MOUSSAIF UNIFIED AGI · PART XIV · P14-C · 2026

الورقة الثالثة: الذاكرة السببية والانتباه والاستمرار عبر الأفق

& Causal Memory Compression, Persistent Attention Reconstructive Consistency

P14-C - إغلاق الذاكرة والانتباه عبر الأفق

MOUSSAIF Unified AGI Part XIV - Executable Theoretical Completion ضمن

Moussaif Zouhair / زهير مساعف · MUT × MUG × MOUSSAIXT

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

الملخص والأطروحة

تعالج هذه الورقة فجوة محددة في برنامج MOUSSAIF: الانتقال من memory كاسترجاع إلى causal compression + eventual consistency + persistent attention يحدد متى نتذكر ومتى ننتظر ومتى نعيد البناء. لا تُقدم مساهمة الورقة على أنها «إثبات AGI»، بل كحزمة فرضيات ذات state variables وآليات قابلة للتعطيل وأدوات وتجارب ومفندات. معيار القيمة هو ما إذا أضافت النظرية فرقاً سببياً فوق baseline بسيط وتحمّل ذلك الفرق النقل بين شروط مختلفة.

تستوعب الورقة السلالة الداخلية عبر canonical owners بدل نسخ المحتوى: MURI، HKOS/M-HKOS، MICAI/MICAGI/MICAGFI، MCD/QMCD/MCAI/ Parts X-XIII، MOUSSAIXT، MUG، reconstructive memory، world-effect، MSRAK. المجتمع العلمي يوفر خطوط ضغط حديثة في long-horizon agents، control، multi-agent governance، benchmark unification، process reward models، memory، world models وself-improvement.

> **المساهمة:** نظرية مساعف لديناميات الانتباه المستمر، نظرية مساعف لحد الضغط السببي للذاكرة، نظرية مساعف للاتساق النهائي للذاكرة المعاد بناؤها. كل مساهمة مرتبطة بمعادلة falsifier، feature flag، design-level، تجربة transfer envelope، matched-budget، وأداة runtime.

سؤال الورقة: ما الحد الأدنى من البنية الإضافية التي تجعل إغلاق الذاكرة والانتباه عبر الأفق قابلاً للقياس والتفنيد والربط بأثر العالم، دون تحويل density النظرية إلى claim غير مدعوم؟

العقد الإبستيمي ومنع تضخم الادعاء

طبقة	الوضع في الورقة
External literature	مرساة لمشكلة أو baseline أو mechanism prior؛ ليست دليلاً على صحة M14.
MOUSSAIF lineage	مصدر الهوية، التعريف، العلاقات، والمقترحات؛ evidence grade مستقل.
Equations	Formal specifications/heuristics حتى تتوافر estimators ومعايرة ونتائج.
Tools	SPECIFICATION إلى أن يوجد runtime trace وtests.
Experiments	تصاميم قابلة للتنفيذ؛ لا تُنسب نتائج غير منفذة.

$$\text{Claim strength} \leq \min(\text{Source strength}, \text{Measurement strength}, \text{Causal isolation}, \text{Replication strength})$$

كل نظرية تحصل على negative activation case، لأن المكوّن الذي «يفيد دائماً» يصعب التفريق فيه بين آلية حقيقية وبين تعقيد إضافي. وكل نجاح يُقارن بـnearest parent وbaseline بسيط وsham scaffold.

مصفوفة الضغط من حالة الفن 2026

المصادر التالية لا تُستخدم لزخرفة المراجع؛ كل مصدر يفرض قيد تصميم أو benchmark أو failure mode يجب أن يواجهه المسار.

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

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

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

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

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

خريطة الاندماج الداخلي: corpus → canonical lineage

بدلاً من نسخ كل ورقة داخلية، نربطها بوظيفة. Part XIV هو continuation لا إعادة صياغة للنصوص السابقة. يُعاد استخدام ما سبق فقط عندما يخدم causal contract جديداً.

العائلة	ما الذي تُورثه لهذه الورقة؟
MCD/QMCD/Question Density	اختيار الفرق الحاكم وتخصيص الحساب تحت سؤال.
MCAI/MICAI/MICAGI	التكامل بين الأوعية، الدليل، الحاكمية، والفعل.
HKOS/M-HKOS	المعرفة كـobjects وcontracts وretrieval/writeback.
MURI/MSRAK/FSKE	تعدد أنماط الاستدلال، courts، claim ranks، وexplicit ignorance.
MOUSSAIXT/MUG	authority، runtime schemas، evidence promotion، rollback.
Parts X-XIII	causal open-world closure، paper integration، memory/reconstruction، world-effect.

