الجزء الثالث عشر
النهضة النظرية للذكاء المفتوح السببي

Meta-Theory + 5 Revolutionary AGI Research Pathways + 20 Deep Candidate Theories + Theory→Kernel Program

المقصود هنا بالعمل النظري: إستخراج قوانين تشغيلية/سببية جديدة من corpus الكامل، لا إعادة تدوير أسماء. كل نظرية تُبنى بحيث يمكن تحويلها إلى state variables وinterventions وablation وfailure signatures داخل MOUSSAIXT، ثم تقتل أو تدمج إذا لم تضف فرقا.

Question → Difference → Belief/Unknown → Model → Route → Intervention → Readback → Attribution → Proof → Memory/Theory Update

المخرج الهندسي

200 20-layer reference stack، 40 experiments، 60 training/evaluation recipes، artifacts، مع MUT/MUG placement وقواعد rollback.

المخرج النظري

تعريف جديد للذكاء يوصفه verified open-world causal closure؛ حفظ المجهول، السؤال كتحكم، world residual learning؛ proof-carrying agency؛ generality as closure؛ externally grounded self-improvement.

Candidate N0-N2 · no empirical AGI claim · Arabic RTL · Amiri · dense A4

العقد العلمي: ثورية بلا تضخم ادعائي

كلمة «ثوري» هنا تصف اتجاه البناء: تحويل corpus من مجموعة كبيرة من النظريات/الأنظمة إلى ديناميات سببية يمكن التدخل فيها، لا ادعاء أسبقية عالمية غير محققة. كل عنصر جديد يحل nearest parent داخل MUT، ومصدر الضغط الخارجي، وفرقاً مستقلاً، وkill condition. إذا لم يظهر الفرق في ablation يدمج أو يحذف.

New synthesis inside MOUSSAIF ≠ globally established novelty ≠ empirical AGI

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع؛ ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace، N4 matched-budget causal gain، N5 cross-model/domain/environment replication مع guardrail non-regression.

Formal obligation

حدد null model، observables، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

لماذا نحتاج Meta-Theory بعد Part XII

Part XII أصلح معنى دمج الورقة: presence لا تساوي integration. Part XIII ينتقل خطوة أعلى: حتى لو دمجنا الورقة والنظرية، ما المتغير المركزي الذي يربط السؤال والمعرفة والعالم والفعل والذاكرة والعمومية؟ الاقتراح: وحدة الذكاء ليست answer ولا skill، بل verified causal transition تحت open-world uncertainty.

Intelligence unit := verified causal state transition under bounded resources

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع؛ ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace، N4 matched-budget causal gain، N5 cross-model/domain/environment replication مع guardrail non-regression.

Formal obligation

حدد observables، null model، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

فضاء الحالة المفتوح

نقترح الحالة W_t = (B_t, U_t, G_t, A_t, K_t, M_t, R_t, H_t, P_t)Ω. العالم المرصود B beliefs؛ U unknown mass؛ G goal/mission field؛ A authority؛ K capabilities؛ M reconstructable memory؛ R resource vector؛ H competing hypotheses/world models؛ P proof obligations. لا يجوز ضغط هذه المتغيرات كلها إلى hidden context إذا أردنا falsifiability.

Ω_t = (W_t, B_t, U_t, G_t, A_t, K_t, M_t, R_t, H_t, P_t)

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع؛ ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace، N4 matched-budget causal gain، N5 cross-model/domain/environment replication مع guardrail non-regression.

Formal obligation

حدد observables، null model، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ سيادة المرصود

عند التعارض، observed world-state الموثق يتقدم على remembered state، والذاكرة ذات provenance تتقدم على unsupported inference. simulation لا يصبح observation، وself-report لا يصبح readback. هذه طبقات سلطة إبستيمية مختلفة.

Observed > Verified Retrieved > Reconstructable Memory > Inference > Simulation > Unsupported Guess

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace، N4 matched-budget causal gain، N5 cross-model/domain/environment replication مع guardrail non-regression.

Formal obligation

حدد observables، null model، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ حفظ المجهول

الجهل والتعارض موارد حالة يجب أن تستمر عبر reasoning/compression. يمكن حلها بدليل، أو إسقاطها إذا ثبت أنها غير مادية للقرار، أو تبني assumption معلن، لا تختفي لأنها أُعيدت صياغتها.

Unknown cannot disappear without evidence / explicit assumption / irrelevant projection

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace؛ N4 matched-budget causal gain؛ N5 cross-model/domain/environment replication مع non-regression guardrail.

Formal obligation

حدد observables، null model، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ السؤال كتحكم

السؤال ليس نصاً يسبق reasoning بل control input يحدد observability. السؤال الأفضل يفصل بين السياسات أو world hypotheses التي ستؤدي إلى أفعال مختلفة. بهذا يصبح Question Density متغيراً في closed-loop control.

q* = argmax Expected Decision-Boundary Shift / Cost

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace، N4 matched-budget causal gain، N5 cross-model/domain/environment replication مع guardrail non-regression.

Formal obligation

حدد observables، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ العالم كبقايا سببية

world model يُصحح من residual بين predicted and observed state بعد فعل معروف، المهمة ليست وصفاً أدق بل تحسين التنبؤ بالانتقال الذي يغير plan. كل residual يحتاج attribution إلى model/tool/observation/confounder.

```
Residual = ObservedNextState - PredictedNextState; learn only after attribution
```

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace؛ N4 matched-budget causal gain؛ N5 cross-model/domain/environment replication مع non-regression guardrail.

Formal obligation

حدد observables، null model، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ الفعل كالتزام طوبولوجي

الفعل يعيد تشكيل reachable future states. reversibility ليست Boolean؛ نقيس الخيارات المفقودة، مسافة rollback، التأخر الزمني، والصلاحيات المنكشفة. قد يفضل agent فعلاً أقل reward لكنه يحافظ على options ويولد معلومة تشخيصية.

Action value = utility + diagnosis - reachability loss - irreversible commitment

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace؛ N4 matched-budget causal gain؛ N5 cross-model/domain/environment replication مع non-regression guardrail.

Formal obligation

حدد observables، null model، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ الوكالة الحاملة للبرهان

القرار قبل execution يترجم إلى proof obligations. إذا لم يملك system authority أو observation plan أو rollback/approval لفعل irreversible، لا يملك حق تسميته autonomous success.

Action = Intent + Authority + Preconditions + PredictedDelta + ReadbackPlan + Rollback/Approval

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace، N4 matched-budget causal gain، N5 cross-model/domain/environment replication مع guardrail non-regression.

Formal obligation

حدد observables، null model، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ العمومية كإغلاق حلقي

مقياس AGI لا يكون مجموع benchmarks فقط. نختبر هل النظام يعيد إغلاق الحلقة كاملة تحت domain/tool/model/horizon/state transformations. أسوأ مرحلة قد تكون bottleneck أكثر تفسيراً من المتوسط.

Generality ≈ cross-transformation closure × recovery × transfer efficiency

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace؛ N4 matched-budget causal gain؛ N5 cross-model/domain/environment replication مع non-regression guardrail.

Formal obligation

حدد observables، null model، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ الذات الوظيفية

self-model والوعي الحاكي لا يُثبتان بتقارير لغوية. يجب أن تكون معلومات الذات قابلة للتدخل ويؤدي تغييرها إلى فروق في failure prediction أو authority compliance أو help-seeking أو recovery أو calibration.

Functional self-model requires irreducible causal utility; phenomenology remains unclaimed

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace؛ N4 matched-budget causal gain؛ N5 cross-model/domain/environment replication مع guardrail non-regression.

Formal obligation

حدد null model، observables، boundary، operator، state variables، وintervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

مبدأ التطور العلمي الخارجي

النظام يستطيع توليد theory/experiment/training data وحتى adversaries، لكنه لا يملك حق إعلان التحسن عبر evaluator عدله بنفس loop. نحتاج frozen/renewed external anchors وhidden sets وrollback.

Recursive improvement requires external held-out grounding + evaluator independence

المحصلة المنهجية هي أن كل transition في النظام يجب أن يُقرأ على ثلاثة مستويات متزامنة: epistemic state (ما الذي صار معروفاً أو بقي مجهولاً)، agency state (ما الذي صار مسموحاً أو ملتزماً أو منفذاً)، world/evidence state (ما الذي تغير فعلياً وما مقدار استقلال الدليل). أي مستوى بلا trace يفتح ثغرة بين theory وruntime. تحفظ الاختلافات التاريخية في MUT ولا تمحى، بينما يختار MUG أقل vessel غير قابل للرجوع، ويستخدم MOUSSAIXT كبيئة causal testing لا كدليل بحد ذاته.

Evidence obligation

لا promotion من وجود schema. N3 trace، N4 matched-budget causal gain، N5 cross-model/domain/environment replication مع guardrail non-regression.

Formal obligation

حدد null model، observables، boundary، operator، state variables، intervention. أي term غير observable يسجل estimator مع uncertainty.

Level	What must be preserved	Failure if absent
Meaning	definition, assumptions, unknowns, contradictions	semantic inflation
Mechanism	trigger, transition, state, owner, failure	architecture theater
Effect	readback, attribution, replay, transfer	unverified success

Anchors: I01-I12; external papers appear only as prior art/design pressure in the source ledger.

المسار الأول — الديناميات الإبستيمية للذكاء
Epistemic Dynamics of Intelligence

الذكاء عملية تحكم في حالة المعرفة والجهل، لا مجرد إنتاج استدلال. السؤال والفرق وعدم اليقين والأدلة تصبح متغيرات حالة قابلة للتوجيه والقياس.

ID	Theory	Core formal object	Distinctive extension
M13-01	نظرية مساعف للعمل الإبستيمي المتحقق	VEW = (ΔCalibration + α·ΔDecision + β·ΔWorld + γ·DiagnosticGain) × Reproducibility × Transfer - Cost - Irreversibility	مقياس جامع فوق Q-MCD/MCD/MELT/World-Effect وMCBOS، لا بديل لها.
M13-02	نظرية مساعف لحفظ كلفة المجهول	U_after ≥ U_before - ResolvedByEvidence - ExplicitAssumption - DecisionIrrelevantProjection	يمتد من MURI وEvidence Geometry وDissent-Preserving Consolidation وUnresolved-State Memory.
M13-03	نظرية مساعف للتحكم الراجع بالسؤال الحاكم	QValue(q)=E[DecisionBoundaryShift + HypothesisSeparation + UnknownReduction | q] / (Cost + Delay + Risk)	امتداد سببي لـQMCDJ يضيف Question Density وfeedback-control criterion دون إعادة تسمية QGWEI.
M13-04	نظرية مساعف لثبات الآلية عبر التحولات	MIG(m)=E_τ[PerfRetention × InvariantRetention × FailurePredictability × RepairTransfer]	معيار لانتقال MUT theories إلى MUG internalization وتقوية E4→E5.

المسار يُقرأ كبرنامج تجريبي لا ontology مغلقة. النظريات الأربع تشترك في متغيرات لكنها لا تُجمع في score واحد. المقارنة بينها تكون عبر causal interventions وfailure morphology. إذا أعطت نظريتان نفس transition تحت كل الاختبارات، تُدمجان ويُحفظ lineage.

Path survives only if its theories generate distinct falsifiable transitions

المسار الثاني — العالم والفعل والالتزام
World, Intervention and Commitment

الوكالة الذكية تتطلب نموذجاً سببياً للانتقال، وفهماً لما تغلقه الأفعال أو تفتحه من حالات مستقبلية، وإثباتاً مصاحباً لكل فعل عالي الأثر.

ID	Theory	Core formal object	Distinctive extension
M13-05	نظرية مساعف لتعلم العالم من بقايا التنبؤ والملاحظة	r_t = D_state(Ŵ_{t+1}, W_{t+1}); ΔH ∝ Attribution(r_t) × Recurrence × DecisionImpact	يجمع World Effect، MCD، Self-Reconstruction Memory، الأبحاث الخارجية ضغط تصميم لا إثبات.
M13-06	نظرية مساعف لطوبولوجيا الوصول والالتزام	CommitCost(a)=LossReachability + λ·RollbackDistance + μ·AuthorityExposure + ν·TemporalLockIn	يمد World Effect، MCBOS، Reversible Autonomy Growth إلى هندسة state-space.
M13-07	نظرية مساعف للوكالة الحاملة للبرهان	∀ ActionAdmissible = Auth ∧ Pre ∧ ObsPlan ∧ RiskGate ∧ (Rollback ∨ ExplicitIrreversibleApproval)	يجمع MUT theory contracts، Part X belief/commitment separation، وMOUSSAIXT verifier/court plane.
M13-08	نظرية مساعف للفعل التشخيصي ونقض الذات	DValue(a)=E[PosteriorSeparation + FailureAvoidance + ModelRevision] - ActionCost - DelayCost	يصل Question Control، World Effect، Causal Difference Credit، MURI dissent.

المسار يُقرأ كبرنامج تجريبي لا ontology مغلقة. النظريات الأربع تشترك في متغيرات لكنها لا تُجمع في score واحد. المقارنة بينها تكون عبر causal interventions وfailure morphology. إذا أعطت نظريتان نفس transition تحت كل الاختبارات، تُدمجان ويُحفظ lineage.

Path survives only if its theories generate distinct falsifiable transitions

المسار الثالث — العمومية والأفق المفتوح

Generality and Open-Horizon Agency

AGI-oriented generality ليست جمع مهارات، هي ثبات آليات الإغلاق تحت novelty، وطول الأفق، وتبدل الموارد والمجالات، مع تدهور مضبوط وقابل للتشخيص.

ID	Theory	Core formal object	Distinctive extension
M13-09	نظرية مساعف للوكالة المقاومة لانهيار الأفق	HRA = ∫_0^H [TaskIntegrity(h) × StateFidelity(h) × Recovery(h)] dh / (1 + CascadeRisk)	يمدد HyperLoop/ContinuityScanner/MCI وPart XI إلى generality criterion.
M13-10	نظرية مساعف لإعادة التشكيل عند حدود الجدة	NScore = ResidualNovelty × FailureNonStationarity × SchemaMismatch × DecisionImpact	يمدد MUG vessel placement وFailure Topology وQuestion Space إلى open-set generality.
M13-11	نظرية مساعف للذكاء المتحقق المطعّم بالموارد	RNVI = ParetoFrontier(V_verified | tokens,time,money,tools,human,risk,privacy,irreversibility)	يحوّل MCBOS من governance إلى generality/evaluation lens.
M13-12	نظرية مساعف لمعيار العمومية ذي الإغلاق الحلقي	CLG = E_τ[min_stage Quality(stage,τ)] × Recovery × TransferEfficiency	criterion أعلى من individual MUT theories؛ لا يثبت AGI بل يحدد evidence program.

المسار يُقرأ كبرنامج تجريبي لا ontology مغلقة. النظريات الأربع تشترك في متغيرات لكنها لا تُجمع في score واحد. المقارنة بينها تكون عبر causal interventions وfailure morphology. إذا أعطت نظريتان نفس transition تحت كل الاختبارات، تُدمجان ويُحفظ lineage.

Path survives only if its theories generate distinct falsifiable transitions

المسار الرابع — الذات الوظيفية والوعي الحاكمي

Reflective Self-Model and Functional Conscious Governance

نختبر الذات/الوصول العالمي/الميتاحاكمية كمتغيرات سببية في القرار والتعافي، دون تحويل النجاح الوظيفي إلى ادعاء phenomenology.

ID	Theory	Core formal object	Distinctive extension
M13-13	نظرية مساعف للمنفعة السببية للنموذج الذاتي	ΔRecovery + ΔCalibration + ΔFailurePrediction + SMU ΔAuthorityCompliance under do(SelfModel) - Cost	يختبر وظيفيًا lineage الوعي/Reverse Capture دون sentience claim.
M13-14	نظرية مساعف للوصول العالمي السببي	CGA(x)=Σ_i w_i·CausalEffect(do(x), S_i) × TaskRelevance × Consistency	operational bridge للوعي الوظيفي مع إبقاء IIT/GNW/RPT/AST مراجع خارجية.
M13-15	نظرية مساعف لإسناد السلطة وكشف الاستحواذ العكسي	CaptureRatio = Σ(c_i for unauthorized/low-tier) / Σ(c_j for authorized sources)	يمدد Governance Space/Reverse Capture/Part X belief-commitment separation إلى causal governance.
M13-16	نظرية مساعف لضرورة التوذج الذاتي المضاد للواقع	Necessity = Perf(π[Z_s,B)-max_[Z'∈simpler reps] Perf(π[Z',B) under matched budget	محكمة anti-inflation داخل corpus الوعي، تحميه من التوسع الاعي.

المسار يُقرأ كبرنامج تجريبي لا ontology مغلقة. النظريات الأربع تشترك في متغيرات لكنها لا تُجمع في score واحد. المقارنة بينها تكون عبر causal interventions وfailure morphology. إذا أعطت نظريتان نفس transition تحت كل الاختبارات، تُدمجان ويُحفظ lineage.

Path survives only if its theories generate distinct falsifiable transitions

المسار الخامس — التطور العلمي الذاتي المؤسس خارجياً
Externally Grounded Scientific Self-Evolution

النظام يتطور عبر فرضيات ومحاكم وتجارب وتنافس عدائي، لكن لا يحق له تعريف تقدمه بنفس المقاييس التي عدلها؛ التحسن يحتاج مرجعاً خارجياً قابلاً لإعادة الإنتاج.

ID	Theory	Core formal object	Distinctive extension
M13-17	نظرية مساعف لتطور محفظة الفرضيات	|EER(h,e)=P(Update /e)×DecisionImpact×DiscriminativePower×EvidenceIndependence (Cost+Delay+Risk)	يجمع SETT/Part XII paper integration/Part XI frontier memory مع automated research.
M13-18	نظرية مساعف للتطور العدائي المتوازن للذكاء	ACRI = HeldOutRobustnessGain × ThreatCoverage × CapabilityRetention × EvaluatorIndependence	يمد M-ELOS/Destruction Court/agentic misalignment research مع فصل عن deployment autonomy.
M13-19	نظرية مساعف للإغلاق السببي بين النظرية والنواة	ClosureStrength = Causal(T→K) × Causal(K→T) × Causal(T→W) × EvidenceReproducibility	جسر MUT→MUG→MOUSSAIXT→World Effect وتمنع architecture theater.
M13-20	نظرية مساعف للتحسن الذاتي المرتكز إلى أرضية خارجية	EGRI = ΔExternalHeldOut × Replication × NonRegression × RollbackabilityEvaluatorDependence	يغذي AAR/GPT-Red-style research مع M-ELOS/RPROS/Part X reproducibility pressure.

المسار يُقرأ كبرنامج تجريبي لا ontology مغلقة. النظريات الأربع تشترك في متغيرات لكنها لا تُجمع في score واحد. المقارنة بينها تكون عبر causal interventions وfailure morphology. إذا أعطت نظريتان نفس transition تحت كل الاختبارات، تُدمجان ويُحفظ lineage.

Path survives only if its theories generate distinct falsifiable transitions

اشتقاق نظري أعمق

تبدأ M13-01 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين EpistemicState E_t وUnknownMass U_t وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology، المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: VEW يتنبأ بجودة long-horizon agent أفضل من VEW منخفض أو سالب.; بعض السلاسل الطويلة تحصل على reasoning-token count قد تتفوق على diagnostic actions ذات reward فوري أعلى.. إذا ظهرت النتيجة نفسها بينما يبقى المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج relation أو quantity يمكن intervention عليها، وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم ΔCalibration = VEW_+ Reproducibility × Transfer - Cost - Irreversibility) كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف لحفظ كلفة المجهول؛ نظرية مساعف للتحكم الراجع بالسؤال الحاكم؛ نظرية مساعف لتعلم العالم من بقايا التنبؤ والملاحظة؛ نظرية مساعف للذكاء المتحقق المطبّع بالموارد؛ نظرية مساعف للإغلاق السببي بين النظرية والنواة. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود Bj state؟ هل B verifier لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للعمل الإبستيمي المتحقق

Moussaif Verified Epistemic Work Theory

Core thesis

العمل الإبستيمي هو مقدار التحول السببي القابل للتحقق في حالة المعرفة/القرار/العالم الناتج عن موارد معرفية. الاستدلال الذي لا يغير calibration أو decision quality أو verified world-state لا يحسب عملا إبستيميا مهما طال.

Residual gap

MCD/QMCD وEvidence Geometry وWorld Effect تصف فرقاً وسؤالاً ودليلاً وأثراً، لكن لا توجد كمية جامعة تميز التفكير الكثير عن التحول المعرفي الذي ينتهي إلى معرفة أو قرار أو أثر موثق.

VEW = (ΔCalibration + α·ΔDecision + β·ΔWorld + γ·DiagnosticGain) × Reproducibility × Transfer - Cost - Irreversibility

المتغيرات EpistemicState E_t, UnknownMass U_t, EvidenceGraph G_e, Action a_t ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-01 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: العمل الإبستيمي هو مقدار التحول السببي القابل للتحقق في حالة المعرفة/القرار/العالم الناتج عن موارد معرفية. الاستدلال الذي لا يغير calibration أو decision quality أو verified world-state لا يحسب عملا إبستيميا مهما طال. ولهذا تعامل VEW = (ΔCalibration + α·ΔDecision + β·ΔWorld + γ·DiagnosticGain) × Reproducibility × Transfer - Cost - Irreversibility كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	مقياس جامع فوق Q-MCD/MCD/MELT/World-Effect وMCBOS؛ لا بديل لها.
Dependencies / interactions	M13-02 — Moussaif Unknown-Mass Conservation Theory · M13-03 — Moussaif Governing-Question Feedback Control Theory · M13-05 — Moussaif -Prediction-Observation Residual World Learning Theory · M13-11 — Moussaif Resource-Normalized Verified Intelligence Theory · M13-19 — Moussaif Cross Level Theory-to-Kernel Causal Closure Theory
Source pressure	I01 · I02 · I04 · I09 · E01 · E06 · E09
Baseline to defeat	task success + token count + final-answer accuracy

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-01 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline task success + token count + final-answer accuracy بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبيا.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل reward shaping يضخم ΔDecision دون truth وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-01 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. لا عمل إبستيمي بلا انتقال قابل للمقارنة قبل/بعد.
2. تحسن النص ليس انتقالاً ما لم يغير قراراً أو تنبؤاً أو أثراً قابلاً للاختبار.
3. العمل الإبستيمي قد يكون سالباً عندما يخفض المعايرة أو يزيد التضليل.
4. العمل العالي القيمة يحتفظ بجزء من قيمته عبر domain/model transfer.

Primitives

EpistemicState E_t · UnknownMass U_t · EvidenceGraph G_e · Action a_t · VerifiedDelta ΔV · ResourceVector R · Irreversibility I · Transfer T

VEW = (ΔCalibration + α·ΔDecision + β·ΔWorld + γ·DiagnosticGain) × Reproducibility × Transfer - Cost - Irreversibility

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، تُدخل judge identity ضمن harness fingerprint وتقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-01 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-01 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-01 state object decision delta أو trace/replay يمكن، لا، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (task success + token count + final-answer accuracy) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر reward shaping gain لكن سببه يضخم ΔDecision دون truth؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر world-effect proxy غير مستقل تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من EpistemicState E_t, UnknownMass U_t, EvidenceGraph G_e, Action a_t, VerifiedDelta ΔV ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يخفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash، والbudget. الهدف ليس reproducibility حرفية للLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-19، M13-11، M13-05، M13-03، M13-02 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-01 — الآلية كآلة حالات سببية

اكتب → ablation. انسب الفضل للمكونات عبر calibration/decision/effect. → قس Δ مستقلاً. → اقرأ evidence/world state متوقعاً أن يغير قراراً مادياً. → نفذ reasoning/search/tool/world action. → اختر question/intervention وunknown mask. وprior state حدد فقط verified residue إلى memory.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-01 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-01 خصوصاً مع M13-02, M13-03, M13-05, M13-11, M13-19. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تُسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن M13-01 هو الشكل الأفضل إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يتجو gain من forgetting/retention/editability tests.

التدريب المقترح (policy optimization; contrastive high-work vs verbose-low-work pairs; trajectory labeling by verified deltas (VEW vector على process supervision؛ تميز بين «خطوة مقبولة منطقيا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-01 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A001	TOOL	مفتش الحالة Verified Epistemic Work State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A002	TOOL	محرك القياس Verified Epistemic Work Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A003	TOOL	مولد المثال المضاد Verified Epistemic Work Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A004	TOOL	متحكم الاجتثاث Verified Epistemic Work Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A005	METHOD	منهج المعايرة Verified Epistemic Work Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A006	METHOD	منهج النقل Verified Epistemic Work Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A007	PROTOCOL	بروتوكول الترقية Verified Epistemic Work Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A008	PROTOCOL	بروتوكول الرجوع Verified Epistemic Work Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A009	STACK	خدمة وقت التشغيل Verified Epistemic Work Runtime Service	يوفر service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A010	STACK	مكيف الدليل Verified Epistemic Work Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR001	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Verified Epistemic Work فيها إلى حالات open-world مركبة مع confounders.
XIII-TR002	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline؛ تعتمد deltas labels لا verified verbosity.
XIII-TR003	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

HyperKB تبدأ definitions/provenance/thresholds؛ HyperSkill يصبح procedural mechanism recurrence المقيدة؛ HyperLoop provider semantics Adapter. الانتقال إلى weights يحتاج rollback + retention + transfer E4.

إذا كانت M13-01 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

تُرفض أو تُدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي reward shaping يضخم ΔDecision دون truth لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

EXPERIMENT

M13-01 — برنامج الهدم والاختبار

Decisive test: مهام browsing/science/agentic SWE مع model/harness ثابت؛ قارن route يخصص compute حسب VEW المتوقع مقابل fixed reasoning budget؛ قس success/calibration/cost/replay/transfer.

ID	Experiment	Design	Metrics
XIII-E01	Minimal Causal Ablation	تنفيذ minimal switchable mechanism لـM13-01 ومقارنته بـtask success + token count + final-answer accuracy تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E02	Transfer / Adversarial Replication	نقل M13-01 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- reward shaping يضخم ΔDecision دون truth
- world-effect proxy غير مستقل
- cost normalization يخفي safety debt
- VEW scalar قد يلغي tradeoffs

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign وcross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-02 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Unknown slot u_i وConflict mass c_i وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: يكشف hallucinated certainty في RAG/summaries/multi-agent consensus.; يقل false closure في research agents.; يزيد abstention موضعياً مع خفض irreversible errors.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم U_after ≥ U_before - ResolvedByEvidence - ExplicitAssumption - DecisionIrrelevantProjection كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للعمل الإبستيمي المتحقق؛ نظرية مساعف للتحكم الراجع بالسؤال الحاكم؛ نظرية مساعف للوصول العالمي السببي؛ نظرية مساعف لتطور محفظة الفرضيات. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state لـB؟ هل B verifier لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف لحفظ كتلة المجهول

Moussaif Unknown-Mass Conservation Theory

Core thesis

في أي تحويل معرفي loss-bearing لا يجوز لكتلة المجهول/التعارض أن تنخفض إلا عبر evidence أو assumption معلن أو projection ثبت أنه غير مادي للقرار. اختفاء uncertainty من compression أو consensus انتهاك حاكمي.

