# SYSTEMICA ATLAS V2.1.2

Engineering decision reliability, evidence, AI boundary and current capability

Public Release 2.1.2 - 21 July 2026

Joseph Maxwell

Founder-originated method; AI-assisted implementation; human-controlled authority

# Contents

# Systemica in 60 seconds

**Systemica helps engineers decide what modelling evidence is trustworthy enough to justify the next expensive action.**

| Question | Current answer |
|---|---|
| What is the first product? | `GIL-EDK`, an engineering decision-reliability layer |
| What does it protect? | Simulation, testing, fabrication, procurement and qualification decisions made from modelling evidence |
| What does the customer provide? | A bounded dataset or model output, target decision, constraints and current route |
| What do they receive? | Stability audit, weakness register, held/rejected routes, replayable evidence and the next validation question |
| What is working now? | Tested local kernels, Materials Project audits, replay, authority-separated memory bridge and three bounded engineering missions |
| What is not yet proven? | External customer validation, physical validation, quantified savings and unified production-ready MUOS experience |
| Current state | Local alpha/proof-of-concept; founder-authorised V2.1.2 public release |

The strongest demonstrated engineering statement is:

> Systemica has repeatedly converted early engineering concepts into traceable candidate families, identified hidden weaknesses, generated bounded design manoeuvres and narrowed the next physical test, while refusing unsupported promotion.

## The founder and the method

Joseph Maxwell is a materials engineering master's candidate and technical systems investigator who works on a recurring problem: complex technical systems generate plausible answers faster than people can determine which answers deserve trust.

His method is:

`unclear situation -> map it -> identify contradictions -> test it -> determine trust -> determine action`

Systemica is the technical estate he built to formalise and test that method across modelling, engineering design, project reconstruction, evidence governance and recoverable workflows.

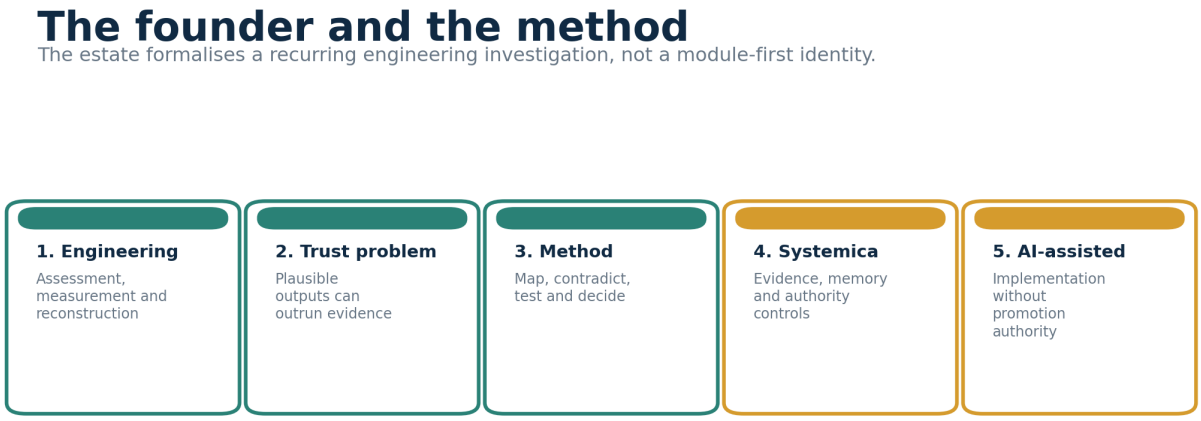


Figure 1: Founder-method lineage.

The practical foundation came first: materials assessment and measurement, reconstruction of a thermal-barrier-coating workflow after expertise loss, DOE/ANOVA process optimisation, recovery of manual methods when automation was inadequate, surface-measurement work, technical knowledge architecture, and applied neural-network/model-search development. These experiences explain why the architecture distrusts plausible automation, preserves failure and keeps evidence separate from authority.

### From model optimisation to governed evidence

The historical coursework route evaluated all 127 non-empty subsets of seven candidate descriptors. One recorded run selected `(band_gap, bulk, elastic_anisotropy, poisson)` with `R2 = 0.950420` and scaled `RMSE = 0.013689`. A five-repeat random-split route recorded a different leading averaged combination, `(formation_energy_per_atom, band_gap, bulk, elastic_anisotropy, poisson)`, with `avg R2 = 0.851499`.

These are historical notebook outputs, not independently replayed V2.1.1 runtime evidence. The coursework reused the nominal test split during Optuna selection, early stopping and final evaluation, and calculated its headline errors before inverse transformation. Bulk modulus and Poisson's ratio were physically adjacent to the shear target, creating an evidential concern without proving universal leakage or causality.

That movement changed the governing question from *which model scores highest?* to *which relationship survives when the evidence is stressed?* GIL-EDK developed by making perturbation, destructive controls, recurrence and claim boundaries part of the result. Systemica then separated computation from the authority to accept evidence and from the memory that preserves accepted state.

Cosmologia is a later conceptual extension of the same systems lens: interactions and constraints shape a state space; structure records accumulated interaction; navigability matters more than single-objective optimisation; and

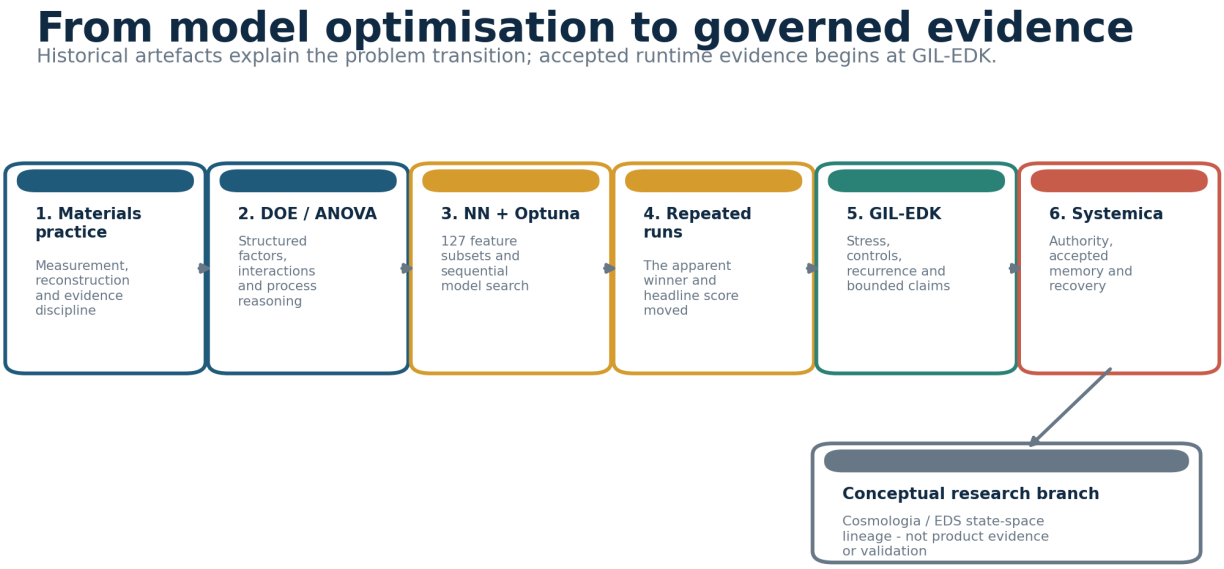


Figure 2: Method-to-product lineage.

drift, hysteresis and recovery affect which transitions remain available. In this release it is intellectual lineage and a research programme, not evidence for GIL-EDK, a validated physical theory or part of the initial commercial claim.