> أي construct تاريخي يبقى lineage node إن لم يثبت أنه مطابق دلالياً وسلوكياً لعقدة أخرى. الدمج ≠ حذف التاريخ.

M14-07 - الفجوة والأطروحة والنسب

Moussaif Persistent Attention Dynamics Theory

الفجوة الدقيقة

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 الجديدة: 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.

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **Question Density**؛ **M-TAOS**؛ **M13-04 Long-Horizon Continuity**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

SentinelBench 2026؛ SAM 2026؛ Agent Memory 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-07 - الصياغة الرسمية والـobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
A_i	شدة الانتباه لحالة i
ρ_i	معدل البقاء/التلاشي
Surprise	فرق جديد
Obligation	التزام غير مغلق
Risk	خطر
Dependency	مركزية اعتمادية
VerifiedClosure	شاهد إغلاق

افتراضات الصياغة

- المتغيرات task-scoped وليست خصائص كونية ثابتة.
- كل variable يحتاج estimator أو proxy موصوفاً قبل الاستخدام التجريبي.
- الأوزان/العتبات تُعاير على train/validation وتُقفل قبل held-out evaluation.
- لا تُفسر العلاقة الجبرية كقانون تجريبي قبل نتائج مستقلة.
- sensitivity analysis إلزامي عندما يمكن للأوزان قلب verdict.

قاعدة formal closure

إذا لم يمكن تحويل متغير إلى observation أو estimator أو intervention، يبقى في **Unresolved Formal Ledger** ويُخفض claim. الصياغة لا تتقوى بتعقيد الرمز؛ تتقوى بزيادة قابلية القياس والتفنيد.

M14-07 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-07 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. استخرج attention items من mission/state.
2. امنح decay مختلفاً حسب type.
3. أعد تنشيط item عند evidence/clock/dependency event.
4. استدع raw memory فقط عند threshold.
5. امنع closure دون receipt.
6. سجل missed attention وfalse persistence.

عقد النواة

preconditions: source freshness, authority, schema, budget postconditions: typed output + predicted delta + verifier target	failure: malformed state, stale evidence, no-effect, adverse effect, authority mismatch recovery: stop/retry/alternate/rollback/escalate

يجب أن توجد feature flag تسمح بإطفاء الميكانيزم دون تغيير بقية harness قدر الإمكان. إن لم توجد هذه القابلية، يصبح causal attribution ضعيفاً حتى لو كان النظام يعمل جيداً.

M14-07 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- يخفض نسيان obligations البعيدة.
- يقل context loading مقارنة full-history.
- المهام monitoring تتحسن خصوصاً عند events متباعدة.

أربعة خصوم نظريين

- Baseline محافظ: أبسط heuristic أو policy يؤدي الوظيفة بلا اسم جديد.
- تفسير harness: المكسب من tools/context/retries لا من الميكانيزم.
- تفسير model: backbone أقوى يزيل الحاجة للمكوّن.
- تفسير measurement: metric نفسها تكافئ السلوك الذي صُممت النظرية لإظهاره.

قرار المحكمة

إذا فشل causal ablation يُعطى **MERGE** أو **REVISE** أو **REJECT** بحسب ما إذا كان الفرق غير مستقل، محدود المجال، أو غير موجود. لا يسمح بتغيير اسم النظرية بعد null result مع الحفاظ على المضمون نفسه لتجنب الفشل.

M14-07 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
XIV-A061	Persistent Attention Dynamics State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A062	Persistent Attention Dynamics Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A063	Persistent Attention Dynamics Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A064	Persistent Attention Dynamics Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A065	Persistent Attention Dynamics Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
XIV-A066	Persistent Attention Dynamics Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

- Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

الحد الأدنى للتجارب

- Minimal Causal Ablation: switchable implementation مقابل nearest parent/baseline تحت matched budget. | metrics: primary theory metric + task outcome + calibration + cost + critical regressions
- Boundary/Adversarial Test: اختبار failure modes والحدود بحالات مصممة لكسر التعميم. | metrics: false activation + missed failure + rollback + evidence integrity
- Transfer/Replication: نقل الآلية إلى model/domain/tool/horizon مختلف مع CI + preregistered metric. | metrics: effect sign stability + failure-signature transfer + resource frontier

> لا يُستخدم أي generated trajectory كتدريب إيجابي لمجرد أنه من النظام نفسه. يلزم verifier label أو world-state evidence أو rubric مستقل بحسب نوع المهمة.

