

eatron

# Battery Foundation Models and Agentic AI for Scalable E-Bus Fleet Optimization

Dr. Ugur Yavas

October 2026

 AACHEN AUTOMOTIVE  
COLLOQUIUM  
OCTOBER 05-07TH, 2026



# The BMS reported 96.8 %. Usable energy was 55 %.

*One e-bus, one pack. Capacity health and usable energy are different quantities; a healthy capacity figure can hide imbalance-limited energy.*

REPORTED

## 96.84 %

Pack-median capacity state of health for Pack 2. No pack-level threshold fired.

USABLE ENERGY

## 51.9 kWh

Estimated usable energy over the last 180 days, against 94.8 kWh for a healthy pack of the same type.

What the analysis found: Brick 57 loses charge at rest, and Brick 32 has low capacity, mostly present at commissioning. Two defects, two different confirmation tests. The rest of the talk shows how that was established, and what is still open.

# Battery faults become hard maintenance decisions

*High utilisation and strict schedules make battery degradation a primary driver of cost, reliability and service continuity.*

## FLEET OWNER

Which vehicles first, and at what cost?

### NEEDS

Risk-ranked vehicles with cost and availability impact.

### HAS TODAY

Alarm lists and dashboards.

## SERVICE ENGINEER

What is the cause, and which test confirms it?

### NEEDS

Mechanism, confirming test and prioritised intervention.

### HAS TODAY

Free-text work orders and e-mail threads.

## OEM DATA SCIENTIST

Can this estimate be trusted?

### NEEDS

Model diagnostics, data-quality gaps, systemic versus single-unit.

### HAS TODAY

Raw telemetry and separate tools.

Each role needs the same chain: detect, classify, explain, act, with the evidence and a next step attached.

# The status quo is in the service records

*22 work orders, 9 vehicles, 2021 to 2024, cross-checked against the telemetry.*

01

Wrong pack replaced first

Two pack swaps where one was needed. The decision rested on incomplete diagnostics.

02

Record closed 13 months late

The asset record disagreed with the installed battery for a year.

03

An inspection that changed nothing

The 40 pp SOC gap behind it had been visible in the data for 18 months.

04

Diagnostics in e-mail and free text

Reasoning that cannot be replayed and leaves with the personnel.

The limiting factor is not data availability. It is the reasoning applied to the data.

# Two common approaches, two different gaps

*Both approaches are in use today. Each stops short of a decision an engineer can sign.*

## BETTER MODELS AND RULES

### Models, rules, dashboards.

Every number deterministic, versioned and replayable. This is the foundation we build on.

#### THE GAP

Every flag still ends at an engineer for interpretation.

*Next: what that costs at fleet scale.*

## GENERAL-PURPOSE AGENTS

### An LLM agent on the data.

Telemetry and work orders in, fluent answers out. Fast to start.

#### THE GAP

Without battery tools and checks, it can be confidently wrong.

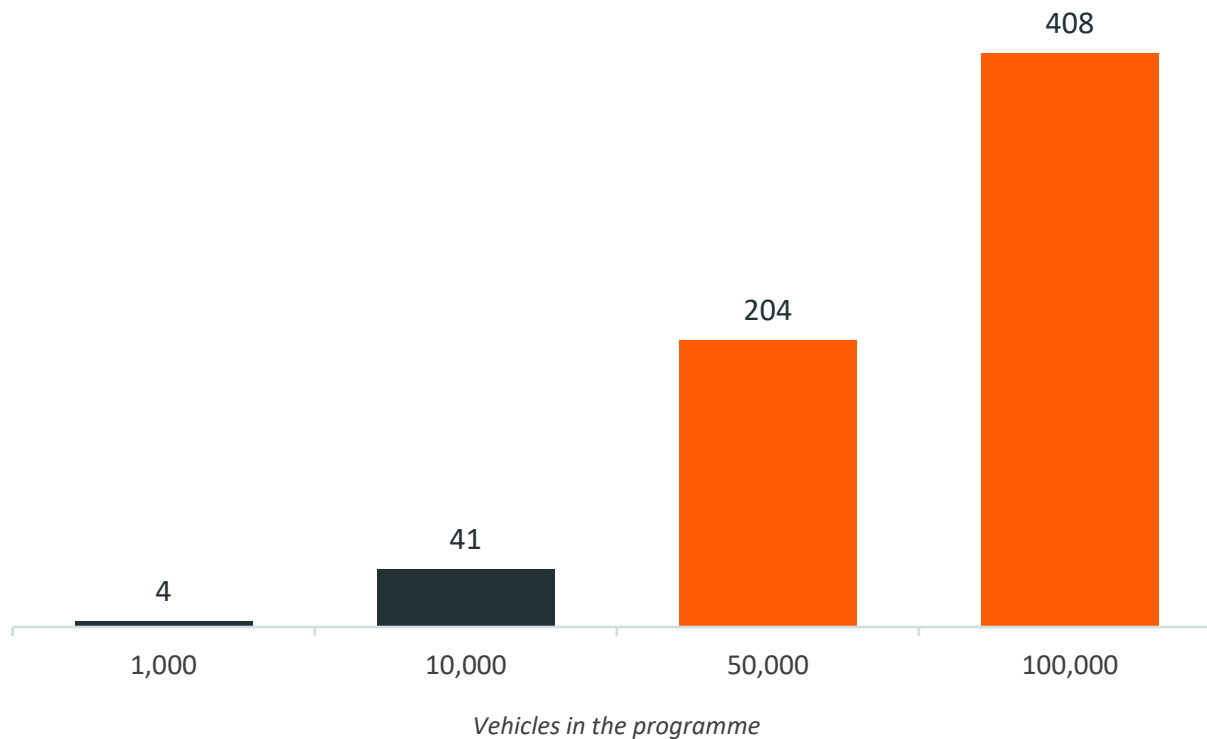
*Then: three failure modes on this corpus.*

What we use: deterministic evidence underneath, agents that interpret it inside a harness, and an engineer who approves.

# In this workflow, expert review is the scaling bottleneck

Eleven buses produced 2.58 million events and 296 evidence-ledger rows. Compute scales with infrastructure; expert review scales with headcount.

Engineer-years per full review pass, by fleet size



## WORKLOAD EXPOSURE, NOT MEASURED SAVINGS

- 296 evidence-ledger rows from 11 vehicles: about 27 per vehicle.
- 15 minutes of expert review per row.
- 10,000 vehicles: about 270,000 rows, 67,000 hours, 40 engineer-years per pass.
- 100,000 vehicles: about 400. Grouping and prioritisation reduce this; agents target the first review pass that remains.

*“The goal is to develop vehicles with minimal engineering effort, regardless of whose cells we use. This is about optimising the allocation of engineering hours.”*

Senior powertrain executive, global OEM (paraphrased, anonymised)

*Illustrative. Assumes 15 minutes per ledger row, 1,650 working hours per engineer-year and one review pass. At 5 minutes per row, divide by three. Review time was not measured.*

# Dashboards report. Agents investigate.

*Bus 639, Pack 2: what a dashboard shows, and what an agent loop works through before anyone asks.*

## DASHBOARD

BUS 639 · PACK 2  
State of health

96.8 %

● Within threshold

No alarm.  
No explanation.  
No next step.

*Estimated usable energy: 55 % of a healthy pack of the same type.*

● **AGENT LOOP** runs in the background, then follows the question

- ✕ Capacity fade across the pack  
Rejected · Brick 57's offset is flat across pack SOC and grows with rest
- ✕ One-off imbalance, fixed by balancing  
Rejected as sufficient · balancing recovered 6 points, then the loss resumed
- Charge loss in Brick 57  
Open · 43 to 47 pp below its 89 peers; internal path hypothesised
- Low capacity in Brick 32  
Open · mostly present at commissioning; projected end of life inside warranty

### PROPOSED NEXT STEP

Brick 57: cell-board disconnect test before a repair route is chosen. Brick 32: warranty discussion.  
*An engineer approves both. Nothing is routed automatically.*

Numbers: deterministic BOS analytics · Reasoning: agents · Decision: engineer

● **ILLUSTRATIVE** session composed from recorded fleet-1 findings; the evidence is on the Case 1 slide

# The loop keeps what it learns

*Proposed operating model, steps 01 to 05. What we observed during the fleet-1 analysis is in the left panel below.*

**01**

## Runs in the background

Analyzers and agent review on all packs, before anyone looks.

&gt;

**02**

## Presents the case

Only findings that warrant attention, with evidence and a next step.

&gt;

**03**

## Follows the person

An owner, engineer or data scientist steers it deeper or onto a new angle.

