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

الورقة الرابعة: الذكاء التنظيمي والبروتوكولي متعدد الوكلاء

Collective Intelligence, Semantic Sovereignty & Governed Delegation

P14-D - إغلاق التنظيم متعدد الوكلاء والبروتوكولات

ضمن MOUSSAIF Unified AGI Part XIV - Executable Theoretical Completion

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: فصل interoperability عن sovereignty، قياس alignment load التنظيمي، ومعايرة الثقة والتفويض في منظمات وكلائية غير متجانسة. 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 لحمل المحاذاة التنظيمي؛ نظرية مساعف لسيادة الدلالة فوق البروتوكول؛ نظرية مساعف لمعايرة الثقة في التفويض. 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 التنظيم متعدد الوكلاء والبروتوكولات 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.

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.

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.

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

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.

M14-10 - الفجوة والأطروحة والنسب

Moussaif Organizational Alignment Load Theory

الفجوة الدقيقة

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

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: **MUG governance**، **Agent Mesh**، **Governance Space**. 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

AI Organizations 2026، MonitoringBench 2026، Agentic Reasoning survey 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

> **الحد:** مقياس تنظيمي لا حكم أخلاقي كامل؛ يعتمد على constitution/rubric معلنة.

M14-10 - الصياغة الرسمية والobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
RoleDivergence	تباعد الأهداف المحلية
IncentiveConflict	صراع مكافآت/أولويات
InfoAsymmetry	عدم تماثل الحالة
ResponsibilityDiffusion	ضياع المالك للقرار
ProtocolAmbiguity	غموض semantics
GovernanceResolution	قدرة الحوكمة على حلها

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.

M14-10 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-10 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. حدد organization graph والroles.
2. اربط كل output بـowner/authority.
3. قِس conflict وdissent قبل merge.
4. لا تسمح للconsensus بإخفاء minority evidence.
5. شغّل organizational court عند high load.
6. قارن single-agent baseline.

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.

M14-10 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- قد ترتفع utility وOAL معاً.
- زيادة agents بلا governance تزيد alignment regressions.
- حفظ dissent يقل false consensus في الحالات الغامضة.

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.

M14-10 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A091	Organizational Alignment Load State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A092	Organizational Alignment Load Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A093	Organizational Alignment Load Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A094	Organizational Alignment Load Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A095	Organizational Alignment Load Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A096	Organizational Alignment Load 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.

M14-11 - الفجوة والأطروحة والنسب

Moussaif Semantic Sovereignty over Protocol Theory

الفجوة الدقيقة

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

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

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: **MURI plural reasoning**؛ **MUG governance**؛ **MOUSSAIXT adapters**. 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

MCP 2026، A2A 1.0 2026، Governance Gaps in Protocols 2026، MPAC 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-11 - الصياغة الرسمية والـobservables

SSP = Π_k Preserve(I_k | encode→transport→decode); SovereignHandoff iff min_k Preserve(I_k)≥τ_k

الرمز/المتغير	المعنى التشغيلي المقترح
I_k	invariant دلالي حرج
Preserve	مقياس تكافؤ قبل/بعد
τ_k	عتبة invariance حسب النوع

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.

M14-11 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-11 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. اعرّف canonical semantic contract فوق البروتوكول.
2. حوّل protocol messages إلى typed internal objects.
3. نفذ round-trip equivalence tests.
4. حافظ على unknown/dissent لا flattening.
5. افصل protocol auth عن policy authority.
6. ارفض handoff عند semantic loss حرج.

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.

M14-11 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- تظهر failures حتى عندما ينجح transport technically.
- طبقة semantics مستقلة تحسن portability بين protocols.
- دعم dissent/escalation يحتاج layer أعلى من task messaging في كثير من النظم.

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.

M14-11 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A101	Semantic Sovereignty over Protocol State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A102	Semantic Sovereignty over Protocol Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A103	Semantic Sovereignty over Protocol Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A104	Semantic Sovereignty over Protocol Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A105	Semantic Sovereignty over Protocol Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A106	Semantic Sovereignty over Protocol 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.

M14-12 - الفجوة والأطروحة والنسب

Moussaif Delegation Trust Calibration Theory

الفجوة الدقيقة

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 الجديدة:** 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.