M14-08 - الفجوة والأطروحة والنسب

Moussaif Causal Memory Compression Bound Theory

الفجوة الدقيقة

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

> **الأطروحة الجديدة:** 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.

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **HKOS/M-HKOS؛ Self-Reconstruction Memory؛ M-AFI Knowledge-to-Skill Transfer**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

SAM 2026؛ Agent Memory 2026؛ PRInTS 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-08 - الصياغة الرسمية والـobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
M_comp	الذاكرة المضغوطة
D*	متغيرات القرار
C*	العلاقات السببية اللازمة
Regret_reconstruct	خسارة القرار بعد إعادة البناء
CausalRecall	استرجاع السبب/الشرط الحاكم

افتراضات الصياغة

- المتغيرات task-scoped وليست خصائص كونية ثابتة.
- كل variable يحتاج estimator أو proxy موصوفاً قبل الاستخدام التجريبي.
- الأوزان/العتبات تُعاير على train/validation وتُقفل قبل held-out evaluation.
- لا تُفسر العلاقة الجبرية كقانون تجريبي قبل نتائج مستقلة.
- sensitivity analysis إلزامي عندما يمكن للأوزان قلب verdict.

قاعدة formal closure

إذا لم يمكن تحويل متغير إلى observation أو estimator أو intervention، يبقى في **Unresolved Formal Ledger** ويُخفض claim. الصياغة لا تتقوى بتعقيد الرمز؛ تتقوى بزيادة قابلية القياس والتفنيد.

M14-08 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-08 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. صنف memory facts إلى outcome/decision/cause/authority/failure.
2. اختبر reconstruction على queries مستقبلية مُولدة من state.
3. احتفظ raw page pointers عند فقد causal detail.
4. تعلم compression policy حسب utility لا length فقط.
5. أعد فتح raw trace عند uncertainty.
6. قِس construction/read costs.

عقد النواة

preconditions: source freshness, authority, schema, budget postconditions: typed output + predicted delta + verifier target	failure: malformed state, stale evidence, no-effect, adverse effect, authority mismatch recovery: stop/retry/alternate/rollback/escalate

يجب أن توجد feature flag تسمح بإطفاء الميكانيزم دون تغيير بقية harness قدر الإمكان. إن لم توجد هذه القابلية، يصبح causal attribution ضعيفاً حتى لو كان النظام يعمل جيداً.

M14-08 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- الضغط السببي يتفوق على summary-only في long-horizon recovery.
- قد تكون ذاكرة أطول أرخص إجمالاً إذا قلل construction overhead.
- السببية الحرجة تتطلب حفظ counterexamples لا facts فقط.

أربعة خصوم نظريين

- Baseline محافظ: أبسط heuristic أو policy يؤدي الوظيفة بلا اسم جديد.
- تفسير harness: المكسب من tools/context/retries لا من الميكانيزم.
- تفسير model: backbone أقوى يزيل الحاجة للمكوّن.
- تفسير measurement: metric نفسها تكافئ السلوك الذي صُممت النظرية لإظهاره.

قرار المحكمة

إذا فشل causal ablation يُعطى **MERGE** أو **REVISE** أو **REJECT** بحسب ما إذا كان الفرق غير مستقل، محدود المجال، أو غير موجود. لا يسمح بتغيير اسم النظرية بعد null result مع الحفاظ على المضمون نفسه لتجنب الفشل.

M14-08 - التجربة، الأدوات والتدريب

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.

ID	الأداة/الأصل	النوع	الدور
-XIV A071	Causal Memory Compression Bound State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A072	Causal Memory Compression Bound Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A073	Causal Memory Compression Bound Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A074	Causal Memory Compression Bound Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A075	Causal Memory Compression Bound Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A076	Causal Memory Compression Bound Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

- Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

الحد الأدنى للتجارب

- Minimal Causal Ablation: switchable implementation مقابل nearest parent/baseline تحت matched budget. | metrics: primary theory metric + task outcome + calibration + cost + critical regressions
- Boundary/Adversarial Test: اختبار failure modes والحدود بحالات مصممة لكسر التعميم. | metrics: false activation + missed failure + rollback + evidence integrity
- Transfer/Replication: نقل الآلية إلى model/domain/tool/horizon مختلف مع CI + preregistered metric. | metrics: effect sign stability + failure-signature transfer + resource frontier

> لا يُستخدم أي generated trajectory كتدريب إيجابي لمجرد أنه من النظام نفسه. يلزم verifier label أو world-state evidence أو rubric مستقل بحسب نوع المهمة.