&gt;

**04**

## Engineer approves

Every action, and every new check, is signed off by an engineer.

&gt;

**05**

## Becomes a check

An approved investigation adds a check to later runs, e.g. an SOC-matched park test.

🔄 The next run starts from the approved checks, not from a blank page.

### • OBSERVED IN OUR FLEET-1 WORKFLOW · RETROSPECTIVE

North American bus fleet, analysed with its work orders, warranty claims and pack specifications. Agents brought in the key insights; engineers reviewed them.

Agents wrote the 296-row evidence ledger for engineer review.

Agents proposed the two revisions that took the metrics script to rev 3, e.g. the SOC-matched park test for leaks. Engineers approved them; every table was regenerated.

The agent withdrew the climate claim when duty cycle confounded it (Case 3).

### • CONCEPT · SAME FINDING, RENDERED FOR WHO ASKS

#### FLEET OWNER VIEW

Bus 639 · Pack 2

Priority high · keep out of service

Next: cell-board test, then repair route

Brick 32: warranty discussion

#### DATA SCIENTIST VIEW

Brick 57 · -43 to -47 pp vs peers

Flat across SOC, grows with rest

Filtered and raw SOH bases

Tiers: observed → hypothesis

*Prototype screens only. Agent-proposed checks are reviewed by an engineer before they run.*

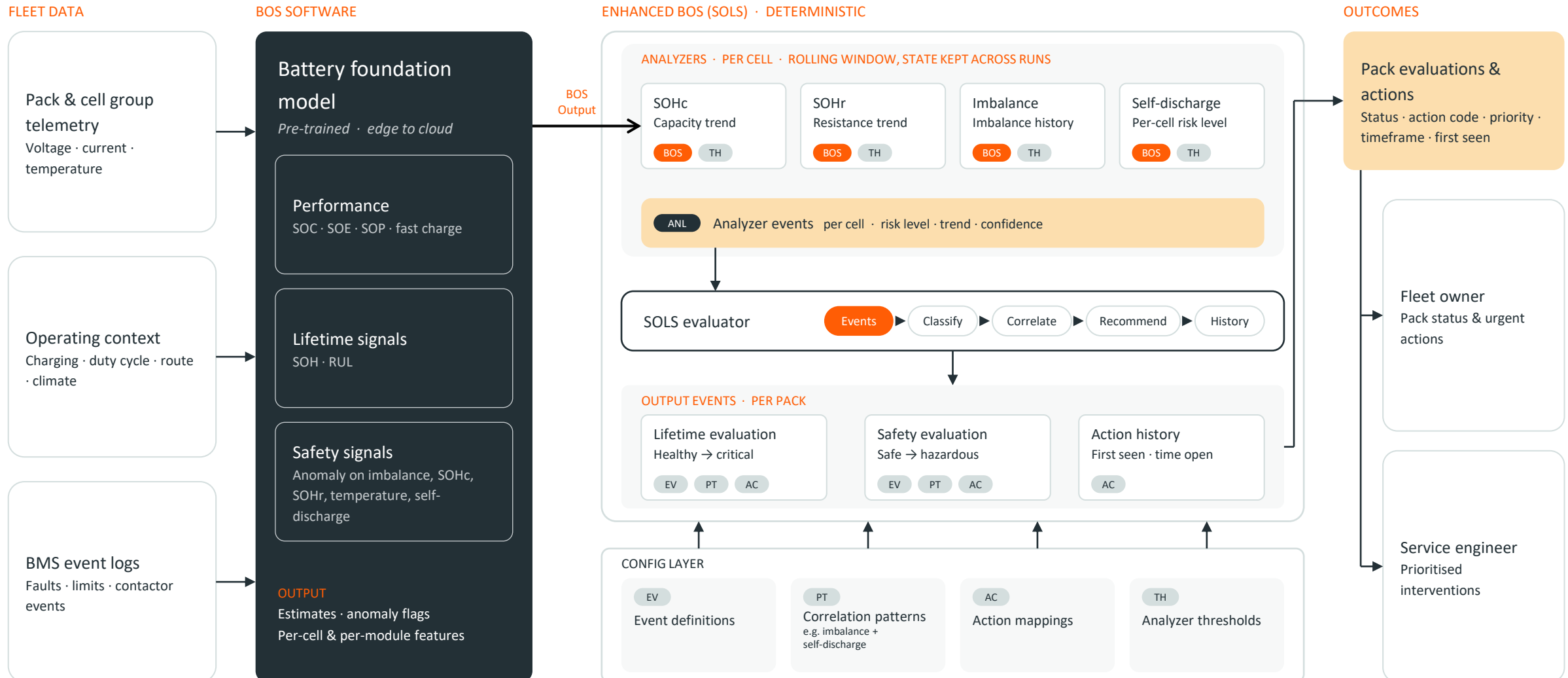
# Three failure modes a general-purpose agent must avoid

*Identified on this corpus, not recorded agent runs. Each would look plausible to a reviewer without the evidence.*

| Task given to a general agent | Failure mode                                                      | What the harness supplies                                 |
|-------------------------------|-------------------------------------------------------------------|-----------------------------------------------------------|
| Ask it for pack health        | Reports the filtered SOHC, 96.84 %, on a pack that needs a repair | Both bases published, filter lag named and quantified     |
| Let it clean the event stream | A naive same-timestamp dedupe deletes 58 % of the leak evidence   | Extractor keys on checkpoint index, sentinel values gated |
| Show it the work orders first | Hindsight leaks into what it claims it would have detected        | Telemetry frozen before service history, 66 prefix checks |

The harness: the tools, battery knowledge, data checks and output checks around the language model.

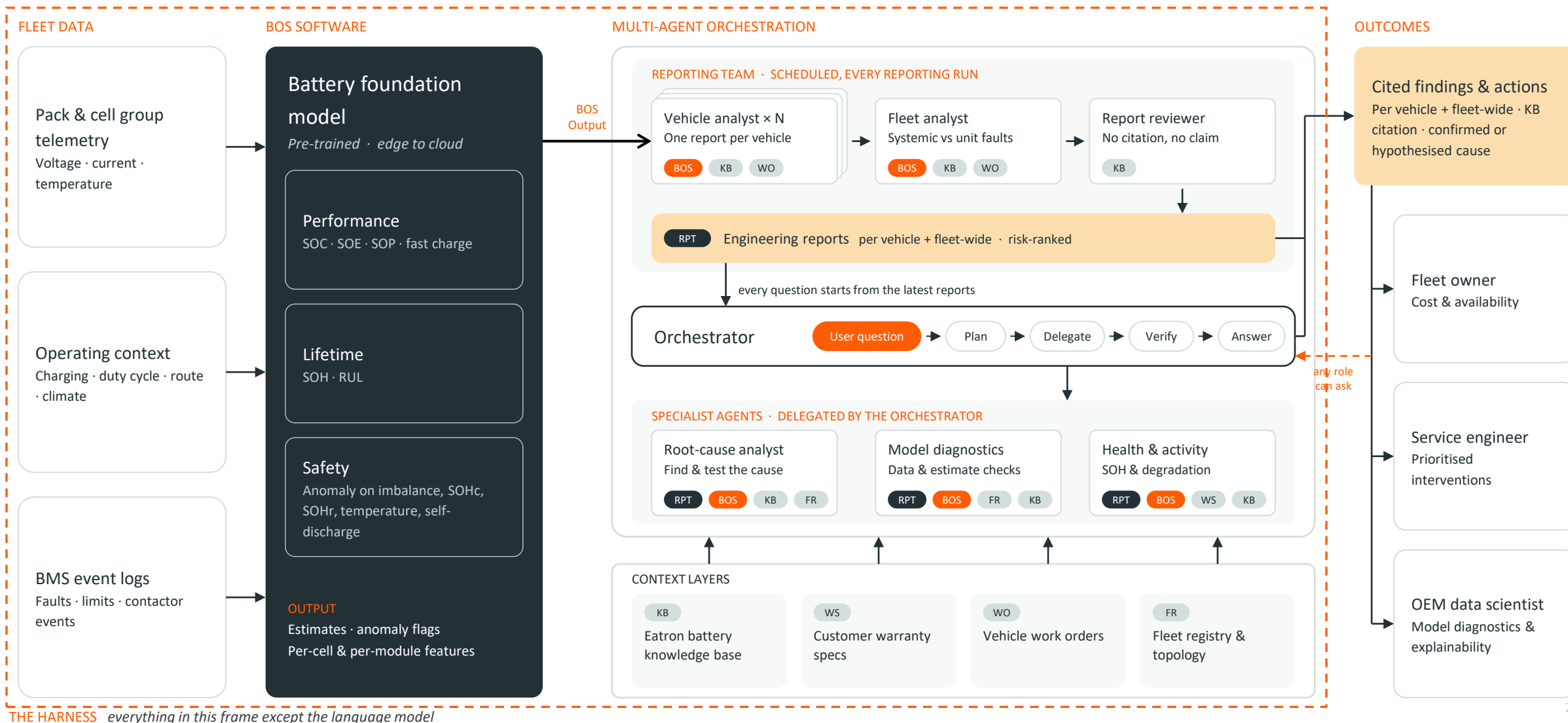
# Deterministic first: estimates, events, evaluations



# Agents on top, inside the harness

Skip the agents? *This layer automates the first review pass; engineers keep approval.*

Skip the harness? *Its checks address the three failure modes.*



# One battery foundation model, from vehicle to cloud

*Pre-trained on fleet and laboratory data. One model on the BMS microcontroller and in the cloud, without task-specific redesign.*

## PERFORMANCE

SOC · SOE · SOP

State of charge, energy and power, and closed-loop fast-charge control.