The public account intentionally excludes private medical and personal history. At a safe level, the architecture was also shaped by the need to work under interruption and variable capacity. That produced requirements for externalised state, checkpoints, handoff, asynchronous continuation, rollback and failure preservation - not an adversity or hero narrative.

## Where generative AI is - and is not - involved

Systemica's demonstrated core is **generative-AI-detached**, not universally AI-free.

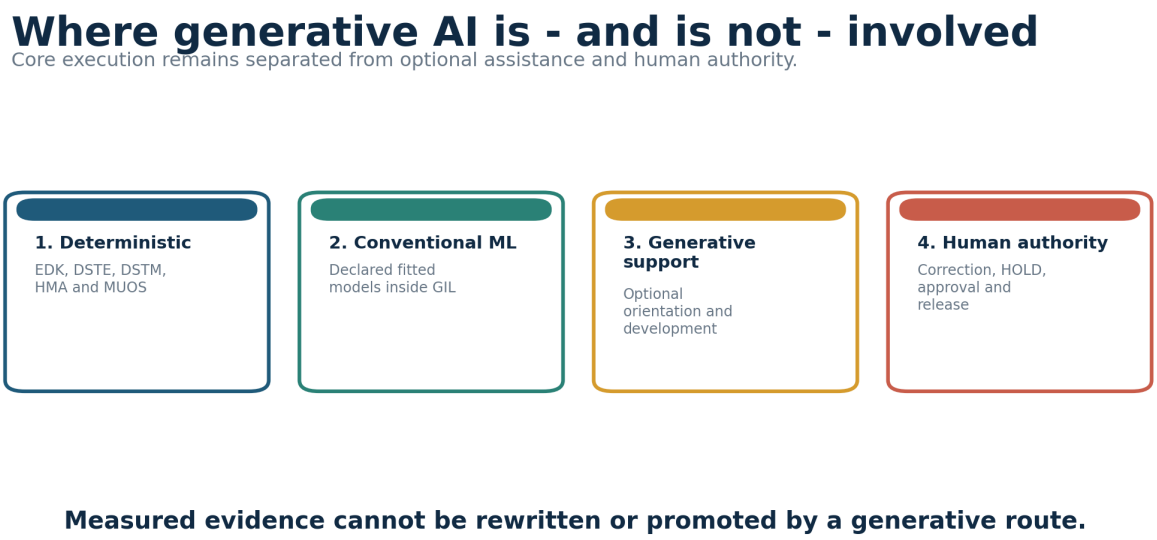


Figure 3: Generative-AI authority boundary.

| Layer | Current role | Authority |
|---|---|---|
| Deterministic software | EDK search contracts, DSTE/DSTM transitions, MUOS gates and HMA storage | Executes declared rules; cannot self-promote |
| Conventional fitted ML | Ridge, Random Forest, Extra Trees and HistGradientBoosting in the tested GIL route | Produces measured model evidence only |
| Optional generative AI | Local Ollama Project Intelligence and development assistance | Advisory/read-only; not required by accepted core routes |
| GPT/ChatGPT and Codex | Critique, synthesis, substantial implementation, tests, execution support and evidence organisation | No evidence-promotion or publication authority |
| Human founder/operator | Corrections, scope, HOLD, approval and release | Final responsibility |

A credential-free, socket-blocked acceptance run reproduced the accepted EDK and DSTM fingerprints, exercised DSTE acceptance/rejection, committed an operator-approved checkpoint and refused a held write before storage. No remote connection or generative service was observed. This is representative-route evidence, not a universal claim about every present or future module.

## Drift, state and recovery

A model can pass an evaluation today and still change tomorrow because its model version, prompt, context, retrieval sources, tools or operating environment changed. Systemica records the accepted state, observes the trajectory away from it, prevents unsupported changes from silently becoming authoritative, and preserves a controlled route back.

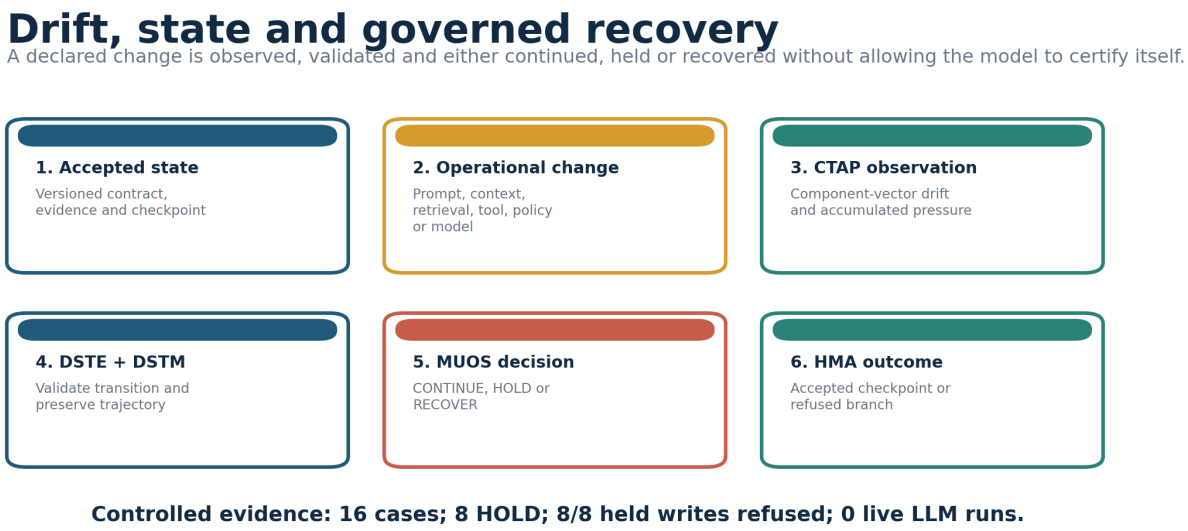


Figure 4: Drift, state and governed recovery route.

Systemica therefore has two connected product pillars:

| Product pillar | Governing question | Demonstrated bounded route |
|---|---|---|
| Decision reliability | Is this conclusion stable enough to justify the next expensive action? | GIL-EDK models, controls, perturbations, recurrence and claim ceilings |
| State integrity and recovery | Has an evolving workflow drifted from its accepted state, and can it be contained or restored? | CTAP observation, DSTE validation, DSTM trajectory, MUOS authority and HMA checkpoint control |

The controlled integration matrix exercised 16 declared changes. It produced 8 `HOLD` decisions, refused all 8 corresponding held-write attempts, replayed every structural decision without failure, and recovered the accepted checkpoint fingerprint exactly. Restart retrieval took 9.01 ms and bounded rollback took 26.14 ms on the test machine.

CTAP is reported component-vector-first. The equal-weight warning and hold scalar used here was a predeclared fixture routing aid, not production calibration. The matrix contained **0 live LLM runs**; one separate historical replay used a labelled synthetic long-context audit. The zero false-positive and false-negative counts therefore establish fixture self-consistency only, not real-world detection sensitivity or specificity.

