Solutions.
People.
Energy.℠

SPE Summit: AI – The Next Evolution of DDAA

9 – 10 June 2026| KUALA LUMPUR, MALAYSIA

LLM-powered Quality gate for Incident and RCA Reports: Field Pilot Results and Practitioner Insights

Alexander Krivosheev
(AI-Solutions Energy)

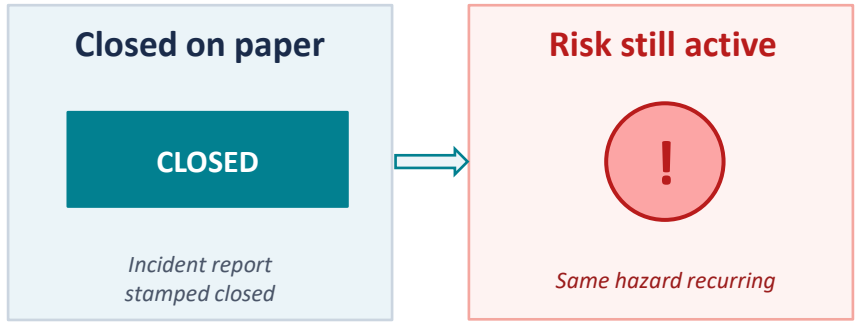


Solutions.
People.
Energy.℠

Weak RCA closes the file, not the risk

Closed report ≠ closed risk: weak causal logic leaves the same hazard live.

- Wrong cause → repeat incidents
- Weak corrective actions → live risk
- Review depth varies under load
- Senior expert time is scarce

