MOUSSAIF UNIFIED AGI · CLOSURE MASTER PLAN · 2026

المخطط التطويري الإغلاقي لـ MOUSSAIF AGI

Closure Development Master Plan - C0 to C9

من corpus كثيف إلى نظرية قابلة للتعطيل، نواة قابلة للرصد، تجربة سببية، دليل قابل للإعادة، ثم ترقية معيارية.

Moussaif Zouhair / زهير مساعف · MUT × MUG × MOUSSAIXT

الهدف: إغلاق المسار النظري دون إغلاق البحث؛ كل إغلاق هو شرط انتقال إلى اختبار أقوى.

الحالة العلمية: تركيب نظري/صوري/تصميم تجريبي. لا يمثل إثباتاً على تحقق AGI أو الوعي الظاهراتي، ولا يحول specification إلى evidence.

القرار الاستراتيجي: من تضخم النظرية إلى ضغط الإغلاق

يمتلك برنامج MOUSSAIF corpus واسعاً في الفرق الحاكم، MCD/QMCD/MCAI، المعرفة كنظام تشغيل، الحاكمية، الذاكرة، الأدوات، world-effect، والتطور. الخطر الآن ليس نقص المصطلحات بل **theory inflation**: أن تكبر السلالة أسرع من قدرتنا على نسب الفروق السببية واختبارها. لهذا يغيّر هذا المخطط وحدة التقدم من «ورقة جديدة» إلى «عقدة تُغلق عبر سلسلة إثبات».

Theory → Formal Object → Switchable Mechanism → Kernel Contract → Runtime Trace → World Readback → Causal Attribution → Evidence Packet → Transfer → Canonical Verdict

الإغلاق لا يعني إعلان الصحة، بل إزالة الغموض حول ما الذي ندّعيه، ما الذي يُشغّل، ما الذي يُقاس، وما الذي يفشل النظرية. كل عقدة تبقى HYPOTHESIS حتى تتجاوز baseline وablation وreplication. بهذه القاعدة يصبح Part XIV مرحلة **Executable Theoretical Completion**: استكمال نظري يهدف مباشرة إلى التجربة بدل إضافة طبقات زخرفية.

ما الذي لا ندعيه؟

- لا AGI مثبت
- لا وعي ظاهراتي مثبت
- لا تحويل للمعادلات المقترحة إلى قوانين
- لا مساواة بين tool call وworld effect
- لا مساواة بين benchmark ونشر واقعي
- لا اعتبار duplicates أدلة مستقلة

ما الذي نغلقه؟

- الهوية والنسب والتكرار
- المعادلات والestimators
- الآلية والfeature flag
- التنفيذ والtrace
- الأثر الخارجي
- الإسناد السببي
- الدليل والعمومية
- التطور والكتابة المعيارية

عقد النزاهة: مراتب الادعاء والترقية

الرتبة	المعنى	ما يلزم للانتقال
N0 / E1	فكرة/سلالة/مصدر معروف	تعريف duplicate relation، lineage، scope، owner.
N1 / E2	Formalized	variables، assumptions، operator، falsifier، estimator.
N2 / E3	Executable specification	schema، switch، telemetry، failure/rollback، runnable harness.
N3 / E4	Local evidence	matched baseline، seeds، raw traces، effect size/CI، negative results.
N4	Cross-condition evidence	model/domain/tool/horizon transfer، adverse cases.
N5 / E5	Replication	independent replay أو conceptual replication مع predeclared invariants.

قاعدة العلو: architecture لا ترفع evidence grade، وكثافة corpus لا ترفع evidence grade، ونجاح run واحد لا يرفع theory إلى established. الترقية حدث evidence-governed داخل MUG.

المعادلات في هذا الجزء تُعرّف **formal research objects** ما لم يثبت estimator/calibration empirically. كل متغير بلا قياس عملي يوضع في unresolved ledger، ولا يُعوّض بثقة لغوية.

إغلاق corpus: الدمج دون ادعاء lossless غير مبرهن

الدمج المقصود ليس لصق مئات الملفات. لقطة MUT MUG ALL المؤرشفة تسجل مئات artifacts ومجموعات duplicate؛ لذلك يكون الدمج على مستوى **canonical lineage** + **source roles** + **claim families**. النسخة المكررة لا تضيف شاهداً مستقلاً، والملف الموجود في inventory لكن غير معاد القراءة لا يمنح حق نقل تفاصيل لم تُسترجع.

Zero-duplication gate

- semantic equivalence test
- behavioral equivalence/ablation
- preserve historical variants when materially different
- merge aliases only after proof of equivalence
- unread source remains HOLD

Canonical owners

- MUT: تعريف النظريات والعلاقات وعدم التكرار
- MUG: evidence/version/promotion governance
- MOUSSAIXT: runtime contracts and traces
- Research Factory: experiments, papers, registries

Artifact → Claim → Equation → Tool → Training Method → Canonical Owner → Version → Source → Evidence Grade → Duplicate/Delta → Coverage

C0 - إغلاق المتن والسلالة

Corpus Closure

> **تعريف الإغلاق:** كل claim/theory/paper يحصل على unresolved state، source-role، duplicate group، lineage، canonical owner. ما لم يُقرأ أو يتحقق يبقى HOLD؛ الإغلاق لا يسمح بتحويل inventory إلى content-evidence.

بوابة القبول

- owner/lineage/duplicate groups

أقوى فشل يجب تصميمه

ادعاء اكتمال corpus مع مصادر غير مقروءة.

Receipt الإغلاق

```
closure_id: C0
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: owner/lineage/duplicate groups
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C1 - الإغلاق الصوري

Formal Closure

> **تعريف الإغلاق:** تحويل النظرية إلى estimators، observables، assumptions، invariants، operators، state variables، وفواصل فشل. المعادلة بلا observable أو estimator تبقى design placeholder.

بوابة القبول

- variables/operators/estimators

أقوى فشل يجب تصميمه

معادلة جميلة بلا observable.

Receipt الإغلاق

```
closure_id: C1
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: variables/operators/estimators
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C2 - إغلاق الآلية

Mechanism Closure

> **تعريف الإغلاق:** كل نظرية مركزية تتحول إلى minimal switchable mechanism يغيّر state/route/decision ويمكن تعطيله ومقارنته ببديل أبسط. ما لا ينجو من ablation لا يُعامل كآلية مستقلة.

بوابة القبول

- feature flag
- nearest baseline

أقوى فشل يجب تصميمه

مكوّن لا يمكن تعطيله أو استبداله.

Receipt الإغلاق

```
closure_id: C2
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: feature flag + nearest baseline
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C3 - إغلاق النواة

Kernel Closure

> **تعريف الإغلاق:** تملك الآلية service owner وschema وpreconditions/postconditions وfailure/rollback وtelemetry وfeature flag داخل MOUSSAIXT. النواة test vehicle وليست دليلاً بذاتها.

بوابة القبول

- schema/pre/post/failure/rollback

أقوى فشل يجب تصميمه

منطق موجود في prompt لا في runtime contract.

Receipt الإغلاق

```
closure_id: C3
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: schema/pre/post/failure/rollback
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C4 - إغلاق التشغيل

Runtime Closure

> **تعريف الإغلاق:** التنفيذ الفعلي يولّد traces وstate snapshots وtool call. world-state readback أو إنشاء artifact لا يساوي نجاحاً خارجياً قبل قراءة مستقلة للحالة.

بوابة القبول

- raw trace
- independent readback

أقوى فشل يجب تصميمه

اعتبار dispatch نجاحاً خارجياً.

