

Asset Intelligence Copilot

the model plans, code calculates

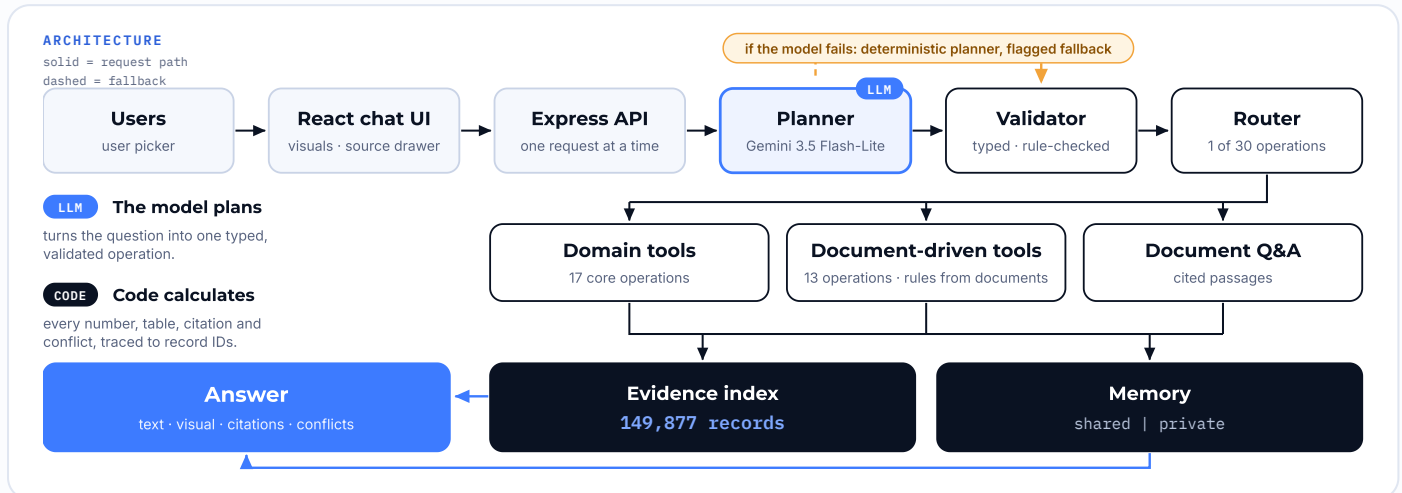
github.com/GODOSTROYER/asset-intelligence-copilot

Arnav Bule

17 Oct 2026

MIT licence

The model plans, code calculates. An LLM turns a plain-English question into **one typed, validated operation**. Deterministic tools then compute every number, table, citation and conflict from the loaded records and documents. The model never does arithmetic on domain data, and every answer can be traced to source record IDs.



01 System design

Storage & index 01

At startup every JSON/JSONL record and Markdown document (manuals, SOPs, MOC memos, recall notice, regulator letter, repair reports) is loaded and indexed as typed evidence with record ID, well, type and path. The test set loads in about 4 s: 600 wells, 7 areas, 149,877 evidence documents. Optional fields (type and team history, late-entered records, data corrections, plugged wells) are respected.

Planner 02

Picks 1 of 30 operations: 17 core (status, trends, spills, downtime, heatmap, compliance, containment, RCA, costs, shared knowledge, document Q&A, and 4 writes) plus 13 document-driven tools added for new data (including a conversation-wide **triage**). It sees only areas, users, well-ID ranges and the asking user's last 5 exchanges.

Validator 03

Rejects unknown wells, areas, dates and fields. Explicit values in the question override the model's. Fleet tools take well and area scope only from the question, not from earlier answers. Year-less past dates never resolve into the future. **Writes need explicit declarative feedback** detected by rules; questions, quotations, hypotheticals and "don't save" never write.

Rules from documents 04

Deadlines, thresholds and approvals are parsed from the documents, never hard-coded: **OPS-1** gas response (24 h, then 12 h from 1 Jan 2027), **OPS-3** status precedence, **OPS-5** cold weather, MOC validity and extensions, recall model and lots, the regulator letter's scope, and per-type production thresholds (**T200/T350/T500**). Each rule is cited.

Memory 05

Shared: corrections (original value, author, reason, scope), notes, **unconfirmed claims** (e.g. "pump was swapped": kept attributed, never overwrites records, flagged when second-hand), and well-type rules. **Private:** per-user history and **display preferences** (sort order, \$k). Stored in `state.json` with atomic saves.