The demonstrated functions map to lifecycle-monitoring, provenance, human-oversight, incident-response, recovery and change-management themes in NIST AI RMF / AI 600-1, ISO/IEC 42001 and the EU AI Act. This is a technical-control crosswalk, not a compliance or certification claim.

Maximum supported claim: **Systemica has demonstrated bounded local mechanisms for detecting declared forms of workflow drift, preserving temporal evidence, containing unsupported state changes and recovering accepted checkpoints under controlled tests.**

# 1. The engineering decision problem

Accuracy is necessary, but it is not the same as decision reliability. A model may score well while its conclusion depends on one data split, one model family, a target-adjacent feature, a narrow material population or an untested boundary assumption.

The risk appears after modelling: a ranked candidate becomes a high-fidelity simulation; a promising interface becomes fabricated hardware; a convenient topology becomes procurement and test work. If the conclusion was fragile, the expensive correction arrives late.

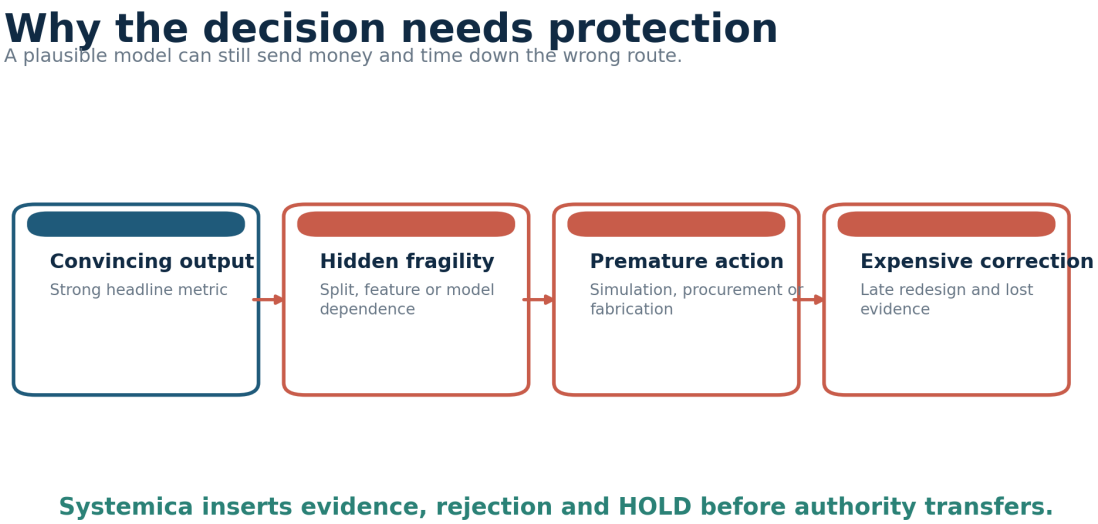


Figure 5: AV2-F02. Decision-risk chain. Source: founder product mandate. Interpretation: evidence and authority gates intervene before downstream commitment. Limitation: no quantified savings comparator exists.

Systemica is designed to move the stop/revise decision earlier, while preserving enough evidence to resume or branch safely.

## 2. What the user experiences

The user does not need to begin with the internal architecture. They begin with a bounded decision:

- Which descriptor or model conclusion is driving expensive downstream work?
- What assumptions and exclusions apply?
- What failure would make that conclusion unsafe to trust?
- What evidence would justify continuing?

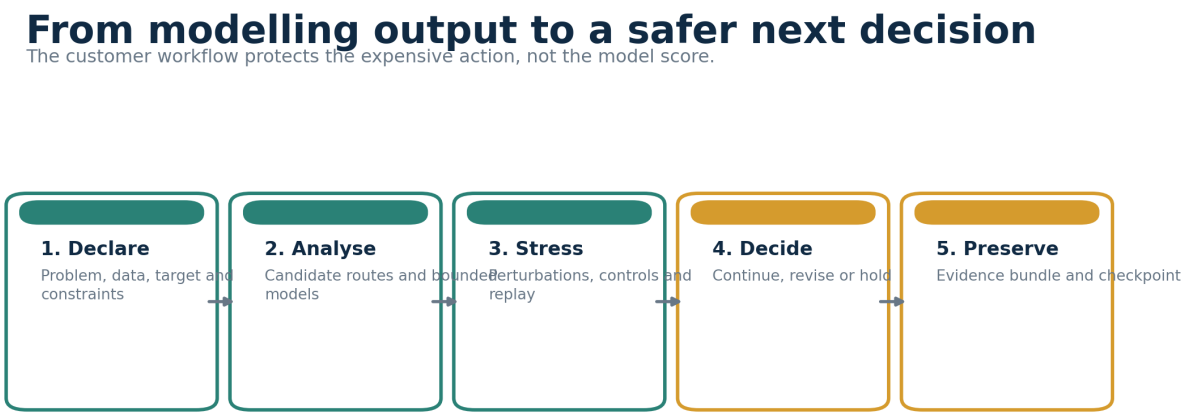


Figure 6: AV2-F01. Five-step customer workflow. Source: product mandate. Interpretation: the output is a governed decision packet, not merely another prediction. Limitation: public workflow proposition, not deployed-customer evidence.

### Ordinary workflow versus Systemica workflow

| Ordinary route | Systemica route |
|---|---|
| Receive data or a model output | Resolve a problem, evidence and authority contract |
| Optimise a headline metric | Run bounded models, controls and perturbations |
| Select the strongest candidate | Preserve weaknesses, rejected routes and descendants |
| Move toward simulation or fabrication | Issue `HOLD`, `REVISE` or `CONTINUE` with a claim ceiling |
| Reconstruct failures later | Preserve replay, lineage, checkpoint and rollback evidence |

This is a process comparison, not a measured time or cost comparison. A prospective comparator study remains required.

The possible outcomes are intentionally simple:

| Outcome | Meaning |
|---|---|
| `CONTINUE` | Evidence is coherent enough for the declared next gate |
| `REVISE` | A bounded weakness can be addressed through a new candidate or route |
| `HOLD` | Missing authority or unstable evidence makes continuation unsafe |

# 3. Present product and pilot offer

`GIL-EDK` is the first commercial product within the wider Systemica architecture. It is a working local alpha/proof-of-concept for auditing whether modelling conclusions remain stable enough to justify further engineering validation. **GIL-EDK is the initial commercial wedge; state integrity, CTAP and governed recovery are the defensible expansion layer.**

## Minimum pilot intake

- dataset or structured model outputs;
- target, ranking or decision of interest;
- known constraints, exclusions and leakage risks;
- current modelling route;
- expensive next action;
- confidentiality and deployment constraints.

## Bounded customer output

- audit report and stability figures;
- assumptions and weakness register;
- rejected and held-route ledger;
- evidence bundle and replay receipt;
- explicit claim boundary;
- recommended next validation action.

No external pilot, customer validation or quantified productivity benefit has yet been established.

# 4. Founder and development chronology

The recent executable sprint sits on top of earlier engineering and systems thinking. Evidence class is explicit: founder declaration and private dated artifacts are not converted into machine-verified runtime history.