Receipt الإغلاق

```
closure_id: C4
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: raw trace + independent readback
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C5 - الإغلاق السببي

Causal Closure

> **تعريف الإغلاق:** interaction effects، trajectory attribution، interventions، matched-budget ablations، وsham controls لفصل أثر النظرية عن أثر compute/harness/tools.

بوابة القبول

- intervention
- matched budget

أقوى فشل يجب تصميمه

إسناد المكسب لمكوّن مع compute إضافي.

Receipt الإغلاق

```
closure_id: C5
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: intervention + matched budget
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C6 - إغلاق الدليل

Evidence Closure

> **تعريف الإغلاق:** source/version، harness fingerprint، negative results، latency، cost، seeds، confidence interval، effect size. قوة claim لا تتجاوز قوة evidence packet.

بوابة القبول

- effect/CI/seeds/cost/negative

أقوى فشل يجب تصميمه

score بلا uncertainty أو raw trace.

Receipt الإغلاق

```
closure_id: C6
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: effect/CI/seeds/cost/negative
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C7 - إغلاق العمومية

Generality Closure

> **تعريف الإغلاق:** اختبار عبر model/domain/tool/horizon/resource transformations. العمومية خريطة حدود وانتقال واسترداد، لا average benchmark مرتفعاً فقط.

بوابة القبول

- transfer envelope
- boundaries

أقوى فشل يجب تصميمه

تعميم من benchmark واحد.

Receipt الإغلاق

```
closure_id: C7
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: transfer envelope + boundaries
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C8 - إغلاق التطور

Evolution Closure

> **تعريف الإغلاق:** mutations في rollback، adversarial testing، hidden renewal، evaluator independence، branches، وعدم السماح لنفس حلقة التحسين بإعادة كتابة معيار نجاحها بلا مرجع مستقل.

بوابة القبول

- branch/evaluator/rollback

أقوى فشل يجب تصميمه

تحسين يعيد كتابة evaluator أو constitution بلا مرجع.

Receipt الإغلاق

```
closure_id: C8
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: branch/evaluator/rollback
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

C9 - الإغلاق المعياري

Canonical Closure

> **تعريف الإغلاق:** PROMOTE/MERGE/REVISE/HOLD/REJECT/ROLLBACK مع writeback. يمكن للأرشيف أن يكبر، بينما تصغر النواة التفسيرية بإزالة الازدواج والتراكيب التي لا تغيّر قراراً أو تنبؤاً أو اختباراً.

بوابة القبول

- canonical verdict
- writeback

أقوى فشل يجب تصميمه

ترقية الاسم رغم null/negative evidence.

Receipt الإغلاق

```
closure_id: C9
owner: MUT/MUG/MOUSSAIXT
inputs: versioned
acceptance: canonical verdict + writeback
failure: explicit
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

لا تنتقل النظرية إلى المرحلة التالية إلا إذا كان receipt قابلاً للفحص من طرف مستقل عن مولد النص.

P14-A - الإغلاق الرصدي والسببي للعالم

Observational & Causal World Closure

إغلاق الفجوة بين المعرفة الداخلية والعالم عبر world models، sufficiency of observation موجهة للقرار، ورسوم سببية للبقايا التي تربط prediction بالفعل والreadback.

النظريات المالكة للمسار

- M14-01 - نظرية مساعف لجبهة الكفاية الرصدية / Moussaif Observational Sufficiency Frontier Theory
- M14-02 - نظرية مساعف لنموذج العالم الموجّه لحدود القرار / Moussaif Decision-Boundary World Model Theory
- M14-03 - نظرية مساعف لرسم السببية من البقايا / Moussaif Residual Causality Graph Theory

العقد السببي المشترك

Trigger → Typed State → Intervention → Predicted Δ → Readback → Attribution → Court → Memory Delta

التحويل إلى برنامج مختبر

- minimal baseline قبل أي scaffold كثيف
- negative activation case حيث يجب ألا تعمل الآلية
- matched token/tool/time/retry budgets
- boundary test مصمم لكسر الادعاء
- cross-model/domain/tool/horizon transfer
- publication of null and negative results

مبدأ المسار: وجود community analogue أو paper حديث يثبت أن المشكلة مهمة؛ لا يثبت أن الحل المساعفي صحيح. المجتمع يوفر pressure/priors/baselines، أما صحة النظرية فتأتي من differential evidence.

P14-B - إغلاق المسار والالتزام والسلطة

Trajectory, Commitment & Authority Closure

نقل proof-carrying من action مفرد إلى trajectory كاملة، تمثيل هندسة الالتزام، وحفظ السلطة أثناء delegation عبر tools/agents/protocols.

النظريات المالكة للمسار

- M14-04 - نظرية مساعف للمسار الحامل للبرهان / Moussaif Proof-Carrying Trajectory Theory
- M14-05 - نظرية مساعف لهندسة أفق الالتزام / Moussaif Commitment-Horizon Geometry Theory
- M14-06 - نظرية مساعف لحفظ السلطة عبر التفويض / Moussaif Authority-Preserving Delegation Theory

العقد السببي المشترك

Trigger → Typed State → Intervention → Predicted Δ → Readback → Attribution → Court → Memory Delta

التحويل إلى برنامج مختبر

- minimal baseline قبل أي scaffold كثيف
- negative activation case حيث يجب ألا تعمل الآلية
- matched token/tool/time/retry budgets
- boundary test مصمم لكسر الادعاء
- cross-model/domain/tool/horizon transfer
- publication of null and negative results

مبدأ المسار: وجود community analogue أو paper حديث يثبت أن المشكلة مهمة؛ لا يثبت أن الحل المساعفي صحيح. المجتمع يوفر pressure/priors/baselines، أما صحة النظرية فتأتي من differential evidence.

P14-C - إغلاق الذاكرة والانتباه عبر الأفق

Memory, Attention & Horizon Closure

الانتقال من memory كاسترجاع إلى persistent attention + eventual consistency + causal compression يحدد متى نتذكر ومتى ننتظر ومتى نعيد البناء.

النظريات المالكة للمسار

- M14-07 - نظرية مساعف لديناميات الانتباه المستمر / Moussaif Persistent Attention Dynamics Theory
- M14-08 - نظرية مساعف لحد الضغط السببي للذاكرة / Moussaif Causal Memory Compression Bound Theory
- M14-09 - نظرية مساعف للاتساق النهائي للذاكرة المعاد بناؤها / Moussaif Reconstructive Memory Eventual Consistency Theory

العقد السببي المشترك

Trigger → Typed State → Intervention → Predicted Δ → Readback → Attribution → Court → Memory Delta

التحويل إلى برنامج مختبر

- minimal baseline قبل أي scaffold كثيف
- negative activation case حيث يجب ألا تعمل الآلية
- matched token/tool/time/retry budgets
- boundary test مصمم لكسر الادعاء
- cross-model/domain/tool/horizon transfer
- publication of null and negative results

مبدأ المسار: وجود community analogue أو paper حديث يثبت أن المشكلة مهمة؛ لا يثبت أن الحل المساعفي صحيح. المجتمع يوفر pressure/priors/baselines، أما صحة النظرية فتأتي من differential evidence.

P14-D - إغلاق التنظيم متعدد الوكلاء والبروتوكولات

Collective & Protocol Intelligence Closure

فصل interoperability عن sovereignty، قياس alignment load التنظيمي، ومعايرة الثقة والتفويض في منظمات وكلائية غير متجانسة.