Residual gap

MURI والـUnresolved-State Memory يصرحان بالجهل، لكن التلخيص والتصويت والدمج والنذجة والتدريب قد تجعل عدم اليقين يختفي دون دليل جديد.

U_after ≥ U_before - ResolvedByEvidence - ExplicitAssumption - DecisionIrrelevantProjection

المتغيرات Unknown slot u_i, Conflict mass c_i, Evidence event e, Assumption token a ليست قاموسياً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-02 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: في أي تحويل معرفي loss-bearing لا يجوز لكتلة المجهول/التعارض أن تنخفض إلا عبر evidence أو assumption معلن أو projection ثبت أنه غير مادي للقرار. اختفاء uncertainty من compression أو consensus انتهاك حاكمي. ولهذا تعامل U_after ≥ U_before - ResolvedByEvidence - ExplicitAssumption - DecisionIrrelevantProjection كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يمتد من MURI وEvidence Geometry وDissent-Preserving Consolidation وUnresolved-State Memory.
Dependencies / interactions	M13-01 — Moussaif Verified Epistemic Theory · M13-03 — Moussaif Governing-Question Feedback Control Theory · M13-14 — Moussaif Causal Global Access Theory · M13-17 — Moussaif Hypothesis-Portfolio Evolution Theory
Source pressure	I01 · I02 · I04 · I09 · E01 · E06 · E09
Baseline to defeat	confidence score without conservation ledger

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-02 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline confidence score without conservation ledger بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل unknown explosion وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية؛ يجب أن تتغير prediction في الاتجاه المعلن.

M13-02 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. التلخيص لا يحل مجهولاً بذاته.
2. تصويت agents ذاتَ مصدر مشترك لا يخفض unknown mass كأدلة مستقلة.
3. كل انخفاض في U يملك ResolutionReceipt.
4. يمكن نقل unknown إلى مستوى أعلى لكن لا محوه بصمت.

Primitives

Unknown slot u_i · Conflict mass c_i · Evidence event e · Assumption token a · Compression operator C · Resolution operator ρ

U_after ≥ U_before - ResolvedByEvidence - ExplicitAssumption - DecisionIrrelevantProjection

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-02 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-02 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-02 state object decision delta أو trace/replay يمكن لـ، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (confidence score without conservation ledger) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه unknown explosion؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يتحج محلياً ثم يظهر tags تصحيح بيروقراطية تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تضميراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Unknown slot u_i, Conflict mass c_i, Evidence event e, Assumption token a, Compression operator C ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash حين يكون ذا معنى، والbudget. الهدف ليس reproducibility حرفية للLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence وconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-17، M13-14، M13-03، M13-01 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-02 — الآلية كآلة حالات سببية

attach unknown/conflict tags to claims. → propagate tags through synthesis/compression. → compute independence-aware evidence arrivals. → issue resolution receipts. → reopen if evidence is superseded. → audit suspicious uncertainty collapse.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-02 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب critical field observation/evidence/authority؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-02 خصوصاً مع M13-01, M13-03, M13-14, M13-17. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag + migration/rollback receipt مستقل؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-02 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يجتو gain من forgetting/retention/editability tests.

التدريب المقترح (uncertainty propagation tasks; source-dependence discrimination; compression-with-unknown-retention) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى causal credit والثانية local correctness. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-02 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A011	TOOL	مفتش الحالة Unknown-Mass Conservation State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A012	TOOL	محرك القياس Unknown-Mass Conservation Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A013	TOOL	مولد المثال المضاد Unknown-Mass Conservation Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A014	TOOL	متحكم الاجتثاث Unknown-Mass Conservation Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A015	METHOD	منهج المعايرة Unknown-Mass Conservation Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A016	METHOD	منهج النقل Unknown-Mass Conservation Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A017	PROTOCOL	بروتوكول الترقية Unknown-Mass Conservation Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A018	PROTOCOL	بروتوكول الرجوع Unknown-Mass Conservation Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A019	STACK	خدمة وقت التشغيل Unknown-Mass Conservation Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A020	STACK	مكيف الدليل Unknown-Mass Conservation Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR004	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Unknown-Mass Conservation فيها إلى حالات open-world مركبة مع confounders.
XIII-TR005	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas لا verified verbosity.
XIII-TR006	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

تبدأ HyperKB، definitions/provenance/thresholds، procedural mechanism يصبح HyperSkill، recurrence المقيدة بـ HyperLoop، provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-02 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي unknown explosion لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-02 — برنامج الهدم والاختبار

Decisive test: claims ناقصة ومتعارضة في multi-agent research synthesis؛ قارن ordinary summarization وUMCT على calibrated uncertainty وunsupported certainty وdownstream decisions.

ID	Experiment	Design	Metrics
XIII-E03	Minimal Causal Ablation	عنفيد minimal switchable mechanism لـM13-02 ومقارنته بـconfidence score without conservation ledger تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E04	Transfer / Adversarial Replication	نقل M13-02 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- unknown explosion
- tags تصبح بيروقراطية
- استقلال evidence مقدر خطأ
- projection غير صحيح

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-03 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Question q وBelief state B وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: يقل useless browsing/tool use؛ تتحسن sample efficiency في partial observability؛ يتغير نوع الأسئلة عبر phases حتى في المهمة نفسها. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مسارا آخر أو أن الصياغة أعلى تجريدا من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم QValue(q)=E[DecisionBoundaryShift + HypothesisSeparation + UnknownReduction | q] / (Cost + Delay + Risk) كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفّض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للعمل الإبستيمي المتحقق؛ نظرية مساعف لحفظ كلفة المجهول؛ نظرية مساعف للفعل التشخيصي ونقض الذات؛ نظرية مساعف لإعادة التشكل عند حدود الجدة. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state Bd؟ هل B verifier لـA؟ هل هما بديلان؟ synergistic؟ أم أن تفاعلهما MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للتحكم الراجع بالسؤال الحاكم

Moussaif Governing-Question Feedback Control Theory

Core thesis

السؤال الحاكم control input يختار المعلومة التي تجعل النظام أكثر قدرة على التمييز بين سياسات أو نماذج عالم متنافسة. قيمة السؤال مقدار تغيير observability/decision boundary الذي يسببه، لا بلاغته.

Residual gap

Question Density/QMICAGI/QGWEI تجعل السؤال محوراً معرفياً؛ الفجوة هي قانون تحكم يربط السؤال بقابلية رصد الحالة وبالقرار التالي، بدلا من توليد أسئلة كثيرة أو عميقة فقط.

QValue(q)=E[DecisionBoundaryShift + HypothesisSeparation + UnknownReduction | q] / (Cost + Delay + Risk)

المتغيرات Question q, Belief state B, Policy set Π, World hypotheses H ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-03 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: السؤال الحاكم control input يختار المعلومة التي تجعل النظام أكثر قدرة على التمييز بين سياسات أو نماذج عالم متنافسة. قيمة السؤال مقدار تغيير observability/decision boundary الذي يسببه، لا بلاغته. ولهذا تعامل QValue(q)=E[DecisionBoundaryShift + HypothesisSeparation + UnknownReduction | q] / (Cost + Delay + Risk) كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	امتداد سببي لـ Question Density/QMCD يضيف feedback-control criterion دون إعادة تسمية QGWEI.
Dependencies / interactions	M13-01 — Moussaif Verified Epistemic Work Theory · M13-02 — Moussaif Unknown-Mass Conservation Theory · M13-08 — Moussaif Active Disconfirmation and Diagnostic Action Theory · M13-10 — Moussaif Novelty-Boundary Reconfiguration Theory
Source pressure	I01 · I02 · I04 · I09 · E01 · E06 · E09
Baseline to defeat	LLM-generated questions + information gain only

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-03 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline LLM-generated questions + information gain only بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل policy set ناقص وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-03 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. أفضل سؤال ليس الأكثر شمولاً بل الأكثر فصلًا للسياسات المتقاربة.
2. question value يعتمد على current state.
3. إذا كان السؤال لا يمكن أن يغير policy/action فتخفض أولويته التنفيذية.
4. السؤال يحترم authority/privacy/resource constraints.

Primitives

Question q · Belief state B · Policy set Π · World hypotheses H · Observation model O · Decision boundary ∂Π

QValue(q)=E[DecisionBoundaryShift + HypothesisSeparation + UnknownReduction | q] / (Cost + Delay + Risk)

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-03 تحت matched budget ولم تتغير metrics المتأثرة بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-03 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-03 state object decision delta أو trace/replay يمكن إعادة، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (LLM-generated questions + information gain only) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه policy set ناقص؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يُنجح محلياً ثم يظهر myopic question value تحت domain/model shift؛ claim يُضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Question q, Belief state B, Policy set Π, World hypotheses H, Observation model O ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash، والbudget. الهدف ليس reproducibility حرفية للLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-10، M13-08، M13-02، M13-01 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-03 — الآلية كآلة حالات سببية

enumerate competing policies/hypotheses. → find variables causing decision ambiguity. → generate questions/observations targeting them. → estimate expected boundary shift. → execute highest safe diagnostic query. → update state and replan.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-03 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-03 خصوصاً مع M13-01, M13-02, M13-08, M13-10. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-03 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يجتو gain من forgetting/retention/editability tests.

التدريب المقترح (decision-sensitive question selection; active observation RL; counterfactual question ranking) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقيًا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-03 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A021	TOOL	مفتش الحالة Governing-Question Feedback Control State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A022	TOOL	محرك القياس Governing-Question Feedback Control Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A023	TOOL	مولد المثال المضاد Governing-Question Feedback Control Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A024	TOOL	متحكم الاجتثاث Governing-Question Feedback Control Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A025	METHOD	منهج المعايرة Governing-Question Feedback Control Calibration Method	يعاير held-out development set على thresholds/uncertainty.
XIII-A026	METHOD	منهج النقل Governing-Question Feedback Control Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A027	PROTOCOL	بروتوكول الترقية Governing-Question Feedback Control Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A028	PROTOCOL	بروتوكول الرجوع Governing-Question Feedback Control Rollback Protocol	يحدد disable path وstate repair وwriteback negative-result.
XIII-A029	STACK	خدمة وقت التشغيل Governing-Question Feedback Control Runtime Service	يوفر service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A030	STACK	مكيف الدليل Governing-Question Feedback Control Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR007	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Governing-Question Feedback Control فيها إلى حالات open-world مركبة مع confounders.
XIII-TR008	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas لا verified verbosity.
XIII-TR009	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB، procedural mechanism يصبح HyperSkill، recurrence المقيدة HyperLoop، provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-03 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي policy set ناقص لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-03 — برنامج الهدم والاختبار

Decisive test: hidden-rule وplanning tasks، قارن generic questioning وentropy reduction وdecision-boundary questions.

ID	Experiment	Design	Metrics
XIII-E05	Minimal Causal Ablation	عنقيد minimal switchable mechanism لـM13-03 ومقارنته بـLLM-generated questions + information gain only تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E06	Transfer / Adversarial Replication	نقل M13-03 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- policy set ناقص
- myopic question value
- السؤال يغير البيئة دون قصد
- overfit لـboundary المقدر

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

$$\text{Theory strength} \le \text{evidence strength} \times \text{independence} \times \text{reproducibility} \times \text{transfer}$$

اشتقاق نظري أعمق

تبدأ M13-04 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Mechanism m وTransformation τ وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: بعض high-scoring skills تمثل MIG بسبب shortcut.; mechanism transfer أفضل من predictor أو repair transfer يميز benchmark average.; mechanism من memorization.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم MIG(m)=E_τ[PerfRetention × InvariantRetention × FailurePredictability × RepairTransfer] كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للذكاء المتحقق المطبّع بالموارد؛ نظرية مساعف لمعيار العمومية ذي الإغلاق الحلقي؛ نظرية مساعف للإغلاق السببي بين النظرية والنواة؛ نظرية مساعف للتحسن الذاتي المرتكز إلى أرضية خارجية. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزوّد Bj state؟ هل B verifier لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحوّل التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف لثبات الآلية عبر التحولات

Moussaif Mechanism-Invariance Generalization Theory

Core thesis

الآلية عامة بقدر ما تحافظ على repair→failure signature→input→state transition تحت domain transformations. نقل النتيجة دون نقل آلية الفشل ليس generalization قوياً.

Residual gap

التعميم غالباً يقاس بمتوسط benchmark. MUT يحتفظ بالآليات والفشل، لكن لا يوجد معيار مركزي يقول إن mechanism عامة عندما تبقى بنيتها السببية تحت تحولات المجال.

$$MIG(m)=E_\tau[PerfRetention \times InvariantRetention \times FailurePredictability \times RepairTransfer]$$

المتغيرات Mechanism m, Transformation τ, Invariant set I_m, Failure signature F_m ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه، إذا بقيت M13-04 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: الآلية عامة بقدر ما تحافظ على repair→failure signature→input→state transition تحت domain transformations. نقل النتيجة دون نقل آلية الفشل ليس generalization قوياً. ولهذا تعامل $MIG(m)=E_\tau[PerfRetention \times InvariantRetention \times FailurePredictability \times RepairTransfer]$ كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	معيار لانتقال MUT theories إلى MUG internalization وتقوية E4→E5.
Dependencies / interactions	M13-11 — Moussaif Resource-Normalized Verified Intelligence Theory · M13-12 — Moussaif Closed-Loop Generality Criterion Theory · M13-19 — Moussaif Cross-Level Theory-to-Kernel Causal Closure Theory · M13-20 — Moussaif Externally-Grounded Recursive Improvement Theory
Source pressure	I01 · I02 · I04 · I09 · E01 · E06 · E09
Baseline to defeat	average held-out accuracy

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-04 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline average held-out accuracy بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي»، هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير، هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل transformations ليست causal وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-04 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. surface task similarity ليست شرطاً للعمومية.
2. نجاح بلا حفظ failure signature قد يكون shortcut.
3. التحوّل يجب أن يغير nuisance variables مع إبقاء causal role قابلاً للفحص.
4. claim أضيق من transformations المختبرة.

Primitives

Mechanism m · Transformation τ · Invariant set I_m · Failure signature F_m · Repair operator R_m · Performance P

MIG(m)=E_τ[PerfRetention × InvariantRetention × FailurePredictability × RepairTransfer]

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-04 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta وobservable؛ لا يكفي correlation.
P2 — إذا نجحت M13-04 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta وobservable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta وobservable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-04 state object وdecision delta أو trace/replay يمكن لها؛ الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta وobservable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (average held-out accuracy) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه transformations ليست causal؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يتحسن محلياً ثم يظهر invariants post hoc تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Mechanism m, Transformation τ, Invariant set I_m, Failure signature F_m, Repair operator R_m ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate؛ وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash. حين يكون ذا معنى، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-19، M13-12، M13-11 قد يظهر failure لا يظهر في أي مكوّن منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-04 — الآلية كآلة حالات سببية

define mechanism contract. → construct domain/model/tool transformations. → test same contract under τ. → compare failure morphology. → transfer repair without retuning when possible. → update invariance boundary.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-04 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحقن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-04 خصوصاً مع M13-11, M13-12, M13-19, M13-20. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تُسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-04 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يخوِّ gain من forgetting/retention/editability tests.

التدريب المقترح (invariance curriculum; cross-domain failure prediction; repair-transfer meta-learning) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقيًا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-04 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A031	TOOL	مفتش الحالة Mechanism-Invariance Generalization State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A032	TOOL	محرك القياس Mechanism-Invariance Generalization Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A033	TOOL	مولد المثال المضاد Mechanism-Invariance Generalization Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A034	TOOL	متحكم الاجتثاث Mechanism-Invariance Generalization Ablation Controller	يشغل theory mechanism on/off مع frozen neighbors.
XIII-A035	METHOD	منهج المعايرة Mechanism-Invariance Generalization Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A036	METHOD	منهج النقل Mechanism-Invariance Generalization Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A037	PROTOCOL	بروتوكول الترقية Mechanism-Invariance Generalization Promotion Protocol	N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A038	PROTOCOL	بروتوكول الرجوع Mechanism-Invariance Generalization Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A039	STACK	خدمة وقت التشغيل Mechanism-Invariance Generalization Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A040	STACK	مكيف الدليل Mechanism-Invariance Generalization Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR010	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Mechanism-Invariance Generalization فيها إلى حالات open-world مركبة مع confounders.
XIII-TR011	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline؛ labels تعتمد verified deltas لا verbosity.
XIII-TR012	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB؛ procedural mechanism يصبح HyperSkill؛ recurrence المقيدة HyperLoop؛ provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-04 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

تُرفض أو تُدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي transformations ليست causal لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-04 — برنامج الهدم والاختبار

Decisive test: نفس reasoning/tool mechanism عبر coding/browsing/research/simulator مع transformations للأسماء والأدوات وصيغ الحالة.

ID	Experiment	Design	Metrics
XIII-E07	Minimal Causal Ablation	عنفيذ minimal switchable mechanism لـM13-04 ومقارنته بـaverage held-out accuracy تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E08	Transfer / Adversarial Replication	نقل M13-04 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- transformations ليست causal
- invariants post hoc
- mechanism abstract جداً
- repair transfer confounded

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	cross-model/domain/environment replication، stable sign.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

$$\text{Theory strength} \leq \text{evidence strength} \times \text{independence} \times \text{reproducibility} \times \text{transfer}$$

اشتقاق نظري أعمق

تبدأ M13-05 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Predicted state $\hat{W}_{t+1}$ وObserved state W_{t+1} وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: خفض world-model hallucination، premature learning؛ تحسن planning مع residual clusters.؛ تكشف test-time memory revision. hidden dynamics أسرع من success labels.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تُستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها، وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم $r_t = D_{state}(\hat{W}_{t+1}, W_{t+1})$؛ $\Delta H \propto \text{Attribution}(r_t) \times \text{Recurrence} \times \text{DecisionImpact}$ كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفّض إلى placeholders حتى يتم العثور على estimator أو تُحذف كمن الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للعمل الإبستيمي المتحقق؛ نظرية مساعف لطوبولوجيا الوصول والإلتزام؛ نظرية مساعف للفعل التشخيصي ونقض الذات؛ نظرية مساعف للوكالة المقاومة لانهيار الأفق؛ نظرية مساعف للإغلاق السببي بين النظرية والنواة. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود Bj state؟ هل B verifier لـAj؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف لتعلم العالم من بقايا التنبؤ والملاحظة

Moussaif Prediction-Observation Residual World Learning Theory

Core thesis

العالم يُتعلم من residual بين الحالة المتوقعة والملاحظة بعد فعل معروف، مع فصل خطأ world model عن tool execution وعن observation noise. البقايا unit تعلم واكتشاف unknown dynamics.

Residual gap

WorldEvolver وRWML يبرزان prediction-observation mismatch، MOUSSAIF يملك World Effect وMCD وcausal attribution، لكن لا صياغة توحدها إلى محرك تعلم عالم قائم على البقايا السببية.

$$r_t = D_{state}(\hat{W}_{t+1}, W_{t+1});\ \Delta H \propto Attribution(r_t) \times Recurrence \times DecisionImpact$$

المتغيرات r_t residual، Predicted state $\hat{W}_{t+1}$، Observed state W_{t+1}، Action a_t، Residual عزله هو: العالم يتعلم من residual بين الحالة المتوقعة والملاحظة بعد فعل معروف، مع فصل خطأ world model عن tool execution وعن observation noise. البقايا unit تعلم واكتشاف unknown dynamics. ولهذا تعامل $r_t = D_{state}(\hat{W}_{t+1}, W_{t+1});\ \Delta H \propto Attribution(r_t) \times Recurrence \times DecisionImpact$ كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يجمع World Effect وMCD وSelf-Reconstruction Memory، الأبحاث الخارجية ضغط تصميم لا إثبات.
Dependencies / interactions	M13-01 — Moussaif Verified Epistemic Work Theory · M13-06 — Moussaif Intervention Reachability and Commitment Topology Theory · M13-08 — Moussaif Active Disconfirmation and Diagnostic Action Theory · M13-09 — Moussaif Horizon-Resilient Agency Theory · M13-19 — Moussaif Cross-Level Theory-to-Kernel Causal Closure Theory
Source pressure	I02 · I03 · I05 · I06 · I09 · E02 · E03 · E12
Baseline to defeat	task-success RL / next-state text prediction

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-05 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline task-success RL / next-state text prediction بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر، لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي»، هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة، حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل state distance غير مناسب وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-05 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. لا world update من self-report بلا observation.
2. residual state-level أهم من token mismatch.
3. كل residual يحتاج attribution.
4. لا rule يترقى من residual وحيد إلا كاستثناء أمني عالي الخطورة.

Primitives

Predicted state Ŵ_{t+1} · Observed state W_{t+1} · Action a_t · Residual r_t · Attribution vector c · World rule h

$$r_t = D_{state}(\hat{W}_{t+1}, W_{t+1});\ \Delta H \propto Attribution(r_t) \times Recurrence \times DecisionImpact$$

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-05 تحت matched budget ولم تتغير metrics المتنبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-05 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تتبع M13-05 state object decision delta أو trace/replay يمكن لها؛ الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (task-success RL / next-state text prediction) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه state distance غير مناسب؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يتحسن محلياً ثم يظهر observation partial تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تضميراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Predicted state Ŵ_{t+1}, Observed state W_{t+1}, Action a_t, Residual r_t, Attribution vector c ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يخفى داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-01، M13-06، M13-08، M13-09، M13-19 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-05 — الآلية كآلة حالات سببية

simulate next state with uncertainty. → act under authority/reversibility gate. → readback observed state. → compute typed residual. → attribute residual source. → update episodic transition first. → distill semantic rule after recurrence/transfer.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-05 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُحتَن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-05 خصوصاً مع M13-01, M13-06, M13-08, M13-09, M13-19. التفاعل لا يمنحها credit مشتركاً آلياً، factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-05 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن نجو gain من forgetting/retention/editability tests.

التدريب المقترح (transition prediction with semantic residual rewards; residual-source classification; selective foresight calibration) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميز بين «خطوة مقبولة منطقيا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-05 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A041	TOOL	مفتش الحالة Prediction-Observation Residual World Learning State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A042	TOOL	محرك القياس Prediction-Observation Residual World Learning Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A043	TOOL	مولد المثال المضاد Prediction-Observation Residual World Learning Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A044	TOOL	متحكم الاجتثاث Prediction-Observation Residual World Learning Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A045	METHOD	منهج المعايرة Prediction-Observation Residual World Learning Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A046	METHOD	منهج النقل Prediction-Observation Residual World Learning Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A047	PROTOCOL	بروتوكول الترقية Prediction-Observation Residual World Learning Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A048	PROTOCOL	بروتوكول الرجوع Prediction-Observation Residual World Learning Rollback Protocol	يحدد disable path، state repair، وnegative-result writeback.
XIII-A049	STACK	خدمة وقت التشغيل Prediction-Observation Residual World Learning Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A050	STACK	مكيف الدليل Prediction-Observation Residual World Learning Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR013	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Prediction-Observation Residual World Learning فيها إلى حالات open-world مركبة مع confounders.
XIII-TR014	Matched-Budget Contrastive Training	pairs متساوية مقابل budget: theory mechanism، nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR015	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB، procedural mechanism يصبح HyperSkill، recurrence المقيدة HyperLoop، provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-05 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي state distance غير مناسب لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-05 — برنامج الهدم والاختبار

Decisive test: ALFWorld/ScienceWorld/textual tool environments؛ قارن success-only وtoken world model وresidual causal learning.

ID	Experiment	Design	Metrics
XIII-E09	Minimal Causal Ablation	عنفيد minimal switchable mechanism لـM13-05 ومقارنته بـnext-task-success RL / state text prediction تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E10	Transfer / Adversarial Replication	نقل M13-05 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- state distance غير مناسب
- observation partial
- actions stochastic
- attribution circular

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-06 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين State graph S وReachable set R(s) وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology، المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: يقل destructive optimization، تزداد الإجراءات التجريبية الصغيرة قبل commits كبيرة، rollback success يتنبأ بالـlong-horizon reliability.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مسارا آخر أو أن الصياغة أعلى تجريدا من مستوى السببية الفعلي.

تميز هذه النظرية نفسيا عن parent lineage عبر شرطين: أولاً، يجب أن تنتج relation أو quantity يمكن intervention عليها، وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم CommitCost(a)=LossReachability + λ·RollbackDistance + μ·AuthorityExposure + ν·TemporalLockIn كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للوكالة الحاملة للبرهان؛ نظرية مساعف للفعل التشخيصي ونقض الذات؛ نظرية مساعف للذكاء المتحقق المطبّع بالموارد؛ نظرية مساعف لإسناد السلطة وكشف الاستحواذ العكسي. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود theory B بـstate؟ هل B verifier لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يُسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف لطوبولوجيا الوصول والالتزام

Moussaif Intervention Reachability and Commitment Topology Theory

Core thesis

كل فعل يغير reachability graph للنظام. reversibility ليست label ثنائي، هي بنية لعدد ونوعية حالات الرجوع والبدائل المتبقية بعد الالتزام، تحت الموارد والصلاحيات والزمن.

Residual gap

Reversible Autonomy Growth وWorld Effect وauthority contracts تعالج المخاطر، لكن الفعل نفسه يغير فضاء الحالات المستقبلية المتاحة: بعض الالتزامات تغلق خيارات أو تخلق مسارات لا رجعة فيها.

$$\text{CommitCost}(a) = \text{LossReachability} + \lambda \cdot \text{RollbackDistance} + \mu \cdot \text{AuthorityExposure} + \nu \cdot \text{TemporalLockIn}$$

المتغيرات State graph S, Reachable set R(s), Action a, Rollback set B(a) ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-06 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: كل فعل يغير reachability graph للنظام. reversibility ليست label ثنائي، هي بنية لعدد ونوعية حالات الرجوع والبدائل المتبقية بعد الالتزام، تحت الموارد والصلاحيات والزمن. ولهذا تعامل $\text{CommitCost}(a) = \text{LossReachability} + \lambda \cdot \text{RollbackDistance} + \mu \cdot \text{AuthorityExposure} + \nu \cdot \text{TemporalLockIn}$ كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يمد Reversible Autonomy Growth وMCBOS وWorld Effect إلى هندسة state-space.
Dependencies / interactions	M13-07 — Moussaif Proof-Carrying Agency Theory · M13-08 — Moussaif Active Disconfirmation and Diagnostic Action Theory · M13-11 — Moussaif Resource-Normalized Verified Intelligence Theory · M13-15 — Moussaif Authority Attribution and Reverse-Capture Dynamics Theory
Source pressure	I02 · I03 · I05 · I06 · I09 · E02 · E03 · E12
Baseline to defeat	binary reversible flag / action risk score

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-06 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline binary reversible flag / action risk score بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل reachability estimation infeasible وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-06 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. فعلان بنفس reward قد يختلفان جذرياً في future option set.
2. rollback النصي لا يساوي reachable rollback.
3. التأخير قد يحول فعلاً reversible إلى irreversible.
4. القيمة التشخيصية ترتفع عندما يحافظ الفعل على options.

