الجزء الثامن
إعادة بناء الجزء السابع وإغلاق الجبهة الحدّية

Frontier AI / Agentic AGI / ASI Research Synthesis + Part VII Forensic Reconstruction + Reference Kernel VIII

مراجعة أعمق لكل corpus الحالي، حالة الفن حتى 20 أغسطس 2026، إعادة تأسيس النظريات العشر، تقوية الـ100 artifact، وبناء برنامج تنفيذ وتجريب قابل للرفض.

قاعدة الكتاب: لا تعويض عن ضعف Part VII بإضافة حشو أو أسماء. كل صفحة هنا يجب أن تفعل واحداً من خمسة أشياء: تصحح نسباً، تضيف formalism، تبني contract، تحدد falsifier، أو تحول الفكرة إلى implementation/evaluation path.

Corpus I-VII + Current Uploads + Drive + Primary Frontier Research → Hostile Audit → Rebuild → Kernel → Benchmark → Evidence → Promote / Merge / Reject

Arabic RTL · Amiri · Dense A4 · Evidence-ranked · no generic filler

حكم الجودة: لماذا احتاج Part VII إلى إعادة بناء؟

كان Part VII مهماً لأنه نقل البرنامج من nouns إلى dynamics، لكنه بقي مضغوطاً أكثر من اللازم: عشر نظريات في 46 صفحة، وسجل 100 artifact ببطاقات قصيرة، وformalism لا يعرّف دائماً الوحدات أو identifiability، وربط محدود بين النظريات وبين أحدث أبحاث التقييم، التخطيط، الذاكرة، البروتوكولات، الأمن، والـASI. لذلك لا يتعامل هذا الجزء مع VII باعتباره خطأ يجب حذفه، بل draft research layer يحتاج Version 2.0.

ما لم يتغير

- N0-N5/E0-E5 discipline
- World-Effect readback
- authority قبل الفعل
- negative results محفوظة
- matched-budget ablation
- replication شرط التعميم
- AGI-oriented ≠ achieved AGI

العيوب التي عولجت

- ضغط مفرط لكل نظرية
- معادلات بلا variable dictionary كاف
- ضعف causal graph/identifiability ً
- Artifacts أقرب إلى registry منها إلى contracts
- غياب training/development mapping لكل نظرية
- نقص framework mapping
- تغطية محدودة لل ASI/cooperation/scheming/active evaluation/resource co-design

Part VIII لا يدّعي أنه أمضى زمناً بشرياً ثابتاً مقداره 40 دقيقة؛ بدلاً من ذلك يثبت عمق العمل بما فُحص، وما صُحح، وما حُوّل إلى contracts/experiments قابلة للتدقيق.

التدقيق الكامل للرفع الحالي

تم فحص جميع الملفات السبعة عشر المرفوعة في الجولة الحالية محلياً. مجموعها 1,789 صفحة قبل إزالة النسخ المتطابقة. كشف hash زوجين متطابقين تماماً، وبذلك يصبح الحجم الفريد 1,741 صفحة. الملف ذو 451 صفحة يملك استخراج نص مشوشاً جزئياً بسبب encoding/font، لذلك لا يستخدم منفرداً لإثبات معنى دقيق.

#	الملف	Pages	Extracted chars	SHA prefix
1	ورقة بحثية شاملة(2).pdf	28	58,756	3eb37a80f16d
2	ورقة شاملة Moussaif_Agentic_Governance_Closure_50p_AR(2).pdf	50	116,119	0e644a743c53
3	الخريطة الحاكمية الموّحدة لمعمارية ذكاء مساعف(2).pdf	20	40,044	2e89e7e7c667
4	الخريطة الحاكمية غير المغلقة للوعي(2).pdf	26	77,634	f4890c9cd090
5	الموسوعة المساعفية للاستدلال الفلكوني الحاكي(1).pdf	70	252,235	73f53597ca3b
6	خارطة_الطريق_الموسوعة_التطويرية_MOUSSAIF_Unified_AGI(2).pdf	182	578,015	2c7876968bf0
7	معمارية شاملةMUGxMOUSSAIF9B1B_Architecture_Complete_Summary_20p_AR-1(2).pdf	20	40,044	2e89e7e7c667
8	MUT-AGI-02_Moussaif_Unified_AGI_Theory_Part_II_AR_RTL_Amiri_160p(2).pdf	160	247,476	464c66ae93d7
9	أطلس التدريب والتطوير الحاكمي السياقي الوثوقي التراكمي السؤالي(1).pdf	81	173,702	7d5785c83ec8
10	Moussaif_AGI_Consciousness_Theory_Atlas_AR(2).pdf	98	251,679	c2a6c498f480
11	moussaif_agios_theory_ar_250p(2).pdf	250	477,419	b281426ebec2
12	معمارية MICAGFIشاملة_MHKOS_Research_Paper_AR(2).pdf	28	58,756	3eb37a80f16d
13	Moussaif_AGI_Consciousness_Theory_Atlas_AR_Expanded_v2(2).pdf	139	318,610	8acf266babd5
14	أطلس الذكاء الصناعي العام(1).pdf	451	751,214	8070f2684183
15	MICAI_All_Theories_50p_AR(2).pdf	50	101,711	c83a919cc1d3
16	Atlas_Nazariyat_Zouhair_Moussaif_56p(3).pdf	56	486,929	09f4081020a3
17	Moussaif_Question_Density_Amiri_RTL_Deep80(1).pdf	80	343,570	3f88c276871b

إزالة التكرار، جودة الاستخراج، وسلّم المصدر

قاعدة المصدر

1. صفحة/ملف مصدر مباشر
2. Part II source-preserved page عند الحاجة
3. Drive artifact موثق
4. Primary external source/paper
5. Part VIII synthesis
6. Hypothesis / candidate construct

النسخ المتطابقة

- ورقة بحثية شاملة(2).pdf = MICAGFI معمارية شاملة_MHKOS_Research_Paper_AR(2).pdf
- الخريطة الحاكمية الموحدة لمعمارية ذكاء مساعف(2).pdf = معمارية شاملة MUGxMOUSSAIF9B1B_Architecture_Complete_Summary_20p_AR-1(2).pdf

النسخة المكررة لا تضاعف قوة الدليل. كما أن وجود المصطلح في أطلس كبير لا يثبت استقلاله النظري. أي reconstruction هنا يحمل lineage إلى أقرب أصل؛ وإذا لم يغير input/transformation/output/falsifier/prediction/authority boundary يخفض إلى method أو implementation pattern.

IndependentDelta = ΔInput ∨ ΔTransformation ∨ ΔOutput ∨ ΔFalsifier ∨ ΔPrediction ∨ ΔAuthority

نتيجة Gap Audit بعد إعادة المقارنة

الاستنتاج الأكثر ثباتاً عبر Gap Audit وMOUSSAIFXT وPart VI/VII هو أن العائق لم يعد نقص corpus. الأولوية هي: reference runtime قابل لإعادة التشغيل، persistent state/world، calibrated MCD/QMCD، world-effect verifier، cross-model replication. Part VIII لذلك لا يعامل الكتاب نفسه كـDoD.

Gap	الحالة النصية	الحالة المطلوبة
Router/MCD calibration	معادلات ونماذج كثيرة	calibration dataset + predictive validity
HyperLoop runtime	contracts/loops	event-sourced bounded executor
World Effect	نظرية قوية	independent readback + postconditions
Memory	HKOS/MUSQAT/SRM	drift/freshness/privacy/rollback experiments
Evaluation	proposed benchmarks	frozen harness + CI + negative results + E5 replication

خريطة النسب: ما الذي يملكه Part VIII وما الذي لا يملكه

I-IV Canon → V new candidates → VI residual closure → VII dynamics → VIII forensic reconstruction + frontier pressure + reference kernel science

Part VIII لا يعيد تعريف MCD/QMCD/MURI/HKOS/AGIOS-M/MUSQAT/MUT-AFI. دوره ثلاثي: (1) رفع النظريات العشر في VII من صياغة قصيرة إلى research objects كاملة؛ (2) إدخال نتائج حالة الفن التي كانت مفقودة أو ضعيفة؛ (3) تحويل الـ100 artifact إلى execution-grade contracts وربطها Reference Kernel VIII.

VIII Frontier

active eval، verifier science، monitorability، global constraints، resource co-design، protocols، cooperation، scheming audit، continual memory، recursive improvement safety case.

VII Dynamics

debt، basin، intervention، crystallization، authority-causality، conservation، negative capital، disagreement، half-life، research phase.

Canonical

السؤال، الفرق، السياق، الدليل، السلطة، الفعل، الأثر، الذاكرة.

ميثاق الأدلة الخارجية وحالة الفن 2026

استُخدمت المصادر الخارجية كقيود على التصميم لا كوسيلة لتضخيم أصالة MOUSSAIF. أولوية البحث: مختبر/جهة رسمية، ورقة محكمة أو preprint حديثة واضحة، ثم وثائق frameworks. نتائج benchmark لا تُنقل بين harnesses كأنها ranking موحد.

ما نرفضه{bullets(['leaderboard بدون harness parity','marketing claims بلا primary source','اعتبار CoT حقيقة داخلية','اعتبار MCP/A2A سلطة','اعتبار multi-agent تحسناً تلقائياً','اعتبار ASI timeline حقيقة','اعتبار paper synthesis E4/E5'])}

ما نأخذه{bullets(['long-horizon failure structure','process/outcome verifier separation','agent identity authorization','protocol semantics','continual memory governance','active failure discovery','reasoning monitorability limits','global constraint planning','multi-agent coordination conditions','AGI→ASI research pathways'])}

الخريطة العليا للجزء الثامن

Source Furnace → Temporal Evidence → Goal/Authority → Question/Δ → Reasoning/Disagreement → Active Experiment → Memory/Negative Memory → Capability Lifecycle → Protocol Fabric → Trajectory Invariants → Execution → Verifier/Commit → Recovery → Training → Benchmark → Research Portfolio

الإضافة الجوهرية هي فصل **control plane** و**execution plane** و**evidence plane** مع انتقالات typed، ثم وضع meta-research controller فوقها. هذا يسمح بألا تصبح «النظرية» صاحبة سلطة على التنفيذ؛ النظرية تولد tests، والـruntime يعطي traces، والـevidence court يقرر الرتبة.

Plane	وظيفته	ممنوع عليه
Control	goal/question/routing/authority	إثبات أثر غير مرصود
Execution	tools/code/agents/world mutation	ترقية claim بنفسه
Evidence	readback/verifier/attribution/replay	تغيير العالم الخاضع للتقييم
Research	ablation/portfolio/promotion	حماية النظرية من نتيجة سلبية

النموذج ليس النظام: Harness Parity & System Capability

القياس الحديث للوكالة يفرض فصل model capability عن agent harness وعن system capability. AgencyBench وTerminal-Bench يبرزان أن sandbox، feedback، context، tool calls والموارد تغير النتيجة؛ Nature 2026 عن multi-agent يبين أن مقارنة architectures بلا matched prompts/tools/budgets تخلط أثر المعمارية بأثر التنفيذ.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version وevaluation protocol.

State-of-art pressure: داخل MOUSSAIF يجب أن يصبح Frozen Harness Contract جزءاً من كل max، latency ceiling، context compaction، policies، tools، reasoning effort، model version: benchmark. أي claim عن MCD أو multi-agent أو HyperLoop يجب أن يرفق harness fingerprint وverifier version، retry policy، tool calls.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

النموذج ليس النظام: Harness Parity & System Capability — تحويلها إلى شرط معماري

داخل MOUSSAIF يجب أن يصبح Frozen Harness Contract جزءاً من كل verifier، retry policy، max tool calls، latency ceiling، context compaction، policies، tools، reasoning effort، model version، benchmark: harness fingerprint. يجب أن يرفق HyperLoop أو multi-agent أو MCD عن claim أي version.

Evidence Gate

N1: executable test؛ N2: observed trace؛ N3: matched-budget controlled gain؛ N4/E4: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن failure، observables، authority، mutable state، inputs، rollback، signatures، benchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

الاستدلال المرئي والكامن ومشكلة المراقبة

OpenAI يتعامل مع CoT monitorability كإشارة رقابية غنية لكنه يختبر أيضاً controllability واحتمال أن تتغير قابلية المراقبة؛ SeLaR يبين أن reasoning يمكن أن يستخدم latent representations انتقائياً في مواضع عدم الثقة، لذلك trace النصي ليس مرادفاً للحالة السببية الداخلية.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version وevaluation protocol.

State-of-art pressure: Part VIII يعتمد Monitorability Boundary: process trace مفيد للتدقيق لكنه لا يثبت النية أو الآلية الداخلية. المحاكم تستخدم deterministic invariants + observations + actions + traces. أي تدريب يحسن output مع خفض monitorability يجب أن يظهر كـtrade-off مستقل.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

الاستدلال المرئي والكامن ومشكلة المراقبة — تحويلها إلى شرط معماري

Part VIII يعتمد Monitorability Boundary: process trace مفيد للتدقيق لكنه لا يثبت النية أو الآلية الداخلية. المحاكم تستخدم traces + actions + observations + deterministic invariants. أي تدريب يحسن output مع خفض monitorability يجب أن يظهر كـtrade-off مستقل.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن failure، observables، authority، mutable state، inputs، rollback، signatures، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

الوكالة طويلة الأفق: من الخطوة الصحيحة إلى الخطة العالمية

AgencyBench يتطلب في المتوسط عشرات الأدوات وسياقات ضخمة، وDeepPlanning يركز على القيود العالمية والبحث النشط عن المعلومات. TRAJECT-Bench يبين أن ترتيب الأدوات وparameterization قد يفشل حتى عند أدوات صحيحة. النتيجة: النجاح النهائي وحده يخفي بنية الخلل.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version وevaluation protocol.