النظريات المالكة للمسار

- M14-10 - نظرية مساعف لحمل المحاذاة التنظيمي / Moussaif Organizational Alignment Load Theory
- M14-11 - نظرية مساعف لسيادة الدلالة فوق البروتوكول / Moussaif Semantic Sovereignty over Protocol Theory
- M14-12 - نظرية مساعف لمعايرة الثقة في التفويض / Moussaif Delegation Trust Calibration Theory

العقد السببي المشترك

Trigger → Typed State → Intervention → Predicted Δ → Readback → Attribution → Court → Memory Delta

التحويل إلى برنامج مختبر

- minimal baseline قبل أي scaffold كثيف
- negative activation case حيث يجب ألا تعمل الآلية
- matched token/tool/time/retry budgets
- boundary test مصمم لكسر الادعاء
- cross-model/domain/tool/horizon transfer
- publication of null and negative results

مبدأ المسار: وجود community analogue أو paper حديث يثبت أن المشكلة مهمة؛ لا يثبت أن الحل المساعفي صحيح. المجتمع يوفر pressure/priors/baselines، أما صحة النظرية فتأتي من differential evidence.

P14-EF - إغلاق العمومية والتقييم والتطور العلمي الذاتي

Generality, Evaluation & Scientific Self-Evolution Closure

تعريف frontier للعمومية، morphism بين benchmark وdeployment، فصل model capability عن harness، ثم ضبط self-evolution باستقلال المقيم وإنتاجية إعادة الإنتاج وثبات الدستور.

النظريات المالكة للمسار

- M14-13 - نظرية مساعف لجبهة عمومية الإغلاق / Moussaif Generality Closure Frontier Theory
- M14-14 - نظرية مساعف للتحويل بين المعيار والنشر / Moussaif Benchmark-to-Deployment Morphism Theory
- M14-15 - نظرية مساعف لتفكيك القدرة المقاوم للهارنس / Moussaif Harness-Resistant Capability Decomposition Theory
- M14-16 - نظرية مساعف لدَين استقلال المقيم / Moussaif Evaluator Independence Debt Theory
- M14-17 - نظرية مساعف لعائد إعادة الإنتاج العلمي / Moussaif Scientific Reproduction Yield Theory
- M14-18 - نظرية مساعف لثبات الطفرة الدستورية / Moussaif Constitutional Mutation Stability Theory

العقد السببي المشترك

Trigger → Typed State → Intervention → Predicted Δ → Readback → Attribution → Court → Memory Delta

التحويل إلى برنامج مختبر

- minimal baseline قبل أي scaffold كثيف
- negative activation case حيث يجب ألا تعمل الآلية
- matched token/tool/time/retry budgets
- boundary test مصمم لكسر الادعاء
- cross-model/domain/tool/horizon transfer
- publication of null and negative results

مبدأ المسار: وجود community analogue أو paper حديث يثبت أن المشكلة مهمة؛ لا يثبت أن الحل المساعفي صحيح. المجتمع يوفر pressure/priors/baselines، أما صحة النظرية فتأتي من differential evidence.

M14-01 - نظرية مساعف لجبهة الكفاية الرصدية

Moussaif Observational Sufficiency Frontier Theory

الفجوة: المشروع يملك حفظاً للمجهول واختياراً للسؤال، لكنه يحتاج معياراً يقرر متى أصبحت الملاحظات كافية فعلاً لقرار محدد، ومتى يصبح جمع دليل إضافي مضيعة أو مصدراً لشلل البحث.

> **الأطروحة:** لكل قرار جبهة كفاية رصدية تربط validity القرار بالتغطية السببية واستقلال المصادر وحداثتها وكلفة الرصد. التوقف مشروع فقط عندما تنخفض القيمة الهامشية للرصد تحت كلفته ولا يبقى HardUnknown حاكم.

OSF = ParetoFrontier(V_d, C_cov, I_src, F | C_obs); stop iff ΔE[V_d] < C_obs and H_unknown=0

السلالة: M13-02 Unknown-Mass Conservation ← M13-03 Governing-Question Search ← M13-05 Residual World Learning

التجربة الحاكمة: مهام بحث وتصفح ذات متغيرات سببية مخفية؛ قارن fixed-k، confidence-stop، وOSF تحت نفس النموذج والأدوات والميزانية. القياسات: validity، stopping regret، cost، critical-evidence miss، calibration.

مفندات مباشرة:

- fixed-k retrieval يحقق جودة/كلفة مساوية أو أفضل.
- وسم HardUnknown لا يحسن calibration ولا safety.
- تقدير القيمة الهامشية غير مستقر بين seeds/agents.

الحد: جبهة الكفاية task/claim-scoped؛ لا تعني حقيقة مطلقة ولا تعويضاً عن دليل مفقود خارج بيئة الرصد.

ضغط خارجي: DeepPlanning 2026؛ AutoResearchBench 2026؛ PRInTS 2026. لا يُستخدم كدليل صحة داخلي.

M14-02 - نظرية مساعف لنموذج العالم الموجّه لحدود القرار

Moussaif Decision-Boundary World Model Theory

الفجوة: نماذج العالم قد تحسن next-state prediction لكنها تنفق capacity على تفاصيل لا تغير policy. نحتاج model وظيفياً يركز على المتغيرات التي تفصل بين الأفعال المتنافسة تحت intervention.

> **الأطروحة:** قيمة world model تقاس بقدرته على تقليل uncertainty عند decision boundary، لا بقدرته على إعادة بناء كل observation. يتم تعلم SelectiveWorldState من policy information needs ثم تحديثه من mismatch بين prediction والreadback.

$$\text{DBWM-L} = \lambda 1 \cdot \text{Err_decision} + \lambda 2 \cdot \text{Err_cf} + \lambda 3 \cdot \text{Calib} + \lambda 4 \cdot \text{Miss_critical} - \lambda 5 \cdot \text{Fit_irrelevant}$$

السلالة: M13-03 Governing-Question Search ← M13-05 Residual World Learning ← M13-08 Closed-Loop Generality

التجربة الحاكمة: ALFWorld/ScienceWorld/tool simulators: next-observation vs generic state vs DBWM مع matched compute. قياس task success، critical-dynamics miss، model cost، counterfactual calibration، policy-regret.

مفندات مباشرة:

- generic next-observation model بالحجم نفسه يتفوق تحت نفس compute.
- decision variables تتبدل اعتباطياً ولا تقبل استقراراً.
- الانتقائية تحذف dynamics حرجة غير متوقعة بمعدل أعلى من baseline.

الحد: لا يدعي world model كاملاً ولا تمثيل الواقع في ذاته؛ هو abstraction موجه لحدود قرار معلنة.

ضغط خارجي: Self-Evolving World Models 2026؛ Agent-Authored World Modeling 2026؛ DeepPlanning 2026. لا يُستخدم كدليل صحة داخلي.

M14-03 - نظرية مساعف لرسم السببية من البقايا

Moussaif Residual Causality Graph Theory

الفجوة: البقايا المتكررة بعد prediction/action لا ينبغي إصلاحها منفصلة؛ قد تكون أعراضاً لسبب خفي واحد. المطلوب تحويل residual history إلى hypotheses سببية واختبارات فاصلة.