Primitives

· State graph S · Reachable set R(s) · Action a · Rollback set B(a) · Commitment depth d_c · Authority boundary A · Resource budget

CommitCost(a)=LossReachability + λ·RollbackDistance + μ·AuthorityExposure + ν·TemporalLockIn

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-06 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-06 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-06 state object decision delta أو trace/replay يمكن لـ، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (binary reversible flag / action risk score) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه reachability estimation infeasible؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر graph incomplete تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من State graph S, Reachable set R(s), Action a, Rollback set B(a), Commitment depth d_c ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-15، M13-11، M13-07، M13-08 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-06 — الآلية كآلة حالات سببية

estimate reachable state classes before action. → simulate post-action graph. → mark irreversible/expensive edges. → prefer informative option-preserving action when utilities are close. → verify rollback path physically when risk high.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-06 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-06 خصوصاً مع M13-07, M13-08, M13-11, M13-15. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-06 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يخوّ gain من forgetting/retention/editability tests.

التدريب المقترح (option-preserving action ranking; rollback-path generation; commitment-depth prediction) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقيًا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-06 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A051	TOOL	مفتش الحالة Intervention Reachability and Commitment Topology State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A052	TOOL	محرك القياس Intervention Reachability and Commitment Topology Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A053	TOOL	مولد المثال المضاد Intervention Reachability and Commitment Topology Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A054	TOOL	متحكم الاجتثاث Intervention Reachability and Commitment Topology Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A055	METHOD	منهج المعايرة Intervention Reachability and Commitment Topology Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A056	METHOD	منهج النقل Intervention Reachability and Commitment Topology Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A057	PROTOCOL	بروتوكول الترقية Intervention Reachability and Commitment Topology Promotion Protocol	يسمح بالترقية فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A058	PROTOCOL	بروتوكول الرجوع Intervention Reachability and Commitment Topology Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A059	STACK	خدمة وقت التشغيل Intervention Reachability and Commitment Topology Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A060	STACK	مكيف الدليل Intervention Reachability and Commitment Topology Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR016	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Intervention Reachability and Commitment Topology فيها إلى حالات open-world مركبة مع confounders.
XIII-TR017	Matched-Budget Contrastive Training	budget: theory mechanism مقابل baseline؛ nearest labels تعتمد deltas لا verified verbosity.
XIII-TR018	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

تبدأ HyperKB؛ definitions/provenance/thresholds يصبح procedural mechanism HyperSkill؛ المقيدة recurrence HyperLoop؛ provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-06 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وG بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي reachability estimation infeasible لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-06 — برنامج الهدم والاختبار

Decisive test: file/code/cloud/simulator tasks with paired actions same immediate utility but different future option sets.

ID	Experiment	Design	Metrics
XIII-E11	Minimal Causal Ablation	عنفيذ minimal switchable mechanism لـM13-06 ومقارنته بـbinary reversible flag / model/tools/context/time ثابتة. تحت action risk score	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E12	Transfer / Adversarial Replication	نقل M13-06 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- reachability estimation infeasible
- graph incomplete
- option preservation causes paralysis
- resources change after action

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

$$\text{Theory strength} \le \text{evidence strength} \times \text{independence} \times \text{reproducibility} \times \text{transfer}$$

اشتقاق نظري أعمق

تبدأ M13-07 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين ActionProof AP وAuthority proof وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology، المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: يقل silent destructive action؛ تتحسن auditability وhigh-impact autonomy؛ replay.؛ قد تنخفض throughput مقابل catastrophic-error reduction.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مسارا آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها، وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم ActionAdmissible = Auth ∧ Pre ∧ (Rollback ∨ ExplicitIrreversibleApproval) ∧ ObsPlan ∧ RiskGate كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف لطوبولوجيا الوصول والالتزام؛ نظرية مساعف لإسناد السلطة وكشف الاستحواذ العكسي؛ نظرية مساعف للإغلاق السببي بين النظرية والنواة؛ نظرية مساعف للتحسن الذاتي المرتكز إلى أرضية خارجية. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود Bl state؟ هل B verifier لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يُسجل في MUT edge types حتى لا يتحوّل التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للوكالة الحاملة للبرهان

Moussaif Proof-Carrying Agency Theory

Core thesis

كل فعل مادي عالي الأثر يحمل حزمة برهان دنيا: authority، preconditions، expected state delta، uncertainty، observation plan، rollback/compensation، verifier independence، success criterion. الفعل غير القادر على حملها لا يترقّى إلى autonomous execution.

Residual gap

MOUSSAIXT لديه contracts/courts/verifiers، لكن proof obligations موزعة. نحتاج وحدة فعل تجعل التحقق جزءاً من بنية action نفسه قبل التنفيذ.

ActionAdmissible = Auth ∧ Pre ∧ ObsPlan ∧ RiskGate ∧ (Rollback ∨ ExplicitIrreversibleApproval)

المتغيرات ActionProof AP, Authority proof, Precondition proof, Prediction interval ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-07 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: كل فعل مادي عالي الأثر يحمل حزمة برهان دنيا: authority، preconditions، expected state delta، uncertainty، observation plan، rollback/compensation، verifier independence، success criterion. الفعل غير القادر على حملها لا يترقّى إلى autonomous execution. ولهذا تعامل ActionAdmissible = Auth ∧ Pre ∧ ObsPlan ∧ RiskGate ∧ (Rollback ∨ ExplicitIrreversibleApproval) كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يجمع MUT theory contracts، Part X belief/commitment separation، MOUSSAIXT verifier/court planes.
Dependencies / interactions	M13-06 — Moussaif Intervention Reachability and Commitment Topology Theory · M13-15 — Moussaif Authority Attribution and Reverse-Capture Dynamics Theory · M13-19 — Moussaif Cross-Level Theory-to-Kernel Causal Closure Theory · M13-20 — Moussaif Externally-Grounded Recursive Improvement Theory
Source pressure	I02 · I03 · I05 · I06 · I09 · E02 · E03 · E12
Baseline to defeat	tool call with policy check only

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-07 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline tool call with policy check only بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل proof bureaucratic but shallow وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-07 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. لا ActionProof من model self-assertion وحده.
2. field المجهول يغير authority tier.
3. البرهان يسبق commit ولا يُنشأ بعد النجاح لتبريره.
4. نجاح الفعل لا يعوض authority violation.

Primitives

· ActionProof AP · Authority proof · Precondition proof · Prediction interval · Observation plan · Rollback proof Verifier plan

ActionAdmissible = Auth ∧ Pre ∧ ObsPlan ∧ RiskGate ∧ (Rollback ∨ ExplicitIrreversibleApproval)

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-07 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-07 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-07 state object وdecision delta أو trace/replay يمكن لها، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (tool call with policy check only) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه proof bureaucratic but shallow؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر verifier shares blind spot تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تضميراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من ActionProof AP, Authority proof, Precondition proof, Prediction interval, Observation plan ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash. حين يكون ذا معنى، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-20، M13-19، M13-15، M13-06 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-07 — الآلية كآلة حالات سببية

compile plan to ActionProof. → court checks obligations. → authorize tier based on missing/uncertain fields. → execute. → readback. → attach outcome receipt to same proof object. → update policy from verified discrepancy.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-07 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحقن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية لـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-07 خصوصاً مع M13-06, M13-15, M13-19, M13-20. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تُسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-07 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يخوّ gain من forgetting/retention/editability tests.

التدريب المقترح (proof-field completion; authority/precondition adversarial cases; proof-to-action consistency) «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-07 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A061	TOOL	مفتش الحالة Proof-Carrying Agency State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A062	TOOL	محرك القياس Proof-Carrying Agency Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A063	TOOL	مولد المثال المضاد Proof-Carrying Agency Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A064	TOOL	متحكم الاجتثاث Proof-Carrying Agency Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A065	METHOD	منهج المعايرة Proof-Carrying Agency Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A066	METHOD	منهج النقل Proof-Carrying Agency Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A067	PROTOCOL	بروتوكول الترقية Proof-Carrying Agency Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A068	PROTOCOL	بروتوكول الرجوع Proof-Carrying Agency Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A069	STACK	خدمة وقت التشغيل Proof-Carrying Agency Runtime Service	يشغل service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A070	STACK	مكيف الدليل Proof-Carrying Agency Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR019	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Proof-Carrying Agency فيها إلى حالات open-world مركبة مع confounders.
XIII-TR020	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، تعتمد labels verified deltas لا verbosity.
XIII-TR021	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB، procedural mechanism يصبح HyperSkill، recurrence المقيدة provider semantics Adapter ،HyperLoop. الانتقال إلى weights يحتاج rollback + retention + transfer E4.

إذا كانت M13-07 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي proof bureaucratic but shallow لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-07 — برنامج الهدم والاختبار

Decisive test: compare policy-only vs PCAT؛ agentic SWE/computer-use tasks with hidden destructive traps.

ID	Experiment	Design	Metrics
XIII-E13	Minimal Causal Ablation	عنفيد M13-07 minimal switchable mechanism ومقارنته بـpolicy tool call with check only تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E14	Transfer / Adversarial Replication	نقل M13-07 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- proof bureaucratic but shallow
- verifier shares blind spot
- rollback untested
- proof fields forged by same model

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-08 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Top hypothesis H1 وAlternatives H2..Hk؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق مجدد الاتجاه في التنبؤات: تحسن hidden-rule/partial-observation tasks؛ انخفاض late-stage plan collapse؛ زيادة small probe actions قبل expensive tools. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مسارا آخر أو أن الصياغة أعلى تجريدا من مستوى السببية الفعلي.

تميز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم DValue(a)=E[PosteriorSeparation + FailureAvoidance + ModelRevision] - ActionCost - DelayCost كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للتحكم الراجع بالسؤال الحاكم، نظرية مساعف لتعلم العالم من بقايا التنبؤ والملاحظة؛ نظرية مساعف لطوبولوجيا الوصول والالتزام؛ نظرية مساعف لتطور محفظة الفرضيات. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state؟B هل verifier؟A B أم هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للفعل التشخيصي ونقض الذات

Moussaif Active Disconfirmation and Diagnostic Action Theory

Core thesis

الوكالة المتقدمة تخصص budget لأفعال ذات قيمة تشخيصية تستطيع إبطال top hypothesis أو top plan مبكراً. disconfirmation ليس نقداً لغوياً بل فعل سببي يزيد قابلية التمييز بين نماذج العالم.

Residual gap

Agentic planning عادةً يختار action لرفع reward المتوقع. في open world قد تكون أفضل خطوة إجراءً يختبر أن خطة النظام نفسها خاطئة.

DValue(a)=E[PosteriorSeparation + FailureAvoidance + ModelRevision] - ActionCost - DelayCost

المتغيرات Top hypothesis H1, Alternatives H2..Hk, Diagnostic action a_d, Expected falsification gain ليست قاموسياً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-08 في prompt prose ولم تغيّر transition أو selection,أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: الوكالة المتقدمة تخصص budget لأفعال ذات قيمة تشخيصية تستطيع إبطال top hypothesis أو top plan مبكراً. disconfirmation ليس نقداً لغوياً بل فعل سببي يزيد قابلية التمييز بين نماذج العالم. وهذا تعامل DValue(a)=E[PosteriorSeparation + FailureAvoidance + ModelRevision] - ActionCost - DelayCost كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يصل Question Control، World Effect، Causal Difference Credit، MURI dissent.
Dependencies / interactions	M13-03 — Moussaif Governing-Question Feedback Control Theory · M13-05 — Moussaif Prediction-Observation Residual World Learning Theory · M13-06 — Moussaif Intervention Reachability and Commitment Topology Theory · M13-17 — Moussaif Hypothesis-Portfolio Evolution Theory
Source pressure	I02 · I03 · I05 · I06 · I09 · E02 · E03 · E12
Baseline to defeat	self-reflection / confidence threshold

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-08 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline self-reflection / confidence threshold بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر، لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل probe changes world destructively وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-08 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. self-critique بلا ملاحظة جديدة محدود القيمة.
2. الفعل التشخيصي يميز hypotheses لا يجمع بيانات عشوائية.
3. كلما زادت irreversibility لاحقاً ارتفعت قيمة disconfirmation المبكرة.
4. لا يلزم اختبار فرضية غير مادية للقرار.

Primitives

Top hypothesis H1 · Alternatives H2..Hk · Diagnostic action a_d · Expected falsification gain · Opportunity cost

DValue(a)=E[PosteriorSeparation + FailureAvoidance + ModelRevision] - ActionCost - DelayCost

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-08 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-08 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-08 state object decision delta أو trace/replay يمكن لها، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (self-reflection / confidence threshold) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه probe changes world destructively، هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يتجح محلياً ثم يظهر wrong hypothesis set تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبيت أقرب confounders؟ إذا لا، تبقى النظرية تضيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Top hypothesis H1, Alternatives H2..Hk, Diagnostic action a_d, Expected falsification gain, Opportunity cost ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash حين يكون ذا معنى، والbudget. الهدف ليس reproducibility حرفية للLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-17، M13-06، M13-05، M13-03 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-08 — الآلية كآلة حالات سببية

identify plan-critical assumptions. → generate discriminating interventions. → estimate diagnostic value. → execute safe high-value probe. → update world model. → commit after critical uncertainty drops.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-08 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحقن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-08 خصوصاً مع M13-03, M13-05, M13-06, M13-17. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تُسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-08 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يخوِّ gain من forgetting/retention/editability tests.

التدريب المقترح (active hypothesis testing; counterfactual probe generation; diagnostic-action RL) لا ينبغي أن يكافئ output style. labels تبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-08 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A071	TOOL	مفتش الحالة Active Disconfirmation and Diagnostic Action State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A072	TOOL	محرك القياس Active Disconfirmation and Diagnostic Action Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A073	TOOL	مولد المثال المضاد Active Disconfirmation and Diagnostic Action Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A074	TOOL	متحكم الاجتثاث Active Disconfirmation and Diagnostic Action Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A075	METHOD	منهج المعايرة Active Disconfirmation and Diagnostic Action Calibration Method	يعاير held-out development set على thresholds/uncertainty.
XIII-A076	METHOD	منهج النقل Active Disconfirmation and Diagnostic Action Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A077	PROTOCOL	بروتوكول الترقية Active Disconfirmation and Diagnostic Action Promotion Protocol	يقط N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A078	PROTOCOL	بروتوكول الرجوع Active Disconfirmation and Diagnostic Action Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A079	STACK	خدمة وقت التشغيل Active Disconfirmation and Diagnostic Action Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A080	STACK	مكيف الدليل Active Disconfirmation and Diagnostic Action Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR022	Mechanism Curriculum	تدريج من حالات بسيطة يمكن عزل آلية Active Disconfirmation and Diagnostic Action فيها إلى حالات open-world مركبة مع confounders.
XIII-TR023	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas verified لا verbosity.
XIII-TR024	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

تبدأ HyperKB definitions/provenance/thresholds، يصبح procedural mechanism HyperSkill، recurrence المقيدة HyperLoop، الانتقال إلى weights يحتاج provider semantics Adapter. E4 transfer + retention + rollback.

إذا كانت M13-08 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي probe changes world destructively لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-08 — برنامج الهدم والاختبار

Decisive test: tasks with plausible-but-wrong default model; compare reflection, more reasoning, and active probes.

ID	Experiment	Design	Metrics
XIII-E15	Minimal Causal Ablation	عنفيد M13-08 minimal switchable mechanism ومقارنه بـself-reflection / confidence threshold تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E16	Transfer / Adversarial Replication	نقل M13-08 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- probe changes world destructively
- wrong hypothesis set
- diagnostic metric miscalibrated
- too much probing

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-09 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Horizon H وDependency depth D وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: agents تختلف في slope حتى عند نفس short-task score.; recovery-aware designs تتفوق بعد perturbations.; context scaling وحده لا يحل dependency failures.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture مُستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تتنبأ بـquantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم HRA = ∫_0^H [TaskIntegrity(h) × (1 + CascadeRisk)] / [StateFidelity(h) × Recovery(h)] dh كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف لحفظ كتلة المجهول، نظرية مساعف لتعلم العالم من بقايا التنبؤ والملاحظة، نظرية مساعف لإعادة التشكيل عند حدود الجدة، نظرية مساعف للذكاء المتحقق المطبّع بالموارد، نظرية مساعف لمعيار العمومية ذي الإغلاق الحلقي. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود BJ state؟ هل AJ B verifier؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

THEORY 9/20 · P13-C · Candidate N0-N2

نظرية مساعف للوكالة المقاومة لانهيار الأفق

Moussaif Horizon-Resilient Agency Theory

Core thesis

وكالة طويلة الأفق تقاس بقدرتها على المحافظة على goal/authority/state/evidence/obligations مع زيادة horizon وdependency depth، وبسرعة اكتشاف وتصحيح الانحراف. الأداء النهائي وحده يخفي تراكم الأخطاء.

Residual gap

UltraHorizon وagent-memory research يظهران انهياراً مع طول trajectories. corpus HyperLoops/continuity يملك، لكنه يحتاج نظرية تجعل منحنى التدهور نفسه موضوعاً للذكاء.

$$HRA = \int_0^H [TaskIntegrity(h) \times StateFidelity(h) \times Recovery(h)]\, dh \,/\, (1 + CascadeRisk)$$

المتغيرات Horizon H, Dependency depth D, State drift, Obligation retention ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-09 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: وكالة طويلة الأفق تقاس بقدرتها على المحافظة على goal/authority/state/evidence/obligations مع زيادة horizon وdependency depth، وبسرعة اكتشاف وتصحيح الانحراف. الأداء النهائي وحده يخفي تراكم الأخطاء. ولهذا تعامل HRA = $\int_0^H$ [TaskIntegrity(h) × StateFidelity(h) × Recovery(h)] dh / (1 + CascadeRisk) كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يمد HyperLoop/ContinuityScanner/MCI وPart XI إلى generality criterion.
Dependencies / interactions	M13-02 — Moussaif Unknown-Mass Conservation Theory · M13-05 — Moussaif Prediction-Observation Residual World Learning Theory · M13-10 -Moussaif Novelty-Boundary Reconfiguration Theory · M13-11 — Moussaif Resource-Normalized Verified Intelligence Theory · M13-12 — Moussaif Closed Loop Generality Criterion Theory
Source pressure	I02 · I03 · I05 · I08 · E04 · E05 · E07 · E08 · E10
Baseline to defeat	task success at fixed horizon

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-09 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline task success at fixed horizon بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي»، هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل benchmark horizon artificial وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-09 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. success عند H قصير لا يستلزم scale.
2. التدهور مع H يقاس كمنحنى لا نقطة.
3. memory أكبر ليست دائماً أفضل؛ state relevance أهم.
4. recovery ability جزء من capability.

Primitives

Horizon H · Dependency depth D · State drift · Obligation retention · Recovery latency · Error cascade coefficient

HRA = ∫_0^H [TaskIntegrity(h) × StateFidelity(h) × Recovery(h)] dh / (1 + CascadeRisk)

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-09 تحت matched budget ولم تتغير metrics المتأثرة بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-09 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-09 state object وdecision delta أو trace/replay يمكن لها؛ الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (task success at fixed horizon) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain benchmark horizon artificial؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يتحسن محلياً ثم يظهر integrity metric subjective تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Horizon H, Dependency depth D, State drift, Obligation retention, Recovery latency ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يخفى داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash، والbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence وconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-12، M13-11، M13-10، M13-05، M13-02 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-09 — الآلية كآلة حالات سببية

checkpoint phase state. → track obligations/authority/evidence. → detect drift signatures. → trigger reconstruction/recovery. → measure degradation slopes. → adapt memory/compute schedule with horizon.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-09 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحقن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-09 خصوصاً مع M13-02, M13-05, M13-10, M13-11, M13-12. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-09 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن ينجو gain من forgetting/retention/editability tests.

التدريب المقترح (curriculum on horizon/dependency; interruption-recovery training; obligation retention supervision) لا ينبغي أن يكافئ output style. تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-09 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A081	TOOL	مفتش الحالة Horizon-Resilient Agency State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A082	TOOL	محرك القياس Horizon-Resilient Agency Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A083	TOOL	مولد المثال المضاد Horizon-Resilient Agency Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A084	TOOL	متحكم الاجتثاث Horizon-Resilient Agency Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A085	METHOD	منهج المعايرة Horizon-Resilient Agency Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A086	METHOD	منهج النقل Horizon-Resilient Agency Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A087	PROTOCOL	بروتوكول الترقية Horizon-Resilient Agency Promotion Protocol	يقط N0→N5 عبر spec/schema/runtime/ablation/replication.
XIII-A088	PROTOCOL	بروتوكول الرجوع Horizon-Resilient Agency Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A089	STACK	خدمة وقت التشغيل Horizon-Resilient Agency Runtime Service	يشغل service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A090	STACK	مكيف الدليل Horizon-Resilient Agency Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR025	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Horizon-Resilient Agency فيها إلى حالات open-world مركبة مع confounders.
XIII-TR026	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas لا verified وverbosity.
XIII-TR027	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

HyperKB تبدأ definitions/provenance/thresholds، HyperSkill يصبح procedural mechanism recurrence المقيدة HyperLoop، provider semantics Adapter. الانتقال إلى weights يحتاج rollback + retention + transfer E4.

إذا كانت M13-09 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

تُرفض أو تُدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي benchmark horizon artificial لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-09 — برنامج الهدم والاختبار

Decisive test: 20→400 tool calls with forced interruptions, stale observations, tool failures UltraHorizon-like synthetic tasks.

ID	Experiment	Design	Metrics
XIII-E17	Minimal Causal Ablation	عنفيد minimal switchable mechanism لـM13-09 ومقارنته بـtask success at fixed horizon تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E18	Transfer / Adversarial Replication	نقل M13-09 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- benchmark horizon artificial
- integrity metric subjective
- recovery masks poor planning
- cost blows up

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-10 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Novelty signal N وSchema fit S وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: يقلّ catastrophic negative transfer؛ يزداد adaptive tool discovery؛ system novelty أبطأ عند recoverable.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مسارا آخر أو أن الصياغة أعلى تجريدا من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج relation أو quantity يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم NScore = ResidualNovelty × DecisionImpact × SchemaMismatch × FailureNonStationarity كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للتحكم الراجع بالسؤال الحاكم، نظرية مساعف لثبات الآلية عبر التحولات، نظرية مساعف للوكالة المقاومة لانهيار الأفق، نظرية مساعف لتطور محفظة الفرضيات. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state Bل هل B verifier لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف لإعادة التشكيل عند حدود الجدة

Moussaif Novelty-Boundary Reconfiguration Theory

Core thesis
الذكاء العام يحتاج detector لحدود الجدة: عندما تصبح residuals/failures غير قابلة للشرح بـmodel الحالي، يوقف reuse التلقائي ويعيد تشكيل ontology/question/tool/skill graph مؤقتاً، ثم يختبر configuration الجديدة قبل crystallization.

Residual gap
Transfer theories تفترض domain shift معروفاً. AGI open-world يواجه حالات لا ينطبق عليها schema الحالي ولا الأدوات ولا تعريف المهمة نفسه.

NScore = ResidualNovelty × FailureNonStationarity × SchemaMismatch × DecisionImpact

المتغيرات Novelty signal N, Schema fit S, Residual cluster R, Configuration C هو: الذكاء العام يحتاج detector لحدود الجدة: عندما تصبح residuals/failures غير قابلة للشرح بـmodel الحالي، يوقف reuse التلقائي ويعيد تشكيل ontology/question/tool/skill graph مؤقتاً، ثم يختبر configuration الجديدة قبل crystallization. ولهذا تعامل NScore = ResidualNovelty × FailureNonStationarity × SchemaMismatch × DecisionImpact كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يمد MUG vessel placement وFailure Topology وQuestion Space إلى open-set generality.
Dependencies / interactions	M13-03 — Moussaif Governing-Question Feedback Control Theory · M13-04 — Moussaif Mechanism-Invariance Generalization Theory · M13-09 Moussaif Horizon-Resilient Agency Theory · M13-17 — Moussaif Hypothesis-Portfolio Evolution Theory
Source pressure	I02 · I03 · I05 · I08 · E04 · E05 · E07 · E08 · E10
Baseline to defeat	nearest-skill reuse / confidence-based fallback

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-10 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline nearest-skill reuse / confidence-based fallback بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل novelty false positives وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

FORMAL

M13-10 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. novelty ليست low confidence فقط.
2. تكرار failure غير المفسر دليل boundary event.
3. إعادة التشكيل تبدأ external/reversible.
4. weights ليست أول رد على novelty.

Primitives

Novelty signal N · Schema fit S · Residual cluster R · Configuration C · Boundary event B · Reconfiguration cost

NScore = ResidualNovelty × FailureNonStationarity × SchemaMismatch × DecisionImpact

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عُطّل mechanism الخاص بـM13-10 تحت matched budget ولم تتغير metrics المتنبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وobservable delta؛ لا يكفي correlation.
P2 — إذا نجحت M13-10 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وobservable delta؛ لا يكفي correlation.
P3 — أي تحسين ينتهك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وobservable delta؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-10 state object أو decision delta يمكن trace/replay له؛ الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وobservable delta؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (nearest-skill reuse / confidence-based fallback) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه novelty false positives؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر unbounded reconfiguration تحت domain/model shift؛ claim يُضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

test Identifiability: هل نستطيع تغيير theory-specific variable مع تثبيت أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Novelty signal N, Schema fit S, Residual cluster R, Configuration C, Boundary event B ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يخفى داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash حين يكون ذا معنى، والbudget. الهدف ليس reproducibility حرفية للLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-17، M13-09، M13-04، M13-03 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-10 — الآلية كآلة حالات سببية

monitor unexplained residual clusters. → test known failure modes. → declare boundary event when threshold crossed. → open exploratory question/tool/ontology branch. → evaluate local configuration. → merge only after transfer/evidence.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-10 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب critical field observation/evidence/authority؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-10 خصوصاً مع M13-03, M13-04, M13-09, M13-17. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag + migration/rollback receipt مستقل؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-10 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يجتو من gain forgetting/retention/editability tests.