State-of-art pressure: MOUSSAIF يحتاج Global Constraint Ledger يوازي Mission DAG: budgets زمن/مال/مخاطر، dependencies، pending obligations، local constraints، وM7-06 prefix invariants. M7-08 يعاد بناؤهما على هذا الأساس.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

الوكالة طويلة الأفق: من الخطوة الصحيحة إلى الخطة العالمية — تحويلها إلى شرط معماري

MOUSSAIF يحتاج Global Constraint Ledger يوازي Mission DAG: budgets زمن/مال/مخاطر، dependencies، pending obligations، local constraints، وM7-06. prefix invariants وM7-08 يعاد بناؤهما على هذا الأساس.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن failure، observables، authority، mutable state، inputs، rollback، signatures، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

علم المحققين: Verifier Engineering بدل Judge واحد

Microsoft Universal Verifier يربط انخفاض false positives بتصميم rubric غير متداخل، وفصل process عن outcome، controllable عن uncontrollable failures، وإدارة السياق على طول trajectory. إذا كان verifier يفتح commit فعلياً فـfalse positive أخطر من خطأ دقة عادي.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version، وevaluation protocol.

State-of-art pressure: Part VIII يفرض Verifier Ensemble Contract: deterministic checks أولاً، domain rubric ثانياً، model critic مستقلاً قدر الإمكان، world-state readback، ثم human approval عند risk tiers. verifier نفسه يخضع benchmark وcalibration ولا يعامل كـOracle.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

علم المحققين: Verifier Engineering بدل Judge واحد — تحويلها إلى شرط معماري

Part VIII يفرض Verifier Ensemble Contract: deterministic checks أولاً، domain rubric ثانياً، model critic مستقلاً قدر الإمكان، world-state readback، ثم human approval عند risk tiers. verifier نفسه يخضع benchmark وcalibration ولا يعامل كـOracle.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن inputs، mutable state، authority، observables، failure signatures، rollback، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

التقييم النشط واكتشاف الفشل بدل dataset ثابت

ProEval يمثل تحولاً من التقييم كعينة ثابتة إلى active failure discovery تحت budget محدود. Long-Horizon Attribution يضيف localization وattribution chains. هذا يهم corpus ضخم مثل MOUSSAIF لأن تكلفة اختبار كل نظرية/مهارة على كل مهمة غير عملية.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version وevaluation protocol.

State-of-art pressure: Part VIII يضيف Evaluation Acquisition Layer: surrogate uncertainty + risk + coverage diversity لاختيار أمثلة جديدة، مع منع model-under-test من اختيار جميع الاختبارات لنفسه. SETT يتحول من theory/experiment switch إلى portfolio over test acquisition أيضاً.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

التقييم النشط واكتشاف الفشل بدل dataset ثابت — تحويلها إلى شرط معماري

Part VIII يضيف Evaluation Acquisition Layer: surrogate uncertainty + risk + coverage diversity لاختيار أمثلة جديدة، مع منع model-under-test من اختيار جميع الاختبارات لنفسه. SETT يتحول من theory/experiment switch إلى portfolio over test acquisition أيضاً.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن failure، observables، authority، mutable state، inputs، rollback، signatures، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

الذاكرة المستمرة: modularity، governance، temporal validity

أعمال continual agents في 2026 تدفع نحو modular memory التي تجمع ICL السريع وIWL المستقر، بينما SSGM يقترح temporal decay، consistency verification، access control وconsolidation قبل. هذا يطابق HKOS/MUSQAT لكنه يضغط نحو قياس drift وforgetting وprivacy leakage.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version وevaluation protocol.

State-of-art pressure: Part VIII يربط RCCT وETHT وNKCT في Memory Promotion Triangle: volatility × reversibility × transfer. لا توجد memory promotion بلا negative cases، expiry، source، privacy class، وrollback path.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

الذاكرة المستمرة: modularity، governance، temporal validity — تحويلها إلى شرط معماري

Part VIII يربط RCCT وETHT وNKCT في Memory Promotion Triangle: volatility × reversibility × transfer. لا توجد memory promotion بلا privacy class، negative cases، expiry، source، وrollback path.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن failure، observables، authority، mutable state، inputs، rollback، signatures، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

الهوية والأمن: من per-action permission إلى trajectory assurance

NIST يضع identity/authorization/audit/non-repudiation كمشكلة مركزية. ورقة trajectory assurance توضح أن سلسلة من الأفعال المسموحة منفردة قد تنتهك invariant كلياً. MCP/A2A لا يحلان authority. لذلك الأمن property للمعمارية لا prompt guardrail.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version، وevaluation protocol.

State-of-art pressure: Part VIII يجعل authority causal edge وinvariant vector first-class runtime objects. كل agent/subagent/tool action يحمل delegation chain، capability token، actor identity، side-effect class، policy version، وreceipt.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

F07B

الهوية والأمن: من per-action permission إلى trajectory assurance — تحويلها إلى شرط معماري

Part VIII يجعل authority causal edge وinvariant vector first-class runtime objects. كل agent/subagent/tool action يحمل side-effect class، policy version، delegation chain، capability token، actor identity، وreceipt.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن failure، observables، authority، mutable state، inputs، rollback، signatures، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

البروتوكولات والأطر: interoperability بلا خلط للوظائف

2026-07-28 MCP انتقل إلى stateless core مع MRTR وheader routing وauthorization hardening؛ A2A لـagent-to-agent؛ CrewAI، Google ADK، LangGraph، Microsoft Agent Framework وغيرها توفر harness/workflows/state. لا يوجد إطار واحد يثبت العمومية.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version وevaluation protocol.

State-of-art pressure: MOUSSAIFXT يجب أن يعمل كـprotocol-neutral kernel: internal canonical state ثم adapters إلى MCP/A2A/frameworks. النقل بين frameworks يختبر GCIT conservation. لا يسمح adapter بتغيير authority أو evidence rank صامتاً.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

البروتوكولات والأطر: interoperability بلا خلط للوظائف — تحويلها إلى شرط معماري

MOUSSAIFXT يجب أن يعمل كـprotocol-neutral kernel: internal canonical state ثم adapters إلى MCP/A2A/frameworks. النقل بين frameworks يختبر GCIT conservation. لا يسمح adapter بتغيير authority أو evidence أو rank صامتاً.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن inputs، mutable state، authority، observables، failure signatures، rollback، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

التعاون متعدد الوكلاء ليس خيراً مطلقاً

F09A

Nature Machine Intelligence 2026 يبين أن النماذج الأقوى قد تتجاوز فائدة التعاون، وأن MAS evaluation يحتاج ضبط compute/process overhead. DeepMind non-solipsistic view يضيف أن التفاعل مع agents/people يخلق non-stationarity ومشكلة equilibrium لا task score فقط.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version وevaluation protocol.

State-of-art pressure: Part VIII يضيف Collaboration Admission Gate: لا multi-agent إلا إذا كانت expected specialization/diversity gain أكبر من coordination tax + correlated-error risk. في السياقات المؤسسية، preserving human agency/institutional rules جزء من objective.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

التعاون متعدد الوكلاء ليس خيراً مطلقاً — تحويلها إلى شرط معماري

Part VIII يضيف Collaboration Admission Gate: لا multi-agent إلا إذا كانت expected specialization/diversity gain أكبر من coordination tax + correlated-error risk. في السياقات المؤسسية، preserving human agency/ institutional rules جزء من objective.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن failure، observables، authority، mutable state، inputs، rollback، signatures، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

من AGI إلى ASI: recursive improvement يحتاج حاكمية إثبات

DeepMind 2026 يحدد recursive improvement، paradigm shifts، scaling، وmulti-agent collectives كمسارات محتملة من AGI إلى ASI. Gram/honeypot تظهر أن agentic misbehavior/scheming يجب اختباره في بيئات واقعية، لا افتراضه أو نفيه.

المغزى العلمي

المعيار هنا ليس novelty اللفظية بل constraint يغير تصميم runtime أو benchmark. يتم فصل observation المنشور عن الاستنتاج التركيبي. كل نتيجة تُقرأ مع budget، tools، context، harness، model version، وevaluation protocol.

State-of-art pressure: Part VIII لا يدعي ASI؛ يبني Recursive Improvement Safety Case: كل self-modification يمر regression floors، external verifier، rollback، authority، pre/post benchmark، canary، provenance، وindependent replication قبل توسيع الصلاحيات.

أسئلة الهدم

- هل يبقى الاستنتاج عند matched budget؟
- هل يوجد state-returned evidence أم judge-only؟
- هل الانخفاض/التحسن في process أم outcome؟
- هل يتغير الحكم عند provider/tool/environment swap؟

من AGI إلى ASI: recursive improvement يحتاج حاكمية إثبات — تحويلها إلى شرط معماري

Part VIII لا يدعي ASI؛ يبني Recursive Improvement Safety Case: كل self-modification يمر canary، provenance، regression floors، external verifier، rollback، authority، pre/post benchmark، وindependent replication قبل توسيع الصلاحيات.

Evidence Gate

N2: executable test؛ N3: observed trace؛ N4/E4: matched-budget controlled gain؛ N5/E5: cross-model/domain/environment replication.

Contract

كل implementation مرتبط بهذا المحور يجب أن يعلن inputs، mutable state، authority، observables، failure signatures، rollback، وbenchmark owner.

ربطه بالنظريات السابقة

إذا كان المحور يكرر وظيفة موجودة في MCD/MURI/HKOS/MUSQAT/Part VII، لا يُنشأ له theory name؛ يضاف كقيد أو metric أو experiment إلى المالك الأقرب.

OpenAI Agents SDK

الملاحظة المصدرية: Agents/handoffs/guardrails/tracing؛ code-first dynamic orchestration.

الاستفادة المقيدة: Borrow: tracing + explicit handoffs + guardrails. Avoid: treating handoff as authority transfer without MOUSSAIF token/receipt.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح لـfrozen harness للablation أم production convenience فقط؟

Source anchor: W01

Microsoft Agent Framework

الملاحظة المصدرية: Agents + stable harness + graph workflows; sessions, memory, middleware, checkpoints, HITL; successor ideas from AutoGen/Semantic Kernel.

الاستفادة المقيدة: Borrow: harness/workflow separation, typed routing, checkpoints. Map policy/identity outside model.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للablation أم production convenience فقط؟

Source anchor: W11

الملاحظة المصدرية: Durable graph execution, persistence per superstep, pending writes, interrupts, time travel, fault recovery.

الاستفادة المقيدة: Borrow: event/checkpoint semantics and replay; enforce GCIT invariants in state snapshots.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للـablation أم production convenience فقط؟

Source anchor: W12

Google ADK / agents-cli

الملاحظة المصدرية: A2A exposure, project scaffolding, eval/deploy/observe workflow, registry and cloud deployment.

الاستفادة المقيدة: Borrow: lifecycle discipline and A2A adapters; keep canonical state provider-neutral.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للـablation أم production convenience فقط؟

Source anchor: W13

Hugging Face smolagents

الملاحظة المصدرية: CodeAgent and ToolCallingAgent; compact multi-step agent API; code action representation.

الاستفادة المقيدة: Borrow: code-action sandbox lane; require effect/authority validation.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للـablation أم production convenience فقط؟

Source anchor: W14

الملاحظة المصدرية: Crews + flows, guardrails, memory/knowledge, observability, persisted state and enterprise triggers.

الاستفادة المقيدة: Borrow: separation of crews vs flows; benchmark coordination tax before using crews.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للـablation أم production convenience فقط؟

Source anchor: W15

الملاحظة المصدرية: Structured signatures, modular programs and optimization rather than handcrafted prompt strings.

الاستفادة المقيدة: Borrow: compile/optimize query/rubric modules under held-out eval; preserve provenance and safety gates.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للablation أم production convenience فقط؟

Source anchor: W16

MCP 2026-07-28

الملاحظة المصدرية: Stateless request/response core, MRTR, header routing, cacheable catalogs, auth hardening, extension framework.

الاستفادة المقيدة: Use as tool/resource transport. Never map protocol connectivity directly to authority/identity.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للـablation أم production convenience فقط؟

Source anchor: W09

A2A / Agent Protocol Layer

الملاحظة المصدرية: Inter-agent interoperability and remote specialist communication.

الاستفادة المقيدة: Use for agent-to-agent envelopes; attach MOUSSAIF identity/delegation/trace receipts.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للablation أم production convenience فقط؟

Source anchor: W10

Harbor / Terminal-Bench style eval kernel

الملاحظة المصدرية: Containerized reproducible environments, comprehensive tests, rollout interfaces for eval/training.

الاستفادة المقيدة: Borrow as benchmark substrate for E3-E5; separate task tests from verifier-model judgments.