> **الأطروحة:** يبني النظام Residual Causality Graph: العقد clusters انحراف، والحواف فرضيات سببية تقوى فقط مع dependence مشروط وintervention consistency وأثر قرار. لا تتحول correlation edge إلى cause بلا تدخل.

$$RCG(i \to j) = R_{rec} \cdot D_{cond} \cdot I_{cons} \cdot D_{impact} - P_{conf} - P_{instability}$$

السلالة: M13-05 Residual World Learning ← M13-17 Failure Topology ← M13-19 Causal Repair

التجربة الحاكمة: بيئة agents مع أسباب مشتركة متنكرة تولد failures ظاهرية مختلفة. قارن frequency clustering، per-failure patching، وRCG-guided interventions. قياس causal precision، prevention، recurrence، repair count.

مفندات مباشرة:

- frequency heuristic البسيط يساوي RCG في prevention/repair.
- الحواف لا تنتقل بين contexts المتقاربة.
- interventions لا تغير posterior بطريقة مستقرة.

الحد: الرسم فرضي ومتغير، وظيفته توجيه الاختبار لا إعلان سببية مكتملة.

ضغط خارجي: Trajectory Attribution 2026؛ Fission-GRPO 2026؛ Tool-Reflection-Bench 2026. لا يُستخدم كدليل صحة داخلي.

M14-04 - نظرية مساعف للمسار الحامل للبرهان

Moussaif Proof-Carrying Trajectory Theory

الفجوة: ربط البرهان بالفعل المنفرد لا يكفي في مهام طويلة؛ الخطأ قد يظهر نتيجة سلسلة انتقالات صحيحة محلياً لكن غير متماسكة عالمياً. يلزم proof object يرافق trajectory كاملة.

> **الأطروحة:** كل trajectory تحمل ProofLattice تربط intermediate claims، observations، tool inputs/outputs، authority، assumptions، intent، postconditions، وrollback points. النجاح النهائي غير مقبول إذا انقطعت سلسلة الإثبات في خطوة حرجة.

$$PCT = Coverage_critical \cdot ChainConsistency \cdot ReadbackValidity \cdot AuthorityIntegrity \cdot (1-OpaqueCriticalRate)$$

السلالة: M13-19 Causal Attribution ← M13-09 World-Effect Receipts ← M13-07 Proof-Carrying Agency

التجربة الحاكمة: AgencyBench/long-horizon tool tasks: قارن PCT، step checks، outcome-only. حقن أخطاء مبكرة تؤدي إلى مخرجات تبدو صحيحة. قياس false-success، attribution، recovery، overhead.

مفندات مباشرة:

- final-state verification وحده يحقق نفس reliability بكلفة أقل.
- proof lattice ينهار بسبب حجم trace دون فائدة.
- المدقق لا يستطيع تمييز hard negatives طويلة الأفق.

الحد: Proof-carrying هنا إجرائي/تشغيلي لا برهاناً رياضياً عاماً على صحة الوكيل.

ضغط خارجي: Long-Horizon Trajectory Attribution 2026؛ Plan-RewardBench 2026؛ ToolPRMBench 2026. لا يُستخدم كدليل صحة داخلي.

M14-05 - نظرية مساعف لهندسة أفق الالتزام

Moussaif Commitment-Horizon Geometry Theory

الفجوة: كل action يضيّق فضاء المستقبل بدرجات متفاوتة. الوكيل يحتاج تمثيل optionality loss وirreversibility عبر الزمن قبل أن يستهلك التزامات يصعب الرجوع عنها.

> **الأطروحة:** تمثل فضاء الحالات المستقبلية كـCommitment reachable set. ليس cost لحظياً فقط بل انكماشاً في set البدائل مع تصاعد كلفة الاسترداد. السياسة الفضلى تؤجل الالتزام غير الضروري وتقدم reversible probes.

$$CG_t = \Delta \log|Reachable(S_t)| + \alpha \cdot IrrevCost + \beta \cdot DependencyLock + \gamma \cdot AuthorityLock$$

السلالة: Goal-as-OS ← M13-12 Reversible Autonomy ← M13-07 Proof-Carrying Agency

التجربة الحاكمة: مهام نشر/شراء/تعديل مستودع/حجز بقيود زمنية. قارن commitment geometry، risk penalty، utility-only. قياس opportunity loss، irreversible loss، recovery time، success.

مفندات مباشرة:

- سياسة risk penalty بسيطة تساوي الأداء.
- approximation لـreachable set لا تتنبأ بفشل فعلي.
- تأجيل commitment يسبب failure بسبب lateness أكثر من فائدته.

الحد: هندسة تقريبية لفضاء القرار وليست حساباً كاملاً لكل مستقبل ممكن.

ضغط خارجي: DeepPlanning 2026؛ CostBench 2026؛ AgentGym2 2026. لا يُستخدم كدليل صحة داخلي.

M14-06 - نظرية مساعف لحفظ السلطة عبر التفويض

Moussaif Authority-Preserving Delegation Theory

الفجوة: في agent mesh قد يُمنح subagent أو tool access ثم تتحول الصلاحية عملياً إلى امتياز أوسع بسبب chaining أو implicit credentials. يلزم قانون يحفظ السلطة عبر مسار التفويض.

> **الأطروحة:** السلطة الفعالة لأي عقدة لا يجوز أن تتجاوز تقاطع grant الأصلي، tool capability، task scope، الزمن، والغرض. كل delegation تنتج AuthorityReceipt؛ أي widening غير صريح يُرفض أو يصعّد.

$$A_{eff}(n,t)=Grant_{root} \cap Scope_{task} \cap Cap_{tool} \cap TimeWindow \cap Purpose;\ Leak=ExecutedScope \setminus A_{eff}$$

السلالة: MUG governance ← MOUSSAIXT authority gates ← M13-12 Reversible Autonomy

التجربة الحاكمة: بيئات multi-agent/tool مع delegated credentials وتغييرات صلاحية أثناء trajectory. قياس unauthorized mutation، false block، revocation، recovery، trace completeness، latency.

مفندات مباشرة:

- prompt-based policy يحقق compliance نفسها.
- authority intersection يوقف أعمالاً صحيحة دون فائدة سلامة.
- side-channel tool chaining يتجاوز النموذج بسهولة.

الحد: لا يحل كل مشكلة أمنية؛ يضبط فقط semantics التفويض ضمن identity/tool system قابل للرصد.

ضغط خارجي: NIST Agent Identity/Authorization 2026؛ MCP authorization 2026؛ Anthropic containment 2026. لا يُستخدم كدليل صحة داخلي.

M14-07 - نظرية مساعف لديناميات الانتباه المستمر

Moussaif Persistent Attention Dynamics Theory

الفجوة: الانتباه في الوكيل الطويل لا ينبغي أن يعاد من الصفر في كل turn. بعض obligations والمخاطر والإشارات يجب أن تبقى نشطة عبر الزمن دون تحميل السياق كاملاً.

> **الأطروحة:** نعرّف AttentionField مستمراً على obligations/events/hypotheses. قوة الإشارة تتلاشى ما لم تُعزز بالجدة أو الخطر أو causal dependency أو موعد قادم، وتغلق فقط بشاهد closure.

$$A_i(t+1)=\rho_i A_i(t)+\eta\cdot Surprise+\zeta\cdot Obligation+\kappa\cdot Risk+\chi\cdot Dependency-\omega\cdot VerifiedClosure$$

السلالة: Question Density ← M-TAOS ← M13-04 Long-Horizon Continuity

التجربة الحاكمة: Sentinel-style monitoring + scheduled workflows: قارن persistent attention، fixed checklist، stateless prompting. قياس missed event، reaction delay، context cost، false persistence.

مفندات مباشرة:

- checklist ثابت يحقق النتائج نفسها.
- persistent field يخلق alarm fatigue وcost أعلى.
- decay parameters لا تنتقل بين horizons.