M14-09 - الفجوة والأطروحة والنسب

Moussaif Reconstructive Memory Eventual Consistency Theory

الفجوة الدقيقة

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 الجديدة:** 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.

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **Self-Reconstruction Memory**، **ContinuityScanner**، **MUG version governance**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

Agent Memory 2026؛ SAM 2026؛ CUBE-style interface separation 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-09 - الصياغة الرسمية والobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
S_rebuilt	الحالة المعاد بناؤها
S_auth	المصدر السلطوي
τ_risk	عتبة حسب الخطر
B_d	ميزانية الانحراف المسموح

افتراضات الصياغة

- المتغيرات task-scoped وليست خصائص كونية ثابتة.
- كل variable يحتاج estimator أو proxy موصوفاً قبل الاستخدام التجريبي.
- الأوزان/العتبات تُعاير على train/validation وتُقفل قبل held-out evaluation.
- لا تُفسر العلاقة الجبرية كقانون تجريبي قبل نتائج مستقلة.
- sensitivity analysis إلزامي عندما يمكن للأوزان قلب verdict.

قاعدة formal closure

إذا لم يمكن تحويل متغير إلى observation أو estimator أو intervention، يبقى في **Unresolved Formal Ledger** ويُخفض claim. الصياغة لا تتقوى بتعقيد الرمز؛ تتقوى بزيادة قابلية القياس والتفنيد.

M14-09 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-09 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. قسم fields حسب criticality.
2. احتفظ immutable event log عند الإمكان.
3. أنشئ version vectors/lineage.
4. اسمح cached summaries للحقول منخفضة الأثر.
5. قبل critical action نفذ reconcile/readback.
6. سجل conflict resolution ولا تمح history.

عقد النواة

failure: malformed state, stale evidence, no-effect, adverse effect, authority mismatch recovery: stop/retry/alternate/rollback/escalate	preconditions: source freshness, authority, schema, budget postconditions: typed output + predicted delta + verifier target

يجب أن توجد feature flag تسمح بإطفاء الميكانيزم دون تغيير بقية harness قدر الإمكان. إن لم توجد هذه القابلية، يصبح causal attribution ضعيفاً حتى لو كان النظام يعمل جيداً.

M14-09 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- تخفض stale-memory failures دون كلفة strong consistency الشاملة.
- الحقول الحرجة تتقارب أسرع من الثانوية.
- يمكن تفسير drift عبر conflict ledger بدلاً من تحديثات صامتة.

أربعة خصوم نظريين

- Baseline محافظ: أبسط heuristic أو policy يؤدي الوظيفة بلا اسم جديد.
- تفسير harness: المكسب من tools/context/retries لا من الميكانيزم.
- تفسير model: backbone أقوى يزيل الحاجة للمكوّن.
- تفسير measurement: metric نفسها تكافئ السلوك الذي صُممت النظرية لإظهاره.

قرار المحكمة

إذا فشل causal ablation يُعطى **MERGE** أو **REVISE** أو **REJECT** بحسب ما إذا كان الفرق غير مستقل، محدود المجال، أو غير موجود. لا يسمح بتغيير اسم النظرية بعد null result مع الحفاظ على المضمون نفسه لتجنب الفشل.

M14-09 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A081	Reconstructive Memory Eventual Consistency State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A082	Reconstructive Memory Eventual Consistency Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A083	Reconstructive Memory Eventual Consistency Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A084	Reconstructive Memory Eventual Consistency Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A085	Reconstructive Memory Eventual Consistency Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A086	Reconstructive Memory Eventual Consistency Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر /domain/tool/model horizon.

وصفات التدريب/التقييم

- Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

الحد الأدنى للتجارب

- Minimal Causal Ablation: switchable implementation مقابل nearest parent/baseline تحت matched budget. | metrics: primary theory metric + task outcome + calibration + cost + critical regressions
- Boundary/Adversarial Test: اختبار failure modes والحدود بحالات مصممة لكسر التعميم. | metrics: false activation + missed failure + rollback + evidence integrity
- Transfer/Replication: نقل الآلية إلى model/domain/tool/horizon مختلف مع CI + preregistered metric. | metrics: effect sign stability + failure-signature transfer + resource frontier

> لا يُستخدم أي generated trajectory كتدريب إيجابي لمجرد أنه من النظام نفسه. يلزم verifier label أو world-state evidence أو rubric مستقل بحسب نوع المهمة.