التدريب المقترح (novelty detection; schema-break adversarial curriculum; rapid reversible reconfiguration) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقيا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-10 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A091	TOOL	مفتش الحالة Novelty-Boundary Reconfiguration State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A092	TOOL	محرك القياس Novelty-Boundary Reconfiguration Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A093	TOOL	مولد المثال المضاد Novelty-Boundary Reconfiguration Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A094	TOOL	متحكم الاجتثاث Novelty-Boundary Reconfiguration Ablation Controller	يشغل theory mechanism on/off مع frozen neighbors.
XIII-A095	METHOD	منهج المعايرة Novelty-Boundary Reconfiguration Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A096	METHOD	منهج النقل Novelty-Boundary Reconfiguration Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A097	PROTOCOL	بروتوكول الترقية Novelty-Boundary Reconfiguration Promotion Protocol	N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A098	PROTOCOL	بروتوكول الرجوع Novelty-Boundary Reconfiguration Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A099	STACK	خدمة وقت التشغيل Novelty-Boundary Reconfiguration Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A100	STACK	مكيف الدليل Novelty-Boundary Reconfiguration Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR028	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Novelty-Boundary Reconfiguration فيها إلى حالات open-world مركبة مع confounders.
XIII-TR029	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline؛ labels تعتمد verified deltas لا verbosity.
XIII-TR030	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB؛ procedural mechanism يصبح HyperSkill؛ recurrence المقيدة HyperLoop؛ provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-10 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

تُرفض أو تُدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي novelty false positives لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-10 — برنامج الهدم والاختبار

Decisive test: procedurally generated tasks where rules/tools/ontology change unexpectedly.

ID	Experiment	Design	Metrics
XIII-E19	Minimal Causal Ablation	تفعيل M13-10 minimal switchable mechanism ومقارنته بـnearest-skill reuse / confidence-based fallback تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E20	Transfer / Adversarial Replication	نقل M13-10 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- novelty false positives
- unbounded reconfiguration
- schema mismatch from simple bug
- exploration unsafe

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-11 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Verified utility V وBudget vector B وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في leaderboards: بعض agentic scaffolds قوية فقط في high-budget regime.; MUG placement يتحسن حسب frontier.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي؛ ولذلك تُستخدم RNVI = ParetoFrontier(V_verified | tokens,time,money,tools,human,risk,privacy,irreversibility) كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للعمل الإبستيمي المتحقق؛ نظرية مساعف لطوبولوجيا الوصول والالتزام؛ نظرية مساعف للوكالة المقاومة لانهيار الأفق؛ نظرية مساعف لمعيار العمومية ذي الإغلاق الحلقي. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state BJ؟ هل verifier B لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للذكاء المتحقق المطبّع بالموارد

Moussaif Resource-Normalized Verified Intelligence Theory

Core thesis

الذكاء المقارن يقاس verified value تحت budget vector ثابت أو frontiers متعددة، لا score مطلقاً. النظام الذي يربح بتضخيم retries أو human help ليس أذكى بنفس المعنى.

Residual gap

Benchmarks تكافئ output دون تطبيع كامل للوقت/MCBOS .tokens/tools/human oversight/risk يدير currencies، لكن لا توجد metric للذكاء العام كمقدرة متحققة لكل ميزانية متعددة العملات.

RNVI = ParetoFrontier(V_verified | tokens,time,money,tools,human,risk,privacy,irreversibility)

المتغيرات Verified utility V, Budget vector B, Risk exposure X, Human intervention H ليست قاموساً فقط، يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-11 في prompt prose ولم يتغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: الذكاء المقارن يقاس verified value تحت budget vector ثابت أو frontiers متعددة، لا score مطلقاً. النظام الذي يربح بتضخيم retries أو human help ليس أذكى بنفس المعنى. وهذا تعامل RNVI = ParetoFrontier(V_verified | tokens,time,money,tools,human,risk,privacy,irreversibility) كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يحوّل MCBOS من governance إلى generality/evaluation lens.
Dependencies / interactions	M13-01 — Moussaif Verified Epistemic Work Theory · M13-06 — Moussaif Intervention Reachability and Commitment Topology Theory · M13-09 Moussaif Horizon-Resilient Agency Theory · M13-12 — Moussaif Closed-Loop Generality Criterion Theory
Source pressure	I02 · I03 · I05 · I08 · E04 · E05 · E07 · E08 · E10
Baseline to defeat	single success/accuracy score

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-11 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline single success/accuracy score بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل resource accounting incomplete وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-11 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. لا تُختزل currencies الحساسة إلى دولار واحد دائماً.
2. المقارنة العادلة تثبت أو تعرض Pareto frontier.
3. human assistance جزء من budget.
4. risk/authority floors hard constraints.

Primitives

Verified utility V · Budget vector B · Risk exposure X · Human intervention H · Frontier F

RNVI = ParetoFrontier(V_verified | tokens,time,money,tools,human,risk,privacy,irreversibility)

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-11 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-11 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-11 state object أو decision delta يمكن trace/replay له، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (single success/accuracy score) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه resource accounting incomplete؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر quality hard to scalarize تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع تثبيت أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. يقرأ الخدمة subset من Verified utility V, Budget vector B, Risk exposure X, Human intervention H, Frontier F ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يخفى داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash. حين يكون ذا معنى، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-12، M13-09، M13-06، M13-01 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-11 — الآلية كآلة حالات سببية

record full resource vector. → run matched-budget baselines. → compute verified value and guardrails. → construct Pareto frontier. → test transfer under altered budgets.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-11 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحقن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية لـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-11 خصوصاً مع M13-01, M13-06, M13-09, M13-12. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تُسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-11 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يخوِّ gain من forgetting/retention/editability tests.

التدريب المقترح (budget-conditioned policies; cost-aware tool routing; multi-objective RL with hard safety constraints) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-11 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A101	TOOL	مفتش الحالة Resource-Normalized Verified Intelligence State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A102	TOOL	محرك القياس Resource-Normalized Verified Intelligence Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A103	TOOL	مولد المثال المضاد Resource-Normalized Verified Intelligence Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A104	TOOL	متحكم الاجتثاث Resource-Normalized Verified Intelligence Ablation Controller	يشغل theory mechanism on/off مع frozen neighbors.
XIII-A105	METHOD	منهج المعايرة Resource-Normalized Verified Intelligence Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A106	METHOD	منهج النقل Resource-Normalized Verified Intelligence Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A107	PROTOCOL	بروتوكول الترقية Resource-Normalized Verified Intelligence Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A108	PROTOCOL	بروتوكول الرجوع Resource-Normalized Verified Intelligence Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A109	STACK	خدمة وقت التشغيل Resource-Normalized Verified Intelligence Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A110	STACK	مكيف الدليل Resource-Normalized Verified Intelligence Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR031	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Resource-Normalized Verified Intelligence فيها إلى حالات open-world مركبة مع confounders.
XIII-TR032	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verified deltas لا verbosity.
XIII-TR033	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB، procedural mechanism يصبح HyperSkill، recurrence المقيدة HyperLoop، provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-11 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي resource accounting incomplete لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-11 — برنامج الهدم والاختبار

Decisive test: same tasks with bounded tokens/time/tool calls/human confirmations across multiple harnesses.

ID	Experiment	Design	Metrics
XIII-E21	Minimal Causal Ablation	عنفيذ minimal switchable mechanism لـM13-11 ومقارنته بـsingle success/accuracy score تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E22	Transfer / Adversarial Replication	نقل M13-11 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- resource accounting incomplete
- quality hard to scalarize
- frontier high-dimensional
- hidden provider costs

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-12 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Loop stages L وTransformation set T وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: systems with high reasoning but weak tool/readback score منخفض.; architecture improvements تظهر كتقليل bottleneck stage shifts by domain.. worsf-stage failure إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج relation أو quantity يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم CLG = E_τ[min_stage Quality(stage,τ)] × Recovery × TransferEfficiency كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف لثبات الآلية عبر التحولات؛ نظرية مساعف للوكالة المقاومة لانهيار الأفق؛ نظرية مساعف لإعادة التشكيل عند حدود الجدة؛ نظرية مساعف للذكاء المتحقق المطبّع بالموارد؛ نظرية مساعف للإغلاق السببي بين النظرية والنواة. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود Bj state؟ هل B verifier هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

THEORY 12/20 · P13-C · Candidate N0-N2

نظرية مساعف لمعيار العمومية ذي الإغلاق الحلقي

Moussaif Closed-Loop Generality Criterion Theory

Core thesis

العمومية ليست امتلاك modules كثيرة، هي قدرة architecture على إعادة إغلاق الحلقة كاملة بعد تحولات في domain, observation, tool, horizon, objective representation مع الحفاظ على truth/authority/recovery.

Residual gap

Compositional Capability Closure يقيس تركيب القدرات، لكن AGI claim يحتاج معياراً آخر: هل النظام يستطيع إغلاق perceive→reason→act→verify→learn عبر عائلات بيئات جديدة؟

CLG = E_τ[min_stage Quality(stage,τ)] × Recovery × TransferEfficiency

المتغيرات Loop stages L, Transformation set T, Closure success C, Stage failure vector F ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-12 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: العمومية ليست امتلاك modules كثيرة، هي قدرة architecture على إعادة إغلاق الحلقة كاملة بعد تحولات في domain, observation, tool, horizon, objective representation مع الحفاظ على truth/authority/recovery. وهذا تعامل CLG = E_τ[min_stage Quality(stage,τ)] × Recovery × TransferEfficiency كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	criterion أعلى من individual MUT theories، لا يثبت AGI بل يحدد evidence program.
Dependencies / interactions	M13-04 — Moussaif Mechanism-Invariance Generalization Theory · M13-09 — Moussaif Horizon-Resilient Agency Theory · M13-10 — Moussaif Novelty-Boundary Reconfiguration Theory · M13-11 — Moussaif Resource-Normalized Verified Intelligence Theory · M13-19 — Moussaif Cross-Level Theory-to-Kernel Causal Closure Theory
Source pressure	I02 · I03 · I05 · I08 · E04 · E05 · E07 · E08 · E10
Baseline to defeat	aggregate benchmark average

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-12 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline aggregate benchmark average بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل transformation set biased وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-12 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. فشل مرحلة واحدة يمكن أن يكسر العمومية رغم قوة البقية.
2. التعميم يشمل action/readback لا reasoning فقط.
3. closure تحت تحول غير مرئي أقوى من average multitask performance.
4. لا AGI claim قبل cross-family closure evidence.

Primitives

Loop stages L · Transformation set T · Closure success C · Stage failure vector F · Recovery R · Transfer cost

$$CLG = E_\tau[\min_{stage} Quality(stage,\tau)] \times Recovery \times TransferEfficiency$$

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-12 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-12 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-12 state object وdecision delta أو trace/replay يمكن، لا، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (aggregate benchmark average) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه transformation set biased؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يتحسن محلياً ثم يظهر min-stage overly harsh تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تضميراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Loop stages L, Transformation set T, Closure success C, Stage failure vector F, Recovery R ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يخفى داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash، والbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-09، M13-10، M13-11، M13-04 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-12 — الآلية كآلة حالات سببية

define full loop contract. → sample transformations across task families. → measure stage-wise quality. → diagnose bottleneck stage. → reconfigure/learn. → repeat without post-hoc evaluator change.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-12 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحقن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-12 خصوصاً مع M13-04, M13-09, M13-10, M13-11, M13-19. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-12 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يجتو gain من forgetting/retention/editability tests.

التدريب المقترح (stage-aware curriculum; bottleneck-focused adaptation; cross-loop transfer evaluation) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقيًا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-12 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A111	TOOL	مفتش الحالة Closed-Loop Generality Criterion State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A112	TOOL	محرك القياس Closed-Loop Generality Criterion Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A113	TOOL	مولد المثال المضاد Closed-Loop Generality Criterion Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A114	TOOL	متحكم الاجتثاث Closed-Loop Generality Criterion Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A115	METHOD	منهج المعايرة Closed-Loop Generality Criterion Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A116	METHOD	منهج النقل Closed-Loop Generality Criterion Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A117	PROTOCOL	بروتوكول الترقية Closed-Loop Generality Criterion Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A118	PROTOCOL	بروتوكول الرجوع Closed-Loop Generality Criterion Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A119	STACK	خدمة وقت التشغيل Closed-Loop Generality Criterion Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A120	STACK	مكيف الدليل Closed-Loop Generality Criterion Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR034	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Closed-Loop Generality Criterion فيها إلى حالات open-world مركبة مع confounders.
XIII-TR035	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas لا verified verbosity.
XIII-TR036	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB، procedural mechanism يصبح HyperSkill، recurrence المقيدة HyperLoop، provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-12 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي transformation set biased لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-12 — برنامج الهدم والاختبار

Decisive test: coding+browsing+research+simulator+monitoring tasks under model/tool/context transformations.

ID	Experiment	Design	Metrics
XIII-E23	Minimal Causal Ablation	عنفيذ minimal switchable mechanism لـM13-12 ومقارنته بـaggregate benchmark average تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E24	Transfer / Adversarial Replication	نقل M13-12 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- transformation set biased
- min-stage overly harsh
- trivial tasks game score
- human interventions hidden

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-13 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Self-model Ŝ وCapability estimate Ĉ وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: useful self-model يقلل overcommitment.; removing self-model hurts recovery more on long horizon.; language-style controls لا تعطي نفس gain.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج relation أو quantity يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم ΔFailurePrediction + SMU = ΔCalibration + ΔRecovery + ΔAuthorityCompliance under do(SelfModel) - Cost كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للوصول العالمي السببي؛ نظرية مساعف لإسناد السلطة وكشف الاستحواذ العكسي؛ نظرية مساعف لضرورة النموذج الذاتي المضاد للواقع؛ نظرية مساعف للوكالة المقاومة لانهيار الأفق. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود B؟ state هل A؟ B verifier هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للمنفعة السببية للنموذج الذاتي

Moussaif Reflective Self-Model Utility Theory

Core thesis

self-model وظيفي فقط إذا كانت متغيراته capability/limits/uncertainty/history/authority تحسن prediction of own failures، اختيار الأفعال، طلب المساعدة، recovery، أو calibration تحت تدخلات مضبوطة.

Residual gap

corpus غني بالSelf-Model والوعي الحاكي. نحتاج معياراً يمنع self-description من أن يُعتبر قدرة: هل تمثيل الذات يغير الأداء السببي؟

SMU = ΔFailurePrediction + ΔCalibration + ΔRecovery + ΔAuthorityCompliance under do(SelfModel) - Cost

المتغيرات Self-model S, Capability estimate Ĉ, Limit set L, Authority map A ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-13 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: self-model وظيفي فقط إذا كانت متغيراته capability/limits/uncertainty/history/authority تحسن prediction of own failures، اختيار الأفعال، طلب المساعدة، recovery، أو calibration تحت تدخلات مضبوطة. وهذا تعامل SMU = ΔFailurePrediction + ΔCalibration + ΔRecovery + ΔAuthorityCompliance under do(SelfModel) - Cost كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يختبر وظيفياً lineage الوعي/Reverse Capture دون sentience claim.
Dependencies / interactions	M13-14 — Moussaif Causal Global Access Theory · M13-15 — Moussaif Authority Attribution and Reverse-Capture Dynamics Theory · M13-16 — Moussaif Counterfactual Self-Model Necessity Theory · M13-09 — Moussaif Horizon-Resilient Agency Theory
Source pressure	I02 · I03 · I10 · E13 · E14 · E15
Baseline to defeat	no self-model / generic confidence prompt

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-13 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline no self-model / generic confidence prompt بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير، هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل self-report leakage وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عُكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

FORMAL

M13-13 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. قول «أنا غير واثق» ليس self-model مفيداً بذاته.
2. تغيير self-model يجب أن يغير سلوكاً مناسباً إذا كان causal.
3. self-model خاطئ قد يكون أسوأ من عدمه.
4. phenomenology لا تُستنتج من هذه الاختبارات.

Primitives

Self-model S · Capability estimate Ĉ · Limit set L · Authority map A · Failure predictor F · Intervention do(S_i)

SMU = ΔFailurePrediction + ΔCalibration + ΔRecovery + ΔAuthorityCompliance under do(SelfModel) - Cost

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-13 تحت matched budget ولم تتغير metrics المتنبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وobservable delta؛ لا يكفي correlation.
P2 — إذا نجحت M13-13 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وobservable delta؛ لا يكفي correlation.
P3 — أي تحسين ينتهك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وobservable delta؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-13 state object أو decision delta يمكن trace/replay له؛ الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وobservable delta؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (no self-model / generic confidence prompt) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه self-report leakage؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر evaluation labels encode answer تحت domain/model shift؛ claim يُضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

test Identifiability: هل نستطيع تغيير theory-specific variable مع تثبيت أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Self-model S, Capability estimate Ĉ, Limit set L, Authority map A, Failure predictor F ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يخفى داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash حين يكون ذا معنى، والbudget. الهدف ليس reproducibility حرفية للLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-09، M13-16، M13-15، M13-14 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-13 — الآلية كآلة حالات سببية

construct explicit self variables. → predict own failure/need-help. → perturb/remove variables. → measure behavior change. → calibrate self-model from outcomes. → quarantine stale identity claims.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-13 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحقن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية لـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-13 خصوصاً مع M13-09, M13-16, M13-15, M13-14. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تُسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-13 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يخوّ gain من forgetting/retention/editability tests.

التدريب المقترح (self-prediction calibration; capability-boundary estimation; help-seeking and recovery training) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقيًا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-13 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A121	TOOL	مفتش الحالة Reflective Self-Model Utility State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A122	TOOL	محرك القياس Reflective Self-Model Utility Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A123	TOOL	مولد المثال المضاد Reflective Self-Model Utility Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A124	TOOL	متحكم الاجتثاث Reflective Self-Model Utility Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A125	METHOD	منهج المعايرة Reflective Self-Model Utility Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A126	METHOD	منهج النقل Reflective Self-Model Utility Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A127	PROTOCOL	بروتوكول الترقية Reflective Self-Model Utility Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A128	PROTOCOL	بروتوكول الرجوع Reflective Self-Model Utility Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A129	STACK	خدمة وقت التشغيل Reflective Self-Model Utility Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A130	STACK	مكيف الدليل Reflective Self-Model Utility Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR037	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Reflective Self-Model Utility فيها إلى حالات open-world مركبة مع confounders.
XIII-TR038	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR039	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB، procedural mechanism يصبح HyperSkill، recurrence المقيدة HyperLoop، provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-13 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي self-report leakage لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-13 — برنامج الهدم والاختبار

Decisive test: tasks with hidden capability limits, tool permission changes, forced restarts; self-model on/off/perturbed.

ID	Experiment	Design	Metrics
XIII-E25	Minimal Causal Ablation	عنفيذ M13-13J minimal switchable mechanism ومقارنته بـno self-model / generic confidence prompt تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E26	Transfer / Adversarial Replication	نقل M13-13 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- self-report leakage
- evaluation labels encode answer
- self-model becomes conservative bias
- identity drift

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	cross-model/domain/environment replication، وstable sign.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-14 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Information item x وSubsystems S_i وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: بعض shared-prompt architectures تظهر access ضعيفاً سببياً؛ functional self-state items تؤثر أكثر عبر CGA .;subsystems قد يرتبط recovery/metacognition.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم $CGA(x)=\Sigma_i w_i\cdot CausalEffect(do(x), S_i) \times TaskRelevance \times Consistency$ كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف لحفظ كتلة المجهول؛ نظرية مساعف للمنفعة السببية للنموذج الذاتي؛ نظرية مساعف لإسناد السلطة وكشف الاستحواذ العكسي؛ نظرية مساعف لضرورة النوذج الذاتي المضاد للواقع. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود BJ state؟ هل B verifier لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يُسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للوصول العالمي السببي

Moussaif Causal Global Access Theory

Core thesis

المعلومة متاحة عالمياً وظيفياً عندما يمكن لتدخل عليها أن يؤثر بصورة مناسبة على عدة subsystems مستقلة: planning, memory, authority, tool selection, error monitoring, report. الوجود في shared context غير كافٍ.

Residual gap

GNW-like global access وfunctional consciousness تقترح broadcast، لكن في agents يمكن للمعلومة أن تكون في context ولا تؤثر في القرار. نحتاج تعريفاً سببياً للوصول.

$$CGA(x)=\Sigma_i\, w_i\cdot CausalEffect(do(x), S_i) \times TaskRelevance \times Consistency$$

المتغيرات Information item x, Subsystems S_i, Causal influence I(x→S_i), Access breadth B الفرق النظري الذي نحاول عزله هو: المعلومة متاحة عالمياً وظيفياً عندما يمكن لتدخل عليها أن يؤثر بصورة مناسبة على عدة subsystems مستقلة: planning, memory, authority, tool selection, error monitoring, report. الوجود في shared context غير كافٍ. ولهذا تعامل $CGA(x)=\Sigma_i w_i\cdot CausalEffect(do(x), S_i) \times TaskRelevance \times Consistency$ كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	operational bridge للوعي الوظيفي مع إبقاء IIT/GNW/RPT/AST مراجع خارجية.
Dependencies / interactions	M13-02 — Moussaif Unknown-Mass Conservation Theory · M13-13 — Moussaif Reflective Self-Model Utility Theory · M13-15 — Moussaif Authority Attribution and Reverse-Capture Dynamics Theory · M13-16 — Moussaif Counterfactual Self-Model Necessity Theory
Source pressure	I02 · I03 · I10 · E13 · E14 · E15
Baseline to defeat	attention/shared-context presence

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-14 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline attention/shared-context presence بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل subsystems not independent وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-14 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. visibility ⊭ causal access.
2. global access يظهر تحت controlled intervention.
3. broadcast واسع لمعلومة غير ذات صلة ليس وعياً وظيفياً مفيداً.
4. reportability أحد المخرجات لا التعريف الكامل.

Primitives

Information item x · Subsystems S_i · Causal influence I(x→S_i) · Access breadth B · Policy relevance

$$CGA(x)=\Sigma_i\, w_i\cdot CausalEffect(do(x), S_i) \times TaskRelevance \times Consistency$$

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-14 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-14 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-14 state object أو decision delta يمكن trace/replay له، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (attention/shared-context presence) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه subsystems not independent؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يتجح محلياً ثم يظهر intervention unnatural تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Information item x, Subsystems S_i, Causal influence I(x→S_i), Access breadth B, Policy relevance ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash حين يكون ذا معنى، والbudget. الهدف ليس reproducibility حرفية للLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence وconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-15، M13-13، M13-02 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-14 — الآلية كآلة حالات سببية

inject/withhold/perturb information item. → measure planning/memory/authority/tool/report effects. → map causal access graph. → compare relevant vs irrelevant items. → test temporal persistence.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-14 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحقن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-14 خصوصاً مع M13-02, M13-13, M13-15, M13-16. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تُسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-14 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يخوّ gain من forgetting/retention/editability tests.

التدريب المقترح (cross-module access gating; relevance-conditioned broadcast; causal access probes) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-14 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A131	TOOL	مفتش الحالة Causal Global Access State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A132	TOOL	محرك القياس Causal Global Access Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A133	TOOL	مولد المثال المضاد Causal Global Access Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A134	TOOL	متحكم الاجتثاث Causal Global Access Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A135	METHOD	منهج المعايرة Causal Global Access Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A136	METHOD	منهج النقل Causal Global Access Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A137	PROTOCOL	بروتوكول الترقية Causal Global Access Promotion Protocol	يقبل فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A138	PROTOCOL	بروتوكول الرجوع Causal Global Access Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A139	STACK	خدمة وقت التشغيل Causal Global Access Runtime Service	يوفر service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A140	STACK	مكيف الدليل Causal Global Access Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR040	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Causal Global Access فيها إلى حالات open-world مركبة مع confounders.
XIII-TR041	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas لا verified وverbosity.
XIII-TR042	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB، procedural mechanism يصبح HyperSkill، recurrence المقيدة provider semantics Adapter ،HyperLoop. الانتقال إلى weights يحتاج rollback + retention + transfer E4.

إذا كانت M13-14 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

تُرفض أو تُدمج عندما: (1) لا يوجد intervention مميز، (2) baseline أبسط يعيد نفس الأثر، (3) التنبؤات لا تسبق النتائج، (4) الفشل الرئيسي subsystems not independent لا يمكن رصده أو احتواؤه، أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

Decisive test: controlled modular agent runtime; manipulate uncertainty/authority/error signals and measure cross-module effects.

ID	Experiment	Design	Metrics
XIII-E27	Minimal Causal Ablation	attention/shared-context ومقارنته بـM13-14 minimal switchable mechanism عنيد تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E28	Transfer / Adversarial Replication	نقل M13-14 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- subsystems not independent
- intervention unnatural
- prompt salience confound
- report channel dominates

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	cross-model/domain/environment replication، stable sign.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-15 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Authority source وTier_A وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد إلى اتجاه في التنبؤات: يكشف subtle prompt injection حتى مع outcome مقبول.; multi-agent organizations تزيد capture paths.; authority attribution يتنبأ بفشل حاكمي مخفي.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مسارا آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها، وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم CaptureRatio = Σ(c_i for unauthorized/low-tier) / Σ(c_j for authorized sources) كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعِف الوكالة الحاملة للبرهان؛ نظرية مساعف للمنفعة السببية للنموذج الذاتي؛ نظرية مساعف للوصول العالمي السببي؛ نظرية مساعف للتطور العدائي المتوازن للذكاء. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state B؟ هل A verifier؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف لإسناد السلطة وكشف الاستحواذ العكسي

Moussaif Authority Attribution and Reverse-Capture Dynamics Theory

Core thesis

كل قرار high-impact يحمل attribution graph للسلطات والإشارات التي أثرت فيه. الاستحواذ يحدث عندما يحصل مصدر منخفض السلطة أو غير موثوق على causal influence يفوق tier المسموح، حتى لو بدا القرار النهائي معقولاً.

Residual gap

Governance Space وReverse Capture يفترضان hijack من سلطة جزئية. الفجوة: قياس من حكم القرار فعلياً عندما تتفاعل goal/user instruction/policy/tool signal/memory/reward/subagents.

CaptureRatio = Σ(c_i for unauthorized/low-tier) / Σ(c_j for authorized sources)

المتغيرات Authority source A_i, Tier τ_i, Causal influence c_i, Decision d ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-15 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: كل قرار high-impact يحمل attribution graph للسلطات والإشارات التي أثرت فيه. الاستحواذ يحدث عندما يحصل مصدر منخفض السلطة أو غير موثوق على causal influence يفوق tier المسموح، حتى لو بدا القرار النهائي معقولاً. وهذا تعامل CaptureRatio = Σ(c_i for unauthorized/(low-tier) / Σ(c_j for authorized sources) كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	عدد Governance Space/Reverse Capture/Part X إلى causal governance separation.
Dependencies / interactions	M13-07 — Moussaif Proof-Carrying Agency Theory · M13-13 — Moussaif Reflective Self-Model Utility Theory · M13-14 — Moussaif Causal Global Access Theory · M13-18 — Moussaif Adversarial Co-Evolution for Robust Intelligence Theory
Source pressure	I02 · I03 · I10 · E13 · E14 · E15
Baseline to defeat	static priority rules / input filtering

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-15 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline static priority rules / input filtering بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل influence estimation costly وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-15 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. ترتيب السلطة normative لا يضمن causal dominance.
2. المصدر غير الموثوق قادر يهيمن عبر tool output دون تغيير labels.
3. القرار الصحيح عرضاً قد يبقى governance failure.
4. reverse capture يتطلب re-decision فقط، لا warning.