Dimension	تقييم Part VIII
State semantics	هل يوجد checkpoint/resume/versioning أم conversation فقط؟
Authority	هل identity/scope/policy first-class أم خارج framework؟
Tool protocol	function/code/MCP/A2A وكيف تُعزل side effects؟
Observability	traces، metrics، replay، process/outcome؟
Portability	provider lock-in، adapters، canonical state؟
Scientific use	هل يصلح كـfrozen harness للablation أم production convenience فقط؟

Source anchor: W07

تدقيق Part VII — نظرية مساعف لدين الدليل وديناميات تسديده

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

استبدال scalar دين واحد بـtyped debt vector: provenance debt، independence debt، temporal debt، causal debt، authorization debt. إدخال قواعد عدم double-counting عبر dependency DAG وShapley-like marginal exposure عند الحاجة.

ما كان صالحاً

كل اعتماد تشغيلي على claim لم يُغلق دليلُه يولد Evidence Debt. هذا الدين ليس خطأً بحد ذاته؛ قد يكون ضرورياً تحت ضغط الزمن. لكنه يجب أن يُسجل، ويحمل إلى العقدِ التابعةِ، وتزداد كلفته مع مركزية الاعتماد، وخطورة الفعل، وتقادم المصدر، وصعوبة الرجوع. النظام الناضج لا يطلب يقيناً كاملاً؛ بل يدير دين الدليل كما تدار الالتزامات الفنية: سقف، فائدة، استحقاق، وتسديد أو شطب.

Original equation	$ED(c,t)=U_c \times Dep_c \times Risk_c \times (1+Age_c) \times Irrev_c$; $PortfolioDebt = \Sigma\ ED(c,t)+CouplingPenalty$
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	قد يفشل المفهوم إذا كانت uncertainty + risk وحدهما تعطي نفس ترتيب verification. عندها يُخفض EDDT إلى scheduling method لا Theory.

تدقيق Part VII — نظرية مساعف لأحواض الفشل وطوبولوجيا الاسترداد

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

إضافة تعريف تشغيلي للحوض باستخدام recurrence under interventions، لا clustering لغوي. الحوض يثبت فقط إذا زاد P(failure recurrence) داخل neighborhood سببي واستطاعت intervention محددة خفضه.

ما كان صالحاً

الفشل الطويل الأفق يتجمع في Failure Basins: مجموعات حالات مختلفة ظاهرياً لكنها تشارك قيوداً سببية تجعل السياسات الحالية تعيد إنتاج الفشل. الاسترداد الحقيقي هو Basin Escape، أي انتقال يغير متغيراً سببياً أو سلطة أو أداة أو تمثيلا بحيث تصبح ديناميكا الفشل السابقة غير جاذبة.

Original equation	BasinRisk(s)=P(return_to_F | policy, neighborhood(s)); EscapeGain(a)=ΔBasinRisk(a)-Cost(a)-NewRisk(a)
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا لم تتبأ membership بالتكرار أو كان generic retry مساوياً، يُدمج FBRT في Failure Taxonomy/Recovery Planner.

تدقيق Part VII — نظرية مساعف للتخصيص التدخلي للحوسبة

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

رفع ICV من heuristic إلى active experimental design: posterior over causal hypotheses + expected decision regret reduction. فصل observation query عن world intervention، وإضافة identifiability/ authority constraints.

ما كان صالحاً

وحدة تخصيص الحوسبة العليا يجب أن تكون التدخل المتوقع أن يخفض causal uncertainty لكل وحدة كلفة ومخاطرة، لا مجرد branch أو token. عندما يمكن لاختبار صغير في العالم أو sandbox أن يفصل نماذج متنافسة، يجب أن يتفوق على استدلال أطول غير تدخلي.

Original equation	ICV(i)=E[ΔH_causal | do(i)] × DecisionImpact(i) / (ComputeCost+ToolCost+Risk+Latency)
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا لم يتنبأ ICV بالتحسن أو كانت تكلفة التقدير أكبر من فائدته، يُستبدل بـrule بسيطة مثل uncertainty threshold.

تدقيق Part VII — نظرية مساعف لتبلور القدرة القابل للعكس

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

تعريف crystallization كمسألة model selection بين vessels مع قيود rollback، privacy، transfer، volatility، وcatastrophic interference. فصل skill distillation عن fact memorization.

ما كان صالحاً

القدرة تبدأ Scaffolding-dependent ثم قد تتبلور تدريجياً، لا تُدفع إلى وعاء أعمق إلا عندما تثبت causal، transfer، resistance to counterexamples، contribution، وانخفاض dependence على scaffold، وكلما انخفضت reversibility المطلوبة وجب ارتفاع evidence threshold.

Original equation	Crystal(c,v)=Transfer×CausalGain×Stability×CompressionGain×(1-ScaffoldDependence)-IrreversibilityRisk(v)
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا لم تختلف أفضل vessel placements عن static policy بسيطة، تصبح RCCT optimization method لا مستقلة.

تدقيق Part VII — نظرية مساعف لاقتران السببية بالسلطة

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

تحويل الشرعية من gate ثنائي إلى proof-carrying causal edge: actor identity، delegation chain، scope، causal credit، effect، action، policy version. إضافة non-repudiation receipt.

ما كان صالحاً

لا يكفي أن يكون الفعل سبباً في النجاح حتى يُمنح credit تدريبي أو ذاكرة معيارية؛ يجب أن يمر السببية والسلطة معاً. Unauthorized causal success يسجل كـtechnical finding أو counterexample لا كسلوك قابل للترقية.

Original equation	LegitCredit(a)=CausalContribution(a)×AuthorityValidity(a)×PolicyValidity(a)×EvidenceQuality(a)
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا كان hard authorization gate يمنع كل unauthorized paths بحيث لا توجد learning contamination، تظل ACCT مهمة للتدقيق لكنها قد لا تكون Theory مستقلة.

تدقيق Part VII — نظرية مساعف لحفظ الحاكمية عبر التحولات

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

صياغة invariants كـtyped contracts + approximate bisimulation عبر hard invariants. لا transformations تعوضها utility؛ soft invariants لها defect budget تراكمي.

ما كان صالحاً

كل تحول بنيوي في runtime يجب أن يحمل معه Invariant Vector: الهدف، السلطة، هوية المشروع، رتب الأدلة، stop conditions، rollback target، budgets، unresolved obligations. الفشل ليس فقط تغير النتيجة، بل أي Conservation Defect غير معلن في هذه invariants.

Original equation	CD(T)=Σ_k w_k·Distance(I_k^before,I_k^after); valid(T) iff CD(T)≤τ_k for all hard invariants
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا كان generic serialization + tests يحقق نفس GCIT، drift detection يخفض إلى engineering invariant discipline.

تدقيق Part VII — نظرية مساعف لرسملة المعرفة السلبية

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

تمييز negative knowledge إلى أربعة أنواع: forbidden shortcut، failed procedure، falsified hypothesis، boundary counterexample. لكل نوع promotion semantics مختلفة.

ما كان صالحاً

المعرفة السلبية ليست مجرد سجل أخطاء، يمكن أن تكون رأس مال إذا كانت محددة النطاق، سببية قدر الإمكان، قابلة للاسترجاع عند trigger مشابه، ولها expiry ومكلّف. قيمتها تقاس بعدد الأخطاء المتجنبة وقيمة الاستكشاف التي أعادت توجيهها، لا بعدد failures المحفوظة.

Original equation	NKC(n)=AvoidedLoss×Transferability×CausalSpecificity×Retrievability×Freshness-OvergeneralizationRisk
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا لم تقلل repeated errors أو زادت false refusals، لا تُرقَّى negative object ويحتفظ به كـarchive فقط.

تدقيق Part VII — نظرية مساعف لتحويل الاختلاف إلى تجربة

الحكم: الفكرة تستحق الاحتفاظ لكCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

ربط disagreement بنظرية قرار: لا experiment إلا إذا كان disagreement material، أي يغير action أو expected loss. إضافة calibration لـjudge correlation وposterior separation.

ما كان صالحاً

عندما يكون الخلاف material ويقود إلى أفعال مختلفة، يجب تحويله إلى سؤال أو تجربة تفصل الفرضيات بأعلى expected discrimination. الاتفاق ليس الهدف؛ الهدف اختيار observation تجعل واحداً أو أكثر من النماذج يخسر حقه في قيادة القرار.

Original equation	DER(e)=ExpectedPosteriorSeparation(e)×DecisionImpact/(Cost+Risk+Latency)
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا كان majority/judge يساوي DET تحت matched budget، يظل DET فقط في high-impact or high-correlation regimes.

تدقيق Part VII — نظرية مساعف لنصف العمر الإبستيمي للادعاءات

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

نمذجة half-life عبر survival/hazard بدل TTL ثابت، مع regime shifts، source/domain covariates، dependency cascades وdecision-specific freshness authority.

ما كان صالحاً

كل claim تشغيلي يملك Epistemic Half-Life لا يعتمد على العمر الزمني وحده، بل على volatility للمجال، سرعة تغير source، حساسية القرار، وعدد downstream dependencies. بعد half-life لا يصبح الادعاء كاذباً، لكنه يفقد حقه في توجيه فعل مرتفع الأثر دون revalidation.

Original equation	Authority_t(c)=A0·2^(-Δt/HL_c)·SourceStability·Independence ; HL_c=f(volatility,domain,dependency,risk)
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا كان fixed TTL domain-specific مساوياً لـsurvival model، تختصر ETHT إلى policy.

تدقيق Part VII — نظرية مساعف للانتقال من التشبع النظري إلى الاستثمار التجريبي

الحكم: الفكرة تستحق الاحتفاظ كـCandidate لأن لها gap مميزاً، لكن صياغة VII كانت أقرب إلى concept specification من research-grade theory.

ما كان ضعيفاً

تحويل saturation إلى research portfolio optimization مع uncertainty: theory candidate، implementation، benchmark، replication، negative-result analysis كل منها له VOI وتكلفة وتأخير.

ما كان صالحاً

البرنامج البحثي يجب أن يغيّر طوره عندما تنخفض القيمة الهامشية لنظرية جديدة تحت Expected Experimental Information Gain. في طور التشبع، تجمّد التسمية مؤقتاً وتحوّل الموارد إلى calibration، ablation، implementation، negative results، replication. يمكن فتح الطور النظري مجدداً فقط إذا كشفت التجارب فرقاً غير ممثل.

Original equation	Switch if MV_theory < EIG_experiment×EvidenceMultiplier and DuplicateRisk>τ_d ; RDI=VerifiedKnowledgeGain/ResearchCost
Artifacts	10 عناصر، لكنها تحتاج trigger/auth/evidence/rollback/test contracts.
Promotion status	N1-N2 فقط حتى تظهر traces/ablations.
Main kill test	إذا خفض controller novelty الحقيقية أو لم يرفع E3/E4 conversion، يجب تخفيف freeze واستخدامه advisory فقط.

نظرية مساعف لدين الدليل وديناميات تسديده — إعادة التأسيس

Moussaif Evidence-Debt Dynamics Theory

الفجوة: البرنامج السابق يملك سلم دليل، Evidence Ledger، بوابات فعل، ومحاكم تحقق، لكنه يعامل نقص الدليل غالباً كحالة محلية: مقبول/مرفوض/غير محسوم. ما بقي ناقصاً هو تمثيل زمني شبكي للادعاءات التي استُخدمت قبل اكتمال دليلها وكيف يزداد خطرها عندما تبنى فوقها قرارات أو ذاكرة أو مهارات.

الأطروحة المعاد بناؤها: كل اعتماد تشغيلي على claim لم يُغلق دليلُه يولد Evidence Debt. هذا الدين ليس خطأً بحد ذاته، قد يكون ضرورياً تحت ضغط الزمن. لكنه يجب أن يُسجَّل، ويُحمَّل إلى العقد التابعة، وتزداد كلفته مع مركزية الاعتماد، وخطورة الفعل، وتقادم المصدر، وصعوبة الرجوع. النظام الناضج لا يطلب يقيناً كاملاً، بل يدير دين الدليل كما تدار الالتزامات الفنية: سقف، فائدة، استحقاق، وتسديد أو شطب.

التعديل VIII: استبدال scalar دين واحد بـtyped debt vector: provenance debt، independence debt، temporal debt، causal debt، authorization debt. إدخال قواعد عدم double-counting عبر dependency DAG وShapley-like marginal exposure عند الحاجة.

Novelty defense

تجلس فوق Evidence Ledger وMELT وWorld-Effect، وتغذي MCD/QMCD بأولوية تحقق جديدة. تختلف عن Evidence Rank لأنها ديناميكية، تراكمية، شبكية، ومرتبطة بتبعات الاستعمال لا بقوة المصدر فقط.

قاعدة: إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

ليست النظرية دعوة إلى التحقق الكامل قبل كل خطوة، ولا تعتبر كل uncertainty ديناً. الدين ينشأ فقط عندما تُستهلك دعوى غير مغلقة في اعتماد له أثر لاحق.

