Evidence & Technical Specimens

From: Joe Maxwell from Joe Maxwell
joesystemica@substack.com

To: joseph.maxwell02@proton.me
joseph.maxwell02@proton.me

On: Tuesday, 26 May 2026 at 0:22:54

Evidence & Technical Specimens

The working artefacts behind Systemica Engineering

JOE MAXWELL
MAY 25

The development of stronger proofing for the engine is ongoing.

The current work includes front-end cleanup, larger sample testing from the Materials Project, runtime proofing, practical papers, images, minimal viable demos and video walkthroughs. These are not all included in this first public specification. They will be finalised and publicised as the runtime becomes cleaner.

This page is a first explainer.

Its purpose is to show what Systemica Engineering is building, what technical artefacts already exist, and how the current evidence supports the direction of the work.

The core idea is simple:

technical claims should earn their confidence before people act on them.

Systemica Engineering is being built around evidence packs: structured outputs that show assumptions, constraints, model behaviour, rejected options, residual risk and the level of confidence a claim has earned.

This is not a claim of certification.

It is not a replacement for domain expertise, high-fidelity simulation, physical testing or peer review.

It is a pre-validation and audit layer.

1. GIL/EDK: Behavioural Feature Auditing

GIL/EDK is the current model-audit runtime.

It was developed around a simple problem in machine learning: feature importance is often treated as explanation too early.

A feature or descriptor may rank highly in one fitted model, but behave unreliably when the data slice, model family, split, perturbation level or audit condition changes. In scientific or engineering settings, that matters because feature interpretation can influence physical follow-up, design decisions or research direction.

The GIL/EDK papers reframe feature importance as a behavioural object:

static rank → perturbation trace → behavioural audit

The descriptor-stability paper states the aim directly: the audit records descriptor drift, descriptor variability, recurrence, negative-control behaviour, tier-removal effects and phase-density occupancy, not to prove causality or physical truth, but to ask which descriptors remain reliable enough under stress to deserve follow-up.

The broader feature-auditing paper makes the same point in general machine-learning terms: a static ranking is a snapshot, while an audit trace is a behavioural object.

Current outputs include:

- descriptor drift and variability tables
- baseline versus audit ranking comparisons
- dynamic perturbation sweeps
- dropout threshold plots
- phase-density projections
- barcode checkpoint tables
- claim-ceiling style action labels

The important output is not just a plot.

It is a decision-support layer.

The barcode states translate model behaviour into action guidance. For example, descriptors may be labelled as invariant_core, stable_influence, conditional_stable, fragile_proxy, volatile_collapse or interaction_dominant, with each state mapped to a different follow-up action.

This makes the audit practical:

- stable descriptors can be retained for cautious follow-up
- conditional descriptors need recurrence or sweep testing
- target-adjacent descriptors can be useful predictors but weak explanations
- fragile proxies should not be over-read
- volatile descriptors should not be used as explanatory evidence

The current materials proof uses a shear-modulus dataset as the origin artefact. It is explicitly treated as proof-of-concept and validation-track evidence, not broad materials-property validation. The paper notes that the current evidence layer does not yet support benchmark validation, physical certification or broad materials-property generality.

That limitation is part of the method.

The runtime is designed to preserve weak outcomes instead of hiding them.

2. TRS-GL: Thermal Runtime Studio

TRS-GL is the engineering-side pre-validation runtime.

It comes from the furnace viewing-cartridge project: a modular viewing cartridge inside a simple insulated furnace wall.

The design problem was bounded. It was not a full furnace redesign. It was a single wall-section problem with a high-risk local feature: a viewing port interrupting an insulation-dominated wall.

The report defines the project around a wall section and removable cartridge, with the design keeping the wall simple while concentrating optical access, sealing, compliance and serviceability into the local cartridge.

That makes it a useful technical specimen.

The practical design logic was:

- keep the main wall insulation-dominated
- isolate the viewing function as a removable subsystem
- separate heat path, load path, sealing and maintenance
- screen non-compliant configurations before expensive validation
- preserve residual risks for later prototype testing

The modelling workflow was deliberately glass-box.