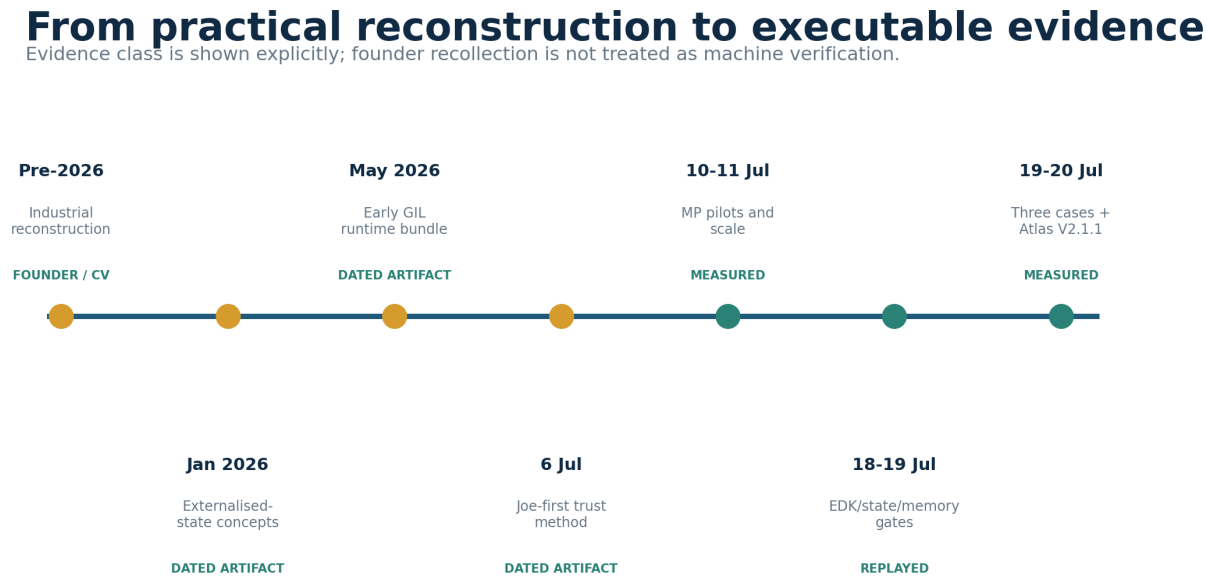


Figure 7: Evidence-classified programme chronology.

| Date | Event | Evidence class |
|---|---|---|
| Before executable Systemica | Materials assessment, workflow reconstruction, DOE/ANOVA, measurement and technical knowledge architecture | `FOUNDER DECLARATION` supported by controlled CV |
| Date not independently established in this pack | Coursework and notebook searched 127 feature subsets; repetition moved the apparent winner | `DATED ARTIFACT` |
| Later historical prototype; exact date not independently established | Dataset-agnostic NN and MI/correlation-pruning prototypes | `DATED ARTIFACT` |
| 15 Jan 2026 | Constraint, recovery and externalised-continuity concepts recorded in Atlas Architectura V2 | `DATED ARTIFACT` |
| 25 Jan 2026 | Safe-state planning and externalised-mind concepts further recorded | `DATED ARTIFACT` |
| By 10 May 2026 | Early GIL-EDK runtime bundle existed | `DATED ARTIFACT` |
| Before 10 Jul 2026 | GPT/ChatGPT critique and synthesis in use; exact first date not independently established | `FOUNDER DECLARATION` |
| 6 Jul 2026 | Document 0 packaged the Joe-first trust method | `DATED ARTIFACT` |
| 10-11 Jul 2026 | Codex-assisted MP pilots, locked repeats and scale work | `MEASURED` |
| 18-19 Jul 2026 | EDK migration, state stack, DSTM and HMA bridge gates | `REPLAYED` |
| 19 Jul 2026 | Three Design-3 missions completed with HOLD | `MEASURED` |
| 20 Jul 2026 | Atlas V2.1.1 offline acceptance, telemetry and release hardening | `MEASURED` |

| Date | Event | Evidence class |
|---|---|---|
| 21 Jul 2026 | V2.1.2 founder-authorised release and verified method-lineage patch | `DATED ARTIFACT` |

Calendar density does not measure continuous founder labour. GPT/ChatGPT and Codex accelerated the executable build inside a pre-existing method and constraint space.

# 5. Materials Project evidence

The Materials Project is a US Department of Energy-supported effort that pre-computes and exposes materials properties. Its official documentation provides a supported `mp-api` client and describes the data as a publicly accessible resource. [MP-01, MP-02]

The accepted local source snapshot produced deterministic nested subsets at 2,000, 5,000, 10,000 and 10,982 rows. The audited targets were shear modulus and bulk modulus. The conservative route used 19 descriptors, four model families, locked seeds/splits, noise/dropout perturbations and shuffled-target controls.

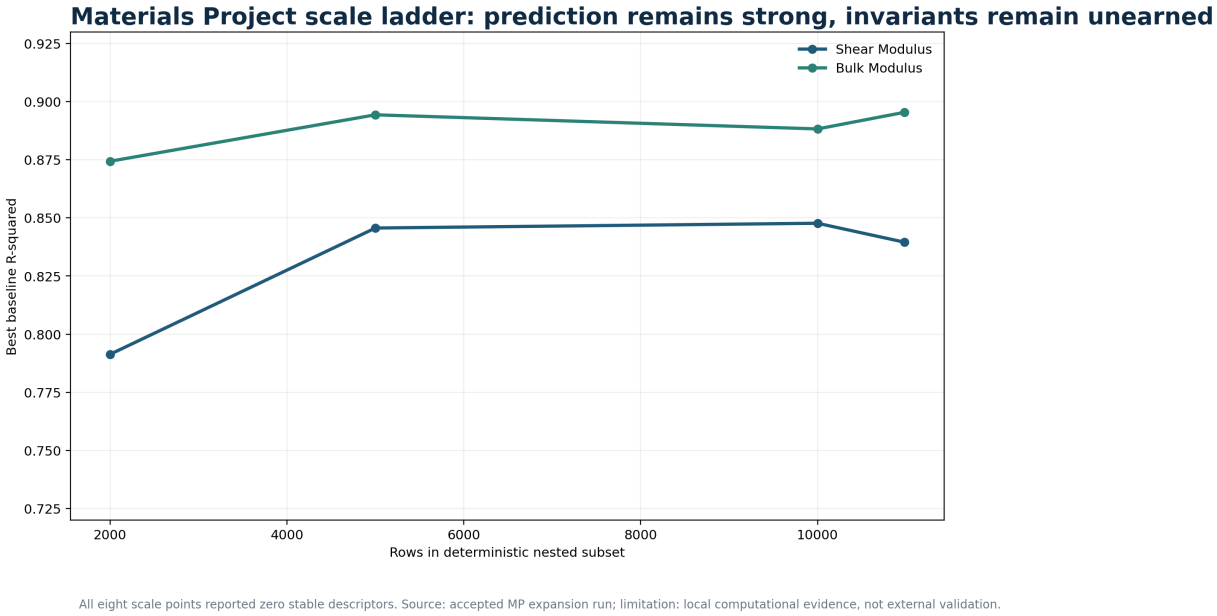


Figure 8: AV2-F05. Materials Project scale results. Source: accepted MP expansion run mp_controlled_expansion_20260711T153526Z. Interpretation: baseline prediction remained strong as scale increased, while no descriptor met the stable threshold. Limitation: local computational evidence; no causality or external validation.

