PRESS + SOVEREIGN TECHNOLOGY STRATEGY | HIGH-TRUST TECHNICAL BINDER

AGI Alpha RSI

Move-37 Breakthrough Dossier (Protocol Binder Template)

A decision-grade binder to recognize, reproduce, stress-test, and brief breakthrough innovations with deterministic evidence.

Document status	FINAL TEMPLATE (institutional, press + sovereign)
Prepared for	Global technology press; sovereign technology leadership; national innovation agencies
System baseline	Runner + Prompt Pack: rr_omni_v7 (Move-37 breakthrough protocol + EIG-scheduled probing)
Confidentiality	Controlled distribution; remove operational deployment details for public release
Document owner	[FILL: organization / program office]

Release date: 2026-01-23 (UTC)

1. Purpose and scope

This binder is the standardized, decision-grade packaging for a Move-37-class breakthrough detected by AGI Alpha RSI. It is designed to be:

- Auditable: every claim is backed by evidence objects, hashes, and deterministic manifests.
- Reproducible: reruns with the same state + seed recreate the result bit-for-bit (or flag nondeterminism).
- Actionable: includes clear go/no-go gates, baseline comparisons, and next-step options.
- Sovereign-ready: structured for high-trust governance, escalation, and strategic accountability.

Use this template whenever the Breakthrough Protocol triggers (see Section 3). For non-breakthrough candidates, use the standard cycle dossier outputs.

2. How to use this binder

Attach the dossier bundle emitted by the runner (rr_omni_v7): the /dossier folder plus the referenced artifacts in run_outputs.zip. Then complete each section below. Sections marked 'MANDATORY' are required for any external briefing.

1. Fill Section 4 (Identity & provenance) directly from run_manifest.json and candidate card JSON.
2. Verify Section 5 (Deterministic novelty distance) using candidates/novelty_distance.jsonl.
3. Verify Section 6 (Baseline advantage) using eval/baseline_comparison.jsonl and eval/eval_results.jsonl.
4. Execute/confirm Section 7 (Advantage persistence) including policy shocks and multi-seed replay results.
5. Complete Section 8 (Stress tests) and Section 9 (Risk & side-effects).
6. Finalize Section 10 (Decision recommendation) and produce a distribution-ready executive brief.

3. Breakthrough trigger definition (Move-37 Protocol)

The Breakthrough Protocol triggers when a candidate demonstrates both (i) high novelty distance and (ii) objective advantage versus a valid baseline, with sufficient credibility and risk controls. The default rr_omni_v7 trigger thresholds are:

Novelty distance (min)	>= 0.9 (see Section 5)
Advantage delta vs baseline (min)	>= 0.15 on metric grade.overall_score (see Section 6)
Risk (max)	<= constraints.max_risk_score (configured); hard gate
Confidence (min)	>= 0.55 (from grading / judge audit)
ECI (min)	>= 0.55 (evidence-backed credibility)

In addition, if novelty_distance >= 0.8, the protocol requires a mandatory unknown to be resolved:

Mandatory unknown id	ADVANTAGE_PERSISTENCE
Applies when novelty_distance >=	0.8
Intent	Ensure the advantage persists under shocks and replays; prevent one-off artifacts.

3.1 Breakthrough protocol checklist (MANDATORY)

- [ ] Novelty distance computed deterministically and recorded.
- [ ] Baseline selected according to policy (incumbent in cell else nearest neighbor).
- [ ] Advantage delta meets breakthrough threshold on the configured metric.
- [ ] At least one executed (deterministic) evidence object supports the advantage claim.
- [ ] ADVANTAGE_PERSISTENCE unknown evaluated via shocks/replays (Section 7).
- [ ] Full dossier bundle assembled with artifact index and hashes.

4. Breakthrough identity and provenance (MANDATORY)

Populate this section directly from run_manifest.json, candidate card JSON, and archive metadata.

Breakthrough name	[FILL: short codename]
Candidate_id	[FILL: candidate_id]
Descriptor cell_key	[FILL: cell_key]
Run_id	[FILL: run_id]
Cycle_index	[FILL: cycle_index]
Lane	[FILL: LHF or PIONEER]
Focus domain	[FILL: focus_domain]
Date/time (UTC)	[FILL: timestamp]
Evidence level	[FILL: EXECUTED / SIMULATED / MIXED]

4.1 One-paragraph statement

[FILL: In one paragraph, state what was discovered, why it matters, and the measurable advantage observed. Avoid speculation; cite evidence object ids.]