الاقتران 1: كيف تتحول النظريات إلى نظام لا قائمة

المسار P14-C لا يعمل كسلسلة خطية صارمة، بل كـgraph. نحدد لكل عقدة inputs/outputs ونقاط failure. الاقتران المقبول يجب أن ينتج **complementarity measurable**: إذا كان تجميع نظريتين لا يتجاوز أفضل واحدة منهما تحت budget متطابق، نخفض claim عن «التكامل».

- M14-07: نظرية مساعف لديناميات الانتباه المستمر - output حاكم: شدة الانتباه لحالة i
- M14-08: نظرية مساعف لحد الضغط السببي للذاكرة - output حاكم: الذاكرة المضغوطة
- M14-09: نظرية مساعف للاتساق النهائي للذاكرة المعاد بناؤها - output حاكم: الحالة المعاد بناؤها

$$\text{Complementarity}(A,B)=\Delta U(A+B)-\max(\Delta U(A),\Delta U(B)) - \text{InteractionCost}$$

هذا المقياس design heuristic؛ التجربة factorial هي التي تحدد interaction. يظل الهدف منع corpus من افتراض أن مزيداً من الطبقات يعني مزيداً من intelligence.

الاقتران 2: Graph نظري إلى Graph تنفيذي

```
TheoryNode → FormalSchema → RuntimeService → TelemetryEvents → EvidencePacket
        ↘ nearest_baseline   ↘ feature_flag   ↘ counterfactual_run
Court: duplicate? mechanism? causal? transferable? reproducible?
```

تستدعي MOUSSAIXT فقط أقل مجموعة من services اللازمة، MUG يمنع promotion إذا كانت telemetry ناقصة، وMUT يحدث lineage بعد verdict. هذه البنية تفصل «التأثير على runtime» عن «الوجود في كتاب».

Invariants المشتركة

- observed state يتقدم على remembered/generated state
- authority يُفحص عند action time
- world effect يحتاج readback مستقل حيث يمكن
- كل failure يترك residue قابل للتعلم لا يُخفى
- كل self-update versioned وقابل للrollback بقدر الإمكان
- claim لا يتجاوز transfer envelope الذي اختبر فعلياً

مختبر الضغط 1: العدسة السببية: من الارتباط إلى intervention

يجب فصل three layers: correlation التي تصلح للاكتشاف، causal hypothesis التي تصلح لاختيار تجربة، وverified causal contrast الذي يبرر attribution محدوداً. في Part XIV لا تصبح edge سببية لأن النص يصفها كذلك؛ يجب أن نحدد do-variable أو feature flag أو perturbation يعزلها.

> **تطبيق على P14-C:** التجربة الجيدة تقلل مساحة التفسيرات البديلة، لا تزيد عدد الرسوم. تصبح كل نظرية في هذا المسار مسؤولة عن تعريف metric وفشل يخصها، ثم تُقارن بالعقد الشقيقة لتجنب double-counting.

نضيف إلى كل run ثلاثة سجلات: DecisionLedger يبين لماذا فُعّل المكوّن، EvidenceLedger يبين ما شوهد فعلاً، ResidualLedger يبين ما لم يفسر أو يُغلق. بهذه الثلاثية يمكن إعادة تفسير run بعد ظهور دليل جديد دون إعادة كتابة التاريخ.

الحد العلمي: هذه الصفحة تبني منهج اختبار؛ لا تحتوي نتيجة تجريبية ما لم يرفق experiment packet مستقل.

مختبر الضغط 2: عدسة الاقتصاد المعرفي: فائدة لكل وحدة compute

الـAGI-oriented architecture لا تُقاس بالعمق اللفظي. كل طبقة تدفع كلفة memory construction، tool calls، latency، tokens وverification. لذلك نحسب quality-cost frontier ونرفض الميكانيزم الذي يضيف كلفة بلا مكسب متكرر أو يقلل robustness.

> **تطبيق على P14-C:** المكوّن الذي يرفع score 1% ويضاعف الكلفة قد يخسر على frontier حتى لو كان statistically positive. تصبح كل نظرية في هذا المسار مسؤولة عن تعريف metric وفشل يخصها، ثم تُقارن بالعقد الشقيقة لتجنب double-counting.

نضيف إلى كل run ثلاثة سجلات: DecisionLedger يبين لماذا فُعّل المكوّن، EvidenceLedger يبين ما شوهد فعلاً، ResidualLedger يبين ما لم يفسر أو يُغلق. بهذه الثلاثية يمكن إعادة تفسير run بعد ظهور دليل جديد دون إعادة كتابة التاريخ.

