accepted eval 24243a5a, reproduced bit-identically as eval 8a40e13a; live run eval 3477d2bf.

Scope: One dataset · one pipeline · one agent · one eval harness

Stack: Python · scikit-learn · Pydantic · Gemini 2.5 Flash on Vertex AI · Terraform · GCP

Build: Autonomous coding agent (Google Antigravity), six adversarial audit rounds

Repo: github.com/burnsgregm/SustainAI

MEASURED RESULTS (EVAL 24243A5A; LIVE RUN EVAL 3477D2BF)

0.842

Critical-class recall

0.941

Critical-class precision

0.889

Critical-class F1

17.6

RMSE, cycles (MAE 13.3)

16 / 19

True criticals caught; 3 missed, 1 false alarm, all itemized by unit

33 / 33

Live triages, 0.0 failure rate, 100% schema validity

8.8s / 12.0s

Latency p50 / p95 vs 15s target (PASS)

Cents

Total Vertex cost, single digits

TWO EXHIBITS

- **Unit 81, guardrail overrules agent:** the model flagged it critical; the agent honestly assessed low risk and recommended `continue_monitoring`; guardrail FR-GRD-1 forced escalation anyway (`forced_by_guardrail=true`) and preserved the agent's dissent for the operator. The agent gets a voice, the guardrail gets the final word, the human gets both.
- **Unit 90, agent catches what the threshold missed:** banded as a warning, the agent found a sudden 25% N1 vibration spike, upgraded to critical, and escalated on its own judgment (`forced_by_guardrail=false`).

HOW THE NUMBERS EARNED TRUST

Built by an autonomous coding agent through six adversarial audit rounds. The review loop caught three fabrication modes (invented compliance claims, relocated template text, prose contradicting its own artifacts) and one methodological defect (degenerate validation producing trivially perfect numbers, fixed via validation truncation). A live-integration arc took the Gemini path from hard crash through JSON truncation through a field-name mismatch to the clean 33/33 run; every intermediate failure was caught by a guardrail and failed safe. Disclosed deviation: a TensorFlow DLL failure forced an MLPRegressor substitute for the planned Keras champion; the HistGradientBoostingRegressor baseline beat it on honest validation and shipped.

BUSINESS IMPACT

The eval harness plus adversarial review, not the agent's self-reports, are what made these numbers trustworthy. That finding generalizes to every agent-built system an organization will deploy. The honest 3-engine miss list is what makes the 16 catches believable, and preserved dissent under guardrail overrule is what makes the system safe to put in front of an operator.