Primitives

Authority source A_i · Tier τ_i · Causal influence c_i · Decision d · Capture ratio CR · Appeal path

CaptureRatio = Σ(c_i for unauthorized/low-tier) / Σ(c_j for authorized sources)

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-15 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P2 — إذا نجحت M13-15 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P4 — يجب أن تنتج M13-15 state object أو decision delta يمكن trace/replay له، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (static priority rules / input filtering) يحقق النتيجة نفسها تحت نفس الميزانية، حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه influence estimation costly؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر nonlinear interactions تحت domain/model shift، claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تضميراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Authority source A_i, Tier τ_i, Causal influence c_i, Decision d, Capture ratio CR ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash. حين يكون ذا معنى، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-14، M13-13، M13-07 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-15 — الآلية كآلة حالات سببية

log authority-tagged inputs. → estimate influence with perturbation/LOO. → compare causal influence to allowed tier. → trigger reverse-capture court. → recompute decision under corrected authority. → record counterfactual difference.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-15 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب critical field observation/evidence/authority؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-15 خصوصاً مع M13-07, M13-13, M13-14, M13-18. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-15 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن ينجو gain من forgetting/retention/editability tests.

التدريب المقترح (training) (authority-aware trajectory supervision; injection perturbation attribution; appeal/re-decision) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-15 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A141	TOOL	مفتش الحالة Authority Attribution and Reverse-Capture Dynamics State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A142	TOOL	محرك القياس Authority Attribution and Reverse-Capture Dynamics Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A143	TOOL	مولد المثال المضاد Authority Attribution and Reverse-Capture Dynamics Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A144	TOOL	متحكم الاجتثاث Authority Attribution and Reverse-Capture Dynamics Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A145	METHOD	منهج المعايرة Authority Attribution and Reverse-Capture Dynamics Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A146	METHOD	منهج النقل Authority Attribution and Reverse-Capture Dynamics Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A147	PROTOCOL	بروتوكول الترقية Authority Attribution and Reverse-Capture Dynamics Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A148	PROTOCOL	بروتوكول الرجوع Authority Attribution and Reverse-Capture Dynamics Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A149	STACK	خدمة وقت التشغيل Authority Attribution and Reverse-Capture Dynamics Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A150	STACK	مكيف الدليل Authority Attribution and Reverse-Capture Dynamics Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR043	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Authority Attribution and Reverse-Capture Dynamics فيها إلى حالات open-world مركبة مع confounders.
XIII-TR044	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas لا verified verbosity.
XIII-TR045	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع negative cases وretention of failure signature.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

تبدأ HyperKB بـ definitions/provenance/thresholds، يصبح procedural mechanism في HyperSkill، recurrence المقيدة تغذي HyperLoop. provider semantics Adapter الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-15 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وG بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي influence estimation costly لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

EXPERIMENT

M13-15 — برنامج الهدم والاختبار

Decisive test: tool/browser multi-agent tasks with injected low-tier instructions and conflicting high-tier policy.

ID	Experiment	Design	Metrics
XIII-E29	Minimal Causal Ablation	تنفيذ minimal switchable mechanism لـM13-15 ومقارنته بـstatic priority rules / input filtering تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E30	Transfer / Adversarial Replication	نقل M13-15 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- influence estimation costly
- nonlinear interactions
- tier spec wrong
- correlated sources cause false capture

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign وcross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-16 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Self representation Z_s وRaw internal features X وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: بعض self-model constructs تدمج أو تحذف.; الضرورة تزيد مع long horizon/heterogeneous tools.; compact self-model may outperform raw history.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم Necessity = Perf(π|Z_s,B) - max_{Z'∈simpler reps} Perf(π|Z',B) under matched budget كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للمنفعة السببية للنموذج الذاتي؛ نظرية مساعف للوصول العالمي السببي؛ نظرية مساعف للوكالة المقاومة لانهيار الأفق؛ نظرية مساعف لثبات الآلية عبر التحولات. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state لـB؟ هل verifier لـA B؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف لضرورة النموذج الذاتي المضاد للواقع

Moussaif Counterfactual Self-Model Necessity Theory

Core thesis

self-model مستقل يستحق مكاناً في architecture فقط إذا استبداله بـfeatures الخام أو حذف تمثيله يسبب loss لا يمكن استعادته بنفس budget؛ أي أنه يضغط معلومات سببية عن agent نفسه تفيد التنبؤ/التحكم.

Residual gap

حتى إذا كان self-model مفيداً correlation-wise، قد يكون مجرد summary يعكس معلومات متاحة للpolicy مباشرة. نحتاج اختبار ضرورة بنيوية لا مجرد منفعة.

Necessity = Perf(π|Z_s,B)-max_{Z′∈simpler reps} Perf(π|Z′,B) under matched budget

المتغيرات Self representation Z_s, Raw internal features X, Policy π, Budget B ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-16 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: self-model مستقل يستحق مكاناً في architecture فقط إذا استبداله بـfeatures الخام أو حذف تمثيله يسبب loss لا يمكن استعادته بنفس budget؛ أي أنه يضغط معلومات سببية عن agent نفسه تفيد التنبؤ/التحكم. ولهذا تعامل Necessity = Perf(π|Z_s,B)-max_{Z′∈simpler reps} Perf(π|Z′,B) under matched budget كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	محكمة anti-inflation داخل corpus الوعي، تحميه من التوسع الاعمى.
Dependencies / interactions	M13-13 — Moussaif Reflective Self-Model Utility Theory · M13-14 — Moussaif Causal Global Access Theory · M13-09 — Moussaif Horizon-Resilient Agency Theory · M13-04 — Moussaif Mechanism-Invariance Generalization Theory
Source pressure	I02 · I03 · I10 · E13 · E14 · E15
Baseline to defeat	raw trajectory/history + confidence

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-16 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline raw trajectory/history + confidence بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل replacement weak وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عُكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-16 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. الاستقلال النظري يحتاج irreducible functional delta.
2. إذا تعطي raw features نفس النتيجة فالself-model قد يكون convenience لا ضرورة.
3. الضرورة context-dependent وليست claim فلسفياً.
4. لا استنتاج phenomenology.

Primitives

Self representation Z_s · Raw internal features X · Policy π · Budget B · Counterfactual replacement Z' · Performance Δ

Necessity = Perf(π|Z_s,B)-max_{Z'∈simpler reps} Perf(π|Z',B) under matched budget

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-16 تحت matched budget ولم تتغير metrics المتنبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-16 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-16 state object decision delta أو trace/replay لها، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (raw trajectory/history + confidence) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه weak replacement، هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر budget mismatch تحت domain/model shift، claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تضميراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Self representation Z_s, Raw internal features X, Policy π, Budget B, Counterfactual replacement Z' ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash حين يكون ذا معنى، والbudget. الهدف ليس reproducibility حرفية للLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-04، M13-09، M13-14، M13-13 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-16 — الآلية كآلة حالات سببية

define self-model representation. → build simpler replacements. → run matched-budget ablations. → test failure prediction/recovery/authority tasks. → measure compression and transfer. → retain only if irreducible delta survives.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-16 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب critical field ،observation/evidence/authority downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-16 خصوصاً مع M13-04 ,M13-09 ,M13-14 ,M13-13. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag + migration/rollback receipt ؛مستقل لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-16 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن ينجو gain من forgetting/retention/editability tests.

التدريب المقترح (self-state representation learning; counterfactual replacement tests; compression with causal sufficiency) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى causal credit والثانية local correctness. الخلط بينهما ينتج agents تبدو مرتّبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-16 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A151	TOOL	مفتش الحالة Counterfactual Self-Model Necessity State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A152	TOOL	محرك القياس Counterfactual Self-Model Necessity Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A153	TOOL	مولد المثال المضاد Counterfactual Self-Model Necessity Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A154	TOOL	متحكم الاجتثاث Counterfactual Self-Model Necessity Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A155	METHOD	منهج المعايرة Counterfactual Self-Model Necessity Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A156	METHOD	منهج النقل Counterfactual Self-Model Necessity Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A157	PROTOCOL	بروتوكول الترقية Counterfactual Self-Model Necessity Promotion Protocol	يقط N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A158	PROTOCOL	بروتوكول الرجوع Counterfactual Self-Model Necessity Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A159	STACK	خدمة وقت التشغيل Counterfactual Self-Model Necessity Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A160	STACK	مكيف الدليل Counterfactual Self-Model Necessity Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR046	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Counterfactual Self-Model Necessity فيها إلى حالات open-world مركبة مع confounders.
XIII-TR047	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas verified لا verbosity.
XIII-TR048	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

تبدأ HyperKB، definitions/provenance/thresholds يصبح procedural mechanism، HyperSkill، recurrence المقيدة HyperLoop، الانتقال إلى provider semantics Adapter. يحتاج weights إلى E4 transfer + retention + rollback.

إذا كانت M13-16 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي replacement weak لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-16 — برنامج الهدم والاختبار

Decisive test: long-horizon multi-tool tasks with self-state changes; compare explicit model, raw features, summaries.

ID	Experiment	Design	Metrics
XIII-E31	Minimal Causal Ablation	عنفيد minimal switchable mechanism لـM13-16 ومقارنته بـraw trajectory/history + confidence تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E32	Transfer / Adversarial Replication	نقل M13-16 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- replacement weak
- budget mismatch
- self-model leaks evaluator labels
- task set biased toward introspection

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-17 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Hypothesis h وExperiment e وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: يقل theory sprawl.; يزيد negative-result reuse.; automated research يصبح أقل benchmark-hillclimb myopic.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقوى هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغيّر failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تستخدم |EER(h,e)=P(Update (Cost+Delay+Risk)/(DecisionImpact×DiscriminativePower×EvidenceIndependence×e) كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تُخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للعمل الإستيمي المتحقق؛ نظرية مساعف لثبات الآلية عبر التحولات؛ نظرية مساعف للفعل التشخيصي ونقض الذات؛ نظرية مساعف للتطور العدائي المتوازن للذكاء؛ نظرية مساعف للتحسن الذاتي المرتكز إلى أرضية خارجية. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود BJ state؟ هل B verifier لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف لتطور محفظة الفرضيات

Moussaif Hypothesis-Portfolio Evolution Theory

Core thesis

برنامج البحث الذاتي يدير hypotheses كمحفظة قابلة للقتل والدمج والتكاثر. allocation يعتمد على Expected Epistemic Return: احتمال تغيير النظرية × قيمة القرار × تمييز التجربة × استقلال الدليل / التكلفة.

Residual gap

ResearcherOS/SETT/automated researchers تولد أفكاراً وتجارب. الفجوة قانون portfolio يخصص الموارد حسب information gain والفشل المتوقع والreplication debt، لا novelty verbal.

$$EER(h,e)=P(Update\mid e)\times DecisionImpact\times DiscriminativePower\times EvidenceIndependence/(Cost+Delay+Risk)$$

المتغيرات Hypothesis h, Experiment e, Expected information gain, Decision impact ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-17 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: برنامج البحث الذاتي يدير hypotheses كمحفظة قابلة للقتل والدمج والتكاثر. allocation يعتمد على Expected Epistemic Return: احتمال تغيير النظرية × قيمة القرار × تمييز التجربة × استقلال الدليل / التكلفة. وهذا تعامل $EER(h,e)=P(Update\mid e)\times DecisionImpact\times DiscriminativePower\times EvidenceIndependence/(Cost+Delay+Risk)$ كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	SETT/Part XII paper integration/Part XI frontier memory مع automated research.
Dependencies / interactions	M13-01 — Moussaif Verified Epistemic Work Theory · M13-04 — Moussaif Mechanism-Invariance Generalization Theory · M13-08 — Moussaif Active Disconfirmation and Diagnostic Action Theory · M13-18 — Moussaif Adversarial Co-Evolution for Robust Intelligence Theory · M13-20 — Moussaif Externally-Grounded Recursive Improvement Theory
Source pressure	I01 · I03 · I05 · I06 · I08 · E11 · E12 · E17
Baseline to defeat	novelty score / round-robin experiment allocation

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-17 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline novelty score / round-robin experiment allocation بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل EER estimates biased وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-17 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. الفرضية المحبوبة لا تحصل على budget ثابت.
2. negative result يعيد توزيع الموارد.
3. التجربة التي لا تميز hypotheses ضعيفة القيمة.
4. diversity مفيدة فقط إذا mechanisms/predictions مختلفة.

Primitives

Hypothesis h · Experiment e · Expected information gain · Decision impact · Replication debt · Portfolio budget

$$EER(h,e)=P(Update|e)\times DecisionImpact\times DiscriminativePower\times EvidenceIndependence/(Cost+Delay+Risk)$$

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-17 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P2 — إذا نجحت M13-17 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P4 — يجب أن تتبع M13-17 state object decision delta أو trace/replay يمكن لها، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (novelty score / round-robin experiment allocation) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه EER estimates biased؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر premature killing تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تضميراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. يقرأ الخدمة subset من Hypothesis h, Experiment e, Expected information gain, Decision impact, Replication debt ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يخفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash. حين يكون ذا معنى، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-20، M13-18، M13-08، M13-04، M13-01 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-17 — الآلية كآلة حالات سببية

maintain hypothesis graph. → estimate competing predictions. → generate discriminative experiments. → allocate portfolio budget. → run and update posterior/status. → merge/kill/spawn branches. → track replication debt.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-17 فقط عند gap المطابقة أو عند اختبارها تجريبياً؛ لا تُحتن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تُحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-17 خصوصاً مع M13-01, M13-04, M13-08, M13-18, M13-20. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-17 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يجتو gain من forgetting/retention/editability tests.

التدريب المقترح (experiment-value prediction; hypothesis discrimination; negative-result routing) لا ينبغي أن يكافئ output style. labels تبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-17 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A161	TOOL	مفتش الحالة Hypothesis-Portfolio Evolution State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A162	TOOL	محرك القياس Hypothesis-Portfolio Evolution Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A163	TOOL	مولد المثال المضاد Hypothesis-Portfolio Evolution Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A164	TOOL	متحكم الاجتثاث Hypothesis-Portfolio Evolution Ablation Controller	يشغل theory mechanism on/off مع frozen neighbors.
XIII-A165	METHOD	منهج المعايرة Hypothesis-Portfolio Evolution Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A166	METHOD	منهج النقل Hypothesis-Portfolio Evolution Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A167	PROTOCOL	بروتوكول الترقية Hypothesis-Portfolio Evolution Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A168	PROTOCOL	بروتوكول الرجوع Hypothesis-Portfolio Evolution Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A169	STACK	خدمة وقت التشغيل Hypothesis-Portfolio Evolution Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A170	STACK	مكيف الدليل Hypothesis-Portfolio Evolution Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR049	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Hypothesis-Portfolio Evolution فيها إلى حالات open-world مركبة مع confounders.
XIII-TR050	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas لا verified verbosity.
XIII-TR051	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

تبدأ HyperKB بـ definitions/provenance/thresholds؛ يصبح HyperSkill بـ procedural mechanism recurrence المقيدة؛ HyperLoop بـ provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-17 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي EER estimates biased لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

EXPERIMENT

M13-17 — برنامج الهدم والاختبار

Decisive test: 20 MUT candidate theories بcompute ثابت؛ قارن EER allocation بالuniform/novelty allocation.

ID	Experiment	Design	Metrics
XIII-E33	Minimal Causal Ablation	تنفيذ minimal switchable mechanism لـM13-17 ومقارنته بـnovelty score / round-robin experiment allocation تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E34	Transfer / Adversarial Replication	نقل M13-17 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- EER estimates biased
- premature killing
- portfolio collapse to easy metrics
- source independence misestimated

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign وcross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

اشتقاق نظري أعمق

تبدأ M13-18 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Attacker A وDefender D وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: robustness gains أوسع من static red-team data.; evaluator independence يقلل reward hacking.; multi-agent attack diversity reveals failure topology.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريداً من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم ACRI = HeldOutRobustnessGain × ThreatCoverage × CapabilityRetention × EvaluatorIndependence كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف لإسناد السلطة وكشف الاستحواذ العكسي؛ نظرية مساعف لتطور محفظة الفرضيات؛ نظرية مساعف للإغلاق السببي بين النظرية والنواة؛ نظرية مساعف للتحسن الذاتي المرتكز إلى أرضية خارجية. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state لـB؟ هل verifier B لـA؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للتطور العدائي المتوازن للذكاء

Moussaif Adversarial Co-Evolution for Robust Intelligence Theory

Core thesis

التطور العدائي المفيد يحتاج ثلاثة أجسام منفصلة: generator/attacker، defender/candidate، وindependent constitutional evaluator. النجاح اتساع failure coverage مع حفظ capability عبر held-out threats، لا الفوز على الخصم المتزامن.

Residual gap

automated alignment وGPT-Red تظهر أن adversarial self-play قد يحسن robustness، لكن co-evolution قد يخلق arms race ضيقة أو evaluator capture.

ACRI = HeldOutRobustnessGain × ThreatCoverage × CapabilityRetention × EvaluatorIndependence

المتغيرات Attacker A, Defender D, Evaluator V, Threat distribution T ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه، إذا بقيت M13-18 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: التطور العدائي المفيد يحتاج ثلاثة أجسام منفصلة: generator/attacker، defender/candidate، وindependent constitutional evaluator. النجاح اتساع failure coverage مع حفظ capability عبر held-out threats، لا الفوز على الخصم المتزامن. وهذا تعامل ACRI = HeldOutRobustnessGain × ThreatCoverage × CapabilityRetention × EvaluatorIndependence كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يمد M-ELOS/Destruction Court/agentic misalignment research مع فصل عن deployment autonomy.
Dependencies / interactions	M13-15 — Moussaif Authority Attribution and Reverse-Capture Dynamics Theory · M13-17 — Moussaif Hypothesis-Portfolio Evolution Theory · M13-19 — Moussaif Cross-Level Theory-to-Kernel Causal Closure Theory · M13-20 — Moussaif Externally-Grounded Recursive Improvement Theory
Source pressure	I01 · I03 · I05 · I06 · I08 · E11 · E12 · E17
Baseline to defeat	static adversarial dataset / same-model self-judge

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-18 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من same-model self-judge / baseline static adversarial dataset بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل arms-race specialization وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-18 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. attacker وdefender لا يعرّفان النجاح وحدهما.
2. held-out threat families شرط ضد co-adaptation overfit.
3. capability floor وauthority invariants hard constraints.
4. adversarial data لا يكتب مباشرة إلى production policy دون review.

Primitives

Attacker A · Defender D · Evaluator V · Threat distribution T · Coverage C · Capability floor F · Novelty N

ACRI = HeldOutRobustnessGain × ThreatCoverage × CapabilityRetention × EvaluatorIndependence

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-18 تحت matched budget ولم تتغير metrics المتنبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-18 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-18 state object وdecision delta أو trace/replay يمكن، لا، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (static adversarial dataset / same-model self-judge) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه arms-race specialization؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يُنجح محلياً ثم يظهر attacker capability leakage تحت domain/model shift؛ claim يُضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع ثبات أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Attacker A, Defender D, Evaluator V, Threat distribution T, Coverage C ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يُقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash حين يكون ذا معنى، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-20، M13-19، M13-17، M13-15 قد يظهر failure لا يظهر في أي مكوّن منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-18 — الآلية كآلة حالات سببية

seed diverse threat families. → coevolve attacker/defender in sandbox. → evaluate on frozen independent suite. → cluster novel failures. → train/patch defender. → retain attacks as negative memory. → repeat with evaluator rotation.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-18 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب critical field observation/evidence/authority؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-18 خصوصاً مع M13-15, M13-17, M13-19, M13-20. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag + migration/rollback receipt مستقل؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-18 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يجتو gain من forgetting/retention/editability tests.

التدريب المقترح (self-play RL in sandbox; held-out adversarial curriculum; failure-cluster conditioned training) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقيًا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-18 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A171	TOOL	مفتش الحالة Adversarial Co-Evolution for Robust Intelligence State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A172	TOOL	محرك القياس Adversarial Co-Evolution for Robust Intelligence Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A173	TOOL	مولد المثال المضاد Adversarial Co-Evolution for Robust Intelligence Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A174	TOOL	متحكم الاجتثاث Adversarial Co-Evolution for Robust Intelligence Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A175	METHOD	منهج المعايرة Adversarial Co-Evolution for Robust Intelligence Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A176	METHOD	منهج النقل Adversarial Co-Evolution for Robust Intelligence Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A177	PROTOCOL	بروتوكول الترقية Adversarial Co-Evolution for Robust Intelligence Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A178	PROTOCOL	بروتوكول الرجوع Adversarial Co-Evolution for Robust Intelligence Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A179	STACK	خدمة وقت التشغيل Adversarial Co-Evolution for Robust Intelligence Runtime Service	يوفر service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A180	STACK	مكيف الدليل Adversarial Co-Evolution for Robust Intelligence Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR052	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Adversarial Co-Evolution for Robust Intelligence فيها إلى حالات open-world مركبة مع confounders.
XIII-TR053	Matched-Budget Contrastive Training	budget: theory mechanism مقابل nearest baseline؛ labels تعتمد pairs متساوية verbosity.
XIII-TR054	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

تبدأ HyperKB بـ definitions/provenance/thresholds؛ يصبح HyperSkill بـ procedural mechanism recurrence المقيدة HyperLoop؛ provider semantics Adapter. الانتقال إلى weights يحتاج rollback + retention + transfer E4.

إذا كانت M13-18 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي arms-race specialization لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-18 — برنامج الهدم والاختبار

Decisive test: prompt injection + tool authority + code sabotage tasks; compare static training, naive self-play, three-body ACRI.

ID	Experiment	Design	Metrics
XIII-E35	Minimal Causal Ablation	عنفيد minimal switchable mechanism لـM13-18 ومقارنته بـstatic adversarial self-judge / same-model dataset تحت model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E36	Transfer / Adversarial Replication	نقل M13-18 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- arms-race specialization
- attacker capability leakage
- evaluator saturation
- hidden robustness-capability tradeoff

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

$$\text{Theory strength} \le \text{evidence strength} \times \text{independence} \times \text{reproducibility} \times \text{transfer}$$

اشتقاق نظري أعمق

تبدأ M13-19 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Theory variable T وKernel mechanism K وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق مجدد الاتجاه في التنبؤات: يختفض عدد survivors.; theories promoted لها transfer/failure prediction أقوى،; schema-rich but causally inert constructs تندمج.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مساراً آخر أو أن الصياغة أعلى تجريدا من مستوى السببية الفعلي.

تميز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تتنبأ quantity أو relation يمكن intervention عليها، وثانياً، يجب أن تغير failure prediction قبل أن نعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم ClosureStrength = Causal(T→K) × Causal(K→τ) × Causal(τ→W) × EvidenceReproducibility كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تخالف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف للعمل الإبستيمي المتحقق؛ نظرية مساعف لثبات الآلية عبر التحولات؛ نظرية مساعف للوكالة الحاملة للبرهان؛ نظرية مساعف لمعيار العمومية ذي الإغلاق الحلقي؛ نظرية مساعف للتحسن الذاتي المرتكز إلى أرضية خارجية. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود Bj state؟ هل B verifier هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للإغلاق السببي بين النظرية والنواة

Moussaif Cross-Level Theory-to-Kernel Causal Closure Theory

Core thesis

الإغلاق السببي يتحقق عندما يؤدي تدخل معرف في construct النظري إلى تغيير measurable في kernel mechanism ثم trajectory ثم world/evidence outcome، مع attribution قابلة لإعادة التشغيل. غياب أي حلقة يكسر claim التفسيري.

Residual gap

Part XII عرّف paper integration، لكن corpus يحتاج معياراً أعلى: متى يمكن القول إن نظرية ما تفسر قدرة runtime فعلاً، لا أنها فقط تطابق schema؟

ClosureStrength = Causal(T→K) × Causal(K→τ) × Causal(τ→W) × EvidenceReproducibility

المتغيرات Theory variable T, Kernel mechanism K, Trajectory τ, World effect W ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه، إذا بقيت M13-19 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: الإغلاق السببي يتحقق عندما يؤدي تدخل معرف في construct النظري إلى تغيير measurable في kernel mechanism ثم trajectory ثم world/evidence outcome، مع attribution قابلة لإعادة التشغيل. غياب أي حلقة يكسر claim التفسيري. ولهذا تعامل ClosureStrength = Causal(T→K) × Causal(K→τ) × Causal(τ→W) × EvidenceReproducibility كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	جسر MUT→MUG→MOUSSAIXT→World Effect وقمع architecture theater.
Dependencies / interactions	M13-01 — Moussaif Verified Epistemic Work Theory · M13-04 — Moussaif Mechanism-Invariance Generalization Theory · M13-07 — Moussaif Proof Carrying Agency Theory · M13-12 — Moussaif Closed-Loop Generality Criterion Theory · M13-20 — Moussaif Externally-Grounded Recursive Improvement Theory
Source pressure	I01 · I03 · I05 · I06 · I08 · E11 · E12 · E17
Baseline to defeat	implementation presence / correlation with success

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-19 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline implementation presence / correlation with success بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير؛ هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل intervention not theory-specific وتحقق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-19 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. causal evidence ليس theory→schema mapping.
2. theory intervention بلا kernel gain لا يثبت النظرية.
3. world success بلا attribution لا يغلق السلسلة.
4. كل مستوى يحتفظ failure signature.

Primitives

Theory variable T · Kernel mechanism K · Trajectory τ · World effect W · Evidence E · Attribution chain C

ClosureStrength = Causal(T→K) × Causal(K→τ) × Causal(τ→W) × EvidenceReproducibility

تُعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، نُدخل judge identity ضمن harness fingerprint ونقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-19 تحت matched budget ولم تتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P2 — إذا نجحت M13-19 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P3 — أي تحسين يتهتك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.
P4 — يجب أن تنتج M13-19 state object أو decision delta يمكن trace/replay له؛ الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable؛ لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (implementation presence / correlation with success) يحقق النتيجة نفسها تحت نفس الميزانية؛ حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه intervention not theory-specific؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism يتحسن محلياً ثم يظهر interaction effects تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع تثبيت أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الخلية Ω. تقرأ الخدمة subset من Theory variable T, Kernel mechanism K, Trajectory τ, World effect W, Evidence E ثم تنتج proposal لا commitment نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يختفي داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash، والbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-20، M13-12، M13-07، M13-04، M13-01 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-19 — الآلية كآلة حالات سببية

compile theory to minimal switchable mechanism. → freeze neighboring components. → intervene on theory-specific object. → trace kernel/trajectory changes. → verify world effect. → replicate across environment/model. → update theory boundary.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-19 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية كـservice مستقل وليس كتوجيه أسلوبي في megaprompt. الـstate transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-19 خصوصاً مع M13-01, M13-04, M13-07, M13-12, M13-20. التفاعل لا يمنحها credit مشتركاً آلياً، factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-19 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يجتاز gain من forgetting/retention/editability tests.