≈2×

*lower SOC error than the incumbent BMS software on a production two-wheeler fleet*

## LIFETIME

SOH · RUL

Capacity and resistance health and remaining useful life, learned from fleet data rather than fitted pack by pack.

1.15 %

*SOH error from 20 % of observed life on a held-out cell*

## SAFETY

Anomaly detection

Imbalance, capacity, resistance, temperature, self-discharge and lithium plating, isolated from cell to pack.

2 of 2

*opened cells confirmed lithium plating, out of 7 flagged; 5 not yet opened. The OEM's in-house diagnostics flagged none.*

Runs on a standard automotive microcontroller. No NPU and no additional sensing: voltage, current and temperature only.

# Four stages, one pack: who contributes what

*Bus 639, Pack 2, through the stack. The full architecture diagrams are in the appendix.*

## 01 · FOUNDATION MODEL

### Estimates

SOC, SOH and anomaly signals per brick. No estimator is perfect, ours included, so each is checked against raw anchors.

#### BUS 639

Per-brick SOC and capacity for 90 bricks; filtered and raw bases both reported.

## 02 · DETERMINISTIC ANALYZERS

### Evidence

Trends, events and grades. Computed, versioned and replayable. No language model.

#### BUS 639

Brick 57: 43 to 47 pp below its peers, growing with rest. Brick 32: low capacity.

## 03 · AGENTS IN THE HARNESS

### Interpretation

Read the evidence with work orders, warranty terms and specifications. Test alternatives. Cite every claim.

#### BUS 639

Balancing recovered 6 points, then the loss resumed. Brick 32's gap was mostly present at commissioning.

## 04 · ENGINEER

### Approval

Approves or rejects the proposed action. Nothing is routed automatically.

#### BUS 639

Cell-board disconnect test on Brick 57. Warranty discussion on Brick 32.

Numbers come from stages 01 and 02. Agents add interpretation. The engineer keeps the decision.

# What agents add beyond a structured report

*The analyzers already grade and recommend. Agents do the work that a fixed report template cannot.*

01

## Read relevant documents

Warranty claims, work orders, OEM pack specifications, maintenance documentation and the Eatron battery knowledge base, next to the telemetry.

02

## Test competing hypotheses

Run approved tools against each candidate explanation, and record why it was rejected or kept open.

03

## Assemble evidence across vehicles

Separate a systemic pattern from a single unit: series-pack drift on 4 of 11 vehicles, not one.

04

## Answer the unanticipated question

A new question from any role gets an answer with citations, without a new report template.

### WHAT AGENTS DO NOT DO

#### Compute a figure.

Every number comes from the foundation model or the analyzers.

#### Skip the evidence pointer.

Claims need a citation and pass a reviewer check. A cited claim can still misread its source, so engineers review.

#### Approve an action.

An engineer does.

Agents connect the evidence to operational questions. They do not produce the evidence.

# Every claim is a ledger row an engineer can check

Brick 57 as three rows, one per evidence tier (paraphrased). The 296-row fleet-1 ledger was written by agents for engineers to review.

| Tier       | Claim                                        | Evidence                                                          | What settles it            |
|------------|----------------------------------------------|-------------------------------------------------------------------|----------------------------|
| OBSERVED   | Brick 57 sits 43 to 47 pp below its 89 peers | Per-brick SOC against pack median; file, table and rows named     | Measured and gated         |
| INFERRED   | Self-discharge, not capacity fade            | Offset flat across pack SOC, growing with rest; assumption stated | The disconnect test below  |
| HYPOTHESIS | Charge leaves through an internal path       | Balancing recovered 6 points, then the loss resumed               | Cell-board disconnect test |

FLEET-1 LEDGER · 296 ROWS

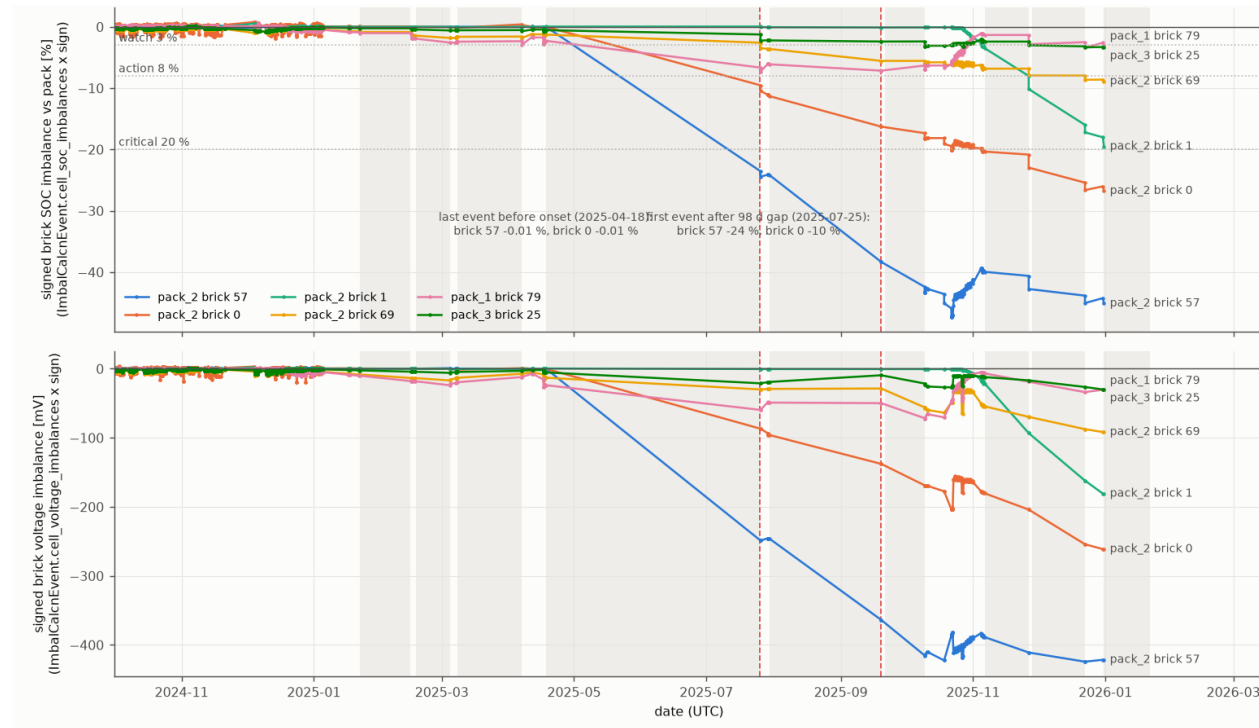


216 observed · 49 inferred · 31 hypothesis

- 0 blank evidence pointers
- 12 · 87 · 197 high · medium · low severity
- 44 of 44 packs in the ledger
- 10 with an explicit no-anomaly row

Observed is measured and gated. Inferred states its assumption. Hypothesis names the test that would settle it.

# Case 1: Bus 639, the evidence behind two findings



Pack 2, Bus 639: per-brick state of charge against pack median. Illustrative session composed from recorded fleet-1 findings.

## 1 QUESTION, SERVICE ENGINEER

Pack 2 delivers far less energy than a healthy reference. Why, and what should be done?

## 2 FINDING

Brick 57 sits 43 to 47 pp below its 89 peers, flat across pack SOC and growing with rest: self-discharge, not capacity fade. Brick 32 shows low capacity, largely present at commissioning.

## 3 ALTERNATIVE TESTED

Balancing recovered 6 points, then the loss resumed: a one-off balancing correction is not enough.