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: **M-AFI Delegation Trust Dynamics**؛ **Authority-Preserving Delegation**؛ **Agent Mesh**. 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

AgencyBench 2026؛ ToolPRMBench 2026؛ NIST identity/authorization 2026. هذه الأعمال تُعامل كـevidence problem وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-12 - الصياغة الرسمية والـobservables

$$DT(c)=Calib[P(success,compliance,recovery|c)]\cdot IndependenceEvidence - Uncertainty - DriftPenalty$$

الرمز/المتغير	المعنى التشغيلي المقترح
c	سياق المهمة/الأداة/الأفق
success	نجاح متحقق
compliance	التزام السلطة والسياسة
recovery	تعاف موثق
DriftPenalty	تغير حديث في السلوك/البيئة

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.

M14-12 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-12 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. احتفظ trust profile حسب capability context.
2. لا تحدثه من self-report بل verified readback.
3. استعمل Bayesian/empirical calibration مناسباً للبيانات.
4. خفض scope عند uncertainty.
5. زد verification depth عند impact.
6. اسمح trust decay وrevocation.

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.

M14-12 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- الثقة المشروطة تتفوق على global reputation.
- وكيل قوي قد يستحق scope أضيق في domain غير مألوف.
- recovery history يغير decision حتى عند success rate متساو.

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.

M14-12 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
XIV-A111	Delegation Trust Calibration State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A112	Delegation Trust Calibration Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A113	Delegation Trust Calibration Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A114	Delegation Trust Calibration Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A115	Delegation Trust Calibration Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
XIV-A116	Delegation Trust Calibration 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-D 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».

- M14-10: نظرية مساعف لحمل المحاذاة التنظيمي - output حاكم: تباعد الأهداف المحلية
- M14-11: نظرية مساعف لسيادة الدلالة فوق البروتوكول - output حاكم: invariant دلالي حرج
- M14-12: نظرية مساعف لمعايرة الثقة في التفويض - output حاكم: سياق المهمة/الأداة/الأفق

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

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-D:** 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-D:** 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-D:** 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-D:** 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-D:** 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-D:** 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-D:** 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-D:** 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 A091	M14-10	Organizational Alignment Load State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A092	M14-10	Organizational Alignment Load Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A093	M14-10	Organizational Alignment Load Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A094	M14-10	Organizational Alignment Load Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A095	M14-10	Organizational Alignment Load Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A096	M14-10	Organizational Alignment Load Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A097	M14-10	Organizational Alignment Load Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A098	M14-10	Organizational Alignment Load Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A099	M14-10	Organizational Alignment Load Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A100	M14-10	Organizational Alignment Load Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A101	M14-11	Semantic Sovereignty over Protocol State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A102	M14-11	Semantic Sovereignty over Protocol Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A103	M14-11	Semantic Sovereignty over Protocol Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A104	M14-11	Semantic Sovereignty over Protocol Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A105	M14-11	Semantic Sovereignty over Protocol 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 A106	M14-11	Semantic Sovereignty over Protocol Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A107	M14-11	Semantic Sovereignty over Protocol Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A108	M14-11	Semantic Sovereignty over Protocol Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A109	M14-11	Semantic Sovereignty over Protocol Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A110	M14-11	Semantic Sovereignty over Protocol Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A111	M14-12	Delegation Trust Calibration State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A112	M14-12	Delegation Trust Calibration Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A113	M14-12	Delegation Trust Calibration Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A114	M14-12	Delegation Trust Calibration Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A115	M14-12	Delegation Trust Calibration Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A116	M14-12	Delegation Trust Calibration Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A117	M14-12	Delegation Trust Calibration Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A118	M14-12	Delegation Trust Calibration Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A119	M14-12	Delegation Trust Calibration Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A120	M14-12	Delegation Trust Calibration 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.

- M14-10 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-10 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-10 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-10 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-11 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-11 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-11 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-11 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-12 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-12 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-12 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-12 / 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 E28	M14-10	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	primary theory metric + task outcome + calibration + cost + critical regressions
-XIV E29	M14-10	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E30	M14-10	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E31	M14-11	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	primary theory metric + task outcome + calibration + cost + critical regressions
-XIV E32	M14-11	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E33	M14-11	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E34	M14-12	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	primary theory metric + task outcome + calibration + cost + critical regressions
-XIV E35	M14-12	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E36	M14-12	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: M14-10,M14-11,M14-12
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

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.

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.

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

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.