4.2 Provenance and determinism

Runner version	rr_omni_v7 (or later)
Random seed policy	[FILL: seeds used; include multi-seed replay seeds]
Inputs digest	[FILL: sha256 inputs digest from run_manifest]
Output hash	[FILL: sha256 of candidate card and key artifacts]
Nondeterminism flags	[FILL: true/false; if true, describe source and mitigation]

5. Deterministic novelty distance (MANDATORY)

Novelty distance is computed deterministically as 1 - max_similarity to the nearest neighbor in the archive/atlas neighbor pool. It is used to (i) maintain systematic pressure toward non-human novelty and (ii) trigger the Breakthrough Protocol.

5.1 Computation spec (rr_omni_v7 default)

Spec id	novelty_distance.v1
Composite similarity	composite_sim = 0.40*descriptor_sim + 0.30*triple_sim + 0.30*text_sim
Novelty distance	novelty_distance = clamp(1 - max(composite_sim), 0, 1)
High novelty threshold	0.8
Breakthrough candidate threshold	0.9

5.2 Neighbor pool and similarity components

Neighbor pool	frontier_cells=True; recent_candidates=True; max_neighbors=50
descriptor_sim	weight=0.4; Axis-wise match ratio over descriptor_cell axes (including optional axes when present).
triple_sim	weight=0.3; Jaccard similarity over normalized causal triples (subject|predicate|object) when available; else 0.
text_sim	weight=0.3; Jaccard similarity over token sets from title+thesis+wedge when available.

5.3 Recorded novelty distance (fill from candidates/novelty_distance.jsonl)

Novelty distance (computed)	[FILL: novelty_distance in [0,1]]
Nearest neighbor id	[FILL: neighbor_candidate_id or cell incumbent id]
Max composite similarity	[FILL: max_similarity]
Similarity breakdown	[FILL: descriptor_sim / triple_sim / text_sim values]
Interpretation	[FILL: what makes this non-human / structurally novel? Keep factual.]

6. Baseline advantage and objective confirmation (MANDATORY)

A Move-37-class claim requires an objective advantage versus a defined baseline, measured on a deterministic metric. Baselines are selected mechanically to avoid narrative bias.

6.1 Baseline selection policy (rr_omni_v7 default)

Enabled	True
Baseline selector	incumbent_elite_in_cell_else_nearest_neighbor
Fallback	nearest_neighbor
Advantage metric	grade.overall_score
Replace threshold (delta)	0.05
Breakthrough threshold (delta)	0.15
Decisions requiring baseline	INSERT, REPLACE, ESCALATE

6.2 Baseline comparison summary (fill from eval/baseline_comparison.jsonl)

Baseline id	Baseline description	Candidate score	Baseline score	Delta	Confidence
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]

Attach: eval/eval_results.jsonl and any executed microbench artifacts supporting the advantage delta.

7. Mandatory unknown: ADVANTAGE_PERSISTENCE (MANDATORY when high novelty)

If novelty_distance >= 0.8, the system requires explicit validation that the observed advantage persists across policy shocks and replay seeds. This prevents one-off artifacts and supports sovereign-grade underwriting.

7.1 Gate definition (rr_omni_v7 default)

Metric	grade.overall_score
Min positive delta	0.05
Min pass rate	0.67
Min shocks/replays	3
Notes	Candidate must maintain positive advantage delta vs baseline across >=min_shocks policy shocks / replays.

7.2 Extra evaluation plan executed (fill from dossier bundle)

Policy shocks	enabled=True; count=3; seeds=[11, 29, 47]
Multi-seed replay	enabled=True; seeds=[101, 202, 303]
Probe budget multiplier	2.0
Require executed evidence	True

7.3 Advantage persistence results

Shock/replay id	Seed	Candidate score	Baseline score	Delta	Pass/Fail	Evidence object id
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]

Decision: [FILL: PASS if deltas remain positive across required shocks/replays; else FAIL or INCONCLUSIVE.]

8. Stress tests, adversarial probes, and robustness (MANDATORY)

Stress tests are designed to expose failure modes, brittleness, distribution shift sensitivity, and hidden dependencies. Prefer deterministic probes and policy shocks. Use this section to document outcomes and residual uncertainty.

8.1 Policy shock suite (fill from P33 outputs and evidence objects)

List the shock scenarios applied (e.g., constraint changes, adversarial inputs, resource limits) and their measured effect on advantage.

Shock id	Scenario description	Delta vs baseline	Outcome	Notes / mitigations
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]

8.2 Adversarial probe schedule (from probe/probe_schedule.jsonl)

Attach the executed probe schedule and summarize the highest expected information gain (EIG) probes and outcomes.