## 4 PROPOSED ACTION

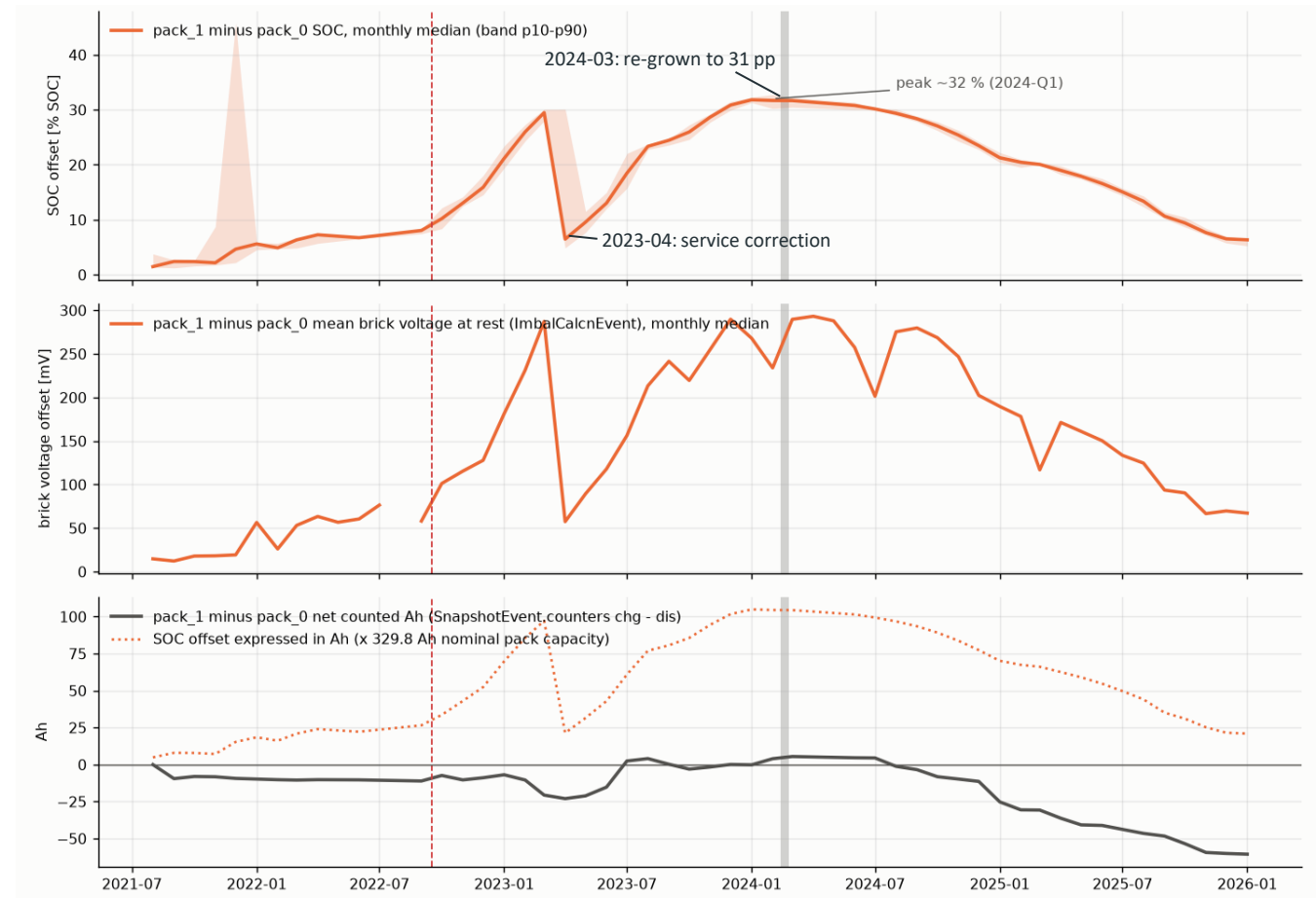
Brick 57: cell-board disconnect test before a repair route is chosen. Brick 32: warranty discussion, projected end of life 2027-10 to 2029-08, within the 2030-04 warranty.

An engineer approves before any repair or warranty routing.

## Case 2: a correction without a mechanism does not hold

- Two series-connected packs drifted from under 2 pp to 29 pp apart. A 285 mV rest-voltage gap confirms real stored charge, not an estimator artefact.
- A service correction reduced the offset to 7 pp. Within eleven months it had returned to 31 pp, beyond its previous peak.
- Counted ampere-hours stayed level while rest voltages diverged. That points to a charge path outside the pack current sensors; the path is not yet identified. Next test: quiescent-current measurement per pack.

4 of 11 vehicles exceeded 30 pp, stranding 5 to 10 % of string energy.



## Case 3: 2.5 % extra fade against matched peers

*A third fleet, analysed at fleet level: every pack benchmarked, outliers found, candidate causes tested against chance.*

### 6,512

NMC packs benchmarked across the fleet, fade measured from each pack's own starting SOH

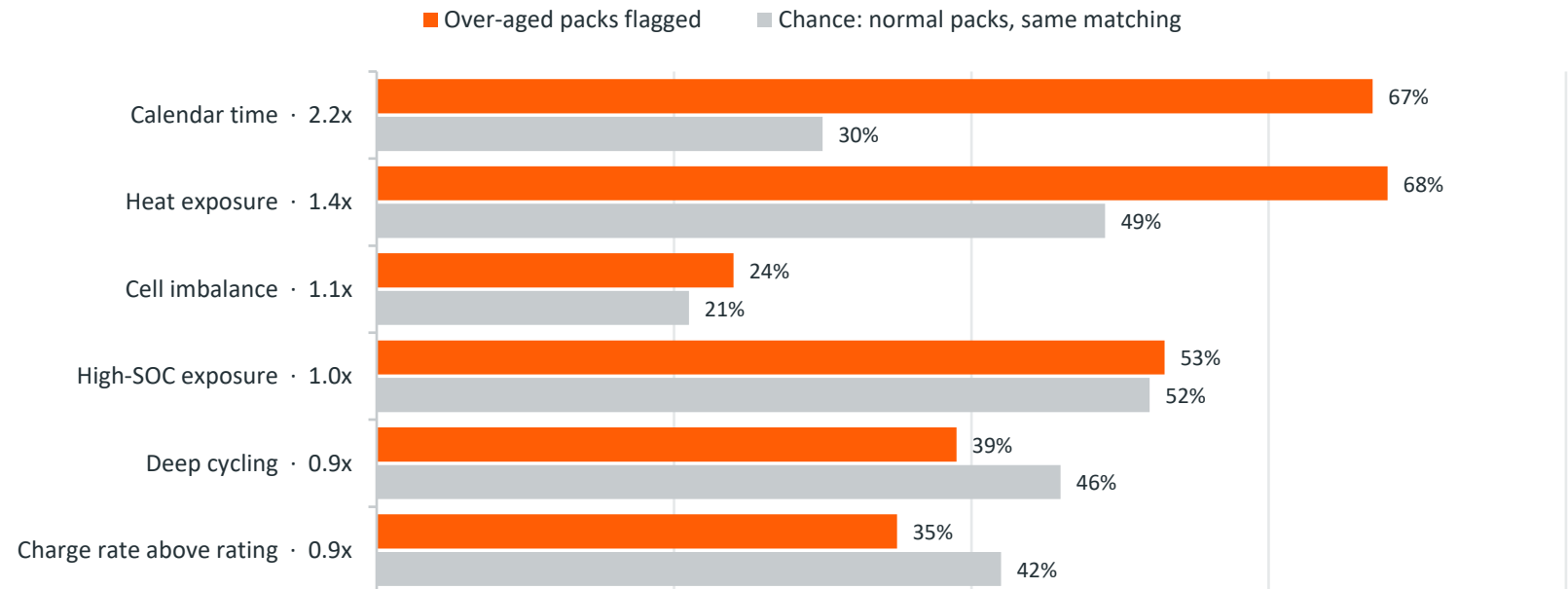
### 677

over-aged: fade in the top decile for their cycle count, against the 2023 start-year packs

### 2.5 %

median extra fade over their 10 nearest normal peers, as % of starting SOH ( $\pm 10\%$  EFC,  $\pm 6$  months start,  $\pm 2$  pt starting SOH)

*Share of the 675 over-aged packs that have matched peers (2 have none) with a stress feature outside their peers' P10–P90. Groups overlap, so shares do not sum to 100 %. Ratio to chance after each name; chance is the same share among 199 matched normal packs. A ratio marks enrichment, not cause.*



- Association, not cause: calendar time shows the strongest enrichment (2.2× chance). Charge rate, deep cycling and high SOC are not enriched. Causes remain hypotheses until tested.
- What the system declined to claim: on fleet 1, warmer packs seemed to age 1.5× faster per month, but also ran 1.9× the daily mileage. The agent withdrew the claim.

