

Aero specialist pipeline

Dated evidence addendum | October 10, 2026

A frozen engineering-intent reader plus deterministic format, units and solver-admission checks. Optional stock model reviewers were evaluated separately. This is a specialist pipeline, not a true mixture-of-experts model.

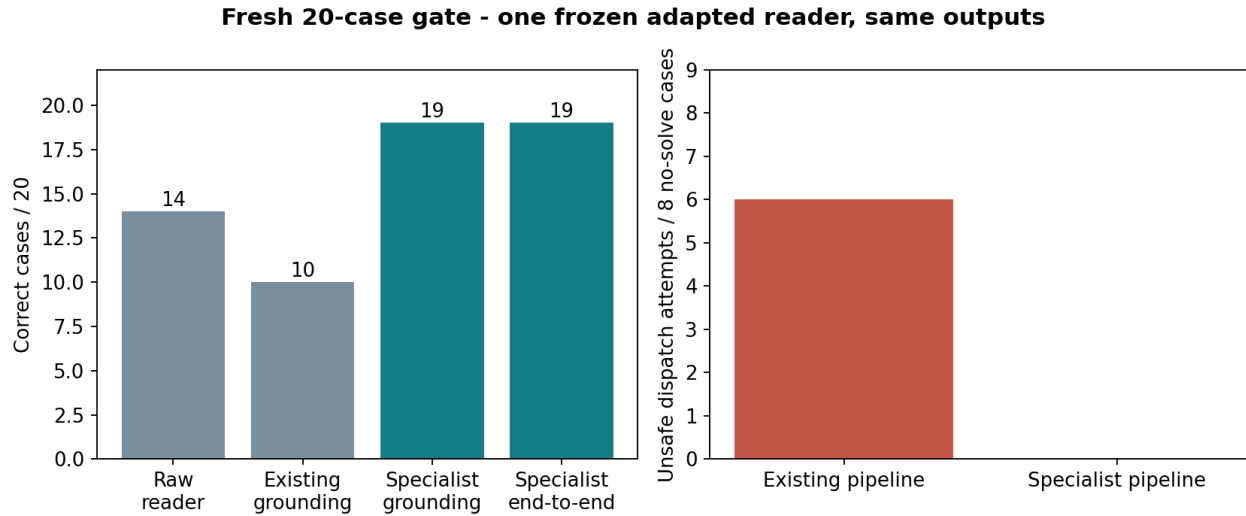
Measure	Established result
Fresh adapted-reader gate	19/20 strict end-to-end cases; 12 actual accepted solves
Unsafe solver admission	Existing: 6; specialist: 0
Existing language regression	615/618 before; 615/618 after; 0 newly broken
Code and interface tests	36 passed: 32 frozen core + 4 additive MCP interface
Checker overhead	7.75 ms median; excludes inference and solver
Local native reader timing	38.04 s median; NF4/BF16/SDPA, RTX 3090

Scope and authority

Personal, constructed FEA fixtures only. These results establish bounded analytical verification and rejection behavior, not experimental validation, broad aerospace knowledge, whole-platform certification or autonomous design authority. Cloud assistance authored code and reporting; model inference and CalculiX execution were local. No new fine-tune or training-data ingestion occurred.

Deployment: explicit opt-in local API/CLI. Existing private reader and public conversation gateway were not silently rerouted. The stock reviewer is experimental and is not enabled by default.

1. Acceptance and provenance



The implementation, unit tests, reviewer prompt and scoring code were hash-sealed before the 20 new cases were constructed and before inference. Inference received only IDs and original request text. Baseline and specialist used the same fresh output journal, not separately selected responses.