[FILL: Summary of top EIG probes, unknown_ids targeted, and what was learned.]

9. Risk, side-effects, and governance (MANDATORY)

A sovereign-grade dossier must surface second-order effects, externalities, and misuse pathways. Document both mitigations and what remains unknown.

9.1 Risk gate summary

Risk score	[FILL: numeric risk_score]
Risk gate decision	[FILL: ALLOW / BLOCK]
Primary risk factors	[FILL: brief list]
Mitigations in place	[FILL: controls / constraints / monitoring]
Residual risks	[FILL: what is not yet mitigated]

9.2 Side-effects and externalities

[FILL: Document possible negative externalities (economic, security, safety, environmental, geopolitical). Include any trade-offs revealed by stress tests.]

9.3 Compliance and dissemination controls

Classification / handling: [FILL]. Export control considerations: [FILL]. Data residency constraints: [FILL].

10. Mechanism and causal atlas extract

This section provides the mechanistic account (as far as is currently interpretable) and the causal atlas integration artifacts. For Move-37-class outcomes, mechanistic interpretability may be partial; the dossier must still be evidence-first.

10.1 Candidate mechanism summary

[FILL: Explain the mechanism in operational terms (inputs -> transformations -> outputs). If mechanistic understanding is incomplete, explicitly list unknowns and how they are being probed.]

10.2 Causal triples and bridge motifs (from atlas/*.jsonl)

Subject	Predicate	Object	Notes / confidence
[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]

10.3 Contradiction and side-effect checks

[FILL: Summarize any contradictions flagged and how they were resolved. Summarize side-effect checks and their status.]

11. Evidence ledger and audit trail (MANDATORY)

This ledger is the canonical 'currency of proof' record. Populate directly from evidence/evidence_objects.jsonl and eci/eci_ledger.jsonl. Every entry must be traceable to artifacts and hashes.

11.1 Evidence objects

Timestamp	Level	Test id / artifact	Outcome	ECI delta	Artifact hash / relpath
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]
[FILL]	[FILL]	[FILL]	[FILL]	[FILL]	[FILL]

11.2 Deterministic audit trail

Attach: run_manifest.json, eval/eval_manifest.jsonl, and logs/stage_log.jsonl. Provide the following summary:

Inputs digest (sha256)	[FILL]
Outputs digest (sha256)	[FILL]
Schema registry digest	[FILL]
Replay instructions	[FILL: exact command / environment spec]
Independent verification	[FILL: who verified, when, and outcome]

12. Decision recommendation (sovereign briefing)

This section converts the breakthrough into decision-grade options. Keep the recommendation grounded in measured advantage, risk posture, and evidence maturity.

12.1 Recommended disposition

Disposition	[FILL: INSERT / REPLACE / ESCALATE / PROMOTE / HOLD]
Rationale (evidence-backed)	[FILL: cite evidence object ids and key metrics]
Required next actions	[FILL: pilots, additional probes, governance review, procurement]
Budget / resources	[FILL: estimate and time horizon]
Decision owner	[FILL]
Decision date	[FILL]

12.2 Strategic option value

[FILL: Describe national / sovereign relevance, second-order impacts, and bridge option value in concrete terms. Avoid hype; focus on scenario-based advantage.]

12.3 External communications posture (press)

[FILL: If release is planned, specify what can be said publicly, what must remain controlled, and what proof artifacts can be shared.]

Appendix A. Dossier bundle index (rr_omni_v7 default)

The runner emits a dossier bundle designed for independent verification. Include the following relpaths (as available):

- run_manifest.json

- probe/probe_schedule.jsonl

- candidates/novelty_distance.jsonl

- eval/baseline_comparison.jsonl

- eval/eval_results.jsonl

- evidence/evidence_objects.jsonl

- eci/eci_ledger.jsonl

- reports/helm_like_summary.md

- eval/eval_manifest.jsonl

Also include: dossier/index.json (if enabled) and any additional executed test artifacts.

Appendix B. Glossary

Term	Definition
ECI	Evidence Contact Index. Credibility in [0,1] updated only via evidence events; simulated evidence is capped.
EIG	Expected Information Gain. Deterministic scheduler for selecting PROBE tests that resolve high-value unknowns cheaply.
MAP-Elites	Quality-diversity archive algorithm storing the best-known solution per descriptor cell.
Novelty distance	Deterministic distance from nearest neighbor in archive/atlas; used to drive non-human exploration.
Move-37	A shorthand for a breakthrough that is both novel and objectively advantageous, even if initially non-intuitive.

End of template.