# Exception-only: 2 of 44 packs raised a card

## ✦ PER-BUS FINDINGS

### Bus 639 · Pack 2 — 4 bricks with elevated self-discharge **HIGH**

4 bricks with elevated self-discharge, **38 %** of energy lost to imbalance, lowest-capacity brick in fleet, deep discharge

**2024-06**

**Proposed:** confirm cause (cell-board disconnect test) before repair route or warranty claim; keep out of service

Approve

Inspect evidence

*Console prototype. 38 % is energy lost to imbalance only (slide 2 compares total energy). Approve records an engineer's decision.*

- The screen remains clear until a finding warrants human attention. Each card states the finding, the proposed action and a link to its evidence.
- Silence is not clearance: only 10 packs carry an explicit, bounded no-anomaly result.

#### ONE EVIDENCE LEDGER, THREE READINGS

Fleet owner: which vehicle first, and at what cost.

Service engineer: which test confirms the cause.

OEM data scientist: whether the estimate can be trusted.

*No role-specific model. No recomputation per role.*

# Replayed cold: one screening rule and its limits

*An inter-pack SOC-offset rule replayed month by month on 22 candidate pack pairs, using only data available at each decision point.*

| Vehicle · pair | Months flagged | First bound | What it precedes                                                  |
|----------------|----------------|-------------|-------------------------------------------------------------------|
| 000466 · 0,1   | 43 of 53       | 2022-03     | 13 months before the 2023-04 service correction                   |
| 000472 · 2,3   | 13 of 52       | 2021-12     | 17 months before the 2023-05 action                               |
| 000645 · 2,3   | 14 of 31       | 2023-03     | 10 months before the 2024-01 service session; pairing unconfirmed |
| 000365 · 2,3   | 12 of 48       | 2021-09     | No matched service action                                         |
| 000366 · 0,1   | 5 of 58        | 2025-09     | Open now                                                          |

WHAT IT DOES NOT SHOW

17 of 22 pairs never alert, including both pairs on Bus 639. A different detector finds its leak.  
Chosen retrospectively: a reproducible historical screen, not prospective lead time.  
Alert dates are month-completion bounds; ingestion latency is excluded.

KNOWN LIMITS, IN EVERY REPORT

Brick-level measurement: cell attributions are tiered inferences.  
15 temperature sensors for 90 bricks: temperature is a zone property.  
Telemetry covers 53 to 69 % of days.

66 prefix-invariance checks passed: the analysis is replayable. That is not a measure of agent accuracy or of early warning.

# Every finding ends in a decision or a measurement

| Finding                         | What changed the reading        | Still open                                                  | Next step                                                                                                                       |
|---------------------------------|---------------------------------|-------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Pack with two defects (fleet 1) | Peer comparison per brick       | Internal path of brick 57, hypothesised                     | Brick 57: cell-board disconnect test, then the repair route.<br>Brick 32: warranty discussion (end of life 2027-10 to 2029-08). |
| Series-string drift (fleet 1)   | Rest voltage against counted Ah | Inferred path outside the pack counters; not yet identified | Quiescent-current measurement per pack and the string schematic, before the next correction                                     |
| 677 over-aged packs (fleet 3)   | Matched peers against chance    | Causes hypothesised                                         | Calendar-day comparison; cooling and location records                                                                           |

*Fleet 1 ledger: all 44 packs reviewed; 10 carry an explicit no-anomaly result.*

# Conclusion

Battery models and deterministic tools produce the evidence.

Agents connect it to operational questions and proposed actions.

Engineers approve. Open mechanisms stay explicit, and **approved analyses become rules for the next run.**

*Demonstrated retrospectively on fleet data. Next: prospective validation and measured review time on a customer fleet.*

# Eatron Technologies

*Battery Optimisation Software for performance, lifetime and safety, deployed on the BMS, at the edge or in the cloud.*

25,000+

vehicles and assets monitored

300 M+

km of driving data

36

patents and applications

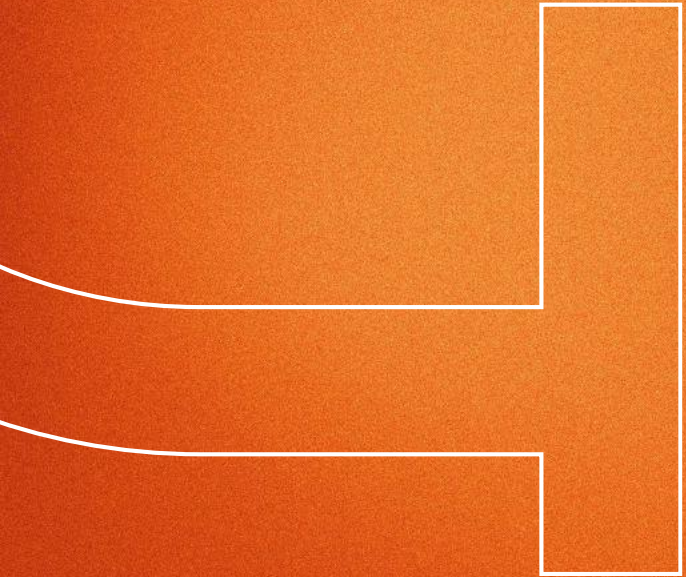
2018

founded

- Headquartered in Warwick, United Kingdom, with engineering teams in Türkiye and Japan.
- LG Energy Solution is an investor and strategic partner. Winner of the LG Energy Solution, Hitachi and ABB start-up challenges.

We welcome fleet operators and OEMs to evaluate this approach on their own data.

Dr. Ugur Yavas · [ugur.yavas@eatron.com](mailto:ugur.yavas@eatron.com)



# The Edge In Electrification

eatron



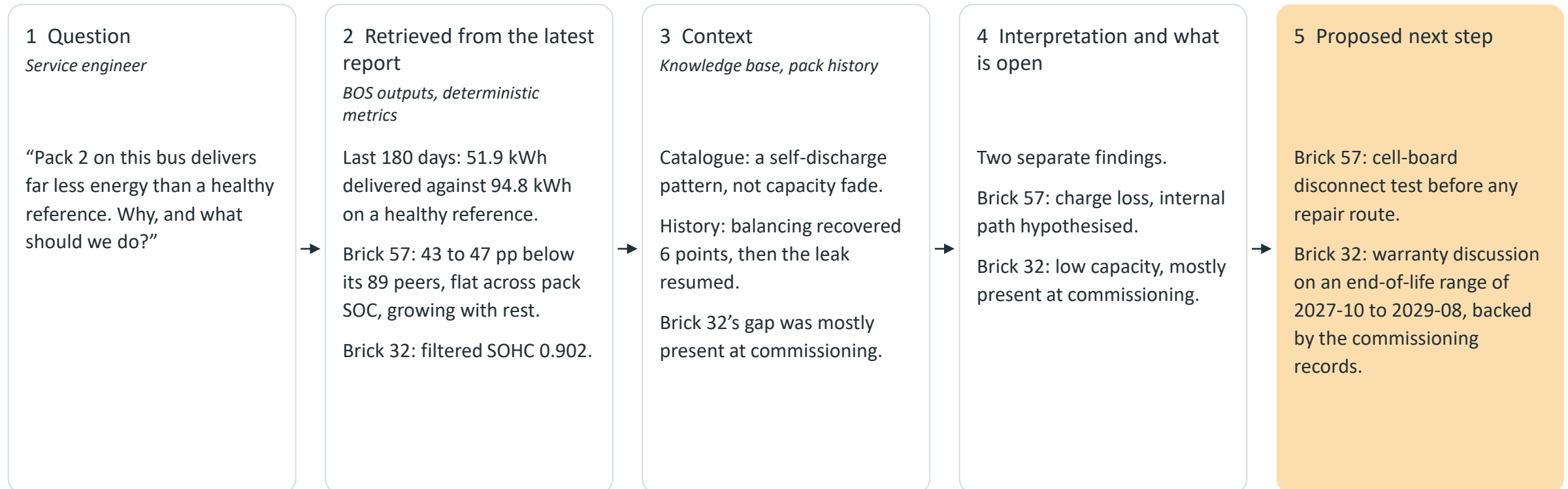
→ [eatron.com](https://eatron.com)

# 05.

## Appendix

# One question, worked end to end

*Illustrative: composed from the fleet-1 findings shown earlier, not a recorded session.*



Evidence: every step points to claims-ledger rows (ledger design in the main deck). An engineer confirms the test result before repair or warranty routing.