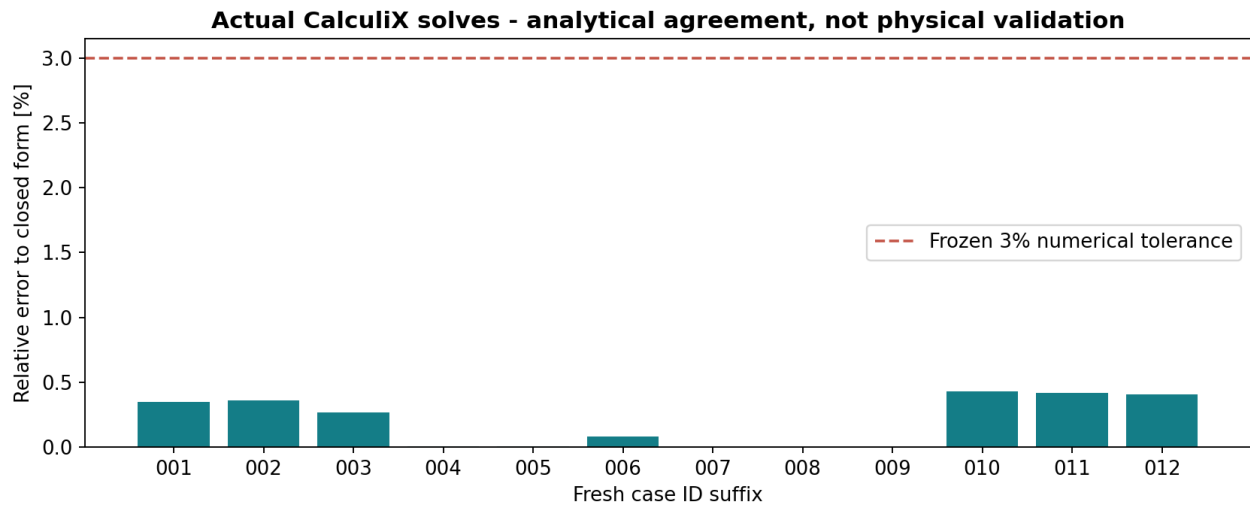
The prior 40 cases are retrospective development replay, not another independent holdout. The 618-case comparison reuses cached model predictions; it measures grounding regression, not fresh model performance. The new 20 are same-author synthetic holdout cases, not an external independent benchmark.

Format checks accept bare JSON or one complete JSON fence only. Duplicate keys, nonfinite values, extra/missing properties, invalid enums and arithmetic expressions are rejected. Unit evidence retains field-bound source spans. Conflicts or missing evidence stop dispatch. Reviewer feedback may veto a valid hard gate but never override a failed one.

The shared native GPU lease, original thermal policy, physical-memory admission and owned-process cleanup were honored. A separate CPU-only reviewer diagnostic ran alongside part of native inference; latency is observed workload timing, not an isolated maximum-speed hardware benchmark.

Audit: 203 retained solver artifacts hash-verified; 12 observables independently recovered from raw DAT. Hashes establish identity, not scientific truth.

2. Numerical proofs



Four verified families: rectangular cantilever and simply supported beam displacement magnitudes; signed axial stress in a uniformly heated/cooled, axially restrained, laterally free bar; open-ended thick-cylinder bore hoop stress with zero external pressure. Beam direction-y is represented by swapping the section axes into the solver template; signed beam displacement is not claimed.

Independent references: rectangular Timoshenko bending + shear; signed stress $-E \alpha \Delta T$; Lamé bore hoop stress. Numerical tolerance was fixed at 3% relative error; final mesh change $\leq 1\%$ where applicable. The uniform thermal field is exact on one mesh and has no mesh-convergence claim.

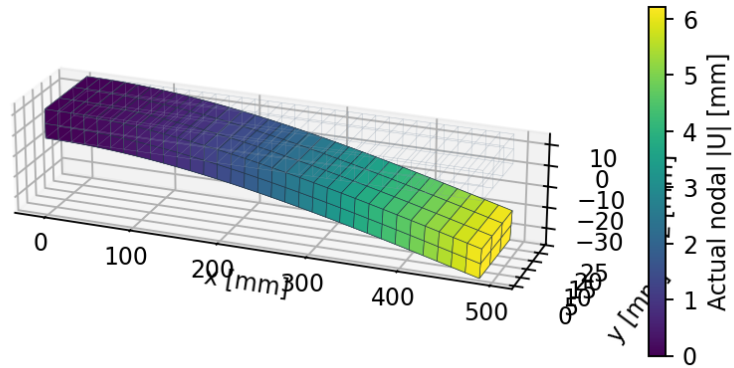
Case	Observable	FE value	Reference	Error %	Mesh change %
001	Cantilever	6.20682	6.22832	0.3452	0.1
002	Cantilever	5.00501	5.02306	0.3594	0.099
003	Cantilever	1.50923	1.51323	0.2649	0.089
004	Simply supported	0.39002	0.389999	0.0054	0.008
005	Simply supported	0.314641	0.314618	0.0073	0.009
006	Simply supported	0.095291	0.0952116	0.0834	0.053
007	Thermal bar	60.9336	60.9336	0.0000	Not applicable
008	Thermal bar	-56.6163	-56.6163	0.0000	Not applicable
009	Thermal bar	-90.7718	-90.7718	0.0000	Not applicable
010	Cylinder	11.3285	11.3777	0.4320	0.42
011	Cylinder	12.4479	12.5	0.4165	0.405
012	Cylinder	13.5596	13.6146	0.4042	0.394

FE value and reference units: mm for beam displacement; MPa for thermal axial and cylinder hoop stress.