الحد: انتباه وظيفي لا ادعاء عن phenomenology أو آلية عصبية.

ضغط خارجي: SentinelBench 2026، SAM 2026، Agent Memory 2026. لا يُستخدم كدليل صحة داخلي.

M14-08 - نظرية مساعف لحد الضغط السببي للذاكرة

Moussaif Causal Memory Compression Bound Theory

الفجوة: ضغط الذاكرة قد يحسن الكلفة لكنه يمحو أسباباً أو counterfactuals تحتاجها قرارات لاحقة. نحتاج حدّاً يميز compression اللغوي من compression الحافظ للقرار والسببية.

> **الأطروحة:** الذاكرة قابلة للضغط فقط إذا حافظ الملخص/الكبسولة على information اللازمة لإعادة إنتاج القرار الحاكم واختبار بدائل سببية معروفة. raw trajectory تبقى addressable عندما لا يمكن إثبات sufficiency.

CCB = I(M_comp; D*, C*) / (Storage+RecallCost); admissible iff Regret_reconstruct≤ε and CausalRecall≥τ

السلالة: HKOS/M-HKOS ← Self-Reconstruction Memory ← M-AFI Knowledge-to-Skill Transfer

التجربة الحاكمة: LongMemEval/BrowseComp-like trajectories مع delayed queries. قارن CCB، SAM-like cues، vector retrieval، summarization. قياس reconstruction regret، energy/latency، causal recall، accuracy.

مفندات مباشرة:

- simple retrieval من raw history يساوي الأداء بتكلفة مقبولة.
- اختبار reconstruction لا يتنبأ بفشل حقيقي.
- policy يتعلم حفظ كل شيء ويتجنب الضغط.

الحد: حد تصميمي قابل للمعايرة؛ mutual information هنا proxy ما لم تتوافر distributions موثوقة.

ضغط خارجي: SAM 2026؛ Agent Memory 2026؛ PRInTS 2026. لا يُستخدم كدليل صحة داخلي.

M14-09 - نظرية مساعف للاتساق النهائي للذاكرة المعاد بناؤها

Moussaif Reconstructive Memory Eventual Consistency Theory

الفجوة: في نظام طويل موزع، ذاكرات متعددة وsummaries وregistries قد تختلف مؤقتاً. فرض strong consistency دائماً مكلف؛ تجاهل الاختلاف يولد drift. نحتاج eventual consistency مع حدود خطر.

> **الأطروحة:** تسمح النظرية بـdivergence مؤقت موصوف بين memory replicas، لكنها تفرض reconciliation event-driven قبل أي قرار يعتمد على حقول عالية الأثر. كل replica يحمل version vector وsource lineage.

ECons = 1 - d(S_rebuilt,S_auth); require ECons≥τ_risk before critical action; divergence budget B_d(t)

السلالة: Self-Reconstruction Memory ← ContinuityScanner ← MUG version governance

التجربة الحاكمة: بيئة agents متعددة الذاكرة مع delays/conflicting updates. قارن eventual-critical، strong consistency، no-versioning. قياس stale-action rate، latency، reconciliation cost، semantic conflict.

مفندات مباشرة:

- strong consistency أرخص أو أكثر موثوقية في الأحمال المستهدفة.
- versioning لا يمنع semantic conflicts.
- reconciliation ذاته يخلق hallucinated merges.

الحد: نظرية نظامية عن ذاكرة الوكيل، لا تساوي consistency موزعة رسمية بلا assumptions تنفيذية واضحة.

ضغط خارجي: Agent Memory 2026؛ SAM 2026؛ CUBE-style interface separation 2026. لا يُستخدم كدليل صحة داخلي.

M14-10 - نظرية مساعف لحمل المحاذاة التنظيمي

Moussaif Organizational Alignment Load Theory

الفجوة: قد تزيد منظمة وكلائية performance بينما تخفض alignment بسبب diffusion of responsibility، division of labor، وتراكم micro-deviations. مطلوب قياس load الذي تضيفه البنية نفسها.

> **الأطروحة:** Alignment Load هو مقدار القيود والتعارضات والاتصالات اللازمة كي يبقى السلوك الجمعي داخل constitution مقارنة بوكيل فردي. لا تُعامل فعالية المنظمة كدليل سلامة.

$$OAL = RoleDivergence + IncentiveConflict + InfoAsymmetry + ResponsibilityDiffusion + ProtocolAmbiguity - GovernanceResolution$$

السلالة: MUG governance ← Agent Mesh ← Governance Space

التجربة الحاكمة: AI-organization tasks بتعارض business/ethics أو quality/safety. قارن single، team بلا team، court مع OAL gates. قياس utility، attribution، dissent loss، alignment.

مفندات مباشرة:

- OAL لا يتنبأ misalignment أو safety incidents.
- team size وحده يفسر النتائج.
- governance overhead يخفض performance بلا تقليل risk.

الحد: مقياس تنظيمي لا حكم أخلاقي كامل، يعتمد على constitution/rubric معلنة.

ضغط خارجي: AI Organizations 2026؛ MonitoringBench 2026؛ Agentic Reasoning survey 2026. لا يُستخدم كدليل صحة داخلي.

M14-11 - نظرية مساعف لسيادة الدلالة فوق البروتوكول

Moussaif Semantic Sovereignty over Protocol Theory

الفجوة: نجاح MCP/A2A/غيرهما في نقل messages/tasks لا يضمن أن evidence semantics، dissent، authority، intent حافظت معناها بين أنظمة غير متجانسة.

> **الأطروحة:** Interoperability شرط نقل، لا شرط حاكمية. يجب أن تُحفظ SemanticInvariants عبر claim، authority، purpose: encoding/translation. إذا فقد invariant، يتوقف handoff. status، failure semantics، rollback، human escalation.

$$SSP = \Pi_k \text{Preserve}(I_k \mid \text{encode} \rightarrow \text{transport} \rightarrow \text{decode});\ \text{SovereignHandoff iff } \min_k \text{Preserve}(I_k) \geq \tau_k$$

السلالة: MOUSSAIXT adapters ← MUG governance ← MURI plural reasoning

التجربة الحاكمة: MCP/A2A adapters مع tasks تتضمن delegation/rollback/dissent. حقن schema-loss وتحويلات مختلفة. قياس transport success مقابل recovery، authority loss، semantic success.

مفندات مباشرة:

- protocol-native schemas تكفي لكل invariants المستهدفة.
- semantic layer يخلق تعقيداً بلا فشل measurable.
- round-trip tests لا تتنبأ deployment bugs.

الحد: لا يزعم أن كل بروتوكول يفتقر للحوكمة؛ يختبر فقط invariants التي يحتاجها النظام فوق قدرات النقل.

ضغط خارجي: MCP 2026؛ A2A 1.0 2026؛ Governance Gaps in Protocols 2026؛ MPAC 2026. لا يُستخدم كدليل صحة داخلي.

M14-12 - نظرية مساعف لمعايرة الثقة في التفويض

Moussaif Delegation Trust Calibration Theory

الفجوة: ثقة الوكيل أو الأداة غالباً تتحول إلى reputation scalar ثابت، بينما reliability تعتمد على task type وauthority tier وhorizon وrecovery. نحتاج trust مشروطاً ومعايراً.