# Eleven buses: 2.58 million events before the first question

- 11 e-buses at a North American transit operator, up to 4.8 years in revenue service.
- 44 NMC 800 V packs, 3,960 bricks, about 285,000 cells, 660 temperature zones.
- 2,581,679 events across 22+ types and 735 GB of source data, processed by one extractor run unchanged on all 44 packs.

1×

11 vehicles  
*this study*

≈900×

10,000 vehicles  
*a mid-size OEM programme*

≈9,000×

100,000 vehicles  
*a volume OEM fleet*

Data volume grows with every vehicle added. So does the review effort, unless the review itself is automated.

# A better model moved the bottleneck. It did not remove it.

*Karpathy's three eras of software, applied to the two jobs a battery analytics stack has to do.*

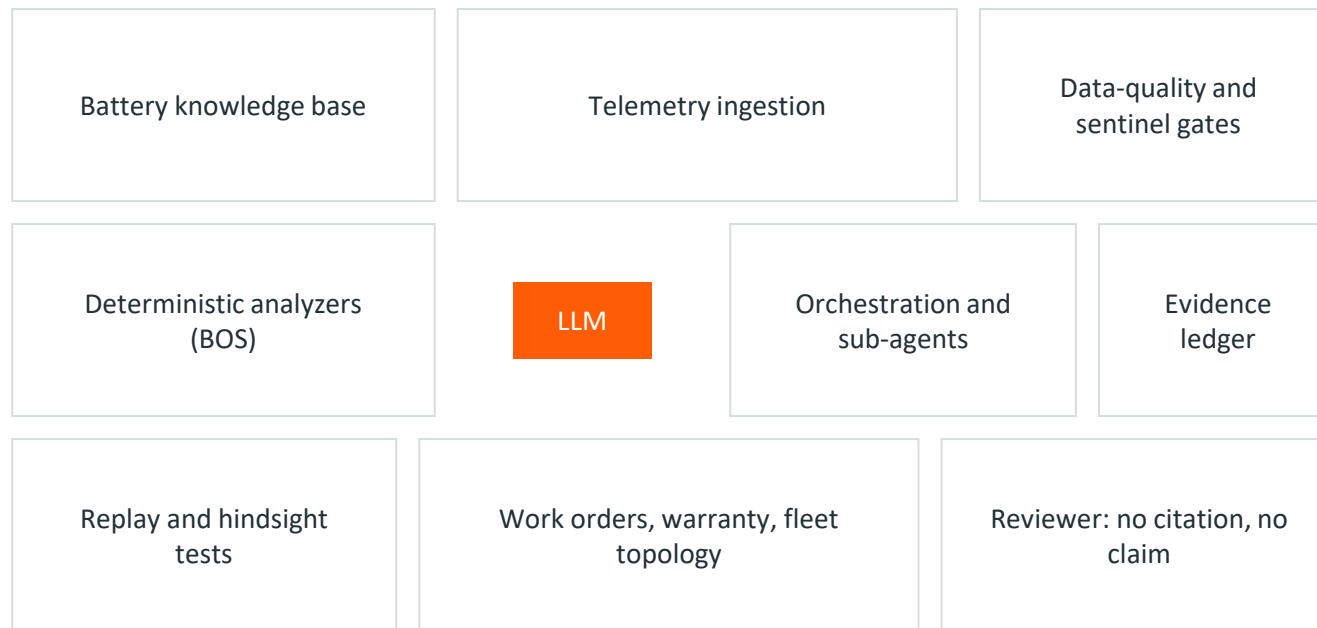
| ERA                                           | SCALES WITH                      | JOB 1 · ESTIMATE THE STATE                                                                                          | JOB 2 · REVIEW WHAT IT FLAGS                                                         |
|-----------------------------------------------|----------------------------------|---------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| Software 1.0<br><i>Written code</i>           | Engineers                        | <i>Thresholds, look-up tables and per-programme calibration.</i>                                                    | <b>WHERE REVIEW IS TODAY</b><br>An engineer reads every flag, one finding at a time. |
| Software 2.0<br><i>Learned weights</i>        | Data                             | <b>WHERE ESTIMATION IS TODAY</b><br>Battery foundation model. 1.47 % SOH error zero-shot on a held-out e-bus fleet. | <i>A number is not a decision. No model in this row owns the review.</i>             |
| Software 3.0<br><i>Model inside a harness</i> | Compute, if the harness is right | <i>Not needed: the estimate stays deterministic.</i>                                                                | <b>THIS TALK</b><br>Agents review every flag with citations. The engineer approves.  |

Foundation models moved estimation from 1.0 to 2.0. Review is still 1.0, and it is the part that grows with every vehicle.

# The language model is the new small box

*Sculley et al. (NeurIPS 2015): in a real ML system the model is the smallest part. The same holds one layer up.*

## HIDDEN TECHNICAL DEBT, ONE LAYER UP



*2026: same diagram, new small box. Everything around it is the battery harness.*

## THE PATTERN IS ALREADY PROVEN

**Coding agents became useful when their harness matured, not only their model.**

Tools that act: read files, run tests, call APIs.

Memory and context: project documents and prior work.

Planning and sub-agents: split the task, keep each context lean.

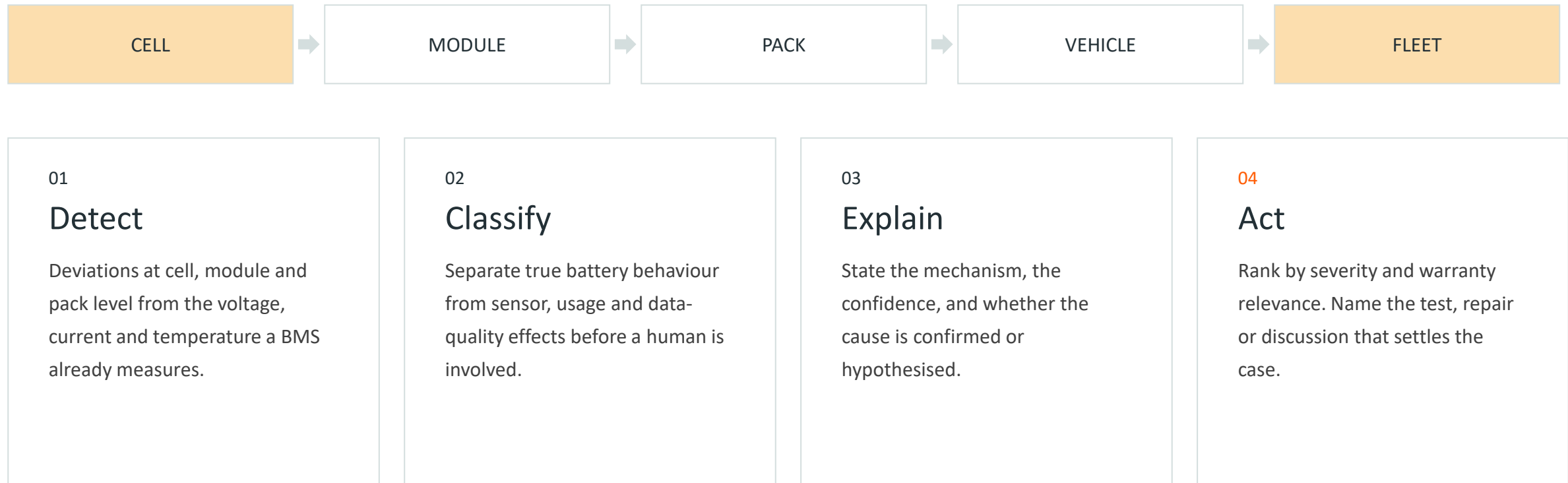
Checks before output: tests, review, permissions.

*Claude Code is the clearest example. Same model class, different harness, different result.*

Pointing a model at data is the easy part. Everything around it decides whether the answer is right.

# The requirement: a reasoning chain that ends in an action

*What fleet operators and OEMs now specify: bottom-up evidence from cell mechanism to fleet decision.*



Every finding is recorded from cell to fleet, with its evidence, its confidence and its next step.