الحد العلمي: هذه الصفحة تبني منهج اختبار، لا تحتوي نتيجة تجريبية ما لم يرفق experiment packet مستقل.

مختبر الضغط 3: عدسة negative activation: متى يجب ألا تعمل النظرية؟

أقوى علامة على mechanism حقيقي هي امتلاك شرط عدم التفعيل. نُنشئ حالات يكون فيها baseline البسيط هو الصحيح ونقيس false activation. هذا يمنع كل نظرية من التحول إلى طبقة إجبارية تُستخدم في كل مهمة لمجرد وجودها.

> **تطبيق على P14-C:** الذكاء الحاكمي يتضمن حق الامتناع عن استخدام تعقيد لا يحتاجه السؤال. تصبح كل نظرية في هذا المسار مسؤولة عن تعريف metric وفشل يخصها، ثم تقارن بالعقد الشقيقة لتجنب double-counting.

نضيف إلى كل run ثلاثة سجلات: DecisionLedger يبين لماذا فُعّل المكوّن، EvidenceLedger يبين ما شوهد فعلاً، ResidualLedger يبين ما لم يفسر أو يُغلق. بهذه الثلاثية يمكن إعادة تفسير run بعد ظهور دليل جديد دون إعادة كتابة التاريخ.

الحد العلمي: هذه الصفحة تبني منهج اختبار، لا تحتوي نتيجة تجريبية ما لم يرفق experiment packet مستقل.

مختبر الضغط 4: عدسة التفاعلات: المكوّن قد لا يعمل منفرداً

قد تظهر قيمة النظرية فقط مع memory أو verifier أو tool contract. بدلاً من إنكار ذلك، نفصل main effect عن interactions بعينة factorial. إذا كان الأثر تفاعلياً نعيد تسمية claim: ليس «M14-X يحسن الأداء» بل «M14-X مع Y تحت Z يغير recovery/quality».

> **تطبيق على P14-C:** الضبط هنا يقوي النظرية لأنه يمنع التعميم الكاذب. تصبح كل نظرية في هذا المسار مسؤولة عن تعريف metric وفشل يخصها، ثم تُقارن بالعقد الشقيقة لتجنب double-counting.

نضيف إلى كل run ثلاثة سجلات: DecisionLedger يبين لماذا فُعّل المكوّن، EvidenceLedger يبين ما شوهد فعلاً، ResidualLedger يبين ما لم يفسر أو يُغلق. بهذه الثلاثية يمكن إعادة تفسير run بعد ظهور دليل جديد دون إعادة كتابة التاريخ.

الحد العلمي: هذه الصفحة تبني منهج اختبار، لا تحتوي نتيجة تجريبية ما لم يرفق experiment packet مستقل.

مختبر الضغط 5: عدسة الزمن الطويل: الخطأ المتأخر أهم من الجواب اللحظي

في trajectories طويلة، صحة خطوة محلية لا تكفي. نقيس recovery and attribution، drift، stale memory، obligation loss، delayed failure. هذه العدسة تربط Part XIV بحوث long-horizon agents التي تكشف انهياراً مع طول المسار.

> **تطبيق على P14-C:** كل mechanism يجب أن يحدد كيف يتغير مع horizon لا أن يفترض stationarity. تصبح كل نظرية في هذا المسار مسؤولة عن تعريف metric وفشل يخصها، ثم تُقارن بالعقد الشقيقة لتجنب double-counting.

نضيف إلى كل run ثلاثة سجلات: DecisionLedger يبين لماذا فُعّل المكوّن، EvidenceLedger يبين ما شوهد فعلاً، ResidualLedger يبين ما لم يفسر أو يُغلق. بهذه الثلاثية يمكن إعادة تفسير run بعد ظهور دليل جديد دون إعادة كتابة التاريخ.

الحد العلمي: هذه الصفحة تبني منهج اختبار، لا تحتوي نتيجة تجريبية ما لم يرفق experiment packet مستقل.

مختبر الضغط 6: عدسة نقل الدليل: من benchmark إلى deployment

نقيد claim بـinvariants التي حُفظت بين بيئة الاختبار والنشر: failure consequences، user feedback، cost، observability، authority، action space. إذا تغيّرت invariants يتغير سقف claim حتى لو كان benchmark score ثابتاً.

> **تطبيق على P14-C:** النشر ليس benchmark أكبر؛ إنه توزيع مختلف للأخطاء والآثار. تصبح كل نظرية في هذا المسار مسؤولة عن تعريف metric وفشل يخصها، ثم تُقارن بالعقد الشقيقة لتجنب double-counting.