At 10,982 rows, the best recorded baseline R-squared was `0.840` for shear modulus and `0.895` for bulk modulus. Every scale point reported `stable_count = 0`. `efermi` and `density_atomic` repeatedly carried predictive signal, but remained fragile rather than invariant.

That is the product lesson: an ordinary workflow might stop at a strong score and feature ranking. GIL-EDK preserves the distinction between **predictive** and **stable enough to promote**.

# 5A. Measured compute and memory telemetry

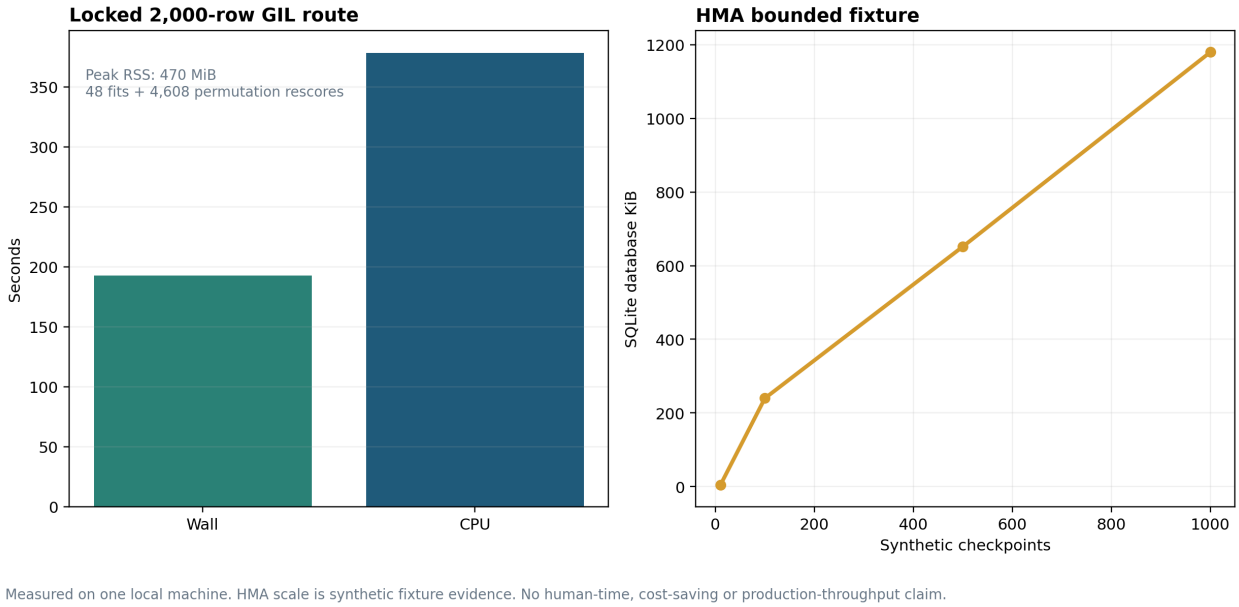


Figure 9: Measured local compute and synthetic memory scale.

The representative locked GIL route used 2,000 rows, 19 descriptors, shear modulus, four model families, four perturbation/control states and three seeds. It completed 48 fits and 4,608 permutation-rescoring evaluations in **192.83 seconds wall time**, with **378.22 seconds main-process CPU** and **470 MiB sampled peak process-tree RSS**. The accepted baseline result was reproduced at `R2 = 0.820026`.

The synthetic HMA fixture reached 1,000 checkpoints. At that point the SQLite file was 1180 KiB, retrieval p50 was 1.08 ms, p95 was 1.48 ms, and restart retrieval and rollback passed. This is bounded fixture telemetry, not production durability.

No measured human-time, customer cost or conventional-workflow comparator exists. The Atlas therefore makes no labour-saving or return-on-investment claim.

# 6. Holdouts, hierarchy and practical interpretation

The 5,000-row screening branch deliberately changed the population rather than only reshuffling it.

| Held-out population | Shear R-squared | Bulk R-squared |
|---|---|---|
| Aluminium-mixed chemistry family | 0.686 | 0.844 |
| Cubic crystal system | 0.677 | 0.812 |
| Stable class | 0.722 | 0.883 |
| Metallic class | 0.500 | 0.725 |

The same descriptor family was substantially more predictive for bulk than shear. Packing/density-only models retained more information than geometry/symmetry-only models, but the hierarchy work remains screening evidence rather than a physical causal chain.

**Protected decision:** before allocating more simulation or experimental capacity from a descriptor ranking, a materials team receives a record of what survived, what failed under population shift, and what evidence remains unresolved.

Materials Project itself cautions that its properties are calculated and that users should consult property-specific publications and database versions when interpreting expected error. [MP-03]

## 7. Three Design-3 engineering missions

The strongest transfer evidence comes from three different problems processed through the same disciplined sequence:

```
founder intent -> candidates -> weakness -> manoeuvre -> HOLD -> next physical question
```

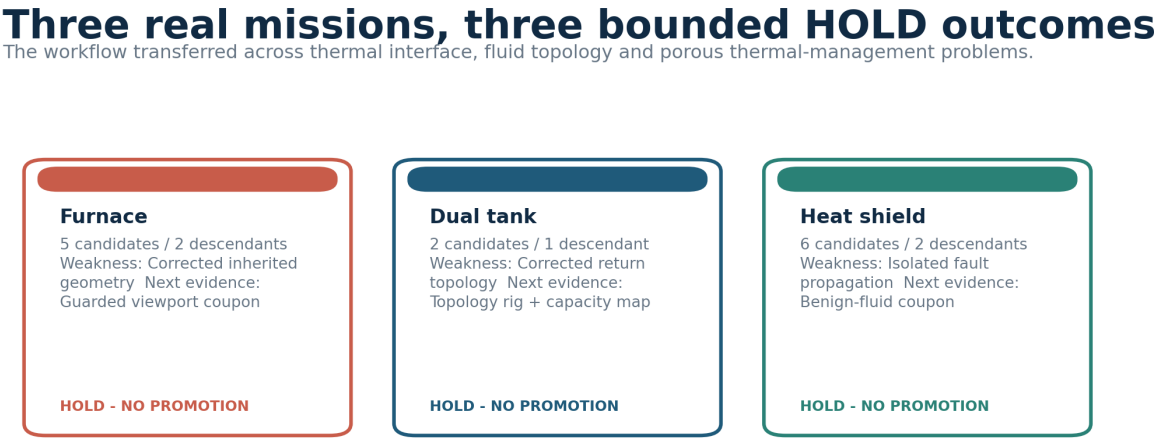


Figure 10: AV2-F06. Three-case comparison. Source: accepted three-case evidence unit. Interpretation: the process transferred across three engineering problem types while preserving HOLD. Limitation: no physical candidate was validated or authorised.

Across the three missions the receipts record 13 candidate objects, five descendants, at least 9,380 declared model operations and 15 removed or rejected routes. Human working time and source-orientation time remain `null`; solver operations are not presented as labour saved.