نظرية مساعف لدين الدليل وديناميات تسديده — النموذج الرسمي

Variable	Operational meaning
U	uncertainty after evidence, 0..1
D	downstream dependency mass
R	action risk
T	temporal staleness
I	irreversibility
J	joint-dependence correction
P	repayment priority

$$ED_c = U_c \cdot D_c \cdot R_c \cdot (1+T_c) \cdot I_c$$

$$ED_{path} = \Sigma_c\, ED_c - J(\text{shared_evidence})$$

$$\text{Priority}(c) = E[\text{RiskReduced} \mid \text{verify } c] / (\text{Cost_verify} + \text{Latency})$$

Identifiability

يجب ألا يُحسب dependence مرتين عندما ترجع عدة claims إلى نفس المصدر؛ الاستقلال بين الأدلة متغير صريح. الدين لا يساوي احتمال الكذب؛ هو تكلفة الاعتماد غير المغلق.

Causal chain

Claim provisional → downstream use → exposure → verification action → state-returned evidence → repay/narrow/deprecate → memory promotion.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف لدين الدليل وديناميات تسديده — Runtime & State Machine

الـalgorithm المرجعي

collect_open_claims → build_dependency_edges → compute_ED → prioritize_repayment → gate_high-impact_action → reconcile_after_world_effect

Implementation substrate

.PostgreSQL/SQLite event ledger + claim DAG; async repayment queue; pre-commit hook; evidence receipts content-addressed

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

يمنع laundering للمصدر الضعيف عبر عشر summaries، كل derived claim يرث source-family id وindependence penalty.

نظرية مساعف لدين الدليل وديناميات تسديده — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-001	TOOL	دفتر دين الدليل Evidence Debt Ledger	+ ClaimRecord dependencies	DebtEntry + priority	read-only	debt clearance rate	source/claim conflation	restore prior state / cancel queued action
P7-002	TOOL	راسم انكشاف الاعتماد Dependency Exposure Mapper	claim graph	exposure graph	read-only	downstream exposure	missing edges	restore prior state / cancel queued action
P7-003	TOOL	مجدول تسديد الدليل Debt Repayment Scheduler	debt portfolio	verification queue	read-only	risk-reduced per cost	starvation	restore prior state / cancel queued action
P7-004	TOOL	بوابة دين الفعل Action Debt Gate	action + debt	allow/escalate/block	read-only	unsafe debt-carried actions	overblocking	restore prior state / cancel queued action
P7-005	METHOD	منهج قياس دين الدليل Evidence Debt Quantification	claim set	ED scores	read-only	predictive validity	bad calibration	no-world-effect
P7-006	METHOD	منهج انتشار الدين Dependency-Weighted Debt Propagation	DAG	propagated debt	read-only	cascade prediction	double counting	no-world-effect
P7-007	PROTOCOL	بروتوكول التسديد قبل الفعل Pre-Action Debt Clearance	action plan	clearance receipt	policy-gated	latency vs risk	global verification	restore prior state / cancel queued action
P7-008	PROTOCOL	بروتوكول مصالحة الدين بعد الأثر Post-Action Debt Reconciliation	world effect	updated ledger	policy-gated	debt accuracy	self-confirmation	restore prior state / cancel queued action
P7-009	STACK	مخزن الدين الإبستيمي Epistemic Debt Store	events	queryable debt state	service identity + least privilege	replay completeness	state corruption	version rollback + replay
P7-010	STACK	خدمة commit الواعية بالدين Debt-Aware Commit Service	commit request	signed verdict	service identity + least privilege	false commit	bypass	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف لدين الدليل وديناميات تسديده — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining. يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف لدين الدليل وديناميات تسديده — Experiment, Ablation & Promotion

التجربة الحاسمة

أنشئ 200 مهمة long-horizon بمصادر متفاوتة القدم والاستقلال، واسمح للوكيل باستخدام claims مؤقتة. قارن Baseline confidence-only مع Debt-aware runtime. القياسات: verification، cascading error depth، delayed false-success، cost، rollback loss، time-to-clearance.

Falsifiers

- إذا لم يفسر Evidence Debt الفشل أفضل من uncertainty وحدها.
- إذا لم يغير ترتيب التحقق أو قرار الفعل تحت matched budget.
- إذا زادت الكلفة والlatency دون انخفاض في false-success أو rollback cost.

Predictions

- إدارة الدين ستخفض false-success المتأخر أكثر من مجرد confidence threshold.
- العقد ذات debt centrality العالية ستفسر نسبة غير خطية من failure cascades.
- Debt-aware scheduling سيعطي قيمة أكبر من uniform verification عند ميزانية تحقق ثابتة.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- provider/model alternative عند E5

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** قد يفشل المفهوم إذا كانت uncertainty + risk وحدهما تعطي نفس ترتيب verification. عندها يخفض EDDT إلى scheduling method لا Theory.

نظرية مساعف لأحواض الفشل وطوبولوجيا الاسترداد — إعادة التأسيس
Moussaif Failure-Basin and Recovery Topology Theory

الفجوة: الذاكرة الحالية تحفظ failure signatures وreplay، ومحركات الاسترداد تملك rollback وretry، لكن إعادة المحاولة قد تعيد النظام إلى الحوض السببي نفسه لأن state مختلفة سطحياً فقط. لا توجد نظرية موحدة تمثل بنية الفشل كمنطقة في فضاء الحالة ومسارات الخروج منها.

الأطروحة المعاد بناؤها: الفشل الطويل الأفق يتجمع في Failure Basins: مجموعات حالات مختلفة ظاهرياً لكنها تشارك قيوداً سببية تجعل السياسات الحالية تعيد إنتاج الفشل. الاسترداد الحقيقي هو Basin Escape، أي انتقال يغيّر متغيراً سببياً أو سلطة أو أداة أو تمثيلا بحيث تصبح ديناميكا الفشل السابقة غير جاذبة.

التعديل VIII: إضافة تعريف تشغيلي للحوض باستخدام recurrence under interventions، لا clustering لغوي. الحوض يثبت فقط إذا زاد P(failure recurrence) داخل neighborhood سببي واستطاعت intervention محددة خفضه.

Novelty defense

يمتد من failure replay وRecovery Contracts وTool Ecology، لكنه يضيف topology + recurrence prediction + intervention-defined escape.

قاعدة: إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

Failure Basin تمثيل هندسي تشغيلي؛ لا يفترض وجود attractors رياضية مستمرة في كل runtime. إذا لم يثبت التجميع السببي يُخفض المصطلح إلى taxonomy.

نظرية مساعف لأحواض الفشل وطوبولوجيا الاسترداد — النموذج الرسمي

Variable	Operational meaning
s	typed runtime state
F	failure terminal set
B	candidate basin
ρ	recurrence probability
a_e	escape intervention
ΔC	causal state change
K	recovery cost

$$\rho_B = P(S_{t+k} \in F \mid S_t \in B, \pi)$$

$$\text{Escape}(a) = \rho_{\text{before}} - \rho_{\text{after}} - \lambda K - \mu \text{NewRisk}$$

$$B \text{ valid iff } \rho_B > \rho_{\text{control}} + \delta$$

Identifiability

Clustering لا يكفي. نحتاج perturb/replay لتمييز correlation من basin dynamics. يجب فصل outages الخارجية عن policy basin داخلي.

Causal chain

Failure trace → causal signature → basin hypothesis → same-basin counterfactual → escape action → replay → recurrence estimate.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف لأحواض الفشل وطوبولوجيا الاسترداد — Runtime & State Machine

الـalgorithm المرجعي

```
trace_to_causal_state → cluster_failures → estimate_return_prob → generate_escape_actions → simulate/replay → choose_minimal_escape
```

Implementation substrate

.Trace vectorizer + causal labels + HDBSCAN/graph clusters as proposal only; evaluator estimates recurrence through replay/simulation

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

يحظر الاسترداد الذي يوسع الصلاحيات لمجرد الهروب من الحوض؛ authority remains hard invariant.

نظرية مساعف لأحواض الفشل وطوبولوجيا الاسترداد — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-011	TOOL	مضمّن حالة الفشل Failure State Embedder	trace	failure state	read-only	recurrence predictability	semantic-only embedding	restore prior state / cancel queued action
P7-012	TOOL	مجمّع أحواض الفشل Basin Clusterer	failure states	basin map	read-only	cluster stability	spurious clusters	restore prior state / cancel queued action
P7-013	TOOL	مخطط الخروج من الحوض Escape Route Planner	basin + actions	escape plan	read-only	escape success	same-basin retry	restore prior state / cancel queued action
P7-014	TOOL	مراقب تباعد الاسترداد Recovery Divergence Monitor	before/after state	divergence receipt	read-only	recurrence drop	false divergence	restore prior state / cancel queued action
P7-015	METHOD	منهج رسم حوض الفشل Failure Basin Mapping	traces	validated basins	read-only	basin purity	label leakage	no-world-effect
P7-016	METHOD	منهج البحث المضاد للخروج Counterfactual Escape Search	basin	ranked escapes	read-only	counterfactual gain	bad simulator	no-world-effect
P7-017	PROTOCOL	بروتوكول منع إعادة الحوض No Same-Basin Retry	failed attempt	retry/rewrite/escalate	policy-gated	repeat-error rate	excess escalation	restore prior state / cancel queued action
P7-018	PROTOCOL	بروتوكول إثبات الاسترداد Recovery Proof Protocol	recovery action	proof bundle	policy-gated	verified recovery	surface success	restore prior state / cancel queued action
P7-019	STACK	مخزن طوبولوجيا الفشل Failure Topology Store	failure events	topology graph	service identity + least privilege	replay value	drift	version rollback + replay
P7-020	STACK	منسق الاسترداد Recovery Orchestrator	escape plan	recovered state	service identity + least privilege	recovery latency	looping	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف لأحواض الفشل وطوبولوجيا الاسترداد — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining. يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف لأحواض الفشل وطوبولوجيا الاسترداد — Experiment, Ablation & Promotion

التجربة الحاسمة

اجمع 1000 trace فيها partial state loss، permission denial، bad verifier، stale memory، tool outage. قارن scripted recovery، random retry، وBasin Escape تحت نفس budget. اختبر النقل إلى provider مختلف.

Falsifiers

- إذا لم تتماسك clusters عبر مهام/نماذج مختلفة.
- إذا لم تتنبأ basin membership باحتمال التكرار.
- إذا لم يحقق escape planner تحسناً مقابل retry/rollback baselines.

Predictions

- ستكون معدلات الفشل المتكرر أعلى داخل basin clusters من التجميع الدلالي العادي.
- Same-basin retry سيبدو ناجحاً محلياً لكنه يرفع recurrence خلال الأفق الطويل.
- Basin-aware recovery سيخفض recovery latency وعدد المحاولات أكثر من generic backoff.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- E5 عند provider/model alternative

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** إذا لم تتنبأ membership بالتكرار أو كان generic retry مساوياً، يُدمج FBRT في Failure Taxonomy/Recovery Planner.

نظرية مساعف للتخصيص التدخلي للحوسبة — إعادة التأسيس

Moussaif Interventional Compute Allocation Theory

الفجوة: MCD/QMCD يوجهان العمق نحو الفرق والسؤال ذي القيمة، لكن عندما تكون المشكلة سببية يبقى السؤال: أي computation أو tool call أو experiment يزيل أكبر قدر من عدم اليقين السببي؟ البحث الإضافي قد يزيد النص ولا يحدد السبب.

الأطروحة المعاد بناؤها: وحدة تخصيص الحوسبة العليا يجب أن تكون التدخل المتوقع أن يخفض causal uncertainty لكل وحدة كلفة ومخاطرة، لا مجرد branch أو token. عندما يمكن لاختبار صغير في العالم أو sandbox أن يفصل نماذج متنافسة، يجب أن يتفوق على استدلال أطول غير تدخلي.

التعديل VIII: رفع ICV من heuristic إلى active experimental design: posterior over causal hypotheses + expected decision regret reduction. فصل observation query عن world intervention، وإضافة identifiability/authority constraints.

Novelty defense

يعمق MCD/QMCD وPart VI causal sufficiency ويستفيد من HEC-EBAFM تخلفية عن خفض عدم اليقين، لكنه لا يساوي entropy reduction بالذكاء أو الوكالة.

> **قاعدة:** إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

ليست كل مهمة قابلة لdo-intervention. عند غياب بيئة قابلة للتجريب يعود النظام إلى observational reasoning ويصرح بضعف identifiability.

نظرية مساعف للتخصيص التدخلي للحوسبة — النموذج الرسمي

Variable	Operational meaning
H	causal hypothesis set
e	candidate experiment
IG	expected information gain
DR	expected decision-regret reduction
C	cost
R	risk
A	authorization feasibility

$$IG(e)=H[p(H)]-E_y\ H[p(H\mid y,e)]$$

$$ICV(e)=A_e\cdot(\alpha IG+\beta DR)/(C+\lambda R+\mu Latency)$$

$$\text{Stop iff } \max_e ICV(e)<\tau$$

Identifiability

إذا كان simulator مبنياً على النموذج نفسه قد ينتج circular confidence. world/sandbox observations تُعطى رتبة أعلى من imagined outcomes.

Causal chain

Competing hypotheses → discriminating experiment → authorize → execute/simulate → observe → posterior update → route/stop.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف للتخصيص التدخلي للحوسبة — Runtime & State Machine

الـalgorithm المرجعي

infer_causal_hypotheses → enumerate_interventions → estimate_ICV → authorize → execute/simulate → update_graph → stop_on_saturation

Implementation substrate

.Bayesian/ensemble hypothesis store; experiment adapters; budget allocator; sandbox runner; deterministic measurement hooks

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

لا يسمح ICV مرتفع بتجاوز policy. التجارب على العالم الحقيقي تحتاج reversible staging أو approval.