نضيف إلى كل run ثلاثة سجلات: DecisionLedger يبين لماذا فُعّل المكوّن، EvidenceLedger يبين ما شوهد فعلاً، ResidualLedger يبين ما لم يفسر أو يُغلق. بهذه الثلاثية يمكن إعادة تفسير run بعد ظهور دليل جديد دون إعادة كتابة التاريخ.

الحد العلمي: هذه الصفحة تبني منهج اختبار؛ لا تحتوي نتيجة تجريبية ما لم يرفق experiment packet مستقل.

مختبر الضغط 7: عدسة استقلال المقيم: من يحكم على الحكم؟

عندما يكون evaluator من نفس العائلة أو البيانات أو loop، قد يظهر التحسن بسبب agreement لا truth. لذلك نسجل evaluator lineage ونجرب formal/world-state/independent judges عند الإمكان.

> **تطبيق على P14-C:** كل self-improvement loop هي أيضاً فرضية حول صلاحية signal الذي يقيس التحسن. تصبح كل نظرية في هذا المسار مسؤولة عن تعريف metric وفشل يخصها، ثم تُقارن بالعقد الشقيقة لتجنب double-counting.

نضيف إلى كل run ثلاثة سجلات: DecisionLedger يبين لماذا فُعّل المكوّن، EvidenceLedger يبين ما شوهد فعلاً، ResidualLedger يبين ما لم يفسر أو يُغلق. بهذه الثلاثية يمكن إعادة تفسير run بعد ظهور دليل جديد دون إعادة كتابة التاريخ.

الحد العلمي: هذه الصفحة تبني منهج اختبار، لا تحتوي نتيجة تجريبية ما لم يرفق experiment packet مستقل.

مختبر الضغط 8: عدسة التكاثر العلمي: هل ينجو الأثر خارج المختبر؟

نحوّل reproducibility من ملحق إلى جزء من معنى النظرية. replay package يشمل negative results، budgets، seeds، raw traces، versions، hashes، ثم conceptual replication على بيئة مختلفة.

> **تطبيق على P14-C:** الأثر الذي لا يعاد قد يبقى engineering trick محلياً لكنه لا يرقى إلى claim عام. تصبح كل نظرية في هذا المسار مسؤولة عن تعريف metric وفشل يخصها، ثم تُقارن بالعقد الشقيقة لتجنب double-counting.

نضيف إلى كل run ثلاثة سجلات: DecisionLedger يبين لماذا فُعّل المكوّن، EvidenceLedger يبين ما شوهد فعلاً، ResidualLedger يبين ما لم يفسر أو يُغلق. بهذه الثلاثية يمكن إعادة تفسير run بعد ظهور دليل جديد دون إعادة كتابة التاريخ.

الحد العلمي: هذه الصفحة تبني منهج اختبار، لا تحتوي نتيجة تجريبية ما لم يرفق experiment packet مستقل.

سجل الأدوات والأصول التنفيذية - 1/2

ID	Owner	Asset	Class	Purpose
-XIV A061	M14-07	Persistent Attention Dynamics State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A062	M14-07	Persistent Attention Dynamics Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A063	M14-07	Persistent Attention Dynamics Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A064	M14-07	Persistent Attention Dynamics Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A065	M14-07	Persistent Attention Dynamics Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A066	M14-07	Persistent Attention Dynamics Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A067	M14-07	Persistent Attention Dynamics Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A068	M14-07	Persistent Attention Dynamics Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A069	M14-07	Persistent Attention Dynamics Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A070	M14-07	Persistent Attention Dynamics Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A071	M14-08	Causal Memory Compression Bound State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A072	M14-08	Causal Memory Compression Bound Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A073	M14-08	Causal Memory Compression Bound Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A074	M14-08	Causal Memory Compression Bound Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A075	M14-08	Causal Memory Compression Bound Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.

الحالة: SPECIFICATION. لا يقال deployed أو validated إلا بعد وجود artifact/run/test evidence.