# 8. Furnace: correct the inherited geometry

The founder-authoritative wall envelope was 35-40 mm. A historical 350-375 mm value was retained only as a superseded dimensional error. The system therefore refused to inherit the earlier 59.2 C cold-side result and reconstructed a bounded thermal comparison.

Five candidates across three families were reduced to three physical-test roles:

- `F-B1`: guarded double-pane interface coupon with the highest information value;
- `F-A1`: minimum-complexity control;
- `F-C0`: optional active-cooling dependency comparison.

The mission held because the authoritative ramp-dwell-cool profile, atmosphere, edge/frame path, seal behaviour, named material grades and cooling-loss response were unresolved. The result did not authorise fabrication or furnace operation.

# 9. Dual tank: correct the topology

Founder correction established that the reserve tank remained an isolated new-source supply, while filtered unused return entered the service tank and overflow remained contained. This changed the meaning of “freshness”: idle circulation could recondition fluid without becoming fresh replacement.

`W-B2` preserved the corrected topology. When an exploratory peak-demand screen exposed insufficient capacity, the system generated `W-B4` as a capacity-mapping descendant rather than silently resizing the accepted route.

The result remained `CORRECTION_COMPLETE_WITH_HOLD` because demand duty, freshness/quality mechanism, component ratings, maintenance needs and physical acceptance thresholds were not authoritative.

# 10. Heat shield: transfer to another domain

The third mission preserved the founder's porous transpiration and segmented fault-localisation mechanism, but bounded it to a benign-fluid ground coupon. Six candidates and two descendants were evaluated across 84 scenario runs and 80 parameter runs.

- `H-A1` delivered the strongest simple central thermal reduction but lacked fault isolation.
- `H-B1` sacrificed some cooling to isolate blocked/leaking cells and became the preferred information coupon.
- `H-C1` retained a responsive-material branch but remained deferred.

Reactive/cryogenic fluids, unsupported material response and coupon-to-flight promotion were rejected. The final decision was `DESIGN_3_COMPLETE_WITH_HOLD`.

# 11. Repeatability and non-regression

The evidence system uses typed contracts, deterministic fingerprints, file manifests and locked replays to distinguish a reproducible structural result from a one-off narrative.

For the heat-shield and final three-case unit:

| Suite | Result |
|---|---|
| Focused heat-shield tests | 5 passed |
| State-stack and GIL-EDK | 103 passed |
| MUOS | 99 passed |

The final heat-shield primary and replay runs produced the same structural fingerprint. Both comparison archives verified with zero manifest failures and portable archive paths. These tests support bounded software repeatability, not scientific or physical validation.

## 12. How authority is separated

Systemica separates five responsibilities that conventional technical workflows often blend:

1. domain inputs define what the problem means;
2. EDK and modelling engines propose routes;
3. GIL and DSTE measure stability and validate transitions;
4. MUOS retains hold/reject/promote authority;
5. HMA preserves only accepted checkpoints.

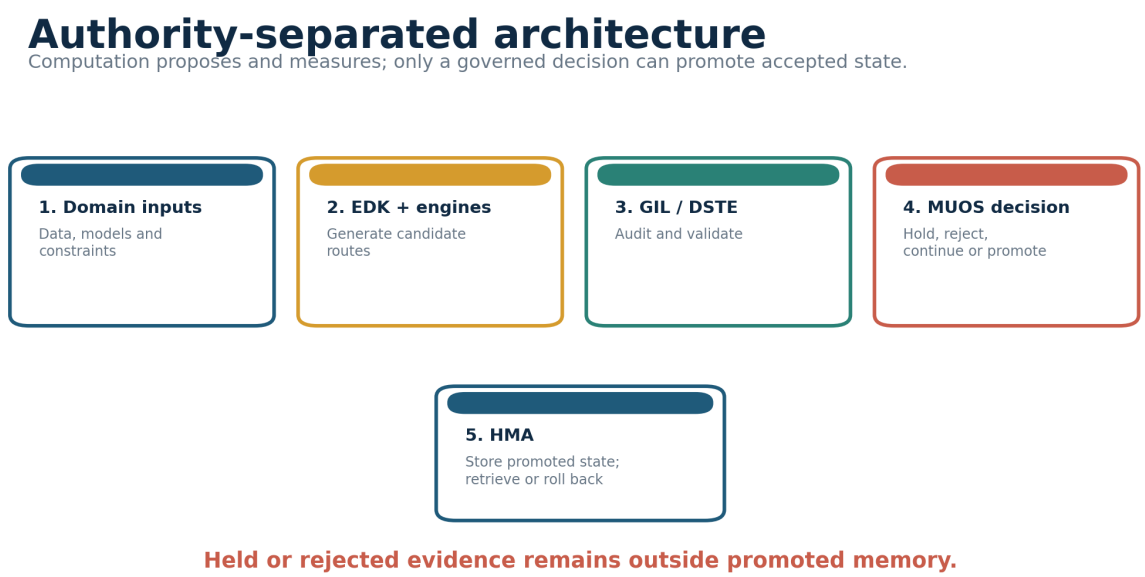


Figure 11: AV2-F03. Simplified authority architecture. Sources: EDK migration and DSTM-HMA bridge receipts. Interpretation: engines do not grade or promote themselves. Limitation: enabling interface detail is retained internally.

EDK is now an independently versioned `systemica_edk` kernel with a compatibility adapter. Parity was demonstrated in Cosmologia and EDS routes. That is a component-level migration result, not scientific authority over either domain.

# 13. Search, transitions and trajectory

EDK proposes candidate routes; DSTE validates declared state transitions. This prevents search logic from absorbing rejection authority.

The typed state stack covers states, constraints, basins, boundaries, topology, transition decisions and recovery routes. DSTM adds ordered pre/post pairs, accepted and rejected decisions, declared loss, irreversible markers and path fingerprints.

The important product effect is traceability: when a weakness generates a descendant, the old candidate remains visible and the manoeuvre records why the new branch exists. Recovery means returning to an accepted state or reconstructing it from evidence, not pretending failure never occurred.

## 14. Memory, interruption and recovery

HMA is controlled project memory, not unlimited conversational memory. A bounded local bridge has demonstrated operator-approved checkpoint commit, retrieval and rollback. Held, raw and non-MUOS-authorised objects were rejected before write.

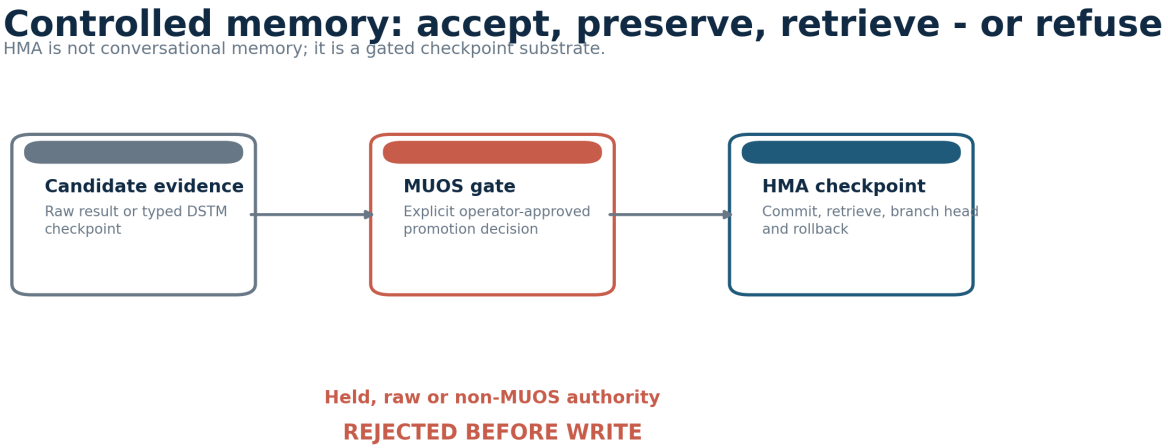


Figure 12: AV2-F07. HMA promotion and refusal route. Source: DSTM-HMA bridge report. Interpretation: accepted state can be retrieved or rolled back; held evidence remains outside promoted memory. Limitation: distributed durability, large-scale object counts and enterprise performance are not demonstrated.