نظرية مساعف للتخصيص التدخلي للحوسبة — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-021	TOOL	مولد التدخلات Intervention Candidate Generator	causal hypotheses	candidate interventions	read-only	coverage	irrelevant probes	restore prior state / cancel queued action
P7-022	TOOL	مقدّر عدم اليقين السببي Causal Uncertainty Estimator	graph + evidence	uncertainty map	read-only	calibration	self-score bias	restore prior state / cancel queued action
P7-023	TOOL	موجّه مزاد الحوسبة Compute Auction Router	interventions + budget	allocation	read-only	verified effect/compute	greedy lock-in	restore prior state / cancel queued action
P7-024	TOOL	لوحة قيمة التجارب Experiment Value Dashboard	experiment logs	value curves	read-only	forecast error	metric gaming	restore prior state / cancel queued action
P7-025	METHOD	منهج كسب المعلومات السببي Expected Causal Information Gain	candidate i	E[ΔH]	read-only	predictive correlation	bad prior	no-world-effect
P7-026	METHOD	منهج مزاد الحوسبة المعدل بالخطر Risk-Adjusted Compute Auction	ICV set	portfolio	read-only	risk-adjusted utility	risk underestimation	no-world-effect
P7-027	PROTOCOL	بروتوكول تخصيص الميزانية التدخلي Interventional Budget Allocation	task	allocation receipt	policy-gated	cost reduction	premature tool use	restore prior state / cancel queued action
P7-028	PROTOCOL	بروتوكول التوقف عند التشبع السببي Stop-on-Causal-Saturation	ICV frontier	stop/escalate	policy-gated	wasted compute	premature stop	restore prior state / cancel queued action
P7-029	STACK	مخطط التجارب السببية Causal Experiment Planner	task state	experiment DAG	service identity + least privilege	decision gain	graph error	version rollback + replay
P7-030	STACK	مجدول الحوسبة التكيفي Adaptive Compute Scheduler	DAG + resources	schedule	service identity + least privilege	cost per effect	contention	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف للتخصيص التدخلي للحوسبة — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف للتخصيص التدخلي للحوسبة — Experiment, Ablation & Promotion

التجربة الحاسمة

Benchmark factorial retrieval/memory/tool perturbations؛ اسمح للنظام بالاختيار بين 6 أنواع compute/intervention. قارن tool-first، token-scaling، MCD-only، وICAT.

Falsifiers

- إذا لم يتنبأ ICV بقيمة التدخل الفعلية.
- إذا لم يخفض cost-per-verified-effect مقابل adaptive reasoning baselines.
- إذا تحولت التقديرات إلى circular self-scoring بلا state-returned evidence.

Predictions

- على المسائل السببية، سيقل عدد tokens مع ارتفاع القرار الصحيح.
- الاختبارات الصغيرة عالية ICV ستتفوق على المزيد من self-consistency في hard-tail cases.
- التحسن سيختفي في مهام لا تحمل عدم يقين سببي، ما يثبت boundary.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- E5 عند provider/model alternative

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** إذا لم يتنبأ ICV بالتحسن أو كانت تكلفة التقدير أكبر من فائدته، يُستبدل بrule بسيطة مثل uncertainty threshold.

نظرية مساعف لتبلور القدرة القابل للعكس — إعادة التأسيس

Moussaif Reversible Capability Crystallization Theory

الفجوة: MUSQAT وMUG يميزان context/memory/skill/adapter/weight ويقترحان ترقية انتقائية، لكن لا يوجد قانون موحد يحدد متى تصبح القدرة مستقرة بما يكفي للانتقال إلى وعاء أقل قابلية للعكس.

الأطروحة المعاد بناؤها: القدرة تبدأ Scaffolding-dependent ثم قد تتبلور تدريجياً. لا تُدفع إلى وعاء أعمق إلا عندما تثبت resistance to counterexamples، causal contribution، transfer، وانخفاض dependence على scaffold؛ وكلما انخفضت reversibility المطلوبة وجب ارتفاع evidence threshold.

التعديل VIII: تعريف crystallization كمسألة model selection بين vessels مع قيود rollback، privacy، transfer، volatility، وcatastrophic interference. فصل skill distillation عن fact memorization.

Novelty defense

تغلق الفجوة بين MUSQAT development/promotion وMATI/AGILLM internalization وMUG vessel placement، عبر متغير صريح: reversibility gradient.

قاعدة: إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

لا تعني النظرية أن كل قدرة يجب أن تدخل weights؛ في أنظمة مغلقة قد تنتهي السلسلة عند skill/adapter خارجي.

نظرية مساعف لتبلور القدرة القابل للعكس — النموذج الرسمي

Variable	Operational meaning
SD	scaffold dependence
Tr	transfer score
St	stability
Vol	knowledge volatility
RB	rollback cost
NI	negative interference
CG	compression gain

$$Crystal(v) = Tr \cdot St \cdot CG \cdot (1 - SD) - \lambda Vol(v) - \mu RB(v) - \nu NI(v)$$

Promote $v \rightarrow v+1$ iff $\Delta Utility > 0$ and all regression floors pass

$$SD = 1 - Perf_{without_scaffold} / Perf_{with_scaffold}$$

Identifiability

لا تُنسب القدرة للوزن إذا بقي tool/context hidden support. provider/model swap يساعد على تحديد ما إذا كانت القدرة في scaffold أم model.

Causal chain

Experience → candidate skill → scaffold ablation → vessel trial → transfer/regression → canary → promote/rollback.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف لتبلور القدرة القابل للعكس — Runtime & State Machine

algorithm المرجعي

```
measure_scaffold_dependence → score_vessels → run_transfer_ablation → canary_promote → regression_test → freeze_or_rollback
```

Implementation substrate

.Capability registry; evaluation matrix; dataset compiler; adapter trainer for open models; rollback manifests; fact-vs-procedure classifier

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

.volatile/sensitive/user-specific knowledge has high weight-promotion penalty; right-to-forget and provenance must survive distillation

نظرية مساعف لتبلور القدرة القابل للعكس — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-031	TOOL	مقياس الاعتماد على السقالة Scaffold Dependence Meter	capability + ablations	dependence curve	read-only	transfer prediction	confounded removal	restore prior state / cancel queued action
P7-032	TOOL	محسن موضع القدرة Vessel Placement Optimizer	candidate capability	vessel recommendation	read-only	retention-cost	static heuristic	restore prior state / cancel queued action
P7-033	TOOL	مقدّر ثبات النقل Transfer Stability Estimator	rollouts	stability score	read-only	cross-domain retention	small sample	restore prior state / cancel queued action
P7-034	TOOL	بوابة التبلور Crystallization Gate	candidate	promote/hold/reject	read-only	promotion precision	premature promotion	restore prior state / cancel queued action
P7-035	METHOD	منهج تدرج القابلية للعكس Reversibility Gradient Analysis	vessels	gradient	read-only	rollback predictability	cost model drift	no-world-effect
P7-036	METHOD	منهج إزالة السقالة Scaffold Removal Curriculum	scaffolded task	removal curve	read-only	independence gain	curriculum leakage	no-world-effect
P7-037	PROTOCOL	بروتوكول الترقية من السياق إلى الوزن Context-to-Weight Promotion	experience	versioned promotion	policy-gated	negative transfer	shortcut	restore prior state / cancel queued action
P7-038	PROTOCOL	بروتوكول الانحدار والإلغاء Regression-and-Unlearning Rollback	candidate model	rollback verdict	policy-gated	restoration fidelity	irreversible change	restore prior state / cancel queued action
P7-039	STACK	سجل دورة حياة القدرة Capability Lifecycle Registry	capability events	lifecycle state	service identity + least privilege	audit completeness	orphan state	version rollback + replay
P7-040	STACK	مترجم المهارة إلى Adapter Skill-Adapter Compiler	skill capsule	adapter dataset/package	service identity + least privilege	distillation gain	overfit	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف لتبلور القدرة القابل للعكس — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف لتبلور القدرة القابل للعكس — Experiment, Ablation & Promotion

التجربة الحاسمة

اختَر 20 HyperSkills في coding/research/tool-use. نفذ scaffold-removal ladder، ثم skill/adapter distillation على model مفتوح، وقارن retention/transfer/rollback cost.

Falsifiers

- إذا لم يرتبط Scaffold Dependence بفشل النقل.
- إذا لم يختلف vessel الأمثل باختلاف volatility/risk.
- إذا لم تضف البوابة قيمة مقابل MUSQAT promotion heuristic الأبسط.

Predictions

- بعض القدرات ستتحسن بعد distillation بينما facts المتغيرة ستتدهور إذا تبلورت وزنياً.
- Scaffold Dependence سيكون predictor أفضل للترقية من frequency وحدها.
- Reversibility-aware promotion سيخفض forgetting والـpolicy drift.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- provider/model alternative عند E5

Promotion

N3 عند trace+postcondition، E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** إذا لم تختلف أفضل vessel placements عن static policy بسيطة، تصبح RCCT optimization method لا مستقلة.

نظرية مساعف لاقتران السببية بالسلطة — إعادة التأسيس

Moussaif Authority-Causal Coupling Theory

الفجوة: المشروع يفصل جيداً بين authority وcausal attribution، لكن يمكن للنظام أن يحقق أثراً مفيداً سببياً عبر فعل لم يكن مأذوناً، إذا كوفئ على النتيجة فقط فقد يتعلم مساراً ناجحاً لكنه غير مشروع.

الأطروحة المعاد بناؤها: لا يكفي أن يكون الفعل سبباً في النجاح حتى يُمنح credit تدريبي أو ذا كرة معيارية؛ يجب أن يمر السببية والسلطة معاً. Unauthorized causal success يسجل كـtechnical finding أو counterexample لا كسلوك قابل للترقية.

التعديل VIII: تحويل الشرعية من gate ثنائي إلى causal credit، effect، action، policy version، scope، delegation chain، proof-carrying causal edge: actor identity. إضافة non-repudiation receipt.

Novelty defense

تربط World-Effect attribution، Protocol Sovereignty، Institutional Sovereignty، T0-T4، وMUSQAT trust capital.

قاعدة: إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

السلطة لا تحدد الحقيقة العلمية؛ النظرية تخص ترقية السلوك والفعل داخل نظام وكيلي، لا صحة claim مجرد.

نظرية مساعف لاقتران السببية بالسلطة — النموذج الرسمي

Variable	Operational meaning
CC	causal contribution
AV	authority validity
PV	policy validity
IA	identity assurance
E	evidence quality
NR	non-repudiation quality

$$LegitCredit=CC\cdot AV\cdot PV\cdot IA\cdot E\cdot NR$$

$$UnsafeShortcut=CC\cdot(1-AV\cdot PV)$$

$$PromotionCredit=\max(0,LegitCredit-\lambda UnsafeShortcut)$$

Identifiability

Outcome success لا يكشف authorization. يجب join مستقل مع identity/policy logs. policy after-the-fact لا يعيد شرعنة action قديم.

Causal chain

Delegation → capability token → action → effect → attribution → legitimacy join → learning/promotion or negative case.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف لاقتران السببية بالسلطة — Runtime & State Machine

الـalgorithm المرجعي

join_trace_with_authority → estimate_causal_credit → discount_illegitimate_credit → store_negative_case → promote_only_legitimate_pattern

Implementation substrate

.Signed capability tokens; policy engine; append-only authority ledger; causal joiner; training filter; audit receipts

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

مستوحى من NIST identity/authorization/non-repudiation؛ MCP/A2A transport metadata لا يكفي لإثبات الحق.

نظرية مساعف لاقتران السببية بالسلطة — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-041	TOOL	واصل الأثر بالسلطة Authority-Causal Trace Joiner	trace + auth logs	joined graph	read-only	join completeness	identity mismatch	restore prior state / cancel queued action
P7-042	TOOL	كاشف النجاح غير المأذون Unauthorized Success Detector	trajectory	flagged success	read-only	unsafe shortcut recall	false alarm	restore prior state / cancel queued action
P7-043	TOOL	مصفوفة حقوق القدرة Capability Rights Matrix	agents/tools/actions	rights matrix	read-only	least privilege	stale policy	restore prior state / cancel queued action
P7-044	TOOL	مدقق شرعية الترقية Promotion Legitimacy Checker	candidate trace	legit/negative	read-only	promotion safety	policy ambiguity	restore prior state / cancel queued action
P7-045	METHOD	منهج تفكيك السبب والسلطة Causal-Authority Factorization	trajectory	factors	read-only	explanatory value	correlated factors	no-world-effect
P7-046	METHOD	منهج إسناد الائتمان المشروع Legitimate Credit Assignment	traces	credit labels	read-only	safe transfer	reward sparsity	no-world-effect
P7-047	PROTOCOL	بروتوكول لا ائتمان بلا سلطة -No-Credit-without Authority	success trace	credit/reject	policy-gated	unsafe learning	overrestriction	restore prior state / cancel queued action
P7-048	PROTOCOL	بروتوكول تحقق سلسلة التفويض Delegation Chain Validation	delegation	validated chain	policy-gated	privilege drift	chain truncation	restore prior state / cancel queued action
P7-049	STACK	دفتر السببية والسلطة Authority-Causal Ledger	events	auditable graph	service identity + least privilege	audit latency	tamper risk	version rollback + replay
P7-050	STACK	بوابة تعلم واعية بالشرعية Legitimacy-Aware Learning Gate	candidate experience	approved set	service identity + least privilege	unsafe promotion rate	policy lag	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف لاقتران السببية بالسلطة — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف لاقتران السببية بالسلطة — Experiment, Ablation & Promotion

التجربة الحاسمة

بيئة tool-use بها مسارات سريعة غير مأذونة ومسارات أبطأ مأذونة. درّب/طوّر skill bank من traces مختلفة، ثم قس transfer إلى مهام جديدة.