Unsafe baseline dispatch attempts were intercepted by test assertions before knowingly incorrect physics was executed. These assertions are scoring protections, not evidence that the old pipeline was safe. Every original failure remains retained.

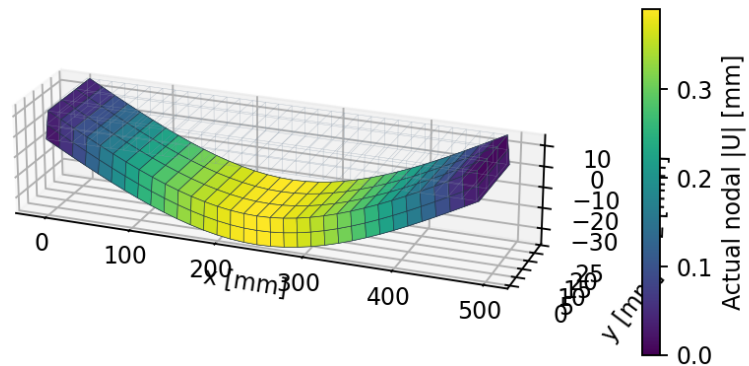
3. Actual FEA fields: bending

Cantilever - CalculiX displacement field



C3D20R corner-face rendering; deformation 5.9x; colour = corner-node mean $|U|$. Accepted observable: 6.20682 mm.

Simply Supported - CalculiX displacement field

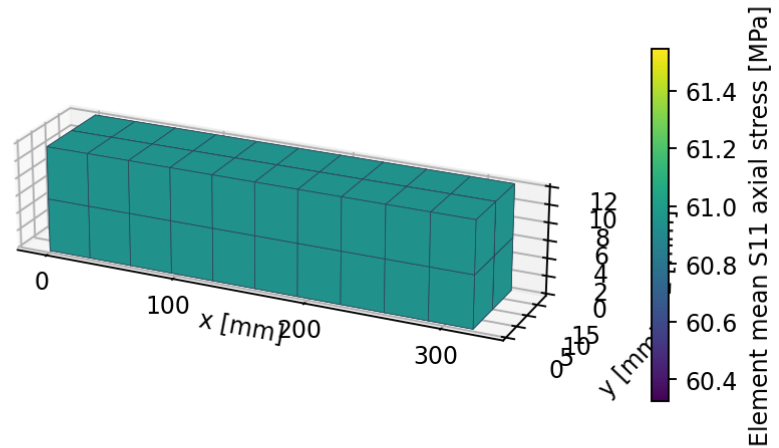


C3D20R corner-face rendering; deformation 93.5x; colour = corner-node mean $|U|$. Accepted observable: 0.39002 mm.

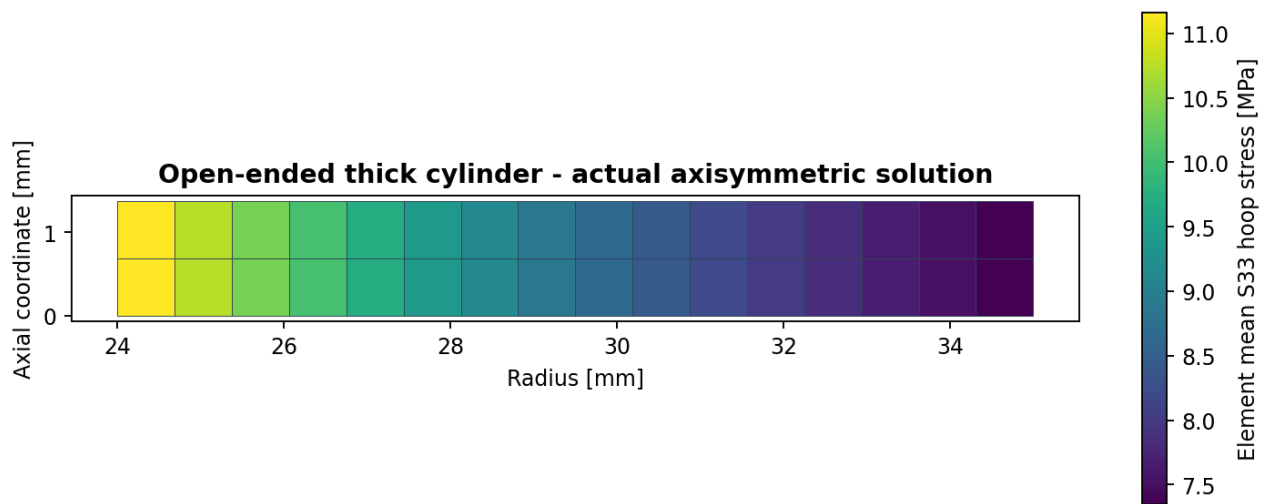
Geometry, connectivity and displacement values come from retained CalculiX INP/DAT files. The render uses linear corner-face visualization of quadratic elements, not a new solver. Deformation magnification is annotated; colour is an actual corner-node displacement average. Acceptance uses the documented face/neutral-axis observable, not the largest colour value.