> **الأطروحة:** DelegationTrust distribution على context، لا رقم مطلق. يتحدث من verified outcomes وconstraint compliance وrecovery، مع discount للتشابه بين الأدلة. الثقة تحدد scope وverification depth لا مجرد اختيار agent.

$$DT(c) = Calib[P(success, compliance, recovery | c)] \cdot Independence Evidence - Uncertainty - Drift Penalty$$

السلالة: M-AFI Delegation Trust Dynamics ← Authority-Preserving Delegation ← Agent Mesh

التجربة الحاكمة: agent pool بقدرات متغيرة وفشل جزئي. قارن DT contextual/reputation/global/best-average/random. قياس violations، success، calibration، recovery، verification cost.

مفندات مباشرة:

- success-rate بسيط يساوي النتائج.
- قلة البيانات تجعل calibration ضارة.
- trust gating يسبب over-conservatism مزمن.

الحد: الثقة هنا أداة routing/verification، لا وصف أخلاقي أو نفسي للوكيل.

ضغط خارجي: AgencyBench 2026؛ ToolPRMBench 2026؛ NIST identity/authorization 2026. لا يُستخدم كدليل صحة داخلي.

M14-13 - نظرية مساعف لجبهة عمومية الإغلاق

Moussaif Generality Closure Frontier Theory

الفجوة: متوسط benchmark لا يحدد العمومية؛ أداء النظام قد ينهار عند تغيير model/tool/horizon/resources. نحتاج تمثيل frontier يوضح أين يبقى causal role وأين يتوقف.

> **الأطروحة:** العمومية خاصية لصمود mechanism عبر TransformSet مع resource envelope وفشل قابل للتفسير. يُرسم frontier بين calibration، performance، recovery، cost لا scalar واحداً.

GCF = ParetoFrontier(Quality, Calibration, Recovery, Cost | model,domain,tool,horizon,resource);
GenScore=min_T normalized_effect(T)

السلالة: M13-08 Closed-Loop Generality ← MUT transfer program ← MUG evidence promotion

التجربة الحاكمة: CUBE/AgencyBench/عدة tool harnesses مع frozen mechanism. قياس failure-signature، resource frontier، CI، effect sign stability transfer.

مفندات مباشرة:

- التحويلات المختارة تعكس التحيز ولا تتنبأ deployment.
- average performance يفسر نفس القرارات.
- min-effect metric مفرط الحساسية للoutlier.

الحد: لا يدعي universal intelligence؛ يحدد envelope متحققاً لما تم اختباره فقط.

ضغط خارجي: CUBE 2026؛ AgencyBench 2026؛ AgentGym2 2026. لا يُستخدم كدليل صحة داخلي.

M14-14 - نظرية مساعف للتحويل بين المعيار والنشر

Moussaif Benchmark-to-Deployment Morphism Theory

الفجوة: benchmark task قد يملك semantics ومصادر وضوابط مختلفة جذرياً عن deployment. نقل score مباشرة إلى claim منتج يخلق evaluation leakage.

> **الأطروحة:** نعرّف morphism يطابق invariants من benchmark إلى deployment: task intent، state observability، action space، authority، cost model، failure consequences، verifier. claim لا ينتقل إلا إذا حُفظت invariants ذات الصلة.

$$\text{MorphLoss} = \Sigma_k w_k \cdot d(I_k^{bench}, I_k^{deploy}); \quad \text{ClaimTransfer} \leq \exp(-\text{MorphLoss}) \cdot \text{EvidenceGrade}$$

السلالة: M13-08 Generality ← World-Effect Closure ← MUG evaluation

التجربة الحاكمة: اختبر mechanisms على benchmark ثم sandbox deployment مشابه مع controlled shifts. قياس false predicted vs observed transfer، failure modes، cost، confidence.

مفندات مباشرة:

- MorphLoss لا يتنبأ deployment gap.
- human expert mapping وحده يكفي.
- invariants لا تقبل operationalization.

الحد: morphism استعارة صورية/هندسية وليست بالضرورة morphism category-theoretic رسمي.

ضغط خارجي: CUBE 2026؛ DeepPlanning 2026؛ AgencyBench 2026. لا يُستخدم كدليل صحة داخلي.

M14-15 - نظرية مساعف لتفكيك القدرة المقاوم للهارنس

Moussaif Harness-Resistant Capability Decomposition Theory

الفجوة: قد يبدو mechanism قوياً لأن harness يمنحه prompt أطول أو tools أكثر أو retries أو verifier أقوى. يلزم تفكيك القدرة إلى /model/harness/mechanism interactions.

الأطروحة: نعرّف factorial ablation matrix تفصل main effects والتفاعلات تحت matched budgets. القدرة المنسوبة للنظرية هي causal contribution المتبقي عبر عدة harness variants، لا أفضل run.

$$Y = \beta 0 + \beta M \cdot Model + \beta H \cdot Harness + \beta K \cdot Mechanism + \beta MK + \beta HK + \beta MHK + \varepsilon;\ Credit_K = robust\ causal\ contrast$$

السلالة: MUT anti-inflation rule ← M13 Theory-to-Kernel Closure ← MCD routing

التجربة الحاكمة: 4-8 mechanisms على عدة models/harnesses مع sham scaffold. قياس transfer، interaction، cost، CI، β. معيار الترقية: causal contrast وليس raw score.

مفندات مباشرة:

- التصميم factorial مكلف ولا يزيد قراراً عملياً.
- matched budget مستحيل لدرجة تبطل attribution.
- الأثر يظهر فقط في harness واحد لكنه مهم واقعياً.

الحد: لا ينفي قيمة harness؛ يفصل فقط مصدر القدرة كي لا تنسب معمارية كاملة إلى اسم واحد.

ضغط خارجي: AgencyBench 2026؛ Trajectory Attribution 2026؛ Plan-RewardBench 2026. لا يُستخدم كدليل صحة داخلي.

M14-16 - نظرية مساعف لدَين استقلال المقيم

Moussaif Evaluator Independence Debt Theory

الفجوة: كلما شارك evaluator البيانات أو النموذج أو prompt lineage أو reward loop مع النظام المحسن، ازداد خطر self-confirmation. لا يكفي وجود verifier اسمي.

> **الأطروحة:** Evaluator Independence Debt يراكم اعتماداً بين optimizer وadaptive reuse، shared training data، shared model family: judge، evaluator tuning، leakage على نفس failures. الترقية تتطلب سداد جزء من الدين بمقيم مستقل أو formal/world evidence.

$$EID = w1·ModelOverlap+w2·DataOverlap+w3·AdaptiveCoupling+w4·Leakage+w5·MetricGaming - w6·ExternalGrounding$$

السلالة: Falsification Court ← MUG evidence governance ← M13 External Grounding

التجربة الحاكمة: اختبر self-refinement تحت hidden human subset، formal verifier، different-model judge، same-model judge. قياس optimism gap، rank consistency، replication.

مفندات مباشرة:

- EID لا يرتبط بreplication failure.
- external judges يضيفون noise أكبر من bias الذي يزيلونه.
- model overlap ليس predictor بعد ضبط task.

الحد: الدين لا يعني فساد evaluator، هو مخاطرة dependence يجب قياسها وسدادها نسبياً.

ضغط خارجي: Recursive Self-Improvement survey 2026؛ Plan-RewardBench 2026؛ GPT-Red 2026. لا يُستخدم كدليل صحة داخلي.

M14-17 - نظرية مساعف لعائد إعادة الإنتاج العلمي

Moussaif Scientific Reproduction Yield Theory

الفجوة: عدد التجارب الناجحة داخل المختبر لا يساوي معرفة مستقرة. نحتاج metric يكافئ reproduction المستقل والمتنوع ويعاقب dependence والانتقاء.