The report describes a deterministic pre-design evaluation using explicit geometry, material properties and boundary conditions. The workflow combines a 1D baseline model, 2D plain-wall parity check, 2D viewport perturbation case, bounded cavity-transport sweep and governing-case selection. It also explicitly states that the workflow is not an optimiser and not a full radiation, transient or stress model.

That is the claim ceiling.

TRS-GL is not intended to replace COMSOL, Ansys, CFD, FEA, certification or physical testing.

It sits before them.

Its purpose is to narrow early design space, expose governing assumptions and reject weak configurations before they absorb time or budget.

3. Furnace Viewing Cartridge: Engineering Evidence Pack

The furnace project is the first complete engineering evidence-pack specimen.

It shows how a design decision can be made traceable.

The workflow is:

1. define the wall and requirement set
2. build a 1D thermal baseline
3. compare against a 2D plain-wall model
4. introduce the local viewport perturbation
5. run cavity sensitivity

6. identify the governing case
7. reject non-compliant thicknesses
8. select the final design thickness
9. record residual risk

The report uses this workflow for elimination rather than optimisation.

The key result is the insulation-thickness decision.

The report records that the cavity-transport condition governs the local safety case and that 350 mm is the minimum compliant final selection against the 60 °C external surface-temperature target.

The poster version makes the engineering logic visible in one page. It records the screened external surface result as 59.2 °C against a 60 °C target, with 350 mm passing the model and 375 mm selected as the practical build thickness.

The useful part is not only the final number.

It is the trace:

- what was assumed
- what was modelled
- what failed
- what passed
- what remained uncertain
- what needs higher-fidelity validation later

That is the evidence-pack format Systemica Engineering is moving toward.

4. Current status

Current technical status:

GIL/EDK

Working proof-of-concept and paper-facing runtime logic exists.

Current evidence supports:

- perturbation-based descriptor auditing

- drift and variability tracking
- dynamic sweep outputs
- barcode decision-support states
- proof-of-concept materials ML case study

Current evidence does not yet support:

- benchmark superiority
- physical causality
- descriptor certification
- DFT or experiment replacement
- broad materials-property generality

TRS-GL

A full engineering case study exists through the furnace wall and viewing-cartridge project.

Current evidence supports:

- deterministic pre-validation workflow
- early elimination of weak configurations
- traceable thermal decision-making
- evidence-pack structure for engineering design

Current evidence does not yet support:

- full transient thermal validation
- stress validation
- manufacturing qualification
- certification
- replacement of high-fidelity simulation or prototype testing

Systemica Engineering

The practical direction is to combine these into a broader evidence-pack workflow:

- model-audit evidence packs

- engineering pre-validation evidence packs
- runtime-generated decision traces
- claim ceilings
- visual summaries
- reproducible technical bundles

The next public proof layer will include cleaner runtime outputs, screenshots, images, MVP demonstrations and video walkthroughs.

5. Pilot direction

The first practical offer is likely to be a bounded model-interpretation audit.

Input:

- small non-sensitive CSV
- descriptor table
- feature-importance output
- early-stage scientific or technical model result

Output:

- stability table
- drift and variability summary
- dynamic perturbation trace
- barcode state/action map
- claim ceiling
- suggested next validation step

The question is not:

what did the model rank highest?

The question is:

which interpretations remain usable when the modelling conditions are stressed?

A second offer is an engineering evidence pack for early design decisions.

Input:

- candidate concept
- requirements
- failure modes
- simple geometry/material assumptions
- existing calculations or model outputs

Output:

- requirement map
- failure-mode response table
- model workflow
- rejected-option trace
- governing-case logic
- residual-risk list
- next validation recommendation

The question is not:

what is the final perfect design?

The question is:

what can be safely eliminated before expensive validation?

Closing

Systemica Engineering is being built to make technical claims easier to inspect.

A useful evidence pack should show:

- what was assumed
- what was tested
- what failed
- what survived
- what remains uncertain

- what level of confidence is justified

The aim is not to make models look more impressive.

The aim is to make the next decision more accountable.

SHARE

LIKE COMMENT RESTACK