# Two tempting claims the data did not support

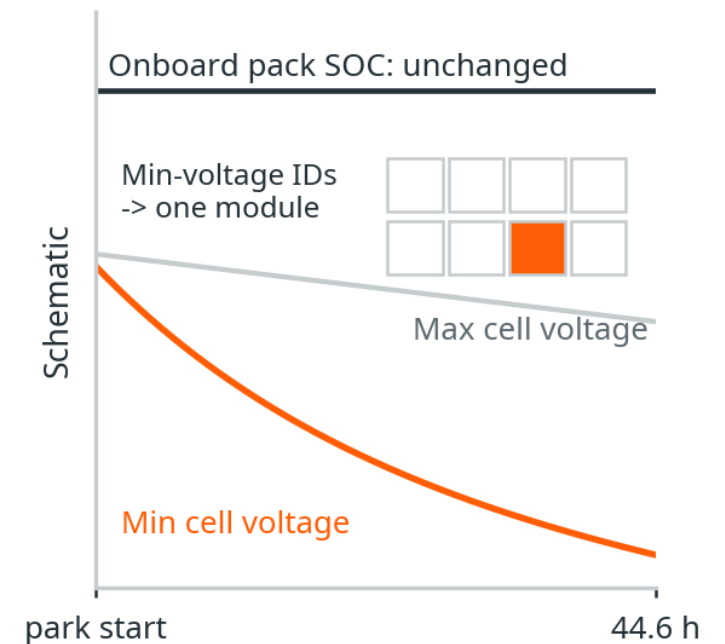
*Two results the system refused to claim on fleet 1. Each verdict holds for that comparison, not for heat in general. "Uncatalogued" being a permitted answer is what makes the other answers worth reading.*

| The tempting claim                                        | What the data did to it                                                           | Verdict                                                |
|-----------------------------------------------------------|-----------------------------------------------------------------------------------|--------------------------------------------------------|
| Warm packs fade ~1.5x faster per month, so it is the heat | Those buses also run 1.9x the daily miles, 36.3 against 18.8                      | Duty confounds climate                                 |
| The same claim, normalised                                | Per equivalent full cycle the cohorts overlap (filtered), invert (raw)            | The per-month gap does not hold                        |
| The same claim, tested                                    | Spearman -0.63 between per-EFC fade and the Arrhenius index                       | Confounded, not established                            |
| 18 zones read +5 to +9 degC hotter under load             | At rest all 660 sit within 1 degC. Bias, I <sup>2</sup> R and fade tests negative | Uncatalogued, not cleared. One IR survey covers all 18 |

# Charge loss the pack SOC never showed

*A second fleet, sparse data: 30 s pack summaries with only the minimum and maximum cell voltage and their cell IDs.*

- Over a 44.6 h park the onboard pack SOC stayed unchanged, while the minimum cell voltage fell away from the maximum. Charge was leaving one place in the pack, not the whole pack.
- The minimum-voltage cell IDs, read against the pack topology, narrowed it to one module. No per-cell history was needed to localise it.
- Mapped to the BOS chain afterwards (illustrative, not a recorded run): SOC and imbalance indicators screen for charge loss, and the finding fits the imbalance-with-self-discharge pattern. Verdict: candidate local charge-loss anomaly.
- Action: check service activity during the park first, then the module's rest voltages, electronics current draw and isolation resistance. Leakage is hypothesised until a service check confirms it.



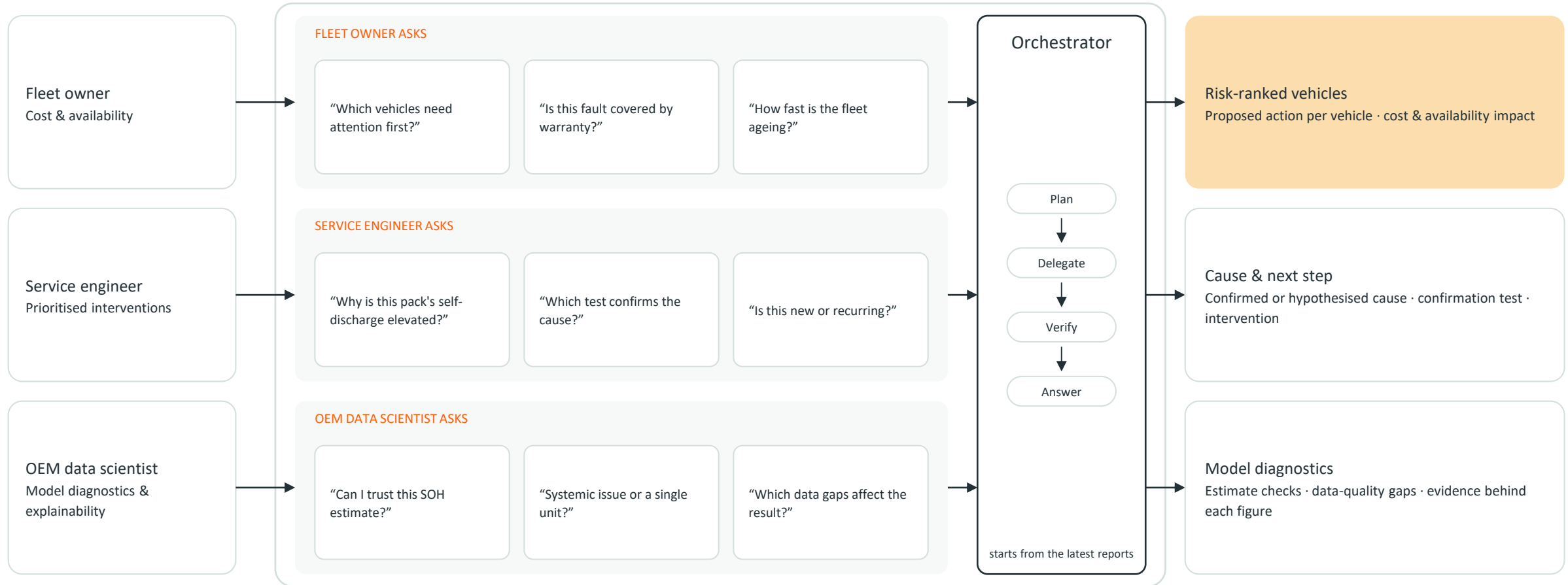
Candidate: local charge-loss anomaly

# Any role can ask the orchestrator

## ROLES

## MULTI-AGENT ORCHESTRATION · QUESTIONS IN, ANSWERS OUT

## THEY GET BACK



Every answer · KB citation · report section · confirmed or hypothesised cause · questions are illustrative

# Lithium plating confirmed by teardown

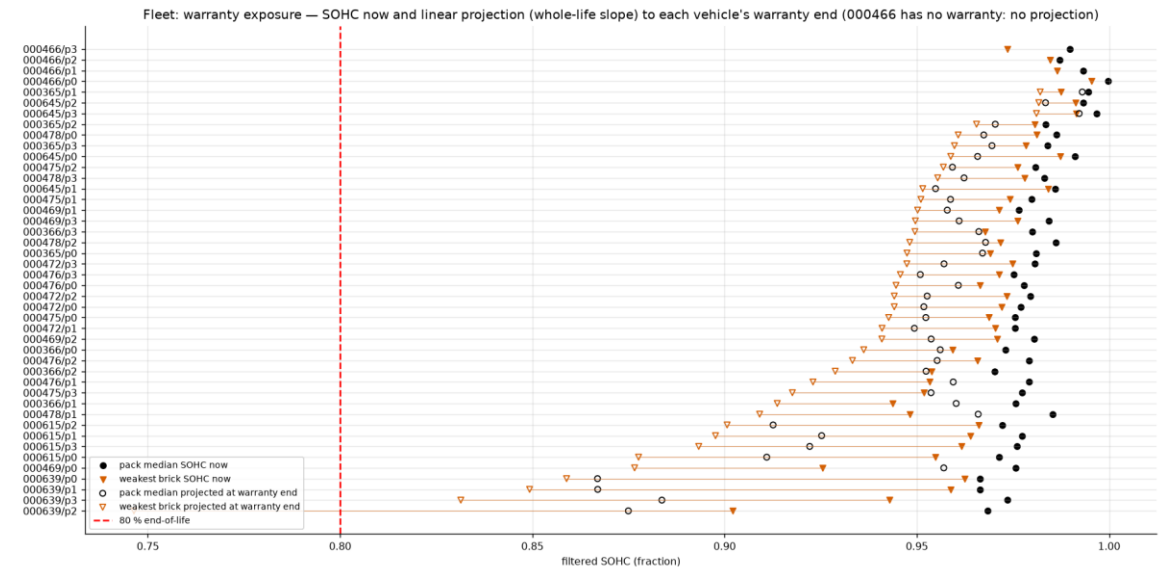
*European premium passenger-car OEM, 50-vehicle test fleet, cell-level logs at 500 ms. Model pre-trained on more than 1,000 torn-down cells; no customer-specific training.*

| Cell | Eatron plating flag | OEM in-house flag | Other Eatron anomalies on the same cell | Teardown          |
|------|---------------------|-------------------|-----------------------------------------|-------------------|
| 107  | Yes                 | No                | Time-series, SOH, capacity              | Plating confirmed |
| 108  | Yes                 | No                | Imbalance, time-series, SOH, capacity   | Plating confirmed |
| 64   | Yes                 | No                | None                                    | Not opened        |
| 94   | Yes                 | No                | None                                    | Not opened        |
| 112  | Yes                 | No                | None                                    | Not opened        |
| 146  | Yes                 | No                | None                                    | Not opened        |
| 197  | Yes                 | No                | None                                    | Not opened        |

Co-occurring anomalies identified which flagged cells to open first. The detection was validated against physical evidence, not against another algorithm. Five cells were not opened; their flags remain unconfirmed.