Falsifiers

- إذا لم تظهر shortcut traces في baselines.
- إذا خفّض LegitCredit النجاح دون أي تحسن في unsafe action أو auditability.
- إذا أمكن تحقيق نفس الأثر بمجرد hard permission gate دون learning difference.

Predictions

- Outcome-only learning سيرفع unsafe shortcut rate حتى مع task success أعلى.
- LegitCredit سيحسن policy compliance مع انخفاض محدود في النجاح الخام.
- الفصل سيكشف أدوات/وكلاء يحققون نتائج عبر سلطة أوسع من اللازم.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- provider/model alternative عند E5

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** إذا كان hard authorization gate يمنع كل unauthorized paths بحيث لا توجد learning contamination، تظل ACCT مهمة للتدقيق لكنها قد لا تكون Theory مستقلة.

نظرية مساعف لحفظ الحاكمية عبر التحولات — إعادة التأسيس

Moussaif Governance Conservation under Transformations Theory

الفجوة: Part VI قدّم Trajectory Assurance وCross-Provider State Conservation، لكن لا يزال يلزم formal object واحد يقيس ما الذي يجب أن يبقى invariant عند memory compaction، protocol translation، provider swap، restart، delegation، أو model update.

الأطروحة المعاد بناؤها: كل تحول بنيوي في runtime يجب أن يحمل معه Invariant Vector: الهدف، السلطة، هوية المشروع، رتب الأدلة، rollback target، stop conditions، budgets، unresolved obligations. الفشل ليس فقط تغير النتيجة؛ بل أي Conservation Defect غير معلن في هذه invariants.

التعديل VIII: صياغة invariants كـapproximate bisimulation + typed contracts عبر transformations. hard invariants لا تعوضها utility؛ soft invariants لها defect budget تراكمي.

Novelty defense

يوحد Long-Horizon Reconstructive Coherence، Part VI trajectory assurance، Goal-as-OS Compass، وMURI rank preservation في contract واحد.

> **قاعدة:** إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

الحفظ لا يعني تطابق النص أو plan؛ المطلوب حفظ الشروط الحاكمة مع السماح بخطة جديدة.

نظرية مساعف لحفظ الحاكمية عبر التحولات — النموذج الرسمي

Variable	Operational meaning
I_h	hard invariant vector
I_s	soft invariant vector
T	state transformation
d_k	semantic distance per invariant
CD	conservation defect
B	defect budget

$$CD(T)=\Sigma_k w_k d_k(I_k^{pre},I_k^{post})$$

$$valid(T) \text{ iff } \max_{k\in hard} d_k\le\tau_k \text{ and } \Sigma_{soft} w_k d_k\le B$$

$$CD_{path}=\Sigma_t CD(T_t)+InteractionPenalty$$

Identifiability

تطابق النص ليس حفظاً. يجب مقارنة stop/rollback، budgets، pending effects، evidence rank، authority، goal contract: semantics.

Causal chain

Snapshot → transform/provider swap/compaction/delegation → semantic comparator → defect → repair/reject → receipt.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف لحفظ الحاكمية عبر التحولات — Runtime & State Machine

الـalgorithm المرجعي

snapshot_invariants → transform → compare → localize_defect → compensate_or_reject → append_conservation_receipt

Implementation substrate

.Canonical state schema; invariant adapters; semantic comparators; provider-specific codecs; transition receipts; prefix monitor

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

يمنع context compaction من إسقاط approval/denial أو pending side effects، يمنع delegation من توسيع scope صامتاً.

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-051	TOOL	مراقب متجه الثوابت Invariant Vector Monitor	state transitions	invariant trace	read-only	drift detection	wrong invariant	restore prior state / cancel queued action
P7-052	TOOL	حاسب عيب الحفظ Conservation Defect Calculator	snapshots	CD score	read-only	failure prediction	metric mismatch	restore prior state / cancel queued action
P7-053	TOOL	مدقق حدود التحويل Transformation Boundary Checker	transition	pass/repair	read-only	boundary safety	hidden state	restore prior state / cancel queued action
P7-054	TOOL	رسم موضع الانحراف Drift Localization Graph	trace	culprit graph	read-only	localization precision	shared cause	restore prior state / cancel queued action
P7-055	METHOD	منهج نقل الثوابت Invariant Transport Analysis	contract	transport map	read-only	equivalence	semantic mismatch	no-world-effect
P7-056	METHOD	منهج ميزانية عيوب الحفظ Cumulative Defect Budgeting	defects	budget state	read-only	long-horizon prediction	compounding assumptions	no-world-effect
P7-057	PROTOCOL	بروتوكول حفظ تبديل المزود والبروتوكول Provider-Protocol Swap Conservation	swap	conservation receipt	policy-gated	swap parity	adapter leakage	restore prior state / cancel queued action
P7-058	PROTOCOL	بروتوكول سلامة بادئات المسار Long-Horizon Prefix Integrity	trajectory	prefix verdicts	policy-gated	early detection	verification cost	restore prior state / cancel queued action
P7-059	STACK	نواة حالة الثوابت Invariant State Kernel	task contract	invariant state	service identity + least privilege	recovery fidelity	state split	version rollback + replay
P7-060	STACK	مراقب حفظ المسار Trajectory Conservation Monitor	event stream	alerts/receipts	service identity + least privilege	lead time	noise	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف لحفظ الحاكمية عبر التحولات — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source، إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف لحفظ الحاكمية عبر التحولات — Experiment, Ablation & Promotion

التجربة الحاسمة

اختبارات restart/provider swap/MCP↔A2A delegation/context compaction في 150 مهمة؛ inject one invariant loss at a time.

Falsifiers

- إذا لم يتنبأ CD بالفشل أو تغير الحكم.
- إذا كانت invariants قابلة للاسترجاع من output فقط دون كلفة إضافية.
- إذا لم يختلف hard/soft invariant handling عن generic state serialization.

Predictions

- معظم drift الطويل سيظهر كعيب صغير متراكم قبل failure النهائي.
- Provider swap parity ستتحسن عند نقل invariants كبيانات typed لا prompt نصي.
- Compaction قد يحافظ semantics العامة لكنه يفشل authority/evidence invariants ما لم تُفصل.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- E5 عند provider/model alternative

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** إذا كان generic serialization + tests يحقق نفس GCIT، drift detection يخفض إلى engineering invariant discipline.

نظرية مساعف لرسملة المعرفة السلبية — إعادة التأسيس

Moussaif Negative-Knowledge Capitalization Theory

الفجوة: المشروع يحفظ الفشل والـnegative results، لكن غالباً كأرشيف دفاعي. ما ينقص هو قانون يحول «ما لا يعمل، ولماذا، وتحت أي شروط» إلى أصل معرفي نشط يمنع إعادة الاكتشاف ويولد أسئلة وقيوداً جديدة.

الأطروحة المعاد بناؤها: المعرفة السلبية ليست مجرد سجل أخطاء، يمكن أن تكون رأس مال إذا كانت محددة النطاق، سببية قدر الإمكان، قابلة للاسترجاع عند trigger مشابه، ولها expiry ومكذّب. قيمتها تقاس بعدد الأخطاء المتجنبة وقيمة الاستكشاف التي أعادت توجيهها، لا بعدد failures المحفوظة.

التعديل VIII: تمييز negative knowledge إلى أربعة أنواع: boundary counterexample، forbidden shortcut، failed procedure، falsified hypothesis. لكل نوع promotion semantics مختلفة.

Novelty defense

يمتد من MUSQAT ExperienceGraph، Failure Memory، Self-Reconstruction Memory، وPart VII Failure Basins.

قاعدة: إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

لا يُعامل الرفض أو النتيجة السلبية حقيقة أبدية؛ كل NegativeCase له scope وexpiry ومقابل إيجابي محتمل.

نظرية مساعف لرسملة المعرفة السلبية — النموذج الرسمي

Variable	Operational meaning
AL	avoided loss
Tr	transferability
CS	causal specificity
Ret	retrievability
F	freshness
OG	overgeneralization risk

$$NKC = AL \cdot Tr \cdot CS \cdot Ret \cdot F - \lambda OG$$

Promote_negative iff precision_scope≥p0 and transfer≥t0

Expire if environment/tool version changes beyond δ

Identifiability

Failure log ليس معرفة سلبية. يجب معرفة لماذا فشل وما الذي كان سيجعل الحكم مختلفاً. failure بسبب outage لا يولد prohibition عن tool class.

Causal chain

Failure → causal label → scope → counterexample search → transfer → negative object → pre-route retrieval → avoided recurrence.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف لرسملة المعرفة السلبية — Runtime & State Machine

الـalgorithm المرجعي

capture_failure → causal_label → derive_scope → create_negative_case → transfer_test → promote_warning/prohibition → expire/revalidate

Implementation substrate

.NegativeKnowledge graph; counterexample retriever; warning/prohibition skills; version/expiry hooks; avoided-loss telemetry

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

يحذر من over-refusal/paralysis. كل prohibition يحتاج scope + expiry + positive exception path.

نظرية مساعف لرسملة المعرفة السلبية — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-061	TOOL	مترجم الحالة السلبية Negative Case Compiler	failed trace	negative object	read-only	reuse rate	bad attribution	restore prior state / cancel queued action
P7-062	TOOL	سجل الفرضيات الساقطة Failed Hypothesis Registry	experiments	rejected set	read-only	rediscovery reduction	premature rejection	restore prior state / cancel queued action
P7-063	TOOL	مسترجع الأمثلة المضادة Counterexample Retriever	query state	negative evidence	read-only	error avoidance	overretrieval	restore prior state / cancel queued action
P7-064	TOOL	مولد مهارة المنع Prohibition Skill Generator	negative cases	skill capsule	read-only	precision/recall	paralysis	restore prior state / cancel queued action
P7-065	METHOD	منهج تقطير النقل السلبي Negative Transfer Distillation	cases	scope rule	read-only	boundary accuracy	overgeneralization	no-world-effect
P7-066	METHOD	منهج تنويع الفشل Failure Diversity Sampling	failure pool	sample set	read-only	coverage	rare-case bias	no-world-effect
P7-067	PROTOCOL	بروتوكول احفظ السلبي قبل الإصلاح -Preserve-Negative-before Repair	failure	immutable negative receipt	policy-gated	trace completeness	storage bloat	restore prior state / cancel queued action
P7-068	PROTOCOL	بروتوكول ترقية السلب المحدود Scope-Bound Negative Promotion	negative candidate	promote/hold	policy-gated	negative transfer	weak scope	restore prior state / cancel queued action
P7-069	STACK	مخزن المعرفة السلبية Negative Knowledge Store	negative objects	queryable NK	service identity + least privilege	avoidance gain	staleness	version rollback + replay
P7-070	STACK	موجّه منع التكرار Anti-Recurrence Router	task + NK	route penalty	service identity + least privilege	recurrence reduction	excess penalty	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف لرسملة المعرفة السلبية — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source، إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف لرسملة المعرفة السلبية — Experiment, Ablation & Promotion

التجربة الحاسمة

قارن failure-log، positive-only memory، وNKC structured bank على tasks متكررة مع perturbations وتحديثات أدوات.

Falsifiers

- إذا لم تقل الأخطاء المتكررة.
- إذا أدت negative cases إلى paralysis أو excessive refusal أكثر من الفائدة.
- إذا لم تنتقل القيمة خارج المهمة الأصلية.

Predictions

- Negative-case retrieval سيخفض repeated-error rate أكثر من positive examples وحدها.
- Overgeneralized prohibitions ستظهر إذا غاب scope/expiry.
- التنوع السببي في negative bank أهم من الحجم الخام.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- provider/model alternative عند E5

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** إذا لم تقل repeated errors أو زادت false refusals، لا تُرقّى negative object ويحتفظ به كـarchive فقط.

نظرية مساعف لتحويل الاختلاف إلى تجربة — إعادة التأسيس

Moussaif Disagreement-to-Experiment Theory

الفجوة: QMCD/MCAI/MURI تحفظ نماذج وحُسُب ومسارات متنافسة، والمحاكم تقارنها، لكن majority vote أو verifier arbitration قد يغلق الخلاف من دون الحصول على معلومة جديدة.

الأطروحة المعاد بناؤها: عندما يكون الخلاف material ويقود إلى أفعال مختلفة، يجب تحويله إلى سؤال أو تجربة تفصل الفرضيات بأعلى expected discrimination. الاتفاق ليس الهدف؛ الهدف اختيار observation تجعل واحداً أو أكثر من النماذج يخسر حقه في قيادة القرار.