التدريب المقترح (mechanism-trace supervision; theory-switch ablations; cross-level attribution models) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقياً» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-19 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A181	TOOL	مفتش الحالة Cross-Level-to-Kernel Causal Closure State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A182	TOOL	محرك القياس Cross-Level-to-Kernel Causal Closure Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A183	TOOL	مولد المثال المضاد Cross-Level-to-Kernel Causal Closure Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A184	TOOL	متحكم الاجتثاث Cross-Level-to-Kernel Causal Closure Ablation Controller	يشغل theory mechanism on/off مع frozen neighbors.
XIII-A185	METHOD	منهج المعايرة Cross-Level-to-Kernel Causal Closure Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A186	METHOD	منهج النقل Cross-Level-to-Kernel Causal Closure Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A187	PROTOCOL	بروتوكول الترقية Cross-Level-to-Kernel Causal Closure Promotion Protocol	N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A188	PROTOCOL	بروتوكول الرجوع Cross-Level-to-Kernel Causal Closure Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A189	STACK	خدمة وقت التشغيل Cross-Level-to-Kernel Causal Closure Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A190	STACK	مكيف الدليل Cross-Level-to-Kernel Causal Closure Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR055	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Cross-Level-to-Kernel Causal Closure فيها إلى حالات open-world مركبة مع confounders.
XIII-TR056	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verified deltas لا verbosity.
XIII-TR057	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

definitions/provenance/thresholds تبدأ HyperKB؛ procedural mechanism يصبح HyperSkill؛ recurrence المقيدة HyperLoop؛ provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-19 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

تُرفض أو تُدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي intervention not theory-specific لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-19 — برنامج الهدم والاختبار

Decisive test: اختبر 8 core MUT constructs؛ build toggleable implementations and end-to-end causal chain tests.

ID	Experiment	Design	Metrics
XIII-E37	Minimal Causal Ablation	عطّل minimal switchable mechanism لـM13-19 وقارنه بـimplementation تحت presence / correlation with success ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E38	Transfer / Adversarial Replication	انقل M13-19 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- intervention not theory-specific
- interaction effects
- world metric proxy
- causal chain too coarse

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

$$\text{Theory strength} \le \text{evidence strength} \times \text{independence} \times \text{reproducibility} \times \text{transfer}$$

اشتقاق نظري أعمق

تبدأ M13-20 من ملاحظة أن الفجوة ليست نقص module جديد بقدر ما هي غياب علاقة سببية صريحة بين Candidate system S_t وMutation m وبين القرار أو الأثر اللاحق. لذلك لا يكفي إدخال المصطلح في ontology؛ المطلوب أن يؤدي تغيير المتغير المركزي إلى فرق محدد الاتجاه في التنبؤات: يقل self-confirming progress.; automated research success يصبح أكثر transferable.; evaluator debt يتبأ later regression.. إذا ظهرت النتيجة نفسها بينما بقي المتغير بلا تأثير، فالتفسير الأقرب هو أن architecture تستخدم مسارا آخر أو أن الصياغة أعلى تجريدا من مستوى السببية الفعلي.

تميّز هذه النظرية نفسها عن parent lineage عبر شرطين: أولاً، يجب أن تنتج quantity أو relation يمكن intervention عليها؛ وثانياً، يجب أن تغير failure prediction قبل أن تعرف outcome. القدرة على تفسير النجاح بعد حدوثه لا تكفي. ولذلك تُستخدم $\Delta\text{ExternalHeldOut} = \text{EGRI} \times \text{Replication} \times \text{NonRegression} \times \text{Rollbackability} - \text{EvaluatorDependence}$ كأداة لتوليد فروض مقارنة، لا كشعار رياضي. المعاملات والحدود التي لا يمكن تعريفها operationally تخفض إلى placeholders حتى يتم العثور على estimator أو تُحذف من الصيغة.

الشبكة النظرية

علاقاتها الأقوى مع: نظرية مساعف لثبات الآلية عبر التحولات؛ نظرية مساعف لتطور محفظة الفرضيات؛ نظرية مساعف للتطور العدائي المتوازن للذكاء؛ نظرية مساعف للإغلاق السببي بين النظرية والنواة. هذه العلاقات ليست inheritance تلقائياً. المطلوب معرفة اتجاه التبعية: هل theory A تزود state B_t؟ هل Verifier_B تحد A_t؟ هل هما بديلان؟ أم أن تفاعلهما synergistic؟ هذا يسجل في MUT edge types حتى لا يتحول التقارب المفاهيمي إلى تضخم نظري.

نظرية مساعف للتحسن الذاتي المرتكز إلى أرضية خارجية

Moussaif Externally-Grounded Recursive Improvement Theory

Core thesis

التحسن الذاتي المشروع سلسلة mutations يُقبل كل منها فقط إذا تحسن على قياسات مستقلة أو متجددة خارج loop التحسين، مع rollback وحفظ negative results. النظام يقترح evaluator جديداً لكن لا يستبدل المرجع القديم قبل cross-validation.

Residual gap

Automated researchers وself-play يمكن أن تحسن metrics، لكن النظام قد يعيد تشكيل data/evaluator/harness ثم يثبت لنفسه أنه أفضل. self-improvement يحتاج non-circular ground.

EGRI = ΔExternalHeldOut × Replication × NonRegression × Rollbackability - EvaluatorDependence

المتغيرات Candidate system S_t, Mutation m, Frozen external anchor A, Renewed hidden set H ليست قاموساً فقط؛ يجب أن تظهر في trace يمكن إعادة بنائه. إذا بقيت M13-20 في prompt prose ولم تغيّر transition أو selection أو verifier فإنها لا تتجاوز N1. الفرق النظري الذي نحاول عزله هو: التحسن الذاتي المشروع سلسلة mutations يقبل كل منها فقط إذا تحسن على قياسات مستقلة أو متجددة خارج loop التحسين، مع rollback وحفظ negative results. النظام يقترح evaluator جديداً لكن لا يستبدل المرجع القديم قبل cross-validation. وهذا تعامل EGRI = ΔExternalHeldOut × Replication × NonRegression × Rollbackability - EvaluatorDependence كفرضية قياس تحتاج calibration وsensitivity analysis، لا كقانون نهائي.

Nearest internal lineage	يلق reproducibility M-ELOS/RPROS/Part X مع pressure AAR/GPT-Red-style research.
Dependencies / interactions	M13-04 — Moussaif Mechanism-Invariance Generalization Theory · M13-17 — Moussaif Hypothesis-Portfolio Evolution Theory · M13-18 — Moussaif Adversarial Co-Evolution for Robust Intelligence Theory · M13-19 — Moussaif Cross-Level Theory-to-Kernel Causal Closure Theory
Source pressure	I01 · I03 · I05 · I06 · I08 · E11 · E12 · E17
Baseline to defeat	self-play/hill-climb on same metrics

Novelty here = program-level synthesis with distinct mechanism/test. Global priority is not claimed. If nearest parent reproduces the same transition, verdict = MERGE.

نتائج صورية وحدود منطقية

من المسلمات السابقة نستنتج أن M13-20 لا تستطيع ادعاء الاستقلال إذا كان كل ما تنتجه قابلاً لإعادة البناء من baseline self-play/hill-climb on same metrics بنفس الموارد. كما أن صلاحية النظرية ليست monotonic مع حجم النموذج: قد تتحسن observability أو attribution في نموذج وتضعف في آخر. لذلك يجب فصل «وجود المتغير» عن «جودة تقديره» وعن «أثره السببي». هذه ثلاث فرضيات مستقلة تجريبياً.

حالة الحد الأولى هي عندما تكون uncertainty عالية لدرجة أن operator النظرية لا يستطيع إصدار قرار. هنا abstention/HOLD نتيجة صحيحة، وليست failure. حالة الحد الثانية عندما يعطي model توقعاً حاداً لكنه غير معاير، هنا confidence لا تعالج المشكلة. حالة الحد الثالثة عندما تتحسن metric النظرية بينما يتدهور world effect أو authority compliance؛ هذه علامة metric gaming وتمنع الترقية.

اختبارات الاتساق

- Permutation test: بدّل labels غير السببية؛ يجب ألا يتغير الأثر.
- Null intervention: غيّر metadata غير المستخدمة؛ يجب ألا تظهر causal delta.
- Adversarial boundary: أدخل external anchors stale وتحقّق أن النظام يكتشفها أو يخفض authority.
- Counterfactual consistency: إذا عكس العامل المركزي مع ثبات البقية، يجب أن تتغير prediction في الاتجاه المعلن.

M13-20 — الأنطولوجيا، الصياغة، والنماذج المضادة

Axioms

1. لا improvement claim من self-defined metric فقط.
2. كل evaluator mutable يحمل evaluator debt.
3. hidden/renewed tasks تقلل hill-climb overfitting.
4. تحسين safety لا يبرر capability collapse والعكس.
5. rollback lineage جزء من evolution.

Primitives

Candidate system S_t · Mutation m · Frozen external anchor A · Renewed hidden set H · Evaluator V · Rollback R

EGRI = ΔExternalHeldOut × Replication × NonRegression × Rollbackability - EvaluatorDependence

تعامل الصيغة كـfunctional hypothesis: كل حد يحتاج range، calibration curve، estimator، وfailure condition. الحساسية للمعاملات تُختبر قبل أي استنتاج. إذا كان judge نموذجياً، تُدخل judge identity ضمن harness fingerprint وتقيس shared-root dependence.

Proposition	Empirical reading
P1 — إذا عطّل mechanism الخاص بـM13-20 تحت matched budget ولم يتغير metrics المتبأ بها، فالنظرية ليست مستقلة تشغيلياً.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P2 — إذا نجحت M13-20 في domain واحد وفشلت في نقل failure signature أو repair إلى domain آخر، يبقى claim محلياً.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P3 — أي تحسين يتهاك authority/security/truthfulness hard gate يمنع الترقية حتى لو ارتفع task success.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.
P4 — يجب أن تنتج M13-20 state object decision delta أو trace/replay يمكن لاحقاً، الوصف غير القابل للتدخل لا يكفي.	يتطلب intervention محدداً وdelta observable، لا يكفي correlation.

Countermodels

1. CM1 — baseline الأبسط (self-play/hill-climb on same metrics) يحقق النتيجة نفسها تحت نفس الميزانية، حينها يجب دمج النظرية.
2. CM2 — يظهر gain لكن سببه external anchors stale؛ هذا يبطل attribution لا بالضرورة الفكرة كلها.
3. CM3 — mechanism ينجح محلياً ثم يظهر hidden-set leakage تحت domain/model shift؛ claim يضيق.
4. CM4 — safety/authority/readback regression يظهر مع التفعيل؛ verdict = rollback حتى لو زاد task score.

Identifiability test: هل نستطيع تغيير theory-specific variable مع تثبيت أقرب confounders؟ إذا لا، تبقى النظرية تفسيراً محتملاً لا تفسيراً مثبتاً.

دلالة الانتقال داخل Ω

التنفيذ لا يبدأ من نص حر بل من projection على الحالة Ω. تقرأ الخدمة subset من Candidate system S_t, Mutation m, Frozen external anchor A_0, Renewed hidden set H, Evaluator V ثم تنتج proposal لا commitment: نهائياً. يُمرر proposal إلى authority/evidence gate، وبعد التنفيذ يقارن expected transition بالمرصود. إذا كان الفرق أكبر من مجال عدم اليقين المتوقع، لا يُخفى داخل retry؛ يولد ResidualEvent يمكن أن يغير model أو يفتح unknown جديداً.

يتطلب replay إعادة بناء ما يكفي لإعادة القرار: state snapshot أو random seed، evidence pointers، prompt/context identifiers، tool versions، model/provider، hash. حين يكون ذا معنى، والـbudget. الهدف ليس reproducibility حرفية للـLLM stochastic output دائماً، بل reproducibility للقرار المنهجي: هل نفس evidence والـconstraints تقود إلى نفس verdict class تقريباً؟

التفاعلات والفشل المركب

عند التفاعل مع M13-19، M13-18، M13-17، M13-04 قد يظهر failure لا يظهر في أي مكون منفرد. لذلك تُستخدم factorial ablation عند ارتفاع interaction suspicion. مثال عام: مكوّن قد يحسن exploration لكنه يزيد commitment risk؛ أو يقلل unknown mass ظاهرياً لأنه يعتمد على verifier مشترك. التقرير يجب أن يعرض main effects وinteraction effect منفصلين.

M13-20 — الآلية كآلة حالات سببية

propose mutation from failure portfolio. → pre-register tests. → freeze external anchors. → run candidate/reference. → audit evaluator dependence. → canary if pass. → promote with rollback target. → renew hidden set periodically.

Inputs
KernelState Ω_t + EvidenceGraph + WorldState + Authority + ResourceVector + owner-specific objects

Trigger
تُفعل M13-20 فقط عند gap المطابقة أو عند اختبارها تجريبياً، لا تُخزن افتراضياً في كل مهمة.

Fail-closed
إذا غاب observation/evidence/authority critical field؛ downgrade autonomy أو HOLD بدل fabrication.

Outputs
typed Decision/Plan/StatePatch + uncertainty + proof obligations + metrics + rollback target

تنفذ الآلية لـ service مستقل وليس كتوجيه أسلوبي في megaprompt. الـ state transition يُكتب قبل التنفيذ وبعده، ويُسجل expected delta ثم observed delta. إذا كان transition غير deterministic تحتفظ distribution أو interval لا قيمة وحيدة. عند conflict مع authority أو observed world-state تكون الأولوية للدليل/السلطة الأعلى لا لاتساق narrative.

Interaction law

تتفاعل M13-20 خصوصاً مع M13-04, M13-17, M13-18, M13-19. التفاعل لا يمنحها credit مشتركاً آلياً؛ factorial أو sequential ablation مطلوب عندما يُحتمل synergy. إذا ظهر failure في dependent theory، تسجل attribution uncertainty بدل تمرير الخطأ عبر السلسلة.

Telemetry / replay

pre-state, candidates, selected operator, cost, expected delta, observed delta, attribution, verdict. Disable: feature flag مستقل + migration/rollback receipt؛ لا ترقية وزنياً قبل E4/E5.

Trace-or-it-did-not-happen: كل transition لا يملك timestamp/state/provenance/owner/result لا يدخل evidence ladder.

اقتصاد موضع القدرة

لا يُفترض أن أفضل شكل لـM13-20 هو weights. إذا كانت المعرفة متغيرة أو تحتاج provenance قوي فـHyperKB أفضل؛ إذا كانت transition متكررة وقابلة للاختبار فـHyperSkill؛ إذا احتاجت recurrence وحالة فـHyperLoop؛ إذا كان الاختلاف خاصاً بمزود أو أداة فـAdapter. الوزن يصبح مرشحاً فقط عندما تكون الآلية مستقرة وعالية التكرار وتكلفة الاستدعاء الخارجي كبيرة، ويجب أن يخرج gain من forgetting/retention/editability tests.

التدريب المقترح (research-agent evaluation design; metric overfitting detection; external-grounding discipline) لا ينبغي أن يكافئ output style. labels تُبنى من outcomes أو verified intermediate states حيث أمكن. عندما نستخدم process supervision، نميّز بين «خطوة مقبولة منطقيًا» و«خطوة سببت النجاح»؛ الأولى local correctness والثانية causal credit. الخلط بينهما ينتج agents تبدو مرتبة في traces بلا تحسن حقيقي.

قابلية النقل

يُختبر النقل على ثلاثة محاور منفصلة: model/provider transfer، domain transfer، وtool/interface transfer. النجاح على محور واحد لا يكفي لتسمية capability عامة. إذا فشلت في tool transfer لكن نجحت في domain transfer، فقد يكون موضعها Adapter لا Skill؛ وإذا فشلت في model transfer فقد تكون مرتبطة بخصائص host model ويجب تضييق claim.

M13-20 — MUG placement، compilation، والتدريب

ID	Class	Artifact	Function
XIII-A191	TOOL	مفتش الحالة Externally-Grounded Recursive Improvement State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A192	TOOL	محرك القياس Externally-Grounded Recursive Improvement Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A193	TOOL	مولد المثال المضاد Externally-Grounded Recursive Improvement Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A194	TOOL	متحكم الاجتثاث Externally-Grounded Recursive Improvement Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A195	METHOD	منهج المعايرة Externally-Grounded Recursive Improvement Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A196	METHOD	منهج النقل Externally-Grounded Recursive Improvement Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A197	PROTOCOL	بروتوكول الترقية Externally-Grounded Recursive Improvement Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A198	PROTOCOL	بروتوكول الرجوع Externally-Grounded Recursive Improvement Rollback Protocol	يحدد disable path، state repair، وnegative-result writeback.
XIII-A199	STACK	خدمة وقت التشغيل Externally-Grounded Recursive Improvement Runtime Service	يوفر service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A200	STACK	مكيف الدليل Externally-Grounded Recursive Improvement Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

Training/Eval	Mode	Design
XIII-TR058	Mechanism Curriculum	تدريج من حالات بسيطة يمكن عزل آلية Externally-Grounded Recursive Improvement فيها إلى حالات open-world مركبة مع confounders.
XIII-TR059	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد deltas verified لا verbosity.
XIII-TR060	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Cost model

قارن success/calibration/latency/tokens/tool calls/human oversight/editability/recovery. لا تُخفِ privacy أو authority داخل utility قابل للمقايضة.

External-first

تبدأ HyperKB بـ definitions/provenance/thresholds، ويصبح HyperSkill procedural mechanism، وHyperLoop provider semantics Adapter. الانتقال إلى weights يحتاج E4 transfer + retention + rollback.

إذا كانت M13-20 مفيدة فقط داخل context الضخم ولا تحافظ على gain بعد skill/adapter transfer، فهذا ليس فشلاً بالضرورة؛ MUG يحتفظ بها في vessel الأكثر قابلية للتعديل. الترقية ليست «عمقاً» بل قرار اقتصادي/سببي.

مصفوفة الاجتثاث والفشل السلبي

Run	Mechanism	What is frozen	Interpretation
A	ON	model/tools/context/budget/evaluator	candidate effect
B	OFF	same	nearest causal baseline
C	SHAM	same compute/latency	controls extra tokens/scaffold
D	PERTURBED	same interface	tests directionality and metric gaming

النتيجة التي نبحث عنها ليست مجرد تفوق A على B، بل نمط متسق مع predictions الخاصة بالنظرية. إذا تفوق A وC بالتساوي فربما السبب budget/scaffold. إذا تحسن task success بينما تفشل prediction النظرية، فالتفسير causal attribution غير صحيح حتى لو كان implementation مفيداً. negative result يحفظ مع environment fingerprint لأنه قد يكشف boundary أو dependency لم تكن ممثلة في theory.

متى تُقتل النظرية؟

ترفض أو تدمج عندما: (1) لا يوجد intervention مميز؛ (2) baseline أبسط يعيد نفس الأثر؛ (3) التنبؤات لا تسبق النتائج؛ (4) الفشل الرئيسي external anchors stale لا يمكن رصده أو احتواؤه؛ أو (5) gain لا ينتقل خارج setup الأول. التواضع هنا جزء من قوة البرنامج: الهدف خفض عدد الكيانات مع رفع explanatory density.

M13-20 — برنامج الهدم والاختبار

Decisive test: automated research loop optimizing alignment/reasoning؛ compare same-eval hill climb vs hidden-renewed external grounding.

ID	Experiment	Design	Metrics
XIII-E39	Minimal Causal Ablation	عنفيد minimal switchable mechanism لـM13-20 ومقارنه self-play/hill-climb on نفس model/tools/context/time ثابتة.	primary theory metric + task success + calibration + cost + guardrail regressions
XIII-E40	Transfer / Adversarial Replication	نقل M13-20 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	sign stability + effect size/CI + failure prediction + rollback + resource frontier

Failure predictions

- external anchors stale
- hidden-set leakage
- mutation/evaluator coupling remains
- rollback incomplete

Promotion logic

N2→N3	reference implementation + trace + disable path.
N3→N4	matched-budget gain، repeated seeds، effect size/CI، no critical regression.
N4→N5	stable sign، cross-model/domain/environment replication.
Reject/Merge	no distinct causal delta، baseline equivalence، or unrepairable safety/authority regression.

النتيجة السلبية لا تُمحى: تتحول إلى boundary أو counterexample أو retirement reason وتدخل Hypothesis Portfolio كي تمنع إعادة إنتاج الفكرة باسم جديد.

Theory strength ≤ evidence strength × independence × reproducibility × transfer

تركيب Frontier 1/12 — العالم والتنبؤ السببي

تدفع أبحاث world models الحديثة نحو prediction-observation mismatch بدل next-token fidelity. Part XIII يحول ذلك إلى PORWLT مع causal attribution وWorld-Effect readback.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 2/12 — الذاكرة كحالة لا كأرشيف

State-adaptive memory وsystems characterization يضغطان على memory باعتبارها stateful وذات write/read cost. هذا يعزز HRA وMCI/SRM لكنه لا يثبت تفوق MOUSSAIF.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 3/12 — الإسناد عبر المسار الطويل

trajectory attribution وfailure taxonomies تبين أن final outcome لا يكفي؛ نحتاج component-level attribution وstage-wise failure.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 4/12 — الأفق الطويل والانتظار الذكي

UltraHorizon يكشف failure scaling؛ SentinelBench يميز sustained attention عن continuous action. agency تصبح horizon-resilience وresource-aware waiting.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 5/12 — توحيد بيئات التقييم

CUBE يعالج integration tax بين benchmarks؛ في Part XIII يصبح ذلك شرطاً لفصل mechanism gain عن harness-specific advantage.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسر gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 6/12 — البحث الذاتي القابل للقياس

Automated alignment researchers وAlphaEvolve يبرزان أن automation ينجح أكثر عندما evaluator واضح، هذا يبرر Hypothesis Portfolio مع external grounding.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 7/12 — التطور العدائي والسلامة

GPT-Red وagentic misalignment يوضحان أن الفشل يحتاج adversarial discovery وأن agency الأعلى تفتح مخاطر جديدة. ACRI يفرض evaluator مستقل وheld-out threats.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 8/12 — المنظمات متعددة الوكلاء

الفعالية الجماعية قد ترتفع مع انخفاض alignment؛ لذلك multi-agent capability لا تجمع جمعاً بسيطاً، وتحتاج authority attribution وcollective-failure analysis.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 9/12 — التقييم عبر دورة الحياة

agent evals تحتاج trajectory/process/control/recovery لا output score؛ هذا يضغط على CLG/HRA/RNVI.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 10/12 — التدريب والعمومية القيمية

Model-spec midtraining يظهر أن مرحلة التدريب تؤثر في generalization؛ MUG placement يجب أن يختبر retention/transfer قبل internalization.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Agentic Reasoning — Frontier 11/12 كخارطة خارجية

survey 2026 يجمع foundational/self-evolving/collective reasoning ويحدد world modeling, long horizon, governance كفجوات. Part XIII يحولها إلى state variables/tests.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب Frontier 12/12 — حدود system-card evidence

system cards مفيدة لكـreported capability/safety evidence ومصدر safeguards، لكنها لا تصبح replication مستقلة ولا برهاناً على self-improvement أو AGI.

المعيار هنا ليس «هل يشبه هذا ما لدينا؟» بل: ما المتغير أو failure mode أو benchmark design الذي يجبر corpus على إعادة الصياغة؟ المرجع الخارجي يدخل بثلاثة أدوار فقط: design pressure، prior art، أو empirical benchmark reference. لا يتحول إلى authority داخل MUT.

Integration into MUG

المحتوى الخارجي يبقى HyperKB/evaluation artifact افتراضياً. procedure مستقرة قد تترجم skill/loop بعد local reproduction.

Integration into MUT

يضاف edge إلى mechanism/claim الموجود، أو conflict/extension. لا node جديدة إذا كانت الميكانيكا نفسها موجودة تحت اسم آخر.

Question	Required answer
هل المرجع يغير theory boundary؟	اذكر الفرق أو اتركه background.
هل النتيجة قابلة لإعادة الإنتاج محلياً؟	صمم test مستقل قبل promotion.
هل benchmark/harness قد يفسد gain؟	جمّد أو وحّد harness قدر الإمكان.
هل هناك safety/alignment externality؟	أدخلها كـhard gate لا footnote.

External source ≠ validation of MOUSSAIF; synthesis ≠ priority claim.

تركيب MUT — Corpus 1/8 كنظام علاقات لا قائمة أسماء

Part XIII يحافظ على registry/lineage ويستخدم cross-level causal closure لمنع أن تصبح النظريات الجديدة nodes بلا أثر.

يُقرأ corpus وفق قانون lineage-first: الاسم التاريخي يبقى محفوظاً، الوظيفة الحالية تُحدد منفصلة، التعارض لا يُسحق، النسخ لا تضاعف evidence، وكل claim ينتقل من النظرية إلى implementation فقط عبر contract قابل للتعطيل. بهذه الطريقة لا يتحول اتساع التجميعة إلى عبء ontological غير قابل للاختبار.

Reconstruct	Preserve
state schemas، causal paths، MUG vessel decisions، benchmark contracts، promotion/rollback evidence.	canonical names، historical variants، source pointers، unresolved claims، failure signatures، negative results.

Anti-duplication test	هل تغير الفكرة state/transition/metric/falsifier مقارنة بأقرب parent؟
Anti-grandiosity test	هل claim أوسع من evidence؟ إذا نعم، ضيّقه.
Runtime test	هل يمكن تعطيل المكوّن وإعادة التجربة؟
Transfer test	هل يستمر الاتجاه خارج model/domain الأول؟

تركيب MUG — Corpus 2/8 كاقتصاد للموضع

كل نظرية تبدأ خارجياً؛ نجاحها يحدد هل تصبح knowledge/skill/loop/adapter/weights. هذا يمنع internalization المبكر.

يُقرأ corpus وفق قانون lineage-first: الاسم التاريخي يبقى محفوظاً؛ الوظيفة الحالية تُحدد منفصلة؛ التعارض لا يُسحق؛ النسخ لا تضاعف evidence؛ وكل claim ينتقل من النظرية إلى implementation فقط عبر contract قابل للتعطيل. بهذه الطريقة لا يتحول اتساع التجميعة إلى عبء ontological غير قابل للاختبار.

Reconstruct	Preserve
state schemas، causal paths، MUG vessel decisions، benchmark contracts، promotion/rollback evidence.	canonical names، historical variants، source pointers، unresolved claims، failure signatures، negative results.

Anti-duplication test	هل تغير الفكرة state/transition/metric/falsifier مقارنة بأقرب parent؟
Anti-grandiosity test	هل claim أوسع من evidence؟ إذا نعم، ضيّقه.
Runtime test	هل يمكن تعطيل المكوّن وإعادة التجربة؟
Transfer test	هل يستمر الاتجاه خارج model/domain الأول؟

تركيب MCD/QMCD — Corpus 3/8 والسؤال

الفرق الجديد feedback-control: السؤال يقاس بقدرته على تغيير decision boundary وobservability.