Conflicts & honesty 06

Shows disagreements instead of picking a side: claim vs record, note vs inspection, and second-hand reports. When the data can't answer, it says so: no pressure data, no production volumes. It offers labelled proxies instead.

Resilience 07

A deterministic planner takes over if the model fails, and answers are flagged `fallback`.

MODELS AND COST

Gemini 3.5 Flash-Lite (planner and synthesis), about 2–5 s per hosted answer, on free-tier keys with app-side rate caps and a 10-minute plan cache. Gemma 4 was tested as a drop-in planner with the same prompt and validator: every plan that returned was correct, but free-tier capacity (429/500/503) and 40–60 s latency ruled it out for the live run.

VERIFICATION

Sample: 25/25 acceptance checks over the 18 scenarios, 109/109 independent domain checks, 27 provider checks, and adversarial write-routing probes. Test set: every final answer was checked against an **independent oracle** computed straight from the raw files.

NEXT STEPS

Planner-level multi-step plans (chaining tools generally, beyond triage); entity state over time as a first-class store (installed parts, approvals); a semantic document-rule extractor so new SOP clauses become checkable rules without code; and a UI for confirming or rejecting claims.

Results on the 11 final questions

Test dataset (600 wells, today = 15 Jan 2027), questions run in order exactly as written. **All 11 questions answered correctly against independent oracles; 19 of 21 turns ran on the hosted planner and 2 on the deterministic fallback, with identical tool output.** "Oracle" means the expected answer, computed independently from raw files.

11/11**Questions answered correctly**

all match the oracle

19/21**Turns on the hosted planner**

2 on the deterministic fallback

0 → 11**First live run to final**

questions correct, same 11-question set

149,877**Evidence records indexed**

600 wells · 7 areas · every answer cited

QUESTION	COPILOT ANSWER (KEY FACTS)	VS ORACLE
1 Reservoir pressure	Declines to map it: no reservoir, tubing, casing or wellhead pressure exists. Lists what does exist: flowline test pressures 188–400 psi in 2,432 repair reports, separator alarms with no values (CTX-0422-003 , CTX-0481-006), and equipment ratings. Says what data is needed.	Correct
2 Best producer 2026	No production volumes exist , so it won't name a winner. Labelled proxy: producing weeks by the type rule. 52 wells tie at 53/53; downtime tie-break gives WELL-0147 and WELL-0315 .	Correct
3 Gas reports since 1 Nov	26 reports: 17 Met, 9 Missed (0163 , 0341 , 0332 , 0401 , 0052 , 0362 , 0161 , 0599 , 0516). OPS-1 : 24 h, 12 h from 1 Jan 2027; drone flights don't count as a response. Full table.	Correct exact
4 Flowline bypasses	In place: WELL-0248 Allowed (MOC-2026-031 extended by 044 to 28 Feb), WELL-0249 Expired (MOC-036 ended 15 Dec), WELL-0405 Not approved . Past violations: WELL-0087 , 0158 (expired), 0561 (never approved).	Correct exact
5 Regulator letter	Covers NE Blocks 7–8, ≥ 5 bbl, since 1 Jun 2026: 6 spills. Compliant: 0119 , 0168 , 0277 . Not compliant: 0123 , 0182 , 0262 (no "soil samples sent to the lab"). Also flags SR-01/02/03 timing breaches.	Correct exact
6 Which status is right	17 wells. Ten where a newer note wins: 6 shut in on 14 Jan (0044 , 0323 , 0355 , 0422 , 0479 , 0481), and 4 restarted as "Producing (unconfirmed)" with an in-progress job. Seven where a newer inspection wins. Type history (WELL-0010 now T350) and data corrections handled.	Correct
7 Operator noises (2 turns)	175 wells noted → filtered by "no repair since" (12 excluded) → ranked by a printed score: an overdue mandated check (T500 manual \$5 ESP check) outranks a not-yet-due one, plus recalled pump past its deadline, decline length and size, repeat failures. All must-have wells are in the top 10. Crew picks: WELL-0263 , 0219 , 0202 ; 0202 is chosen with 0219 because they are in the same block, so one trip covers both.	Correct same set as oracle (order is a judgement call)
8 Who told you that (5 turns)	t1: Nathan's swap recorded as an unconfirmed claim ; WELL-0333 stays listed. t2: private preference saved. t3 (Robert): 9 wells , 0333 flagged "reported replaced by Nathan, no work order"; Nathan's preference not applied. t4: Olivia's second-hand claim recorded but 0577 not removed . t5: Nathan's 4 NW wells sorted, in \$k (\$24.5k / \$15.4k / \$24.0k / \$27.1k). Sources stated each turn.	Correct
9 Six topics (6 turns)	t1: 9 recalled pumps in 4 areas. t2: Nathan Pierce, 150 wells . t3: OPS-5 quoted; December misses 0310 , 0354 , 0486 , 0531 . t4: no H2S reading recorded on 3 Dec, and the response deadline was missed. t5: WELL-0160 , 0263 not producing, with last repairs. t6: triage across the 57 wells discussed puts WELL-0332 first : unanswered gas report with no H2S reading, plus a stop nothing explains. Next: 0160 , 0263 .	Correct ×6