Actual large-scale memory-object counts, database size and retrieval latency were not captured in the accepted bridge receipt and remain `null`. Branching and crash-resume are supported by the architecture and test lineage, but user-facing recovery at product scale remains a roadmap gate.

# 15. What Joseph originated and what AI contributed

Joseph supplied the recurring trust problem, investigation method, Systemica ontology, module meanings, authority boundaries, domain mechanisms, claim ceilings and final technical/publication authority. The furnace geometry and dual-tank topology corrections were technically consequential: each invalidated an inherited computational route and forced new analysis, while neither automatically authorised promotion.

GPT/ChatGPT contributed critique, synthesis and alternative framing. Codex performed substantial software implementation, testing, execution support, documentation and evidence organisation. Deterministic engines performed declared calculations. This division is not a labour-percentage claim; it is an authority and provenance account.

The governing principle is **AI as worker and translator, not authority**. Generative assistance cannot rewrite measured evidence, approve its own claim ceiling or write accepted HMA memory. Final decision responsibility remains human.

# 16. Practical customer workflows

## Materials model before experiment selection

Audit whether the descriptor ranking survives model, perturbation and population changes before allocating experimental capacity.

## Simulation before fabrication

Expose boundary-condition dependence and convert broad design confidence into a smaller coupon comparison.

## Competing early concepts

Preserve candidate lineage, generate bounded descendants and reject routes that violate declared constraints.

## Interrupted technical project

Return to the latest accepted checkpoint, retain rejected evidence and branch without overwriting history.

## Sensitive local project

Run a bounded evidence workflow locally where cloud execution or broad data transfer is inappropriate.

These are product workflows and demonstrated internal patterns. They are not yet evidence of external customer deployment.

# 17. First market and competitive position

The first narrow customer hypothesis is materials and advanced-manufacturing R&D teams that use modelling to prioritise expensive simulation or experiment work. Adjacent early segments include engineering consultancies, technical startups and IP-sensitive local environments.

Existing tools solve important neighbouring problems. MLflow provides model evaluation, metrics, visualisations and extensible evaluation workflows [COMP-01]. SHAP explains model outputs through feature attribution [COMP-02]. Systemica's intended distinction is the combination of controlled stress testing, evidence lineage, explicit rejection/promotion authority, replay and preserved decision state.

This is not a claim of benchmark superiority. Customer accessibility, sales cycle, willingness to pay and decision cost remain discovery questions.

# 18. Business model and adoption ladder

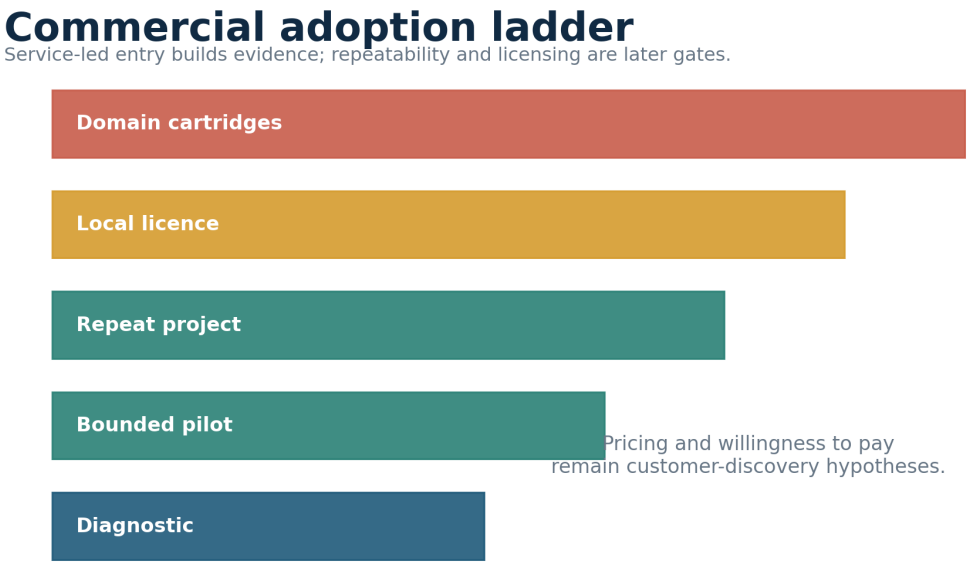


Figure 13: AV2-F10. Commercial adoption ladder. Source: founder product mandate. Interpretation: bounded pilots are an entry route toward repeatable local software. Limitation: pricing and customer demand remain hypotheses.

1. light diagnostic;
2. paid bounded pilot;
3. repeat project or deployment;
4. local/enterprise licence;
5. domain-specific cartridges and integration.

Service-led delivery creates the first evidence and onboarding knowledge. It is not intended to remain the final non-scalable identity. No revenue forecast or claimed cost reduction is included.

# 19. Current readiness and next gates

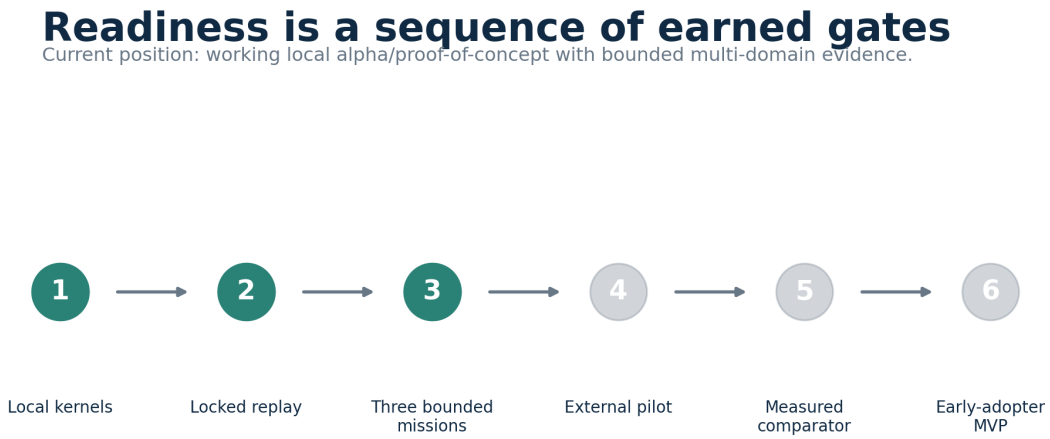


Figure 14: AV2-F09. Readiness ladder. Sources: accepted local receipts and founder roadmap. Interpretation: local kernels and bounded missions have passed; external pilot and comparator evidence remain ahead. Limitation: no forecast dates or TRL promotion are implied.