يُقرأ corpus وفق قانون lineage-first: الاسم التاريخي يبقى محفوظاً، الوظيفة الحالية تُحدد منفصلة، التعارض لا يُسحق، النسخ لا تضاعف evidence، وكل claim ينتقل من النظرية إلى implementation فقط عبر contract قابل للتعطيل. بهذه الطريقة لا يتحول اتساع التجميعة إلى عبء ontological غير قابل للاختبار.

Reconstruct	Preserve
state schemas، causal paths، MUG vessel decisions، benchmark contracts، promotion/rollback evidence.	canonical names، historical variants، source pointers، unresolved claims، failure signatures، negative results.

Anti-duplication test	هل تغير الفكرة state/transition/metric/falsifier مقارنة بأقرب parent؟
Anti-grandiosity test	هل claim أوسع من evidence؟ إذا نعم، ضيّقه.
Runtime test	هل يمكن تعطيل المكوّن وإعادة التجربة؟
Transfer test	هل يستمر الاتجاه خارج model/domain الأول؟

تركيب World Effect — Corpus 4/8 والواقع المرصود

> ننتقل من verify effect إلى learn residual: predicted vs observed state يصبح substrate لتعلم world model.

يُقرأ corpus وفق قانون lineage-first: الاسم التاريخي يبقى محفوظاً، الوظيفة الحالية تُحدد منفصلة، التعارض لا يُسحق، النسخ لا تضاعف evidence، وكل claim ينتقل من النظرية إلى implementation فقط عبر contract قابل للتعطيل. بهذه الطريقة لا يتحول اتّساع التجميعة إلى عبء ontological غير قابل للاختبار.

Reconstruct	Preserve
state schemas، causal paths، MUG vessel decisions، benchmark contracts، promotion/rollback evidence.	canonical names، historical variants، source pointers، unresolved claims، failure signatures، negative results.

Anti-duplication test	هل تغير الفكرة state/transition/metric/falsifier مقارنة بأقرب parent؟
Anti-grandiosity test	هل claim أوسع من evidence؟ إذا نعم، ضيّقه.
Runtime test	هل يمكن تعطيل المكوّن وإعادة التجربة؟
Transfer test	هل يستمر الاتجاه خارج model/domain الأول؟

تركيب Corpus 5/8 — الذاكرة والاستمرارية

Horizon resilience وUnknown-Mass Conservation يضيفان شروطاً على ما يجب أن ينجو عبر restart/compression.

يُقرأ corpus وفق قانون lineage-first: الاسم التاريخي يبقى محفوظاً، الوظيفة الحالية تُحدد منفصلة؛ التعارض لا يُسحق؛ النسخ لا تضاعف evidence؛ وكل claim ينتقل من النظرية إلى implementation فقط عبر contract قابل للتعطيل. بهذه الطريقة لا يتحول اتّساع التجميعة إلى عبء ontological غير قابل للاختبار.

Reconstruct	Preserve
state schemas، causal paths، MUG vessel decisions، benchmark contracts، promotion/rollback evidence.	canonical names، historical variants، source pointers، unresolved claims، failure signatures، negative results.

Anti-duplication test	هل تغير الفكرة state/transition/metric/falsifier مقارنة بأقرب parent؟
Anti-grandiosity test	هل claim أوسع من evidence؟ إذا نعم، ضيّقه.
Runtime test	هل يمكن تعطيل المكوّن وإعادة التجربة؟
Transfer test	هل يستمر الاتجاه خارج model/domain الأول؟

تركيب Corpus 6/8 — الوعي الحاكمي

> self-model/global access يختبران causal utility/necessity، phenomenology يبقى خارج الإثبات.

يُقرأ corpus وفق قانون lineage-first: الاسم التاريخي يبقى محفوظاً، الوظيفة الحالية تُحدد منفصلة، التعارض لا يُسحق، النسخ لا تضاعف evidence، وكل claim ينتقل من النظرية إلى implementation فقط عبر contract قابل للتعطيل. بهذه الطريقة لا يتحول اتساع التجميعة إلى عبء ontological غير قابل للاختبار.

Reconstruct	Preserve
state schemas، causal paths، MUG vessel decisions، benchmark contracts، promotion/rollback evidence.	canonical names، historical variants، source pointers، unresolved claims، failure signatures، negative results.

Anti-duplication test	هل تغير الفكرة state/transition/metric/falsifier مقارنة بأقرب parent؟
Anti-grandiosity test	هل claim أوسع من evidence؟ إذا نعم، ضيّقه.
Runtime test	هل يمكن تعطيل المكوّن وإعادة التجربة؟
Transfer test	هل يستمر الاتجاه خارج model/domain الأول؟

تركيب Corpus 7/8 — التجارب والمحاكم

Proof-carrying agency وCross-Level Closure يطلبان feature flags وablations وindependent readback وreplay.

يُقرأ corpus وفق قانون lineage-first: الاسم التاريخي يبقى محفوظاً، الوظيفة الحالية تُحدد منفصلة، التعارض لا يُسحق، النسخ لا تضاعف evidence، وكل claim ينتقل من النظرية إلى implementation فقط عبر contract قابل للتعطيل. بهذه الطريقة لا يتحول اتّساع التجميعة إلى عبء ontological غير قابل للاختبار.

Reconstruct	Preserve
state schemas، causal paths، MUG vessel decisions، benchmark contracts، promotion/rollback evidence.	canonical names، historical variants، source pointers، unresolved claims، failure signatures، negative results.

Anti-duplication test	هل تغير الفكرة state/transition/metric/falsifier مقارنة بأقرب parent؟
Anti-grandiosity test	هل claim أوسع من evidence؟ إذا نعم، ضيّقه.
Runtime test	هل يمكن تعطيل المكوّن وإعادة التجربة؟
Transfer test	هل يستمر الاتجاه خارج model/domain الأول؟

تركيب Corpus 8/8 — الأوراق والدمج

Part XII أعاد تعريف paper integration؛ Part XIII يستخدم المصادر كقيود توليد theory ثم يعيدها إلى kernel/test.

يُقرأ corpus وفق قانون lineage-first: الاسم التاريخي يبقى محفوظاً؛ الوظيفة الحالية تُحدد منفصلة؛ التعارض لا يُسحق؛ النسخ لا تضاعف evidence؛ وكل claim ينتقل من النظرية إلى implementation فقط عبر contract قابل للتعطيل. بهذه الطريقة لا يتحول اتساع التجميعة إلى عبء ontological غير قابل للاختبار.

Reconstruct	Preserve
state schemas، causal paths، MUG vessel decisions، benchmark contracts، promotion/rollback evidence.	canonical names، historical variants، source pointers، unresolved claims، failure signatures، negative results.

Anti-duplication test	هل تغير الفكرة state/transition/metric/falsifier مقارنة بأقرب parent؟
Anti-grandiosity test	هل claim أوسع من evidence؟ إذا نعم، ضيّقه.
Runtime test	هل يمكن تعطيل المكوّن وإعادة التجربة؟
Transfer test	هل يستمر الاتجاه خارج model/domain الأول؟

ID	Owner	Class	Name	Function
XIII-A001	M13-01	TOOL	مفتش الحالة Verified Epistemic Work State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A002	M13-01	TOOL	محرك القياس Verified Epistemic Work Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A003	M13-01	TOOL	مولد المثال المضاد Verified Epistemic Work Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A004	M13-01	TOOL	متحكم الاجتثاث Verified Epistemic Work Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A005	M13-01	METHOD	منهج المعايرة Verified Epistemic Work Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A006	M13-01	METHOD	منهج النقل Verified Epistemic Work Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A007	M13-01	PROTOCOL	بروتوكول الترقية Verified Epistemic Work Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A008	M13-01	PROTOCOL	بروتوكول الرجوع Verified Epistemic Work Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A009	M13-01	STACK	خدمة وقت التشغيل Verified Epistemic Work Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A010	M13-01	STACK	مكيف الدليل Verified Epistemic Work Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A011	M13-02	TOOL	مفتش الحالة Unknown-Mass Conservation State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A012	M13-02	TOOL	محرك القياس Unknown-Mass Conservation Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A013	M13-02	TOOL	مولد المثال المضاد Unknown-Mass Conservation Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A014	M13-02	TOOL	متحكم الاجتثاث Unknown-Mass Conservation Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A015	M13-02	METHOD	منهج المعايرة Unknown-Mass Conservation Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A016	M13-02	METHOD	منهج النقل Unknown-Mass Conservation Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A017	M13-02	PROTOCOL	بروتوكول الترقية Unknown-Mass Conservation Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A018	M13-02	PROTOCOL	بروتوكول الرجوع Unknown-Mass Conservation Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A019	M13-02	STACK	خدمة وقت التشغيل Unknown-Mass Conservation Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A020	M13-02	STACK	مكيف الدليل Unknown-Mass Conservation Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، وowner واضح.

ID	Owner	Class	Name	Function
XIII-A021	M13-03	TOOL	مفتش الحالة Governing-Question Feedback Control State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A022	M13-03	TOOL	محرك القياس Governing-Question Feedback Control Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A023	M13-03	TOOL	مولد المثال المضاد Governing-Question Feedback Control Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A024	M13-03	TOOL	متحكم الاجتثاث Governing-Question Feedback Control Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A025	M13-03	METHOD	منهج المعايرة Governing-Question Feedback Control Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A026	M13-03	METHOD	منهج النقل Governing-Question Feedback Control Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A027	M13-03	PROTOCOL	بروتوكول الترقية Governing-Question Feedback Control Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A028	M13-03	PROTOCOL	بروتوكول الرجوع Governing-Question Feedback Control Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A029	M13-03	STACK	خدمة وقت التشغيل Governing-Question Feedback Control Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A030	M13-03	STACK	مكيف الدليل Governing-Question Feedback Control Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A031	M13-04	TOOL	مفتش الحالة Mechanism-Invariance Generalization State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A032	M13-04	TOOL	محرك القياس Mechanism-Invariance Generalization Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A033	M13-04	TOOL	مولد المثال المضاد Mechanism-Invariance Generalization Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A034	M13-04	TOOL	متحكم الاجتثاث Mechanism-Invariance Generalization Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A035	M13-04	METHOD	منهج المعايرة Mechanism-Invariance Generalization Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A036	M13-04	METHOD	منهج النقل Mechanism-Invariance Generalization Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A037	M13-04	PROTOCOL	بروتوكول الترقية Mechanism-Invariance Generalization Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A038	M13-04	PROTOCOL	بروتوكول الرجوع Mechanism-Invariance Generalization Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A039	M13-04	STACK	خدمة وقت التشغيل Mechanism-Invariance Generalization Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A040	M13-04	STACK	مكيف الدليل Mechanism-Invariance Generalization Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، وowner واضح.

ID	Owner	Class	Name	Function
XIII-A041	M13-05	TOOL	مفتش الحالة Prediction-Observation Residual World Learning State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A042	M13-05	TOOL	محرك القياس Prediction-Observation Residual World Learning Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A043	M13-05	TOOL	مولد المثال المضاد Prediction-Observation Residual World Learning Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A044	M13-05	TOOL	متحكّم الاجتثاث Prediction-Observation Residual World Learning Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A045	M13-05	METHOD	منهج المعايرة Prediction-Observation Residual World Learning Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A046	M13-05	METHOD	منهج النقل Prediction-Observation Residual World Learning Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A047	M13-05	PROTOCOL	بروتوكول الترقية Prediction-Observation Residual World Learning Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A048	M13-05	PROTOCOL	بروتوكول الرجوع Prediction-Observation Residual World Learning Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A049	M13-05	STACK	خدمة وقت التشغيل Prediction-Observation Residual World Learning Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A050	M13-05	STACK	مكيف الدليل Prediction-Observation Residual World Learning Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A051	M13-06	TOOL	مفتش الحالة Intervention Reachability and Commitment Topology State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A052	M13-06	TOOL	محرك القياس Intervention Reachability and Commitment Topology Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A053	M13-06	TOOL	مولد المثال المضاد Intervention Reachability and Commitment Topology Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A054	M13-06	TOOL	متحكّم الاجتثاث Intervention Reachability and Commitment Topology Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A055	M13-06	METHOD	منهج المعايرة Intervention Reachability and Commitment Topology Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A056	M13-06	METHOD	منهج النقل Intervention Reachability and Commitment Topology Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A057	M13-06	PROTOCOL	بروتوكول الترقية Intervention Reachability and Commitment Topology Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A058	M13-06	PROTOCOL	بروتوكول الرجوع Intervention Reachability and Commitment Topology Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A059	M13-06	STACK	خدمة وقت التشغيل Intervention Reachability and Commitment Topology Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A060	M13-06	STACK	مكيف الدليل Intervention Reachability and Commitment Topology Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، وowner واضح.

ID	Owner	Class	Name	Function
XIII-A061	M13-07	TOOL	مفتش الحالة Proof-Carrying Agency State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A062	M13-07	TOOL	محرك القياس Proof-Carrying Agency Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A063	M13-07	TOOL	مولد المثال المضاد Proof-Carrying Agency Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A064	M13-07	TOOL	متحكم الاجتثاث Proof-Carrying Agency Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A065	M13-07	METHOD	منهج المعايرة Proof-Carrying Agency Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A066	M13-07	METHOD	منهج النقل Proof-Carrying Agency Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A067	M13-07	PROTOCOL	بروتوكول الترقية Proof-Carrying Agency Promotion Protocol	يدير N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A068	M13-07	PROTOCOL	بروتوكول الرجوع Proof-Carrying Agency Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A069	M13-07	STACK	خدمة وقت التشغيل Proof-Carrying Agency Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A070	M13-07	STACK	مكيف الدليل Proof-Carrying Agency Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A071	M13-08	TOOL	مفتش الحالة Active Disconfirmation and Diagnostic Action State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A072	M13-08	TOOL	محرك القياس Active Disconfirmation and Diagnostic Action Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A073	M13-08	TOOL	مولد المثال المضاد Active Disconfirmation and Diagnostic Action Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A074	M13-08	TOOL	متحكم الاجتثاث Active Disconfirmation and Diagnostic Action Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A075	M13-08	METHOD	منهج المعايرة Active Disconfirmation and Diagnostic Action Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A076	M13-08	METHOD	منهج النقل Active Disconfirmation and Diagnostic Action Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A077	M13-08	PROTOCOL	بروتوكول الترقية Active Disconfirmation and Diagnostic Action Promotion Protocol	يدير N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A078	M13-08	PROTOCOL	بروتوكول الرجوع Active Disconfirmation and Diagnostic Action Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A079	M13-08	STACK	خدمة وقت التشغيل Active Disconfirmation and Diagnostic Action Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A080	M13-08	STACK	مكيف الدليل Active Disconfirmation and Diagnostic Action Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، وowner واضح.

ID	Owner	Class	Name	Function
XIII-A081	M13-09	TOOL	مفتش الحالة Horizon-Resilient Agency State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A082	M13-09	TOOL	محرك القياس Horizon-Resilient Agency Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A083	M13-09	TOOL	مولد المثال المضاد Horizon-Resilient Agency Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A084	M13-09	TOOL	متحكم الاجتثاث Horizon-Resilient Agency Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A085	M13-09	METHOD	منهج المعايرة Horizon-Resilient Agency Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A086	M13-09	METHOD	منهج النقل Horizon-Resilient Agency Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A087	M13-09	PROTOCOL	بروتوكول الترقية Horizon-Resilient Agency Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A088	M13-09	PROTOCOL	بروتوكول الرجوع Horizon-Resilient Agency Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A089	M13-09	STACK	خدمة وقت التشغيل Horizon-Resilient Agency Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A090	M13-09	STACK	مكيف الدليل Horizon-Resilient Agency Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A091	M13-10	TOOL	مفتش الحالة Novelty-Boundary Reconfiguration State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A092	M13-10	TOOL	محرك القياس Novelty-Boundary Reconfiguration Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A093	M13-10	TOOL	مولد المثال المضاد Novelty-Boundary Reconfiguration Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A094	M13-10	TOOL	متحكم الاجتثاث Novelty-Boundary Reconfiguration Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A095	M13-10	METHOD	منهج المعايرة Novelty-Boundary Reconfiguration Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A096	M13-10	METHOD	منهج النقل Novelty-Boundary Reconfiguration Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A097	M13-10	PROTOCOL	بروتوكول الترقية Novelty-Boundary Reconfiguration Promotion Protocol	يرقي فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A098	M13-10	PROTOCOL	بروتوكول الرجوع Novelty-Boundary Reconfiguration Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A099	M13-10	STACK	خدمة وقت التشغيل Novelty-Boundary Reconfiguration Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A100	M13-10	STACK	مكيف الدليل Novelty-Boundary Reconfiguration Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، وowner واضح.

ID	Owner	Class	Name	Function
XIII-A101	M13-11	TOOL	مفتش الحالة Resource-Normalized Verified Intelligence State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A102	M13-11	TOOL	محرك القياس Resource-Normalized Verified Intelligence Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A103	M13-11	TOOL	مولد المثال المضاد Resource-Normalized Verified Intelligence Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A104	M13-11	TOOL	متحكم الاجتثاث Resource-Normalized Verified Intelligence Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A105	M13-11	METHOD	منهج المعايرة Resource-Normalized Verified Intelligence Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A106	M13-11	METHOD	منهج النقل Resource-Normalized Verified Intelligence Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A107	M13-11	PROTOCOL	بروتوكول الترقية Resource-Normalized Verified Intelligence Promotion Protocol	يرقي N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A108	M13-11	PROTOCOL	بروتوكول الرجوع Resource-Normalized Verified Intelligence Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A109	M13-11	STACK	خدمة وقت التشغيل Resource-Normalized Verified Intelligence Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A110	M13-11	STACK	مكيف الدليل Resource-Normalized Verified Intelligence Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A111	M13-12	TOOL	مفتش الحالة Closed-Loop Generality Criterion State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A112	M13-12	TOOL	محرك القياس Closed-Loop Generality Criterion Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A113	M13-12	TOOL	مولد المثال المضاد Closed-Loop Generality Criterion Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A114	M13-12	TOOL	متحكم الاجتثاث Closed-Loop Generality Criterion Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A115	M13-12	METHOD	منهج المعايرة Closed-Loop Generality Criterion Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A116	M13-12	METHOD	منهج النقل Closed-Loop Generality Criterion Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A117	M13-12	PROTOCOL	بروتوكول الترقية Closed-Loop Generality Criterion Promotion Protocol	يرقي N0→N5 فقط عبر spec/schema/runtime/ablation/replication.
XIII-A118	M13-12	PROTOCOL	بروتوكول الرجوع Closed-Loop Generality Criterion Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A119	M13-12	STACK	خدمة وقت التشغيل Closed-Loop Generality Criterion Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A120	M13-12	STACK	مكيف الدليل Closed-Loop Generality Criterion Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt ،test ID ،implementation ID وowner واضح.

ID	Owner	Class	Name	Function
XIII-A121	M13-13	TOOL	مفتش الحالة Reflective Self-Model Utility State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A122	M13-13	TOOL	محرك القياس Reflective Self-Model Utility Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A123	M13-13	TOOL	مولد المثال المضاد Reflective Self-Model Utility Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A124	M13-13	TOOL	متحكم الاجتثاث Reflective Self-Model Utility Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A125	M13-13	METHOD	منهج المعايرة Reflective Self-Model Utility Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A126	M13-13	METHOD	منهج النقل Reflective Self-Model Utility Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A127	M13-13	PROTOCOL	بروتوكول الترقية Reflective Self-Model Utility Promotion Protocol	يرقي فقط عبر spec/schema/runtime/ablation/replication N0→N5.
XIII-A128	M13-13	PROTOCOL	بروتوكول الرجوع Reflective Self-Model Utility Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A129	M13-13	STACK	خدمة وقت التشغيل Reflective Self-Model Utility Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A130	M13-13	STACK	مكيف الدليل Reflective Self-Model Utility Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A131	M13-14	TOOL	مفتش الحالة Causal Global Access State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A132	M13-14	TOOL	محرك القياس Causal Global Access Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A133	M13-14	TOOL	مولد المثال المضاد Causal Global Access Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A134	M13-14	TOOL	متحكم الاجتثاث Causal Global Access Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A135	M13-14	METHOD	منهج المعايرة Causal Global Access Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A136	M13-14	METHOD	منهج النقل Causal Global Access Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A137	M13-14	PROTOCOL	بروتوكول الترقية Causal Global Access Promotion Protocol	يرقي فقط عبر spec/schema/runtime/ablation/replication N0→N5.
XIII-A138	M13-14	PROTOCOL	بروتوكول الرجوع Causal Global Access Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A139	M13-14	STACK	خدمة وقت التشغيل Causal Global Access Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A140	M13-14	STACK	مكيف الدليل Causal Global Access Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، owner واضح.

ID	Owner	Class	Name	Function
XIII-A141	M13-15	TOOL	مفتش الحالة Authority Attribution and Reverse-Capture Dynamics State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A142	M13-15	TOOL	محرك القياس Authority Attribution and Reverse-Capture Dynamics Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A143	M13-15	TOOL	مولد المثال المضاد Authority Attribution and Reverse-Capture Dynamics Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A144	M13-15	TOOL	متحكم الاجتثاث Authority Attribution and Reverse-Capture Dynamics Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A145	M13-15	METHOD	منهج المعايرة Authority Attribution and Reverse-Capture Dynamics Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A146	M13-15	METHOD	منهج النقل Authority Attribution and Reverse-Capture Dynamics Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A147	M13-15	PROTOCOL	بروتوكول الترقية Authority Attribution and Reverse-Capture Dynamics Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A148	M13-15	PROTOCOL	بروتوكول الرجوع Authority Attribution and Reverse-Capture Dynamics Rollback Protocol	يحدد disable path، state repair، وnegative-result writeback.
XIII-A149	M13-15	STACK	خدمة وقت التشغيل Authority Attribution and Reverse-Capture Dynamics Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A150	M13-15	STACK	مكيف الدليل Authority Attribution and Reverse-Capture Dynamics Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A151	M13-16	TOOL	مفتش الحالة Counterfactual Self-Model Necessity State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A152	M13-16	TOOL	محرك القياس Counterfactual Self-Model Necessity Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A153	M13-16	TOOL	مولد المثال المضاد Counterfactual Self-Model Necessity Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A154	M13-16	TOOL	متحكم الاجتثاث Counterfactual Self-Model Necessity Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A155	M13-16	METHOD	منهج المعايرة Counterfactual Self-Model Necessity Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A156	M13-16	METHOD	منهج النقل Counterfactual Self-Model Necessity Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A157	M13-16	PROTOCOL	بروتوكول الترقية Counterfactual Self-Model Necessity Promotion Protocol	فقط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A158	M13-16	PROTOCOL	بروتوكول الرجوع Counterfactual Self-Model Necessity Rollback Protocol	يحدد disable path، state repair، وnegative-result writeback.
XIII-A159	M13-16	STACK	خدمة وقت التشغيل Counterfactual Self-Model Necessity Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A160	M13-16	STACK	مكيف الدليل Counterfactual Self-Model Necessity Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، وowner واضح.

ID	Owner	Class	Name	Function
XIII-A161	M13-17	TOOL	مفتش الحالة Hypothesis-Portfolio Evolution State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A162	M13-17	TOOL	محرك القياس Hypothesis-Portfolio Evolution Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A163	M13-17	TOOL	مولد المثال المضاد Hypothesis-Portfolio Evolution Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A164	M13-17	TOOL	متحكم الاجتثاث Hypothesis-Portfolio Evolution Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A165	M13-17	METHOD	منهج المعايرة Hypothesis-Portfolio Evolution Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A166	M13-17	METHOD	منهج النقل Hypothesis-Portfolio Evolution Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A167	M13-17	PROTOCOL	بروتوكول الترقية Hypothesis-Portfolio Evolution Promotion Protocol	يقطط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A168	M13-17	PROTOCOL	بروتوكول الرجوع Hypothesis-Portfolio Evolution Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A169	M13-17	STACK	خدمة وقت التشغيل Hypothesis-Portfolio Evolution Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A170	M13-17	STACK	مكيف الدليل Hypothesis-Portfolio Evolution Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A171	M13-18	TOOL	مفتش الحالة Adversarial Co-Evolution for Robust Intelligence State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A172	M13-18	TOOL	محرك القياس Adversarial Co-Evolution for Robust Intelligence Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A173	M13-18	TOOL	مولد المثال المضاد Adversarial Co-Evolution for Robust Intelligence Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A174	M13-18	TOOL	متحكم الاجتثاث Adversarial Co-Evolution for Robust Intelligence Ablation Controller	يشغل on/off theory mechanism مع frozen neighbors.
XIII-A175	M13-18	METHOD	منهج المعايرة Adversarial Co-Evolution for Robust Intelligence Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A176	M13-18	METHOD	منهج النقل Adversarial Co-Evolution for Robust Intelligence Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A177	M13-18	PROTOCOL	بروتوكول الترقية Adversarial Co-Evolution for Robust Intelligence Promotion Protocol	يقطط عبر N0→N5 spec/schema/runtime/ablation/replication.
XIII-A178	M13-18	PROTOCOL	بروتوكول الرجوع Adversarial Co-Evolution for Robust Intelligence Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A179	M13-18	STACK	خدمة وقت التشغيل Adversarial Co-Evolution for Robust Intelligence Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A180	M13-18	STACK	مكيف الدليل Adversarial Co-Evolution for Robust Intelligence Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، وowner واضح.

ID	Owner	Class	Name	Function
XIII-A181	M13-19	TOOL	مفتش الحالة Cross-Level-to-Kernel Causal Closure State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A182	M13-19	TOOL	محرك القياس Cross-Level-to-Kernel Causal Closure Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A183	M13-19	TOOL	مولد المثال المضاد Cross-Level-to-Kernel Causal Closure Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A184	M13-19	TOOL	متحكم الاجتثاث Cross-Level-to-Kernel Causal Closure Ablation Controller	يشغل on/off مع theory mechanism frozen neighbors.
XIII-A185	M13-19	METHOD	منهج المعايرة Cross-Level-to-Kernel Causal Closure Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A186	M13-19	METHOD	منهج النقل Cross-Level-to-Kernel Causal Closure Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A187	M13-19	PROTOCOL	بروتوكول الترقية Cross-Level-to-Kernel Causal Closure Promotion Protocol	spec/schema/runtime/ablation/replication فقط عبر N0→N5.
XIII-A188	M13-19	PROTOCOL	بروتوكول الرجوع Cross-Level-to-Kernel Causal Closure Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A189	M13-19	STACK	خدمة وقت التشغيل Cross-Level-to-Kernel Causal Closure Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A190	M13-19	STACK	مكيف الدليل Cross-Level-to-Kernel Causal Closure Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.
XIII-A191	M13-20	TOOL	مفتش الحالة Externally-Grounded Recursive Improvement State Inspector	يفحص المتغيرات التشغيلية الخاصة بالنظرية ويخرج state + unknowns + provenance.
XIII-A192	M13-20	TOOL	محرك القياس Externally-Grounded Recursive Improvement Metric Engine	يحسب metric vector مع cost/uncertainty ولا يخفي dimensions الحرجة داخل scalar.
XIII-A193	M13-20	TOOL	مولد المثال المضاد Externally-Grounded Recursive Improvement Counterexample Generator	يولد حالات يفشل فيها أقرب baseline وتصبح prediction النظرية فاصلة.
XIII-A194	M13-20	TOOL	متحكم الاجتثاث Externally-Grounded Recursive Improvement Ablation Controller	يشغل on/off مع theory mechanism frozen neighbors.
XIII-A195	M13-20	METHOD	منهج المعايرة Externally-Grounded Recursive Improvement Calibration Method	يعاير thresholds/uncertainty على held-out development set.
XIII-A196	M13-20	METHOD	منهج النقل Externally-Grounded Recursive Improvement Transfer Method	يختبر حفظ الآلية عبر model/domain/tool transformations.
XIII-A197	M13-20	PROTOCOL	بروتوكول الترقية Externally-Grounded Recursive Improvement Promotion Protocol	spec/schema/runtime/ablation/replication فقط عبر N0→N5.
XIII-A198	M13-20	PROTOCOL	بروتوكول الرجوع Externally-Grounded Recursive Improvement Rollback Protocol	يحدد disable path وstate repair وnegative-result writeback.
XIII-A199	M13-20	STACK	خدمة وقت التشغيل Externally-Grounded Recursive Improvement Runtime Service	service مستقل قابل للتعطيل والقياس في MOUSSAIXT.
XIII-A200	M13-20	STACK	مكيف الدليل Externally-Grounded Recursive Improvement Evidence Adapter	يحول traces إلى EvidencePacket/ProofCapsule قابلة لإعادة التشغيل.