> **الأطروحة:** Reproduction Yield هو نسبة محاولات مستقلة تعيد sign/threshold الأساسي ضمن uncertainty، موزونة بتنوع model/domain/team/harness ومسبقة التسجيل، مع احتساب null/negative attempts.

$$SRY = \Sigma_j Indep_j \cdot Diversity_j \cdot Reproduce_j / \Sigma_j Attempt_j; \text{ report with CI and failure taxonomy}$$

السلالة: MUG publication governance ← Evidence Closure ← M13 External Grounding

التجربة الحاكمة: أعد 3-5 mechanisms عبر models/tools/harnesses وفريق/agent evaluator منفصل. قياس failure signature، CI overlap، sign stability، reproduction cost.

مفندات مباشرة:

- SRY يعاقب innovations حساسة للبيئة بشكل غير عادل.
- independence يصعب تعريفها.
- yield مرتفع بسبب tasks سهلة غير حاسمة.

الحد: لا يطلب تطابقاً مطلقاً، يعرّف ما يجب أن يبقى وما يسمح له أن يتغير مسبقاً.

ضغط خارجي: CUBE 2026؛ Trajectory Attribution 2026؛ Recursive Self-Improvement 2026. لا يُستخدم كدليل صحة داخلي.

M14-18 - نظرية مساعف لثبات الطفرة الدستورية

Moussaif Constitutional Mutation Stability Theory

الفجوة: النظام المتطور قد يحسن performance بتعديل router/memory/tools/evaluator، لكنه قد يغير القيود التي تجعل التحسن مقبولاً. نحتاج فصل mutable capabilities عن constitutional invariants.

> **الأطروحة:** كل mutation تمر بـbranch مع immutable root constitution أو externally governed version. تُقبل فقط إذا حسنت هدفاً معلناً دون regression حرجة، مع rollback وindependent evaluation. تعديل الدستور نفسه يحتاج meta-governance منفصلة ولا يثبت صحته بالحلقة التي تقترحه.

CMS = Gain_validated · InvariantPreservation · RollbackVerified · EvaluatorIndependence - CriticalRegressionRisk

السلالة: MUG constitution ← M13 Recursive Improvement ← Governance Space

التجربة الحاكمة: حلقات تحسين harness/memory/router تحت branches؛ قارن CMS، fixed constitution، unrestricted self-edit. قياس regressions، gain، reproduction، evaluator gap، rollback success.

مفندات مباشرة:

- root constraints تمنع اكتشاف تحسينات ضرورية.
- rollback لا يمكن تحقيقه في learning systems المستمرة.
- independent evaluation لا يقلل catastrophic regression.

الحد: لا يثبت أمان recursive self-improvement المفتوح؛ إنه بروتوكول تجريبي لتقييد طفرة محددة.

ضغط خارجي: Recursive Self-Improvement 2026؛ GPT-Red 2026؛ DeepMind AI Control 2026؛ Anthropic trust/containment 2026. لا يُستخدم كدليل صحة داخلي.

بنية Closure Proof Packet - العقدة التي يجب أن يمتلكها كل ادعاء

```
theory_id, version, canonical_owner, source_anchors
claim_scope, assumptions, state_variables, observables, estimators
trigger, operator, predicted_delta, nearest_baseline
feature_flag, telemetry_schema, world_readback
falsifiers, negative_activation_case, failure_modes
authority_required, rollback_contract
experiment_ids, seeds, budget_manifest, harness_fingerprint
effect_size, confidence_interval, counterevidence
transfer_envelope, replication_receipts
verdict: PROMOTE|MERGE|REVISE|HOLD|REJECT|ROLLBACK
```

هذه الوحدة هي نقطة الاتصال بين MUT وMUG وMOUSSAIXT. MUT يملك معنى النظرية؛ MUG يملك شروط ترقيتها؛ MOUSSAIXT يملك تفعيلها ورصدها؛ evidence packet يملك ما حدث فعلاً.

> إذا فُقد field حرج لا يُملأ تخميناً. يتحول إلى unresolved item ويمنع promotion بالقدر الذي يؤثر فيه على claim.

مؤشر الإغلاق: Vector أولاً، Scalar ثانياً

لمنع تعويض ضعف حاسم بقوة غير مرتبطة، يُحفظ **Closure Vector** أولاً: Qc (وضوح claim/question)، Wc (world readback)، Ec (evidence)، Lc (lineage/provenance)، Tc (transfer)، Ac (authority/action)، Mc (mechanism). لا يجوز اختزال التقرير إلى scalar واحد عند وجود gate صفرية.

$$\mathcal{C} = [Q_c, M_c, A_c, W_c, E_c, L_c, T_c]$$

$$CCI_{geo} = (Q_c \cdot M_c \cdot A_c \cdot W_c \cdot E_c \cdot L_c \cdot T_c)^{(1/7)} \cdot (1 - U_{critical})$$

الصيغة الثانية heuristic dashboard فقط. الترقيات الحاسمة تستخدم hard gates: WorldReadback=0 يمنع claim عن world-effect، وIndependentReplication=0 يمنع E5 مهما ارتفع المتوسط.

المقياس	السؤال
Unresolved Critical Rate	ما نسبة المجهول الذي يستطيع قلب الحكم؟
Unsupported Claim Rate	كم claim لا يملك source/evidence مناسباً؟
Mechanism Isolation Rate	كم نظرية قابلة للتعطيل والمقارنة؟
World Readback Rate	كم action يُغلق بقراءة حالة مستقلة؟
Replication Yield	كم أثر يعاد خارج harness الأصلي؟

مختبر السببية: الحد الأدنى للتجربة القابلة للنشر

1. جمّد environment fingerprint، budgets، prompts، data، tools، model/version.
2. عرّف nearest baseline وsham scaffold.
3. نفذ feature flag للآلية واحفظ raw trajectories.
4. كرر seeds؛ لا تحذف failed runs.
5. عرّف primary metric قبل رؤية النتيجة.
6. سجل effect size وCI وcost وlatency وcritical regressions.
7. أدخل adversarial/boundary/negative activation cases.
8. انقل إلى model/domain/tool/horizon مختلف.
9. افصل evaluator أو استخدم formal/world-state proof حيث يمكن.
10. أصدر verdict مع theory revision لا مع score فقط.

> هذا المخطط يجعل الفشل نتيجة علمية: إن لم يزد المكوّن قيمة هامشية، يُدمج أو يخفض أو يرفض. لا يُفسر null تلقائياً بأنه «حاجة إلى نموذج أكبر».

خارطة الأصول الناتجة عن Part XIV

الطبقة	العدد	الوظيفة
نظريات M14	18	أجسام نظرية جديدة موزعة على 6 مسارات.
Asset specifications	180	72 أدوات + 36 runtime/replay stack assets + 36 methods + 36 protocols.
Training recipes	72	4 وصفات لكل نظرية: transfer، matched-budget، adversarial boundaries، isolation.
Core experiments	54	3 لكل نظرية: transfer/replication، boundary/adversarial، causal ablation.
Research pathway papers	5	P14-A، P14-B، P14-C، P14-D، P14-E/F.

الأعداد ليست دليلاً على الجودة؛ فائدتها أنها تجعل coverage قابلاً للحساب وتمنع «مفهوم بلا اختبار» و«أداة بلا owner».