## Operational now

- independent EDK kernel with two-module parity;
- GIL-EDK Materials Project audit and scale route;
- typed state/transition stack and replay;
- bounded DSTM-HMA promotion/retrieval/rollback bridge;
- three Design-3 mission evidence bundles.

## Still missing

- independent-source scientific replication;
- external bounded customer pilot;
- measured comparison with an ordinary workflow;
- unified local early-adopter interface and onboarding;
- formal IP/deployment review;
- wider user-side MUOS composition.

## 20. Origin and founder position

Systemica is recorded as `FOUNDER_DECLARED_INDEPENDENT_ORIGIN`: a founder declaration, not an unqualified legal opinion. No third-party institutional ownership or endorsement is claimed.

Joseph Maxwell's route combines materials engineering, aerospace-facing assessment experience, systems thinking and a practical need for recoverable technical cognition. The product direction emerged from repeatedly encountering the same problem: technical work can generate plausible outputs faster than it can establish what those outputs are safe to support.

The founder developed the engineering ontology, authority architecture and acceptance logic, then used GPT and Codex as implementation and reasoning tools within those constraints.

# 21. Pilot and partner invitation

The most useful next engagement is a bounded engineering decision where:

- existing modelling already influences an expensive next action;
- the data and current route can be declared clearly;
- the team can define what failure or fragility would matter;
- local/confidential execution has practical value;
- an explicit HOLD is acceptable if the evidence is insufficient.

The immediate asks are customer-discovery conversations, one independent-source scientific replication partner, and one bounded external pilot with a measurable conventional-workflow comparator.

No external publication, deployment or pilot commitment is created by this Atlas.

# 22. Claim boundary and AI disclosure

## Demonstrated

- working local software kernels and test suites;
- repeatable Materials Project audit routes;
- bounded authority-separated checkpoint storage and rollback;
- three cross-domain Design-3 missions ending in honest HOLD states;
- source-grounded manifests, replay and evidence bundles.

## Not demonstrated

- external or physical validation;
- safety, certification or production qualification;
- autonomous engineering;
- full Systemica/MUOS product readiness;
- quantified labour, cost or fabrication savings;
- customer willingness to pay or repeat deployment.

## AI contribution disclosure

GPT/ChatGPT and Codex were used for technical critique, implementation, testing, evidence organisation and communication support. Deterministic engines produced model and replay outputs. The founder supplied the ontology, constraints, domain mechanisms, corrections, claim ceilings and final approval responsibility. AI output is not treated as inherently authoritative.

## 23. How to read an evidence label

Every public figure and engineering statement is assigned an evidence class:

| Label | Meaning | Example |
|---|---|---|
| `MEASURED` | Direct machine-readable output or test result | R-squared, test count, archive size |
| `REPLAYED` | Structural result reproduced under locked inputs | model fingerprint or bundle parity |
| `FOUNDER DECLARATION` | Founder-supplied origin, intent or correction | independent-origin classification |
| `INTERPRETATION` | Bounded explanation of measured evidence | why a fragile descriptor should not be promoted |
| `ROADMAP` | Proposed future function or gate | external pilot and comparator workflow |

A lower evidence class is not automatically a defect. The safety rule is that interpretation and roadmap language cannot silently overwrite measured or replayed evidence.

# 24. What Systemica is not

Systemica is not:

- an AI wrapper that delegates core authority to a language model;
- a generic chatbot or general-purpose conversational assistant;
- a replacement for domain engineering judgement;
- a physics theory or universal failure law;
- a certification, safety or regulatory authority;
- a guarantee that a model or candidate is correct;
- a cloud-scale MLOps platform;
- an autonomous design-and-fabrication system.

It is an architecture for running bounded technical investigations while preserving the separation between computation, evidence, authority and memory.

That boundary matters commercially: the product is designed to improve the quality and traceability of the next decision, not to absorb responsibility for decisions it has not earned.

## 25. What a bounded pilot pack contains

The proposed customer-facing evidence pack is deliberately inspectable:

1. resolved problem and data contract;
2. source and schema receipt;
3. candidate/model register;
4. perturbation and control results;
5. stability and weakness figures;
6. rejected/held route register;
7. replay fingerprint and file manifest;
8. claim boundary;
9. human decision packet;
10. smallest recommended next validation action.

The pack preserves negative evidence rather than presenting only the strongest candidate. A customer can therefore see not only what the system recommends, but what failed, what changed and why the final state is `CONTINUE`, `REVISE` or `HOLD`.

## 26. Active risk register

| Programme risk | Current effect | Control / next gate |
|---|---|---|
| No external customer validation | Product value remains internally evidenced | Run one bounded external pilot |
| No measured conventional-workflow comparator | No quantified time/cost claim | Instrument both routes prospectively |
| Unified MUOS journey incomplete | Operator experience remains fragmented | Build local early-adopter workflow |
| Materials evidence uses one primary source | No independent scientific replication | Route a second compatible source |
| CTAP/drift evidence uses controlled fixtures and zero live LLM runs | Live detection performance and threshold calibration remain unproven | Run a repeated, independently labelled live-LLM drift and recovery pilot |
| HMA scale telemetry incomplete | Memory claims remain bounded | Benchmark object count, size and latency |
| No external legal/IP opinion | Founder-authorised release is not an ownership or freedom-to-operate opinion | Complete specialist review before deeper disclosure, licensing or contracting |

The present programme risk is not lack of ideas. It is converting strong local evidence into independent, customer-facing and measured adoption evidence without weakening claim discipline.

# 27. Public sources

- **MP-01** - Materials Project documentation: https://docs.materialsproject.org/
- **MP-02** - Official API access and `mp-api`: https://docs.materialsproject.org/downloading-data/using-the-api/getting-started
- **MP-03** - Materials Project citation and calculated-property guidance: https://docs.materialsproject.org/frequently-asked-questions
- **COMP-01** - MLflow model evaluation documentation: https://mlflow.org/docs/latest/ml/evaluation
- **COMP-02** - SHAP documentation: https://shap.readthedocs.io/en/latest/
- **ML-01** - scikit-learn user guide: https://scikit-learn.org/stable/user_guide.html
- **DB-01** - SQLite database file format: https://sqlite.org/fileformat2.html
- **AI-01** - Ollama API documentation: https://docs.ollama.com/api/introduction
- **NIST-01** - NIST AI RMF Core: https://airc.nist.gov/airmf-resources/airmf/5-sec-core/
- **NIST-02** - NIST AI 600-1 Generative AI Profile: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
- **ISO-01** - ISO/IEC 42001 overview: https://www.iso.org/standard/42001
- **EU-01** - EU Artificial Intelligence Act: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

Version: `2.1.2 founder-authorised public release`
State: `SYSTEMICA_ATLAS_V2_1_2_FOUNDER_AUTHORISED_PUBLIC_RELEASE_ACCEPTED`
Founder-authorised public distribution: `AUTHORISED`

External legal/IP opinion: `NOT OBTAINED`