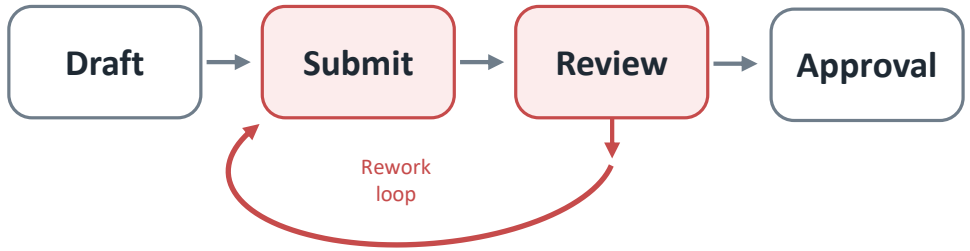


Solutions.
People.
Energy.℠

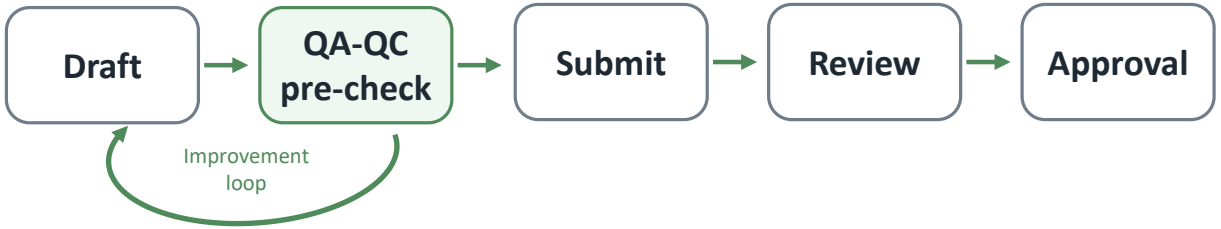
A 24×7 pre-submission quality gate

Expert reviewers should not be the first reliable quality filter

Pain to address



Target state



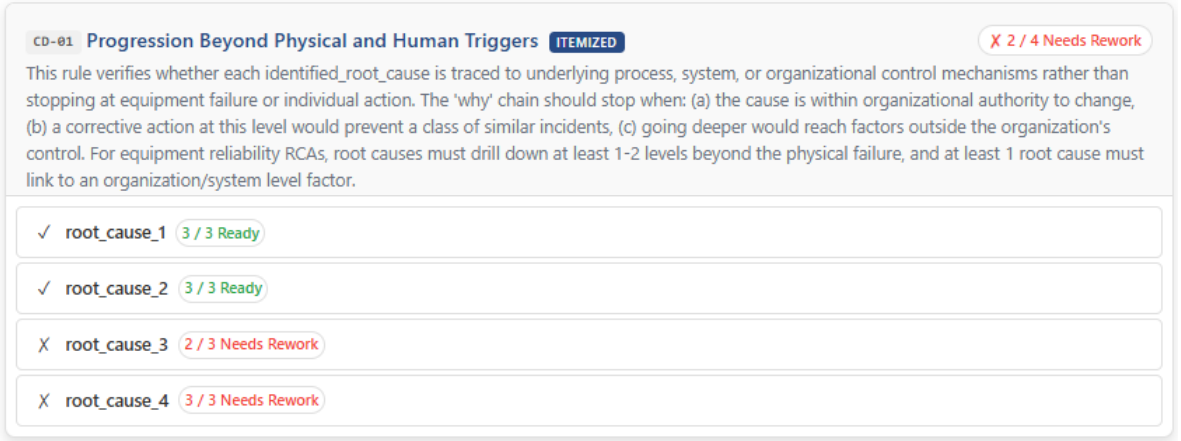
Improvement loop before expert review

- Author uploads the draft report
- Receives feedback and improvement recommendations
- Adjusts report before submission

Solutions.
People.
Energy.℠

Output the author can act on

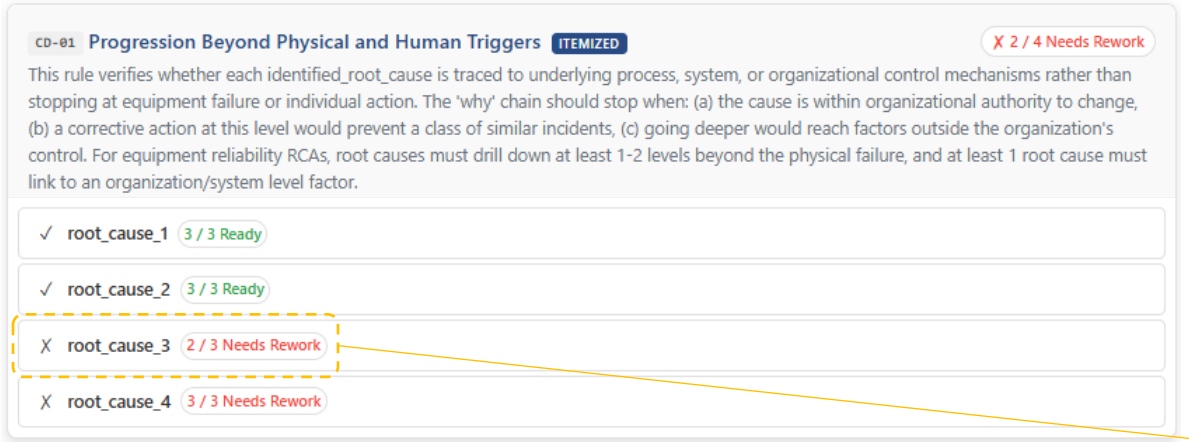
- Verdict: ready / refinement / rework
- Quoted evidence from report
- Specific improvement guidance