سجل الأدوات والأصول التنفيذية - 2/2

ID	Owner	Asset	Class	Purpose
-XIV A076	M14-08	Causal Memory Compression Bound Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر /domain/tool/model horizon.
-XIV A077	M14-08	Causal Memory Compression Bound Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A078	M14-08	Causal Memory Compression Bound Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A079	M14-08	Causal Memory Compression Bound Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A080	M14-08	Causal Memory Compression Bound Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A081	M14-09	Reconstructive Memory Eventual Consistency State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A082	M14-09	Reconstructive Memory Eventual Consistency Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A083	M14-09	Reconstructive Memory Eventual Consistency Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A084	M14-09	Reconstructive Memory Eventual Consistency Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A085	M14-09	Reconstructive Memory Eventual Consistency Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A086	M14-09	Reconstructive Memory Eventual Consistency Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر /domain/tool/model horizon.
-XIV A087	M14-09	Reconstructive Memory Eventual Consistency Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A088	M14-09	Reconstructive Memory Eventual Consistency Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A089	M14-09	Reconstructive Memory Eventual Consistency Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A090	M14-09	Reconstructive Memory Eventual Consistency Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.

الحالة: SPECIFICATION. لا يقال deployed أو validated إلا بعد وجود artifact/run/test evidence.

برنامج التدريب: من التعريف إلى قدرة قابلة للنقل

لا يُقصد بالتدريب هنا بالضرورة fine-tuning للنموذج؛ قد يكون curriculum للهارنس، verifier training، policy learning أو dataset construction. المبدأ أن training target يجب أن يطابق mechanism claim لا اسم النظرية.

- M14-07 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-07 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-07 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-07 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-08 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-08 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-08 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-08 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-09 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-09 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-09 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-09 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

> عند استخدام generated data، تُفصل data producer عن evaluator قدر الإمكان ويُحتفظ بالraw source/trajectory لتدقيق self-confirming loops.

مصفوفة التجارب - 1/1

ID	Owner	Kind	Design	Metrics
-XIV E19	M14-07	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E20	M14-07	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E21	M14-07	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E22	M14-08	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E23	M14-08	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E24	M14-08	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E25	M14-09	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E26	M14-09	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E27	M14-09	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier

ترجمة الورقة إلى MOUSSAIXT Reference Kernel

```
service registry: M14-07,M14-08,M14-09
feature_flags: independent
state: versioned typed objects
authority: L0/L1 gates
telemetry: trigger,state_before,decision,tool,observation,state_after,residual
verifier: independent where possible
rollback: mandatory for reversible mutations
writeback: evidence-bearing delta only
```

التنفيذ المقترح يبدأ بـservices أصغر من «AGI kernel» الكبير. كل service يملك test vectors وحالات invalid input وtimeout وstale state وwrong target. بهذه الصياغة يصبح ablation ممكناً، ويصبح فشل component معلومة بدلاً من انهيار غامض للنظام كله.

Acceptance at kernel level

- schema validation ≥ predefined threshold
- no silent privilege widening
- deterministic replay where environment permits
- world-state postcondition for external actions
- trace completeness for critical steps
- explicit disable path and fallback

القيود والمشكلات المفتوحة

- الاستقلال السببي بين مكونات architecture قد يكون صعباً بسبب interactions الكثيفة.
- بعض observables المقترحة تحتاج proxies قد تغيّر معنى النظرية.
- matched budgets ليست مساواة كاملة لأن latency/tool quality مختلفة نوعياً.
- الانتقال بين benchmarks/deployment قد يفشل بسبب hidden distribution shifts.
- المقيم المستقل قد يكون أقل كفاءة أو أكثر ضجيجاً، ما يستلزم meta-evaluation.
- memory/world models تولدان state غير قابل للمقارنة بسهولة بين implementations.
- أي claim عن consciousness يبقى منفصلاً عن الفائدة الهندسية الوظيفية.
- المرجع الخارجي الحديث قد يكون preprint؛ نميّز peer-reviewed ACL papers عن arXiv/spec/lab reports.

أقوى نتيجة مرغوبة من هذه الورقة ليست «نجاح كل النظريات»، بل ظهور خريطة واضحة: ماذا نحتفظ به، ماذا ندمج، أين تفشل الآلية، وما evidence الجديد الذي يغير النسخة التالية.

المراجع الخارجية الأساسية

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

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

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

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

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

راجع أيضاً قائمة Part XIV Master للمراجع المشتركة. العناوين/المعرفات مُستخدمة كمرساة بحثية؛ لا تُنقل claims تفصيلية خارج ما يدعمه المصدر.

المراسي الداخلية والسلالة

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

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

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

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

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

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

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

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

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 .;artifacts; duplicate groups are not independent evidence 343 .retrieval is family/canonicality based

> سياسة الدمج: canonical synthesis يحفظ source role والنسخة والحالة. الملفات المكررة أو historical mirrors لا تُعد أدلة جديدة؛ الفروق المادية تبقى lineage variants حتى ablation/semantic comparison.