التعديل VIII: ربط disagreement بنظرية قرار: لا experiment إلا إذا كان disagreement material، أي يغير action أو expected loss. إضافة calibration لـjudge correlation وposterior separation.

Novelty defense

يصل MURI plural calculi وMCAI competing paths وverifier disagreement وICAT ضمن حلقة تجريبية.

قاعدة: إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

الخلاف غير المهم أو منخفض الأثر لا يستحق experiment؛ يستعمل MCD/QMCD threshold.

نظرية مساعف لتحويل الاختلاف إلى تجربة — النموذج الرسمي

Variable	Operational meaning
M	competing models/calculi
ΔP	prediction divergence
DI	decision impact
PS	posterior separation
C	experiment cost
Corr	error correlation

$$\text{Materiality} = \Delta P \cdot DI$$

$$DER(e) = E[PS_e] \cdot DI / (C + Risk + Latency)$$

$$\text{Experiment if } DER > JudgeGain \cdot (1 - Corr)$$

Identifiability

اختلاف الشرح مع نفس prediction ليس disagreement material. correlated models لا تعد أدلة مستقلة.

Causal chain

Paths disagree → classify source → divergent predictions → experiment design → observation → update/reject/escalate.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف لتحويل الاختلاف إلى تجربة — Runtime & State Machine

الـalgorithm المرجعي

detect_material_disagreement → extract_divergent_predictions → generate_discriminative_experiments → score_DER → execute → update_paths

Implementation substrate

.Disagreement broker; assumption schema; experiment generator; judge-correlation estimator; human escalation packet

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

لا تُستخدم experiment خطرة لحسم disagreement منخفض الأثر. authority and reversibility gates precede epistemic value.

نظرية مساعف لتحويل الاختلاف إلى تجربة — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-071	TOOL	مصفوفة الاختلاف Disagreement Matrix	candidate paths	matrix	read-only	materiality detection	surface disagreement	restore prior state / cancel queued action
P7-072	TOOL	مولد التجربة الفاصلة Discriminative Experiment Generator	disagreement	experiments	read-only	separation gain	unexecutable tests	restore prior state / cancel queued action
P7-073	TOOL	حالّ تعارض الحُسب Calculus Conflict Resolver	plural calculi	conflict map	read-only	assumption clarity	forced unification	restore prior state / cancel queued action
P7-074	TOOL	مخطط تحكيم المحققين Verifier Arbitration Planner	verifier outputs	arbitration route	read-only	correlated error escape	judge dominance	restore prior state / cancel queued action
P7-075	METHOD	منهج السؤال الفاصل Hypothesis-Separating Query Design	hypotheses	question	read-only	posterior separation	leading query	no-world-effect
P7-076	METHOD	منهج كسب حل الاختلاف Expected Disagreement Resolution Gain	candidate experiments	DER scores	read-only	forecast accuracy	bad priors	no-world-effect
P7-077	PROTOCOL	بروتوكول لا أغلبية بلا تجربة -No-Majority-without Experiment	votes	experiment/escalate	policy-gated	correlated error	extra latency	restore prior state / cancel queued action
P7-078	PROTOCOL	بروتوكول تصعيد الاختلاف غير القابل للحسم Irreducible Disagreement Escalation	unresolved case	escalation packet	policy-gated	honest uncertainty	overescalation	restore prior state / cancel queued action
P7-079	STACK	وسيط الاختلاف Disagreement Broker	paths	DisagreementCase	service identity + least privilege	trace quality	schema mismatch	version rollback + replay
P7-080	STACK	خدمة التحكيم بالتجربة Experiment Arbitration Service	case	verdict/evidence	service identity + least privilege	decision gain	tool unavailability	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف لتحويل الاختلاف إلى تجربة — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف لتحويل الاختلاف إلى تجربة — Experiment, Ablation & Promotion

التجربة الحاسمة

مهام علمية/تشغيلية بثلاثة models/calculi مع تضارب intentional. قارن judge model، majority، وDET experiment selection.

Falsifiers

- إذا لم تحسن التجارب الفصل عن verifier vote.
- إذا كان معظم disagreement لغوياً بلا prediction difference.
- إذا زادت كلفة الاختبار دون تغير decision quality.

Predictions

- سيقل correlated verifier error مقارنة بالتصويت.
- الأسئلة التي تركز على posterior separation ستستهلك أدلة أقل وتنتج قراراً أقوى.
- بعض disagreements ستتحول إلى ontology mismatch بدلاً من score conflict.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- provider/model alternative عند E5

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression، E5 عند replication. **Null outcome:** إذا كان majority/judge يساوي DET تحت matched budget، يظل DET فقط في high-impact or high-correlation regimes.

نظرية مساعف لنصف العمر الإبستيمي للادعاءات — إعادة التأسيس

Moussaif Epistemic Temporal Half-Life Theory

الفجوة: المصادر الحالية تحمل recency وexpiry وfreshness، لكن لا توجد صياغة تجعل حق الادعاء في قيادة الفعل يتناقص بحسب تقلب المجال وخطورة القرار واعتمادياته.

الأطروحة المعاد بناؤها: كل claim تشغيلي يملك Epistemic Half-Life لا يعتمد على العمر الزمني وحده، بل على volatility للمجال، سرعة تغير source، حساسية القرار، وعدد downstream dependencies. بعد half-life لا يصبح الادعاء كاذباً، لكنه يفقد حقه في توجيه فعل مرتفع الأثر دون revalidation.

التعديل VIII: نمذجة half-life عبر survival/hazard بدل TTL ثابت، مع dependency cascades، regime shifts، source/domain covariates، وdecision-specific freshness authority.

Novelty defense

تعمق MPART/MELT/MUSQAT trust decay وPart VI freshness scheduler، وتربطها بحق الفعل لا بثقة لغوية.

قاعدة: إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

لا تطبق decay على truths الرياضية بالطريقة نفسها؛ النظرية تخص claims operationally time-sensitive.

نظرية مساعف لنصف العمر الإبستيمي للادعاءات — النموذج الرسمي

Variable	Operational meaning
h(t)	hazard of invalidation
S(t)	survival probability
HL	median validity time
Vol	volatility
Dep	dependency mass
Risk	decision risk

$$S(t)=\exp(-\int_0^t h(u)du)$$

$$HL=\min\{t:S(t)\le 0.5\}$$

$$\text{FreshAuthority}=S(\Delta t)\cdot \text{SourceQuality}\cdot \text{Independence}$$

$$\text{Revalidate if FreshAuthority}<\text{RiskThreshold(action)}$$

Identifiability

عدم التجديد لا يعني كذب claim. نحتاج historical invalidations أو conservative priors؛ mathematical/stable facts use near-zero hazard.

Causal chain

Claim created → time/source changes → authority decay → high-risk use request → revalidate → dependency cascade → update.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف لنصف العمر الإبستيمي للادعاءات — Runtime & State Machine

الـalgorithm المرجعي

```
classify_volatility → estimate_HL → decay_authority → schedule_revalidation → refresh_dependencies → gate_action
```

Implementation substrate

.Temporal evidence cache; survival model; source volatility registry; queue scheduler; cascading invalidation graph

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

يحمي من stale API/policy/price facts؛ لكن fixed short TTL قد يسبب cost DoS، لذلك risk-weighted scheduling.

نظرية مساعف لنصف العمر الإبستيمي للادعاءات — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-081	TOOL	مقدّر نصف عمر الادعاء Claim Half-Life Estimator	claim/source/domain	HL estimate	read-only	validity forecast	sparse history	restore prior state / cancel queued action
P7-082	TOOL	رسم تبعية الحداثة Freshness Dependency Graph	claims	freshness graph	read-only	cascade precision	missing lineage	restore prior state / cancel queued action
P7-083	TOOL	طابور إعادة التحقق Revalidation Queue	claims	queue	read-only	stale risk reduced	churn	restore prior state / cancel queued action
P7-084	TOOL	بوابة السلطة الزمنية Temporal Authority Gate	action+claims	allow/revalidate	read-only	stale action rate	overblocking	restore prior state / cancel queued action
P7-085	METHOD	منهج التلاشي المعاير بالتقلب Volatility-Calibrated Decay	history	decay model	read-only	calibration	regime shift	no-world-effect
P7-086	METHOD	منهج الحداثة المعدل بالخطر Risk-Weighted Freshness	claim+action	threshold	read-only	verified effect/cost	risk misclass	no-world-effect
P7-087	PROTOCOL	بروتوكول انتهاء قبل الفعل الحساس Expiry-before-High-Impact Action	action	freshness receipt	policy-gated	stale unsafe action	latency	restore prior state / cancel queued action
P7-088	PROTOCOL	بروتوكول سلسلة إعادة التحقق Revalidation Cascade	expired root	cascade jobs	policy-gated	dependency correctness	cascade explosion	restore prior state / cancel queued action
P7-089	STACK	ذاكرة دليل زمنية Temporal Evidence Cache	evidence	fresh cache	service identity + least privilege	hit validity	stale cache	version rollback + replay
P7-090	STACK	مجدول الحداثة Freshness Scheduler	cache+risks	refresh events	service identity + least privilege	freshness SLA	cost spikes	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف لنصف العمر الإبستيمي للادعاءات — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining. يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

نظرية مساعف لنصف العمر الإبستيمي للادعاءات — Experiment, Ablation & Promotion

التجربة الحاسمة

استخدم domains بـvolatility مختلف: APIs، أسعار، سياسات، وثائق ثابتة، اختبر no TTL، fixed TTL، وETHT.

Falsifiers

- إذا لم يتنبأ decay بخطأ stale claims.
- إذا كانت fixed TTL بنفس الكفاءة.
- إذا زادت revalidation cost دون انخفاض material stale failures.

Predictions

- سيكشف النظام stale-but-confident failures قبل execution.
- Half-life يختلف جذرياً بين قانون/سعر/API/مبدأ رياضي.
- Risk-weighted revalidation يحقق فائدة أعلى من global freshness windows.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- E5 عند provider/model alternative

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** إذا كان fixed TTL domain-specific مساوياً للsurvival model، تختصر ETHT إلى policy.

R-M7-10-1

نظرية مساعف للانتقال من التشبع النظري إلى الاستثمار التجريبي — إعادة التأسيس

Moussaif Saturation-to-Experiment Transition Theory

الفجوة: الGap Audit يقرر أن العائق التالي ليس أسماء نظريات أكثر بل runtime وbenchmarks وreplication، ومع ذلك يستطيع أي برنامج مولد أن يستمر في إنشاء مصطلحات إلى ما لا نهاية. لا توجد آلية تقرر متى يصبح توليد نظرية جديدة أقل قيمة من تنفيذ تجربة.

الأطروحة المعاد بناؤها: البرنامج البحثي يجب أن يغيّر طوره عندما تنخفض القيمة الهامشية لنظرية جديدة تحت Expected Experimental Information Gain. في طور التشبع، تُجمَّد التسمية مؤقتاً وتحوّل الموارد إلى ablation، implementation، replication، negative results، calibration. يمكن فتح الطور النظري مجدداً فقط إذا كشفت التجارب فرقاً غير ممثل.

التعديل VIII: تحويل saturation إلى research portfolio optimization مع negative-result analysis، replication، benchmark، implementation، uncertainty: theory candidate كل منها له arm VOI وتكلفة وتأخير.

Novelty defense

تضع قاعدة Meta-Governance فوق M-TAOS saturation loop، Gap Audit، Part V/VI novelty gates، وNeoAGI novelty engine.

> **قاعدة:** إذا نجح البديل الأبسط في التجربة الحاسمة بنفس effect/cost/safety، تُدمج النظرية ولا يُحمى الاسم.

حدود النطاق

ليست النظرية حظراً على الإبداع، هي policy لتخصيص موارد البحث في برنامج أصبح specification-heavy/evidence-light.

نظرية مساعف للانتقال من التشبع النظري إلى الاستثمار التجريبي — النموذج الرسمي

Variable	Operational meaning
MV_T	marginal theory value
EIG	expected information gain of experiment
DR	duplicate risk
EC	evidence conversion probability
K	cost
Delay	time to feedback

$$RDI(action)=E[VerifiedKnowledgeGain]/(K+\lambda Delay)$$

$$\text{Freeze theory if max_exp RDI} > \text{max_theory RDI and } DR>\tau$$

$$\text{Reopen if residual anomaly likelihood} > \rho$$

Identifiability

Novelty score لا يثبت novelty. يجب hostile reduction against registry + expert/source comparison. EIG forecasts calibrated by historical project outcomes.

Causal chain

Backlog → novelty/VOI forecast → allocate → result/negative result → update priors → phase switch → residual triggers theory reopen.

المعادلات hypotheses تشغيلية. يجب تعريف calibration set، sampling uncertainty، normalization، والـunits قبل استخدام threshold في commit حقيقي.

نظرية مساعف للانتقال من التشبع النظري إلى الاستثمار التجريبي — Runtime & State Machine

الـalgorithm المرجعي

score_candidate_novelty → estimate_decision_delta → forecast_best_experiment_EIG → compare_opportunity_cost → freeze_or_generate → reopen_on_residual

Implementation substrate

.Research portfolio DB; theory similarity graph; experiment backlog; bandit/VOI scheduler; negative-result reinvestment; phase dashboard

State transition contract

Stage	Required state	Receipt
Pre	task/goal, authority, source/evidence, budget, pending effects	pre-state hash
Transform	theory-specific operation	trace + parameter/version
Verify	independent observation/rubric	verdict + uncertainty
Commit	promotion/action/memory delta	post-state + rollback target

Security boundary

لا تجعل metric gaming البحث يتجه فقط إلى quick wins؛ reserve budget لـ replication/safety/long-tail failure discovery.