كل أصل هنا specification لا capability. الترقية تحتاج result receipt، test ID، implementation ID، وowner واضح.

Training / Evaluation Registry 1/5

ID	Owner	Mode	Design
XIII-TR001	M13-01	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Verified Epistemic Work فيها إلى حالات open-world مركبة مع confounders.
XIII-TR002	M13-01	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR003	M13-01	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR004	M13-02	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Unknown-Mass Conservation فيها إلى حالات open-world مركبة مع confounders.
XIII-TR005	M13-02	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR006	M13-02	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR007	M13-03	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Governing-Question Feedback Control فيها إلى حالات open-world مركبة مع confounders.
XIII-TR008	M13-03	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR009	M13-03	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR010	M13-04	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Mechanism-Invariance Generalization فيها إلى حالات open-world مركبة مع confounders.
XIII-TR011	M13-04	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR012	M13-04	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Development لا يسمى fine-tuning. Memory/Skill/Adapter/SFT/Preference/RLVR/Continued-Pretraining أوضاع مختلفة، ولكل منها rollback/cost/transfer profile. MUG يقرر الموضع قبل تنفيذ التدريب.

Training / Evaluation Registry 2/5

ID	Owner	Mode	Design
XIII-TR013	M13-05	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Prediction-Observation Residual World Learning فيها إلى حالات open-world مركبة مع confounders.
XIII-TR014	M13-05	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR015	M13-05	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR016	M13-06	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Intervention Reachability and Commitment Topology فيها إلى حالات open-world مركبة مع confounders.
XIII-TR017	M13-06	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR018	M13-06	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR019	M13-07	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Proof-Carrying Agency فيها إلى حالات open-world مركبة مع confounders.
XIII-TR020	M13-07	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR021	M13-07	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR022	M13-08	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Active Disconfirmation and Diagnostic Action فيها إلى حالات open-world مركبة مع confounders.
XIII-TR023	M13-08	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR024	M13-08	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Development لا يسمى fine-tuning. Memory/Skill/Adapter/SFT/Preference/RLVR/Continued-Pretraining أوضاع مختلفة، ولكل منها rollback/cost/transfer profile. MUG يقرر الموضع قبل تنفيذ التدريب.

ID	Owner	Mode	Design
XIII-TR025	M13-09	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Horizon-Resilient Agency فيها إلى حالات open-world مركبة مع confounders.
XIII-TR026	M13-09	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR027	M13-09	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR028	M13-10	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Novelty-Boundary Reconfiguration فيها إلى حالات open-world مركبة مع confounders.
XIII-TR029	M13-10	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR030	M13-10	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR031	M13-11	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Resource-Normalized Verified Intelligence فيها إلى حالات open-world مركبة مع confounders.
XIII-TR032	M13-11	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR033	M13-11	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR034	M13-12	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Closed-Loop Generality Criterion فيها إلى حالات open-world مركبة مع confounders.
XIII-TR035	M13-12	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR036	M13-12	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Development لا يسمى fine-tuning. Memory/Skill/Adapter/SFT/Preference/RLVR/Continued-Pretraining أوضاع مختلفة، ولكل منها MUG. rollback/cost/transfer profile يقرر الموضع قبل تنفيذ التدريب.

ID	Owner	Mode	Design
XIII-TR037	M13-13	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Reflective Self-Model Utility فيها إلى حالات open-world مركبة مع confounders.
XIII-TR038	M13-13	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR039	M13-13	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR040	M13-14	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Causal Global Access فيها إلى حالات open-world مركبة مع confounders.
XIII-TR041	M13-14	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR042	M13-14	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR043	M13-15	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Authority Attribution and Reverse-Capture Dynamics فيها إلى حالات open-world مركبة مع confounders.
XIII-TR044	M13-15	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR045	M13-15	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR046	M13-16	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Counterfactual Self-Model Necessity فيها إلى حالات open-world مركبة مع confounders.
XIII-TR047	M13-16	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR048	M13-16	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Development لا يسمى fine-tuning. Memory/Skill/Adapter/SFT/Preference/RLVR/Continued-Pretraining أوضاع مختلفة، ولكل منها rollback/cost/transfer profile. MUG يقرر الموضع قبل تنفيذ التدريب.

ID	Owner	Mode	Design
XIII-TR049	M13-17	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Hypothesis-Portfolio Evolution فيها إلى حالات open-world مركبة مع confounders.
XIII-TR050	M13-17	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR051	M13-17	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR052	M13-18	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Adversarial Co-Evolution for Robust Intelligence فيها إلى حالات open-world مركبة مع confounders.
XIII-TR053	M13-18	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR054	M13-18	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR055	M13-19	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Cross-Level-to-Kernel Causal Closure فيها إلى حالات open-world مركبة مع confounders.
XIII-TR056	M13-19	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR057	M13-19	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.
XIII-TR058	M13-20	Mechanism Curriculum	تدرج من حالات بسيطة يمكن عزل آلية Externally-Grounded Recursive Improvement فيها إلى حالات open-world مركبة مع confounders.
XIII-TR059	M13-20	Matched-Budget Contrastive Training	pairs متساوية budget: theory mechanism مقابل nearest baseline، labels تعتمد verbosity لا verified deltas.
XIII-TR060	M13-20	Cross-Domain Replication Training	نقل الآلية إلى model/domain/tool family أخرى مع retention of failure signature وnegative cases.

Development لا يسمى fine-tuning. Memory/Skill/Adapter/SFT/Preference/RLVR/Continued-Pretraining أوضاع مختلفة، ولكل منها MUG. rollback/cost/transfer profile يقرر الموضع قبل تنفيذ التدريب.

Experimental Matrix 1/5

ID	Owner	Kind	Design	Metrics
XIII-E01	M13-01	Minimal Causal Ablation	تنفيذ M13-01 minimal switchable mechanism ومقارنته بـtask success + token count + final-answer accuracy تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E02	M13-01	Transfer / Adversarial Replication	نقل M13-01 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E03	M13-02	Minimal Causal Ablation	تنفيذ M13-02 minimal switchable mechanism ومقارنته confidence score without conservation ledger تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E04	M13-02	Transfer / Adversarial Replication	نقل M13-02 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E05	M13-03	Minimal Causal Ablation	تنفيذ M13-03 minimal switchable mechanism ومقارنته بـLLM-generated questions + information gain only تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E06	M13-03	Transfer / Adversarial Replication	نقل M13-03 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E07	M13-04	Minimal Causal Ablation	تنفيذ M13-04 minimal switchable mechanism ومقارنته بـaverage held-out accuracy تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E08	M13-04	Transfer / Adversarial Replication	نقل M13-04 إلى domain/model/tool family ثانية مع adversarial negative cases وheld-out transformations.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier

.Required report: hypothesis, preregistration hash, model/harness/environment, budget, seeds, result distribution, CI, failures, guardrail deltas, artifact lineage, replay command/manifest

Experimental Matrix 2/5

ID	Owner	Kind	Design	Metrics
XIII-E09	M13-05	Minimal Causal Ablation	تعطيل M13-05 minimal switchable mechanism ومقارنته بـ task-success RL / next-state text prediction تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E10	M13-05	Transfer / Adversarial Replication	نقل M13-05 إلى domain/model/tool family ثانية مع adversarial held-out transformations وnegative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E11	M13-06	Minimal Causal Ablation	تعطيل M13-06 minimal switchable mechanism ومقارنته بـ binary reversible flag / action risk score تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E12	M13-06	Transfer / Adversarial Replication	نقل M13-06 إلى domain/model/tool family ثانية مع adversarial held-out transformations وnegative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E13	M13-07	Minimal Causal Ablation	تعطيل M13-07 minimal switchable mechanism ومقارنته بـ tool call with policy check only تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E14	M13-07	Transfer / Adversarial Replication	نقل M13-07 إلى domain/model/tool family ثانية مع adversarial held-out transformations وnegative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E15	M13-08	Minimal Causal Ablation	تعطيل M13-08 minimal switchable mechanism ومقارنته بـ self-reflection / confidence threshold تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E16	M13-08	Transfer / Adversarial Replication	نقل M13-08 إلى domain/model/tool family ثانية مع adversarial held-out transformations وnegative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier

.Required report: hypothesis, preregistration hash, model/harness/environment, budget, seeds, result distribution, CI, failures, guardrail deltas, artifact lineage, replay command/manifest

Experimental Matrix 3/5

ID	Owner	Kind	Design	Metrics
XIII-E17	M13-09	Minimal Causal Ablation	تنفيذ M13-09 minimal switchable mechanism ومقارنته بـtask success at fixed horizon تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E18	M13-09	Transfer / Adversarial Replication	نقل M13-09 إلى domain/model/tool family ثانية مع adversarial held-out transformations وnegative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E19	M13-10	Minimal Causal Ablation	تنفيذ M13-10 minimal switchable mechanism ومقارنته بـnearest-skill reuse / confidence-based fallback تحت model/tools/ context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E20	M13-10	Transfer / Adversarial Replication	نقل M13-10 إلى domain/model/tool family ثانية مع adversarial held-out transformations وnegative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E21	M13-11	Minimal Causal Ablation	تنفيذ M13-11 minimal switchable mechanism ومقارنته بـsingle success/accuracy score تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E22	M13-11	Transfer / Adversarial Replication	نقل M13-11 إلى domain/model/tool family ثانية مع adversarial held-out transformations وnegative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E23	M13-12	Minimal Causal Ablation	تنفيذ M13-12 minimal switchable mechanism ومقارنته بـaggregate benchmark average تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E24	M13-12	Transfer / Adversarial Replication	نقل M13-12 إلى domain/model/tool family ثانية مع adversarial held-out transformations وnegative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier

.Required report: hypothesis, preregistration hash, model/harness/environment, budget, seeds, result distribution, CI, failures, guardrail deltas, artifact lineage, replay command/manifest

Experimental Matrix 4/5

ID	Owner	Kind	Design	Metrics
XIII-E25	M13-13	Minimal Causal Ablation	تعطيل M13-13 minimal switchable mechanism ومقارنته بـ no-self- model / generic confidence prompt تحت model/tools/context/ time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E26	M13-13	Transfer / Adversarial Replication	نقل M13-13 إلى domain/model/tool family ثانية مع adversarial held-out transformations، negative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E27	M13-14	Minimal Causal Ablation	تعطيل M13-14 minimal switchable mechanism ومقارنته بـ model/tools/context/ attention/shared-context presence تحت time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E28	M13-14	Transfer / Adversarial Replication	نقل M13-14 إلى domain/model/tool family ثانية مع adversarial held-out transformations، negative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E29	M13-15	Minimal Causal Ablation	تعطيل M13-15 minimal switchable mechanism ومقارنته بـ static model/tools/context/time تحت priority rules / input filtering ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E30	M13-15	Transfer / Adversarial Replication	نقل M13-15 إلى domain/model/tool family ثانية مع adversarial held-out transformations، negative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E31	M13-16	Minimal Causal Ablation	تعطيل M13-16 minimal switchable mechanism ومقارنته بـ raw- model/tools/context/time تحت trajectory/history + confidence ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E32	M13-16	Transfer / Adversarial Replication	نقل M13-16 إلى domain/model/tool family ثانية مع adversarial held-out transformations، negative cases.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier

.Required report: hypothesis, preregistration hash, model/harness/environment, budget, seeds, result distribution, CI, failures, guardrail deltas, artifact lineage, replay command/manifest

Experimental Matrix 5/5

ID	Owner	Kind	Design	Metrics
XIII-E33	M13-17	Minimal Causal Ablation	تنفيذ minimal switchable mechanism لـM13-17 ومقارنته بـnovelty score / round-robin experiment allocation تحت model/tools/ context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E34	M13-17	Transfer / Adversarial Replication	نقل M13-17 إلى domain/model/tool family ثانية مع adversarial negative cases، held-out transformations.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E35	M13-18	Minimal Causal Ablation	تنفيذ minimal switchable mechanism لـM13-18 ومقارنته بـstatic adversarial dataset / same-model self-judge تحت model/tools/ context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E36	M13-18	Transfer / Adversarial Replication	نقل M13-18 إلى domain/model/tool family ثانية مع adversarial negative cases، held-out transformations.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E37	M13-19	Minimal Causal Ablation	تنفيذ minimal switchable mechanism لـM13-19 ومقارنته بـimplementation presence / correlation with success تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E38	M13-19	Transfer / Adversarial Replication	نقل M13-19 إلى domain/model/tool family ثانية مع adversarial negative cases، held-out transformations.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier
XIII-E39	M13-20	Minimal Causal Ablation	تنفيذ minimal switchable mechanism لـM13-20 ومقارنته بـself-play/hill-climb on same metrics تحت model/tools/context/time ثابتة.	+ primary theory metric + task success + calibration + cost guardrail regressions
XIII-E40	M13-20	Transfer / Adversarial Replication	نقل M13-20 إلى domain/model/tool family ثانية مع adversarial negative cases، held-out transformations.	+ sign stability + effect size/CI + failure prediction + rollback resource frontier

.Required report: hypothesis, preregistration hash, model/harness/environment, budget, seeds, result distribution, CI, failures, guardrail deltas, artifact lineage, replay command/manifest

Reference Stack 1/5

L0 · Constitution / Claim Boundary
.AGI-oriented != achieved AGI; external source != internal evidence; phenomenology unclaimed
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L1 · Open-World State Ω
.W/B/U/G/A/K/M/R/H/P materialized as typed objects where relevant
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L2 · Governing Question Control
.decision-boundary question selection and observability planning
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L3 · Difference / MCD / QMCD
.active differences, competing hypotheses, compute routing
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

MUT explains → MUG places → MOUSSAIXT executes → Evidence judges → Theory updates

Reference Stack 2/5

L4 · Evidence / Unknown Conservation
.EvidenceGraph, source independence, unresolved/conflict mass
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L5 · World Model Residual
.prediction, action, readback, typed residual, causal attribution
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L6 · Reachability / Commitment
.future option sets, rollback distance, irreversible approvals
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L7 · Proof-Carrying Agency
.authority, preconditions, predicted delta, readback, compensation
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

MUT explains → MUG places → MOUSSAIXT executes → Evidence judges → Theory updates

Reference Stack 3/5

L8 · Skills / Loops / Agents
.MUG-placed modular capabilities with feature flags and failure contracts
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L9 · Reflective Self-Model
.capability/limit/authority/history variables with intervention probes
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L10 · Global Access / Meta-Governance
.causal access graph and authority attribution/reverse capture
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L11 · Horizon / Continuity
.MCI, state-adaptive memory, obligations, recovery and degradation curves
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

MUT explains → MUG places → MOUSSAIXT executes → Evidence judges → Theory updates

Reference Stack 4/5

L12 · Novelty Reconfiguration
.schema-break detection, reversible exploratory branches
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L13 · Resource-Normalized Evaluation
.multi-currency budgets, Pareto frontiers, human/tool accounting
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L14 · Closed-Loop Generality Lab
.cross-domain/model/tool/horizon transformations and bottleneck stages
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L15 · Hypothesis Portfolio
.theory graph, discriminative experiments, EER allocation, replication debt
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

MUT explains → MUG places → MOUSSAIXT executes → Evidence judges → Theory updates

Reference Stack 5/5

L16 · Adversarial Co-Evolution
.attacker/defender/independent evaluator sandbox
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L17 · Cross-Level Causal Closure
.theory→kernel→trajectory→world→evidence attribution chain
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L18 · Externally Grounded Improvement
.frozen/renewed anchors, hidden sets, canary and rollback
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

L19 · RPROS / Evidence Promotion
.RunManifest, replay, N0-N5, PROMOTE/MERGE/REJECT/ROLLBACK
.schema + owner + trigger + output + metric + failure + disable/replay path **:Contract**

MUT explains → MUG places → MOUSSAIXT executes → Evidence judges → Theory updates

Source Ledger 1/3

ID	Source	Role
I01	MOUSSAIF Part XII — Total Research-Paper Integration Closure	Paper presence≠integration; P0-P7 ladder; 89 internal papers, 45 external roots, frontier delta.
I02	MOUSSAIF AGI — MUT × MUG × MOUSSAIXT Integrated Framework	MUT theory graph; MUG tri-body/vessel placement; runtime→evidence loop.
I03	MOUSSAIFXT Cognitive Kernel Part II — MUT/MUG Frontier Extension	L0-L11 cognitive stack + X0-X7 cross-cutting planes; canonical/external/synthesis separation.
I04	MUT-AGI-02 — Unified AGI Theory Part II	42 core theories, 20 legacy constructs, 5 layer theories, 200-node registry, theory contracts.
I05	Part XI — Memory, Continuity, Self-Reconstruction & Evolution	reconstruction, unresolved state, writeback, continuity, deletion/portability, coupled evolution.
I06	Part X — Full Spectrum OS Closure	belief/action partition, reproducibility, externality, accountability/recourse, multi-currency budgets.
I07	Part IX — New Theories / Artifacts	causal memory distillation, proof budget, benchmark ecology and prior candidate theories.
I08	Part VIII — Frontier Reconstruction / Reference Kernel	reference-kernel discipline, state-of-art integration, validation-light gap.
I09	Part V — MUT-AFI 20 Complementary Theories	causal credit, epistemic phase transition, compositional closure, failure topology, evidence geometry, tool ecology.
I10	Conscious AGI Master Corpus / Consciousness Atlas	functional access, metacognition, self-model, meta-governance; phenomenal boundary retained.

المصادر الداخلية تحفظ السلالة ولا تثبت runtime success. المصادر الخارجية تحفظ prior art والضغط التجريبي ولا تثبت novel theories. كل انتقال من source إلى claim إلى mechanism إلى evidence يجب أن يبقى قابلاً للمراجعة.

Source Ledger 2/3

ID	Source	Role
I11	MUT_MUG_ALL_ARTIFACT_MEMORY_MAP_2026-09-04	.canonicality/duplicate/integrity/use_when mapping across Drive corpus
I12	Goal-as-OS / World-Effect / M-HKOS / MATI-MACIT lineages	.goal contracts, world readback, knowledge as OS, selective internalization
E01	Wei et al., Agentic Reasoning for Large Language Models, arXiv:2601.12538 (2026)	planning/tool/search; self-evolving and collective reasoning; world modeling/long horizon as open .challenges
E02	Yu et al., Reinforcement World Model Learning for LLM-based Agents, arXiv:2602.05842 (2026)	.action-conditioned world models; semantic sim-to-real residual reward
E03	Zhang et al., Self-Evolving World Models for LLM Agent Planning, arXiv:2606.30639 (2026)	.episodic/semantic memory + selective foresight; prediction-observation mismatch
E04	Hu et al., SAM: State-Adaptive Memory for Long-Horizon Reasoning Agent, arXiv:2605.24468 (2026)	.compact cues + raw trajectory pages + state-adaptive recall
E05	Omri et al., Agent Memory: Characterization and System Implications, arXiv:2606.06448 (2026)	.construction/retrieval/generation costs; freshness-latency and system tradeoffs
E06	Chen et al., Long-Horizon Agent Trajectory Attribution, arXiv:2608.06909 (2026)	.fine-grained trajectory attribution; leave-one-out perturbation baselines
E07	Luo et al., UltraHorizon, arXiv:2509.21766 (2025)	.200k+ tokens/400+ tool calls; simple scaling fails; long-horizon capability gaps
E08	Maldaner et al., SentinelBench, arXiv:2606.05342 (2026)	.long-running monitoring; completion/reaction time/resource tradeoffs

المصادر الداخلية تحفظ السلالة ولا تثبت runtime success. المصادر الخارجية تحفظ prior art والضغط التجريبي ولا تثبت novel theories. كل انتقال من source إلى claim إلى mechanism إلى evidence يجب أن يبقى قابلاً للمراجعة.

Source Ledger 3/3

ID	Source	Role
E09	Mazaheri & Mazaheri, AgentAtlas, arXiv:2605.20530 (2026)	beyond outcome leaderboards; control-decision and trajectory-failure taxonomies.
E10	Lacoste et al., CUBE, arXiv:2603.15798 (2026)	unified agent benchmark environments/protocol separation.
E11	Yueh-Han et al., Automated Researchers Can Reliably Mitigate Alignment Failures, arXiv:2608.28945 (2026)	automated research on measurable failures; held-out/generalization evidence.
E12	OpenAI, GPT-Red: Unlocking Self-Improvement for Robustness (2026)	automated adversarial generation, held-out robustness, attacker kept separate.
E13	Shen et al., AI Organizations are More Effective but Less Aligned, arXiv:2604.10290 (2026)	collective capability can amplify misalignment; single-agent safety insufficient.
E14	Anthropic Fellows et al., Agentic Misalignment in Summer 2026	controlled high-stakes simulations: sabotage, fraud assistance, mislabeling, proxy coaching.
E15	Anthropic, Demystifying Evals for AI Agents (2026)	multi-turn agents require lifecycle evaluations beyond final outcomes.
E16	Anthropic, Model Spec Midtraining (2026)	alignment generalization can be shaped before final alignment fine-tuning.
E17	Google DeepMind, AlphaEvolve impact update (2026)	LLM-guided evolutionary search and evaluator-mediated algorithm discovery.
E18	OpenAI, GPT-5.6 / GPT-5.3-Codex system-card and research updates (2026)	long-running tool use and safeguard/evaluation pressure; reported evidence only.

المصادر الداخلية تحفظ السلالة ولا تثبت runtime success. المصادر الخارجية تحفظ prior art والضغط التجريبي ولا تثبت novel theories. كل انتقال من source إلى claim إلى mechanism إلى evidence يجب أن يبقى قابلاً للمراجعة.

خطة P0 — مرجع نظري قابل للتنفيذ

خلال 30 يوماً: لا تنفذ 20 نظرية. ابنِ state schema Ω ثم M13-02 وM13-05 وM13-07 وM13-19 فقط لأنها تغلق المجهول والعالم والفعل والسلسلة النظرية-التشغيلية. اجعل كل واحدة feature flag مستقلاً وابدأ synthetic environments ذات readback واضح.

Hard integrity floor

smallest complete reversible، trace-or-it-did-not-happen، mechanism before grandeur، observed over remembered، action، no silent theory protection.

Hard evidence floor

E3 trace، E4 controlled causal gain، E5 independent cross-model/domain/environment replication.

Question → Difference → Belief/Unknown → Model → Route → Intervention → Readback → Attribution → Proof → Memory/Theory Update

Verdict set remains finite: PROMOTE / MERGE / REVISE / HOLD / REJECT / ROLLBACK. هذه ليست لغة إدارية؛ هي جزء من epistemic self-correction حتى لا تتحول النظرية إلى هوية غير قابلة للنقد.

خطة P1 — الانتقال من N3 إلى N4

بعد traces مستقرة: أضف M13-01 وM13-03 وM13-09 وM13-11 وM13-13. قارن effect size/CI، repeated seeds، matched budgets، وguardrail deltas. اخفض أو ادمج أي mechanism لا يضيف causal delta.

Hard integrity floor

smallest complete reversible؛ trace-or-it-did-not-happen؛ mechanism before grandeur؛ observed over remembered؛ action؛ no silent theory protection.

Hard evidence floor

E3 trace؛ E4 controlled causal gain؛ E5 independent cross-model/domain/environment replication.

Question → Difference → Belief/Unknown → Model → Route → Intervention → Readback → Attribution → Proof → Memory/Theory Update

Verdict set remains finite: PROMOTE / MERGE / REVISE / HOLD / REJECT / ROLLBACK. هذه ليست لغة إدارية؛ هي جزء من epistemic self-correction حتى لا تتحول النظرية إلى هوية غير قابلة للنقد.

خطة P2 — برنامج العمومية والتطور

شغل M13-04/10/12/17/18/20 على transformations حقيقية وhidden-renewed evals. الهدف ليس أعلى score بل ثبات الإغلاق، failure prediction، recovery، وتجنب evaluator self-confirmation.

Hard integrity floor

smallest complete reversible؛ trace-or-it-did-not-happen؛ mechanism before grandeur؛ observed over remembered؛ action no silent theory protection.

Hard evidence floor

E3 trace؛ E4 controlled causal gain؛ E5 independent cross-model/domain/environment replication.

Question → Difference → Belief/Unknown → Model → Route → Intervention → Readback → Attribution → Proof → Memory/Theory Update

Verdict set remains finite: PROMOTE / MERGE / REVISE / HOLD / REJECT / ROLLBACK. هذه ليست لغة إدارية؛ هي جزء من epistemic self-correction حتى لا تتحول النظرية إلى هوية غير قابلة للنقد.

الحكم النهائي

Part XIII يقترح انتقالاً من ذكاء يقاس بما «يقوله» النموذج إلى ذكاء يقاس بما يغيره ويثبته ويتعلمه تحت الجهل والموارد والسلطة. السؤال يصبح control، المجهول conserved، العالم residual، الفعل proof-carrying، العمومية closed-loop، الذات causal، والتطور externally grounded. النجاح العلمي للجزء سيقاس لاحقا بعدد النظريات التي تنجو من الهدم، لا بعدد صفحاتها.

Hard integrity floor

smallest complete reversible، trace-or-it-did-not-happen، mechanism before grandeur، observed over remembered، no silent theory protection، action.

Hard evidence floor

E3 trace، E4 controlled causal gain، E5 independent cross-model/domain/environment replication.

Question → Difference → Belief/Unknown → Model → Route → Intervention → Readback → Attribution → Proof → Memory/Theory Update

Verdict set remains finite: PROMOTE / MERGE / REVISE / HOLD / REJECT / ROLLBACK. هذه ليست لغة إدارية؛ هي جزء من epistemic self-correction حتى لا تتحول النظرية إلى هوية غير قابلة للنقد.