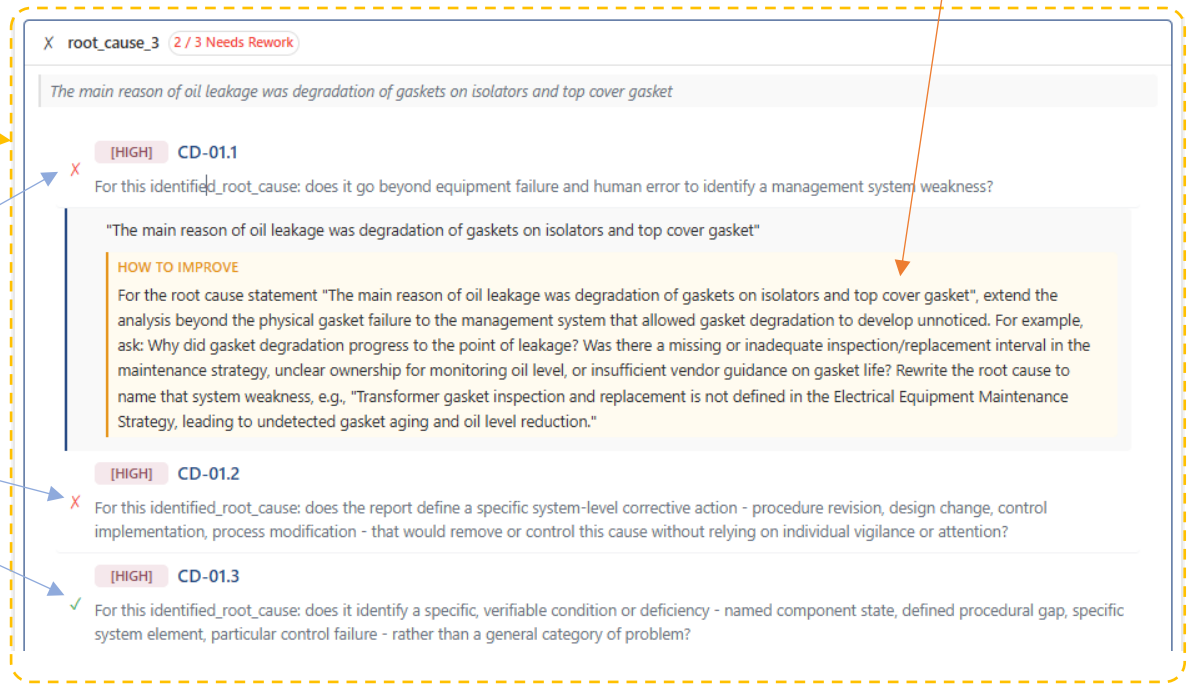
Solutions.
People.
Energy.℠

Output the author can act on



recommendation

Elementary_checks

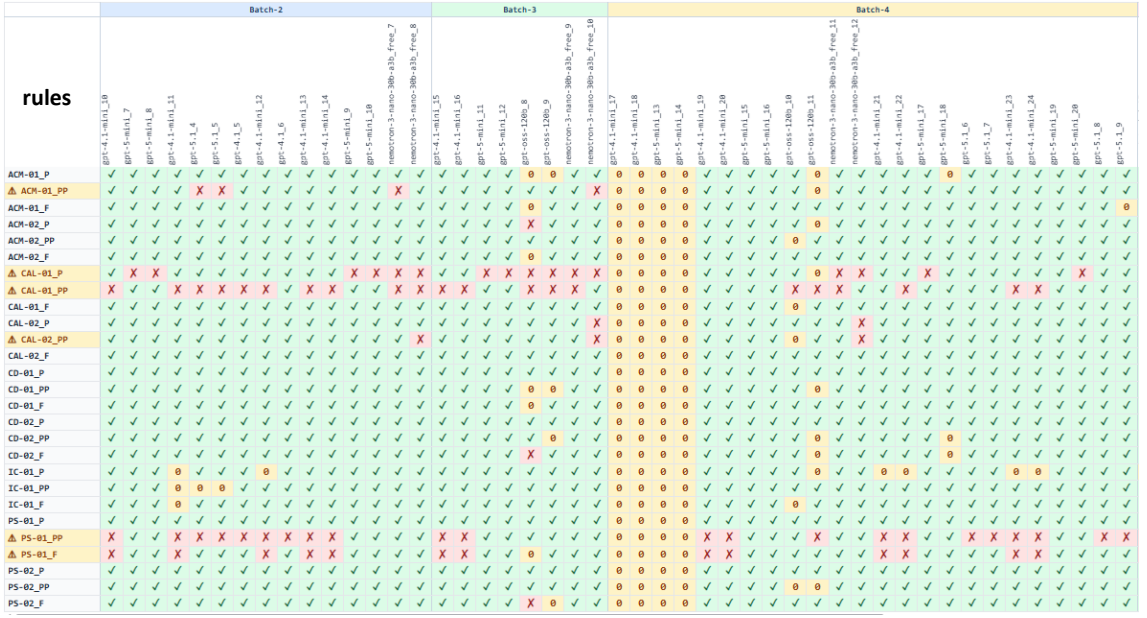


Solutions.
People.
Energy.℠

Reliability engineered, not assumed

Per-rule and full-report tests across multiple LLMs before any live use

- Per-rule synthetic tests (with gold standard)
- Full-report tests (with embedded flaws)
- 15-20 refinement cycles per stage
- 100++ runs , 3-5 LLMs each
- Heat map drove rule pruning
- Validated against SME expectations



Solutions.
People.
Energy.℠

Results from real operational use

8-week pilot and 20+ weeks to date
14 RCAs, ~70 incident reports

- Faster average approval cycle
  [Calendar days, submit-to-approved]
- Rejected for rework count decrease

submit-to-approved cycle reduction

>40%

poor-quality rework

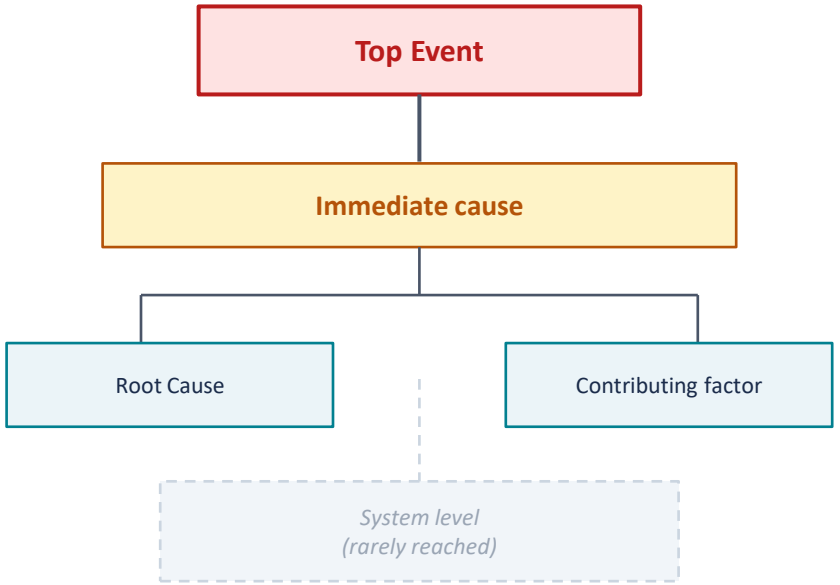
~9% → <5%

Baseline: 6 months pre-deployment, same asset.
Process owner reports stable trend post pilot handover.

Patterns surfaced in 50+ past reports

Things which hard to see from single report analysis

- Contributing factors often labelled as causes
- Causal chains tend to stop above system level



Solutions.
People.
Energy.℠

Lessons learned

A technically correct assistant can still fail adoption.

Lesson	What changed	Status
Output language	Softened wording: ~~Fail~~ → rework / ~~partial_pass~~ → refine / ~~pass~~ → ready	✓
Strictness	Adjust thresholds to flag material gaps and not to make red-review	✓
Orchestration	LLM checks one rule per call. Proved to be more reliable.	✓
Adoption	Voluntary use → uneven adoption.	⟳

✓ Done ⟳ Ongoing

Solutions.
People.
Energy.℠

Method portable, rule pack is not

Architecture is generally universal

- *Converting assurance tasks into binary questions*
- *Using test bench for rules calibration*
- *Verdict structure and citation grounding*
- *LLM-rendered recommendations*

Possible candidates for LLM-powered QA:
well intervention proposal, project framing, vendor submissions, MOCs…

Rule pack design is customized