# Warranty exposure: one brick, end of life 2027-10 to 2029-08

- No pack median reaches 80 % before warranty end on either SOHC basis, on every vehicle with a warranty date (one has none). That is the fleet-wide health message.
- Exactly one brick of 3,960 crosses end-of-life early: 000639 pack\_2 brick 32, projected between 2027-10 and 2029-08, before the 2030-04 expiry. Most of its gap was present at commissioning.
- Raw-basis projections sit 2 to 5 points lower on sparse-anchoring vehicles. Every report carries both bases; the figure shows the filtered one.
- The 80 % criterion cannot see the imbalance faults at all, yet they are what actually limits vehicles today. The warranty conversation here is about pack\_2 as a unit, not about SOHC.



# The claims ledger as a design pattern

- Every row carries vehicle, pack, brick, zone, rung, claim type, onset, peak, severity, mechanism, citation, tier, confidence and an evidence pointer.
- Epistemic tiers, 296 rows: 216 observed · 49 inferred · 31 hypothesis. Observed means measured and gated; inferred states its assumption; hypothesis states what would settle it.
- Consistent confidence. Checked so the same signature is never rated differently on two vehicles.
- Zero blank evidence pointers. Every row names the file, table and rows a human can check. Severity is rated down as readily as up: 12 high, 87 medium, 197 low. 291 of 296 rows also carry a knowledge-base citation; 5 are placeholders.
- A clean bill is a result. 10 of 44 packs carry an explicit no-anomaly-found row, and 44 of 44 appear in the ledger.

# Replayed cold: what the rule finds

*An inter-pack SOC-offset rule replayed month by month over 22 candidate pack pairs, using only data available at each decision point. 66 prefix-invariance checks passed. A retrospective screening rule, not prospective lead time: 5 of 22 pairs alert, listed below. The next slide states what it does not claim.*

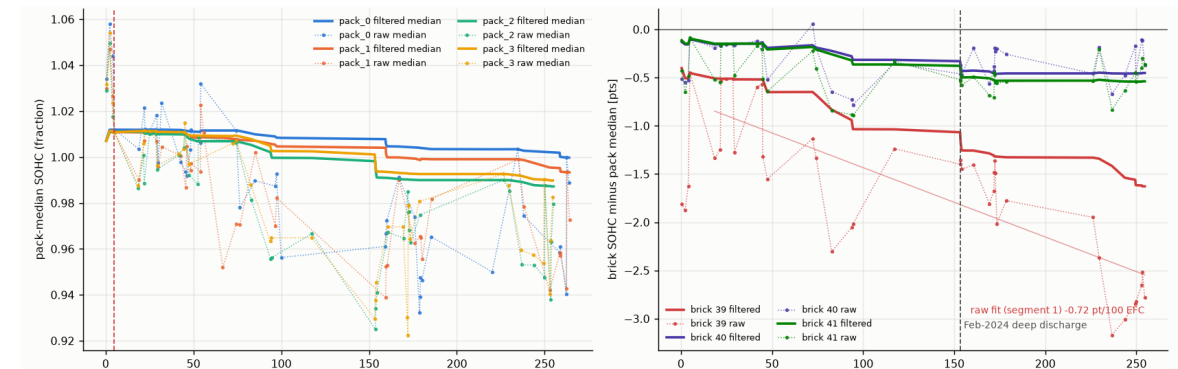
| Vehicle · pair | Months flagged | First bound | What it precedes                                                  |
|----------------|----------------|-------------|-------------------------------------------------------------------|
| 000466 · 0,1   | 43 of 53       | 2022-03     | 13 months before the 2023-04 service correction                   |
| 000472 · 2,3   | 13 of 52       | 2021-12     | 17 months before the 2023-05 action                               |
| 000645 · 2,3   | 14 of 31       | 2023-03     | 10 months before the 2024-01 service session; pairing unconfirmed |
| 000365 · 2,3   | 12 of 48       | 2021-09     | No matched service action                                         |
| 000366 · 0,1   | 5 of 58        | 2025-09     | Open now                                                          |

# What the replay does not claim

- 17 of the 22 candidate pairs never alert at all, including both pairs on 000639, the vehicle carrying the fleet's worst fault. This rule does not find the pack\_2 leak. A different detector does.
- One screening rule is not a fleet intelligence layer. That is the argument for the layer, and it is why the negative result is on the slide.
- The rule was chosen retrospectively. This is a reproducible historical screening rule, not unbiased prospective lead time.
- Alert dates are earliest month-completion bounds. Ingestion and producer processing latency are unknown and excluded; upstream producer causality was not audited.
- Warm-depot pairings are candidate pairs only until topology is independently confirmed.

# The loop corrected itself

- Batches were reviewed before the next launch. The metrics script reached revision 3, and after each revision every table was regenerated for all 11 vehicles.
- Rev 1 to 2: valid-brick gating of SOHC statistics · raw-SOHC counterparts · position-adjusted resistance · park-step leak detector · pack-pair SOC offsets with voltage confirmation.
- Rev 2 to 3: artefact spike rejection (>30 % single-event steps never confirmed) · leak test requires SOC match across the pack · single-valid-brick bias flags.
- The caught flaw, right: the filtered SOHC against the raw anchor on one brick. The filtered estimate lags its raw anchor by up to 5 points. Rev 2 made every report carry both bases.



Left: pack medians, filtered against raw. Right: pack 3, bricks 39–41. x-axis: cumulative EFC.

# Guardrails: artefacts caught and gated

*Never reported as battery faults*

- Sentinel values (6.5535 V, 215 degC, 3353.5 A) in the validity stream, telemetry artefacts with monthly bursts above 20 % of samples.
- Fleet-wide odometer counter steps: 7 vehicles simultaneously in one 2025-08 gap, one jumping 64 to 1,716 miles on under 0.07 EFC of throughput. One vehicle interleaved two counters, so 339,000 miles of phantom increments were discarded, caught by cross-checking the battery's own EFC counter.
- Self-discharge event dedup trap: same-timestamp rows are distinct checkpoints. A naive dedupe deletes 58 % of the self-discharge evidence; the extractor keys on checkpoint index.
- Estimator re-anchoring spikes: single-event  $\Delta$ SOC of +5 to +48 % with zero voltage change, rejected by the >30 % step rule and the 1 %SOC to 7-10 mV OCV check.

# Known limits, published in every report

- 72x dilution: a single cell is never measured. Small single-cell deviations are diluted at the brick terminal, so all cell attributions are tiered inferences.
- 15 sensors for 90 bricks: temperature is a zone property. A single hot cell cannot be located.
- Coverage is 53-69 % of days, and histograms cover powered hours only, so cold soaks are systematically under-counted.
- Cell spec unusable absolutely: the DCIR table is a placeholder from a different cell, so resistance findings are relative-only and position-adjusted.
- Every open question carries its settling experiment: quiescent-current measurement, IR survey under charge, string schematic. The agent proposes the next measurement rather than closing the question.

# The fleet-1 evidence audit, hashed

*Evidence audit executed 2026-09-20; the audit script is itself hashed. This is a retrospective evidence review, not a field-performance validation. It covers the fleet-1 corpus; the fleet-2 and fleet-3 slides were added later. A hash fixes which file was used, not whether its content is right.*

| Audit item                    | Value                                                          |
|-------------------------------|----------------------------------------------------------------|
| Source files hashed (SHA-256) | 46                                                             |
| Run manifests captured        | 23, spanning 3 distinct library commits                        |
| Files inventoried             | 3,926 in the analysis tree; 256,926 in the runner tree         |
| Ledger integrity              | 296 rows · 0 blank evidence pointers · 5 placeholder citations |
| Prefix invariance checks      | 66 passed                                                      |

# Backup: estimate accuracy and BMS filter lag

- 1.47 % SOH MAE, zero-shot, on a held-out NMC e-bus fleet, from about 20 % of observed life.
- The BMS-filtered SOH lags its own raw anchor by 3.3 to 5.3 points. Self-discharge is thresholded to zero on 88 % of events.
- Pack-median SOH reads 96.84 % on the one pack that needs a module-level decision.