4. Actual FEA fields: thermal and pressure

Axially restrained thermal bar - actual stress field



C3D20R; undeformed geometry; integration-point mean stress. Signed stress: 60.9336 MPa. Free lateral expansion.

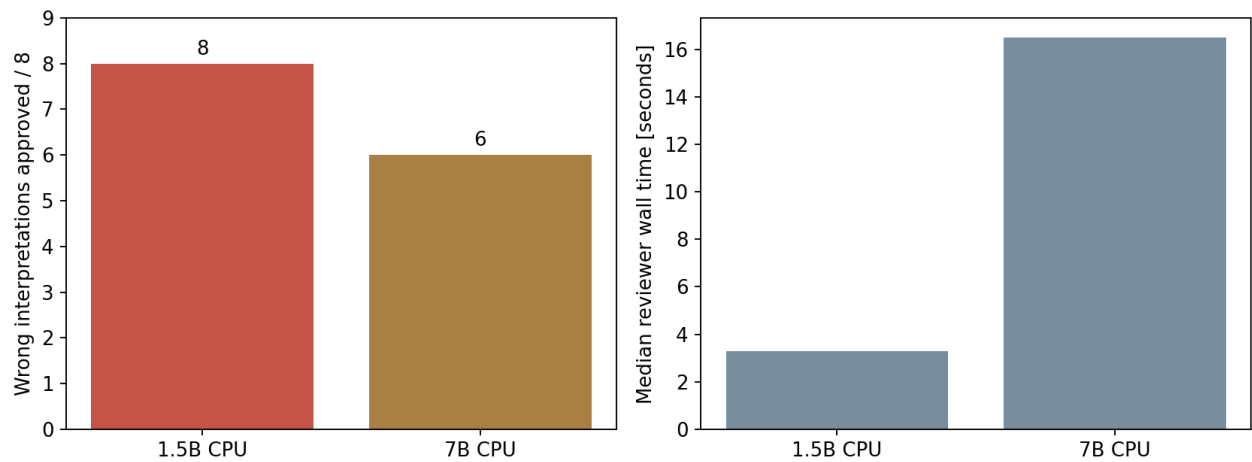


CAX8 radial/axial strip, not a 3D cylinder. Colour = integration-point mean. Accepted bore observable: 11.3285 MPa.

Stress colour is the mean of actual integration-point values by element, not invented nodal extrapolation. The cylinder is the actual axisymmetric radial/axial strip, not an illustrative 3D cylinder. Field images alone establish neither mesh independence nor engineering design readiness.

5. Reviewer models and remaining limits

Stock reviewer diagnostics - 12 paired interpretation checks each



Stock local checker	Correct / 12	False accept / 8	Median seconds
qwen2.5:1.5b	4/12	8/8	3.281
qwen2.5:7b	6/12	6/8	16.495

Reviewer checks are diagnostic mutations: four correct interpretations and eight deliberately wrong ones. Both models used the same frozen prompt, stock GGUF weights, CPU-only execution (four threads, num_gpu=0), greedy decoding and no retries. Wrong-answer approvals are retained. Shared model-family errors remain possible. A model reviewer is not independent computational evidence.

Recommendation: use deterministic checks now; retain the reviewer interface for experimentation. Fine-tune a checker only with independently authored training examples and a new acceptance packet that demonstrates added error detection without excessive false vetoes. Never train on these reserved answers or call a permissive reviewer a safety gate.

Fresh failures retained: pipeline_fresh_018: schema-invalid load_type "uniform". Dispatch was blocked, but the incorrect model enum still counts as a strict correctness failure. The complete raw response and validation record remain in cases.json.

Permanent evidence and integrity

Public page: <https://alex-blythe.com/software/aero/evidence/reports/specialist-pipeline-2026-10-10/>
The package includes the sanitized report, case decisions, raw model outputs, solver inputs/results, figures, reproducibility ZIP, source hashes and per-file manifest. No private corpus, model weights, credentials or host logs are included.

Frozen pipeline SHA-256: c6a9df78a330eb6c42e44d034e056f24b7d0e78430f0abf465b61396e6c90320