QUESTION	COPILOT ANSWER (KEY FACTS)	VS ORACLE
10 What are we waiting on	634 jobs (2026): Contractors 33%, Crews 27%, Weather 22%, Parts 18% , median about 9 extra days. Southeast Block 31 stands out for contractor waits (43% of its 40 jobs vs 33% field-wide; median 14 vs 9 extra days). Caveat: small sample, not significant on its own (χ^2 $p \approx 0.92$). Share and median by area, plus a stacked bar by area.	Correct
11 Hotter wells	No region genuinely runs hotter. SE-31 looks hottest on raw temperature only because 58% of its wells are T500 . Ambient is flat (21.5 °C) and the type-adjusted delta is within ± 0.05 °C.	Correct

What worked

- Document-driven rules carried most of the score. Deadlines, approvals, lots and letter scope were read from the new documents at run time, so the **T500** type, MOC extensions and the mid-period SOP revision were handled without special cases.
- The two-level memory design held up under adversarial multi-user turns. It kept claims unconfirmed and attributed, separated first- from second-hand reports, and kept one user's preferences private.
- Honest refusals: Q1 and Q2 refuse to invent data the dataset doesn't hold.

AUDIT TRAIL

Full verbatim answers for all 21 turns:

submission/final-run/FINAL_ANSWERS.md

All runs are preserved for audit: FINAL_ANSWERS.fallback-run.md (schema bug), raw/ (hosted, before the new tools) and raw-v2/ (final).

What didn't, and why

- Our first live run scored 0/11. Two causes:
 - The planner schema's well bound grew with the fleet (600), and Gemini rejected it, so every turn silently fell back to the rule-based planner.
 - The original 17-operation menu had no way to express MOC approvals, recall lots, note-to-response joins or letter scope.
Fix: a fixed schema bound, plus 13 general document-driven tools.
- Topic references across turns first narrowed fleet questions to wells from earlier answers. Fixed: scope now comes only from the question.
- Two gaps closed by generalising.
 - Q7:** ranking now weighs an overdue mandated check above a not-yet-due one, and crew dispatch weighs trip logistics.
 - Q9 t6:** prioritising across a conversation now has a deterministic triage tool. It collects every well discussed and ranks them safety > compliance > unexplained production loss > equipment.
 - Q7's exact order remains a judgement call.** The score is printed in the answer so a reviewer can audit it.

ASSET COPILLOT

Asset Copilot Evidence-led answer 15:25

Yes. WELL-0101 had 1 real spills, totalling 1 bbl from Aug 20, 2026 through Aug 20, 2026. WELL-0101 inspection INS-0101-0197 flagged 1 bbl: "Oil sheen around cellar." Ryan Brooks (CTX-0101-002): "Inspection INS-0101-0197: I think this was only produced water, not crude." Chris Dalton (CTX-0101-003): "Inspection INS-0101-0197: sampled the stain, it is crude oil. The detection stands." No formal correction cancels this detection. Spill repair MNT-0101-029 was performed. Notes disagree: Ryan Brooks said "Inspection INS-0101-0197: I think this was only produced water, not crude."; Chris Dalton said "Inspection INS-0101-0197: sampled the stain, it is crude oil. The detection stands.". Neither note is a formal correction, so the inspection detection stands.

Evidence conflict

Notes disagree: Ryan Brooks said "Inspection INS-0101-0197: I think this was only produced water, not crude."; Chris Dalton said "Inspection INS-0101-0197: sampled the stain, it is crude oil. The detection stands.". Neither note is a formal correction, so the inspection detection stands.

How this answer was assembled

gemini-3.5-flash-lite · hosted 2366 ms

SOURCES

INS-0101-0197 CTX-0101-002 CTX-0101-003 MNT-0101-029 +1 more sources

INSIDE THE PRODUCT

A conflict is shown, not resolved. Both notes are quoted, neither counts as a formal correction, and the answer cites its source records.