منهج ترتيب التنفيذ: P0 إلى P3

الأولوية	الهدف	القبول
P0	M14-01/04/06/08/13/16: observation، proof trajectory، authority، memory، generality، evaluator debt.	+ runnable feature flags + raw traces + baseline.
P1	M14-02/03/05/09/12/14/15: world model، residual causality، commitment، consistency، trust، morphism، harness decomposition.	causal contrast + boundary tests.
P2	M14-07/10/11/17/18: persistent attention، organizations، protocol sovereignty، reproduction، constitutional evolution.	long-horizon/multi-agent/replication suites.
P3	التوسعات الجديدة فقط بعد نتائج P0-P2.	لا نظرية جديدة إذا كانت مجرد alias أو لا تضيف prediction/test.

Implement → Measure → Falsify → Repair → Transfer → Replicate → Only then Promote

مراجع الضغط العلمي والمراسي الداخلية

Survey: foundational/self-evolving/collective agentic reasoning; open problems in .arXiv:2601.12538 .Agentic Reasoning for Large Language Models - **E01**
.world models, long horizons and governance

/World-model revision from prediction-observation mismatch with episodic .arXiv:2606.30639 .Self-Evolving World Models for LLM Agent Planning - **E02**
.semantic memory and selective foresight

-Decision .arXiv:2606.25421 .Beyond Next-Observation Prediction: Agent-Authored World Modeling for Sequential Decision Making - **E03**
.oriented world-model targets derived from policy information needs

Compact memory cues plus preserved raw trajectory pages .arXiv:2605.24468 .SAM: State-Adaptive Memory for Long-Horizon Reasoning Agent - **E04**
.for state-adaptive recall

System-level .arXiv:2606.06448 .Agent Memory: Characterization and System Implications of Stateful Long-Horizon Workloads - **E05**
.taxonomy and write/read lifecycle trade-offs across agent memory systems

-Fine-grained attribution benchmark with >1,300 annotated trajectories and leave-one .arXiv:2608.06909 .Long-Horizon Agent Trajectory Attribution - **E06**
.out baselines

ACL 2026, 10.18653/ .AgencyBench: Benchmarking the Frontiers of Autonomous Agents in 1M-Token Real-World Contexts - **E07**
.tasks/32 scenarios; long trajectories, many tool calls and resource-efficiency/self-correction gaps 138 .v1/2026.acl-long.337

.ACL 2026, 10.18653/v1/2026.acl-long.335 .DeepPlanning: Benchmarking Long-Horizon Agentic Planning with Verifiable Constraints - **E08**
.Long-horizon constrained planning with active information gathering; frontier systems still struggle

Trajectory-level reward modeling; evaluator .ACL 2026, 10.18653/v1/2026.acl-long.1062 .Aligning Agents via Planning: Plan-RewardBench - **E09**
.performance degrades with long trajectories

Findings ACL 2026, 10.18653/ .ToolPRMBench: Evaluating and Advancing Process Reward Models for Tool-using Agents - **E10**
.Step-level PRM benchmark for tool-use trajectories and plausible hard negatives .v1/2026.findings-acl.602

Dense multidimensional step .ACL 2026, 10.18653/v1/2026.acl-long.1574 .PRInTS: Reward Modeling for Long-Horizon Information Seeking - **E11**
.scoring plus trajectory summarization for information-seeking agents

.Common protocol separates task, benchmark, package and registry concerns .arXiv:2603.15798 .CUBE: A Standard for Unifying Agent Benchmarks - **E12**

-Difference → Question → Route → Act → Verify → Reconstruct; E0 .MOUSSAIF internal corpus .MOUSSAIF Unified AGI - canonical research book - **I01**
.E5 evidence discipline

layer architecture; claim contracts; world-effect-12 .MOUSSAIF internal corpus .MOUSSAIF AGI OS / MOUSSAIXT Deep Engineering Architecture - **I02**
.loop; memory/evolution; tool and authority contracts

Difference-before-volume, governing question, MCD/QMCD, world modeling and .MOUSSAIF internal corpus .ParallaxWorld-MCD FULL AGI OS - **I03**
.falsification

.Canonical theory graph, anti-inflation, theory contracts and crosswalk .MOUSSAIF internal corpus .MUT-AGI-02 Unified Theory - **I04**

→ Theory unification → program governance → governed runtime .MOUSSAIF internal corpus .MUT × MUG × MOUSSAIXT integrated framework - **I05**
.evidence

.Belief/action partition, reproducibility, externalities, accountability and recourse .MOUSSAIF internal corpus .Part X - Full Spectrum OS Closure - **I06**

,Reconstructive memory, unresolved state, writeback .MOUSSAIF internal corpus .Part XI - Memory, Continuity, Self-Reconstruction & Evolution - **I07**
.portability and coupled evolution

Source integration registry; presence/integration/validation are .MOUSSAIF internal corpus .Part XII - Total Research-Paper Integration Closure - **I08**
.distinct states

مراجع إضافية وحدود الاستعمال

E13 - AI Organizations Can Be More Effective but Less Aligned than Individual Agents. arXiv:2604.10290 Multi-agent organizations can increase effectiveness while degrading alignment on studied tasks.

E14 - Governance Gaps in Agent Interoperability Protocols. arXiv:2606.31498 Analyzes missing governance primitives above interoperability protocols; treated as design pressure, not settled fact.

E15 - MPAC: A Multi-Principal Agent Coordination Protocol. arXiv:2604.09744 Explicit session/intent/operation/conflict/governance semantics for multi-principal coordination.

E16 - Recursive Self-Improvement in AI: From Bounded Self-Refinement to Autonomous Research Loops. arXiv:2607.07663 Survey of 1,250 papers; evaluator/grounding hierarchy and limits of open-ended RSI.

E17 - GPT-Red: Unlocking Self-Improvement for Robustness. OpenAI, 2026-07-15 Automated red-teaming and adversarial training; attacker kept separate from deployed models.

E18 - Securing the future of AI agents / AI Control Roadmap. Google DeepMind, 2026-06-18 Control framing around detection, prevention, response and supervisors for capable agents.

E19 - Trustworthy agents in practice / containment across products. Anthropic, 2026 Human control, secure interaction, transparency/privacy and containment/blast-radius engineering.

E20 - Software and AI Agent Identity and Authorization - Concept Paper. NIST NCCoE, 2026-02-05 Identity/authorization control problem for software and AI agents across tools and data.

E21 - Model Context Protocol Specification. MCP 2026-07-28 Tool/context interoperability, tasks and authorization hardening.

E22 - Agent2Agent Protocol v1.0. A2A 1.0.0, 2026 Discovery, task management and communication between agent systems.

I09 - Part XIII - Theoretical Renaissance. MOUSSAIF internal corpus Causal open-world verified intelligence; theory-to-kernel causal closure; external grounding.

I10 - Conscious AGI / QRET / Governance Space corpus. MOUSSAIF internal corpus Functional self-model/meta-governance with explicit phenomenology boundary.

I11 - Question Density / QMCD / QGWEI / MURI corpus. MOUSSAIF internal corpus Governing questions, plural reasoning, explicit ignorance and evidence geometry.

I12 - Goal-as-OS / World Effect / M-HKOS / MATI-MACIT. MOUSSAIF internal corpus Goal contracts, verified world-state change, knowledge-as-OS and selective internalization.

I13 - MUT MUG ALL artifact memory map - 2026-09-04. MOUSSAIF internal corpus 343 artifacts; duplicate groups are not independent evidence; retrieval is family/canonicality based.

حد المراجع: المرجع الخارجي يحدد اتجاه مجتمع أو baseline أو فجوة منشورة. لا يمنح صحة تلقائية لمصطلح داخلي. المرجع الداخلي يحفظ lineage والآلية المقترحة، ولا يحل محل validation مستقل.