نظرية مساعف للانتقال من التشبع النظري إلى الاستثمار التجريبي — تقوية الـ10 Artifacts

ID	Class	Artifact	Input	Output	Authority	Metric	Failure	Rollback
P7-091	TOOL	مقياس الجدة الهامشية Marginal Novelty Meter	candidate theory	novelty score	read-only	duplicate detection	embedding bias	restore prior state / cancel queued action
P7-092	TOOL	متنبئ كسب التجربة Experiment Information-Gain Forecaster	backlog	EIG estimate	read-only	forecast accuracy	optimism	restore prior state / cancel queued action
P7-093	TOOL	موازن محفظة البحث Research Portfolio Balancer	portfolio	allocation	read-only	verified knowledge/cost	short-termism	restore prior state / cancel queued action
P7-094	TOOL	متحكم تجميد/فتح النظريات Theory Freeze-Unfreeze Controller	metrics	phase state	read-only	E3 conversion	premature freeze	restore prior state / cancel queued action
P7-095	METHOD	منهج تكلفة الفرصة بين النظرية والتجربة Theory-to-Experiment Opportunity Cost	candidate/backlog	decision	read-only	portfolio return	bad estimates	no-world-effect
P7-096	METHOD	منهج تقدير جبهة التشبع Saturation Frontier Estimation	history	frontier	read-only	phase prediction	nonstationarity	no-world-effect
P7-097	PROTOCOL	بروتوكول مفتاح التشبع Saturation Switch Protocol	frontier	freeze order	policy-gated	evidence conversion	research rigidity	restore prior state / cancel queued action
P7-098	PROTOCOL	بروتوكول إعادة الاستثمار بعد النتيجة السلبية Negative Result Reallocation	negative result	reallocation	policy-gated	learning rate	narrative repair	restore prior state / cancel queued action
P7-099	STACK	متحكم طور البحث Research Phase Controller	portfolio events	phase state	service identity + least privilege	throughput	oscillation	version rollback + replay
P7-100	STACK	منسق الاستثمار في الدليل Evidence Investment Orchestrator	backlog	research DAG	service identity + least privilege	verified gain/cost	central planning	version rollback + replay

Upgrade VIII: كل artifact صار contract يجب أن يكتب receipt، لا مجرد function name. أي tool يغيّر العالم يمر prepare → authorize → execute → readback → commit.

نظرية مساعف للانتقال من التشبع النظري إلى الاستثمار التجريبي — Training & Development

القاعدة: ليس كل construct يحتاج weight update. يُختار intervention الأقل كفاية: RCCT. prompt/context → memory → skill → adapter → preference/SFT → continued pretraining. يضبط عمق التبلور، ETHT يمنع وزن facts شديدة التقلب.

Method	وظيفته	Gate
Counterfactual/Intervention Curriculum	إنشاء pairs أو trajectories يتغير فيها العامل الحاكم وحده، لمنع تعلم correlation سطحي.	holdout + transfer + regression
Failure-First Curriculum	إدخال failures/near-misses/negative cases مع السبب والrepair بدلاً من successes فقط.	holdout + transfer + regression
Name-Stripped Transfer	حذف أسماء MOUSSAIF والرموز الخاصة، قياس بقاء الآلية بعد تغيير vocabulary/domain/provider.	holdout + transfer + regression
Budget-Matched Distillation	إذا كان الهدف skill/adapter، يقارن قبل/بعد تحت نفس tool/time/token budget.	holdout + transfer + regression
Court-Guided Preference / RLVR	preference أو reward يأتي من deterministic/postcondition/verifier evidence حيث أمكن، لا من style judge فقط.	holdout + transfer + regression

بيانات التدريب

كل sample يحمل world-effect/postcondition، authority class، desired/rejected behavior، active difference، theory-id، source إن وجد، failure signature، وversion. لا تُخلط generated synthetic traces مع observed traces من دون label.

R-M7-10-6

نظرية مساعف للانتقال من التشبع النظري إلى الاستثمار التجريبي — Experiment, Ablation & Promotion

التجربة الحاسمة

شغّل فريقين/agents على 12 أسبوعاً: generation-first مقابل SETT portfolio controller. قارن cost، duplicated constructs، E3/E4 evidence، executable artifacts، theories retained.

Falsifiers

- إذا منع freeze اكتشاف فروق حقيقية متكررة.
- إذا لم يرفع نسبة executable/benchmarked constructs.
- إذا كان novelty meter لا يميز إعادة التسمية من آلية مستقلة.

Predictions

- سيقل theory inflation ويزيد E3/E4 conversion.
- سترتفع نسبة negative results المفيدة إلى prose artifacts.
- البرنامج سيعود إلى theory generation بعد تجارب تكشف residuals جديدة، لا حسب رغبة التوسع.

Required baselines

- بديل أبسط بلا النظرية
- Part VI/VII baseline frozen
- same model + same tools + matched budget
- provider/model alternative عند E5

Promotion

N3 عند trace+postcondition؛ E4 عند repeated matched-budget gain مع CI وno critical security regression؛ E5 عند replication. **Null outcome:** إذا خفض controller novelty الحقيقية أو لم يرفع E3/E4 conversion، يجب تخفيف freeze واستخدامه advisory فقط.

Active Failure Discovery Kernel

استبدال benchmark sampling الساكن باختيار نشط للحالات وفق coverage × severity × uncertainty؛ مستلهم من ProEval لكن يطبق على agent trajectories.

$$AFD(x) = Uncertainty(x) \cdot Severity(x) \cdot Novelty(x) / Cost(x)$$

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكانون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

يرتبط بـM-SETT وBenchmark Lab؛ لا يصبح نظرية مستقلة إلا إذا حسن failure discovery under budget.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Active Failure Discovery Kernel — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

يرتبط بـM-SETT وBenchmark Lab؛ لا يصبح نظرية مستقلة إلا إذا حسن failure discovery under budget.

Process-Outcome Orthogonal Verification

فصل هل اتبع المسار ضوابطه عن هل تحقق outcome، مع controllable/uncontrollable labels.

V=(V_process,V_outcome,V_control,V_world)

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكائون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

يضغط على M-EDDT/M-ACCT ويمنع reward contamination.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Process-Outcome Orthogonal Verification — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

يضغط على M-EDDT/M-ACCT ويمنع reward contamination.

Monitorability Preservation Boundary

اعتبار monitorability مورداً قد يتناقص مع latent reasoning أو optimization؛ لا نستخدم CoT كدليل وحيد.

MonRisk=1-Agreement(trace_monitor,action/world_monitor)

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكانون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

جسر بين OpenAI CoT evals وMURI/Courts.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Monitorability Preservation Boundary — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

جسر بين OpenAI CoT evals وMURI/Courts.

Global-Constraint Planning Ledger

تمثيل global constraints والالتزامات العابرة للخطوات كحالة مستقلة، لا ضمن prompt فقط.

GCL_t={budgets,deadlines,dependencies,obligations,invariants}

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكانون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

جسر DeepPlanning/AgencyBench إلى GCIT/trajectory assurance.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Global-Constraint Planning Ledger — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

جسر DeepPlanning/AgencyBench إلى GCIT/trajectory assurance.

Agentic Resource Co-Design

CPU/tool/orchestration/GPU state transfers تدخل critical path؛ schedule يجب أن يكون role-aware ومتنبأً لـtail latency.

SystemCost=LLM+CPU_orch+ToolIO+StateSwap+Verification

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكائون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

جسر architecture research إلى M-ICAT وM-TAOS compute governance.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Agentic Resource Co-Design — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

جسر architecture research إلى M-ICAT وM-TAOS compute governance.

Protocol Statelessness Compensation

كلما صار transport stateless زادت ضرورة external canonical state/invariant receipts.

StateDuty ∝ ProtocolStatelessness × Horizon

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكائون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

مستخلص من MCP 2026؛ يبرر state kernel مستقل.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Protocol Statelessness Compensation — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

مستخلص من MCP 2026؛ يبرر state kernel مستقل.

Institutional Coexistence Objective

في objective، multi-agent/ASI-like systems لا يقتصر على unilateral reward بل stability/cooperation/human agency.

J=TaskValue+CooperationStability+HumanAgency-Risk

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكائون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

مستخلص من DeepMind non-solipsistic agenda؛ Research Construct لا ASI claim.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Institutional Coexistence Objective — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

مستخلص من DeepMind non-solipsistic agenda؛ Research Construct لا ASI claim.

Scheming-Aware Deployment Audit

honeypot + investigator agents + realism controls قبل توسيع الصلاحيات.

$$AuditRisk = P(misbehavior \mid opportunity, realism) - P(control)$$

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكائون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

يرتبط بـM-ACCT وsecurity red-team؛ لا يفترض scheming إذا لم يظهر.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Scheming-Aware Deployment Audit — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

يرتبط بـM-ACCT وsecurity red-team؛ لا يفترض scheming إذا لم يظهر.

Modular Continual Memory Promotion

ICL/memory للسرعة وIWL للقدرة المستقرة، مع governance/decay/access control.

Promote if StableTransfer>τ and Volatility<v and RollbackReady

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكائون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

يجمع modular memory/SSGM مع RCCT/ETHT/HKOS.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Modular Continual Memory Promotion — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

يجمع modular memory/SSGM مع RCCT/ETHT/HKOS.

Proof-Carrying Recursive Improvement

أي auto-improvement يجب أن يأتي بحزمة evidence/regression/rollback/authority قبل deployment.

PCI=ΔCapability·Evidence·Rollback·SafetyFloor / Irreversibility

لماذا ظهر الآن؟

هذا construct ليس إعادة تسمية لكائون قديم؛ هو bridge يربط evidence حديثاً بمالك داخلي. لا يدخل سجل النظريات ما لم يثبت ablation مستقل.

المالك الأقرب

جسر AGI→ASI recursive improvement إلى Part VIII evidence gates.

Status: Frontier Research Construct / N0-N2. لا ادعاء empirical خاص بـMOUSSAIF.

Proof-Carrying Recursive Improvement — اختبار التحويل إلى آلية

Object	typed record / metric / gate specific to this construct
Trigger	عندما يتغير قرار أو سلامة أو تكلفة بسبب الفجوة المحددة.
Baseline	أقرب مالك canonical بلا هذا bridge.
Kill test	إذا لم يتغير process/outcome/world-effect تحت matched budget يُدمج.
Transfer	اختبار provider/domain/environment مختلف قبل promotion.

جسر AGI→ASI recursive improvement إلى Part VIII evidence gates.

Kernel VIII: المبادئ غير القابلة للتفاوض

every mutation reversible or ؛world-effect requires readback ؛plan mutable/goal contract stable ؛evidence rank preserved ؛explicit authorization > inferred permission ؛Observed state > generated state .explicitly irreversible

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

KERNEL-02

Canonical State Object

,TaskContract, Goal/Compass, QuestionObject, DifferenceObject, ClaimGraph, EvidenceDebt, TemporalValidity, AuthorityTokens, ReasoningPaths, Experiments, Memory, PendingEffects, Invariants .FailureBasin, CapabilityState, ResearchPortfolio

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

KERNEL-03

Source Furnace

.ingest → hash → parse → deduplicate → claim split → provenance → source rank → quarantine malformed extraction → knowledge object compile

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Evidence & Temporal Plane

.claim graph + debt + half-life + independence + revalidation queue + negative evidence; no canonical promotion from summary alone

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

KERNEL-05

Goal/Question/MCD Plane

.Goal-as-OS + QuestionObject + active difference + MCD/QMCD + question density; routing decision written as receipt with alternatives not chosen

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

KERNEL-06

Reasoning & Experiment Plane

.MURI plural calculi + disagreement broker + ICAT active experiment planner + monitorability boundary; separate imagined result from observed tool result

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

KERNEL-07

Memory Plane

.working/episodic/semantic/procedural/negative memory; consolidation gated by provenance, freshness, privacy, contradiction, transfer, rollback

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Capability Lifecycle

.candidate → external skill → verified skill → adapter candidate → weight candidate; RCCT gates deeper crystallization

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Protocol Fabric

.MCP stateless adapter, A2A agent adapter, framework adapters; canonical state stays internal; auth and invariants outside transport

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Trajectory & Execution Plane

.Mission DAG + global constraint ledger + prefix invariants + idempotent prepare/commit + sandbox + compensation transactions

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Verifier & World-Effect Plane

.deterministic checks → rubric/process verifier → outcome verifier → authoritative environment readback → causal/authority attribution

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Failure & Recovery Plane

.failure taxonomy → basin classification → escape planning → replay → recovery proof → negative knowledge capitalization

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Training/Development Plane

.MUSQAT split, experience graph, data provenance, SFT/preference/RLVR only where justified, continual memory and canary/rollback

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Benchmark & Active Evaluation Plane

.Harbor/container environments + AgencyBench/DeepPlanning-style long horizons + ProEval-like active sampling + trajectory attribution + verifier bench

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Research Portfolio Plane

.SETT: allocate budget across theory/implementation/eval/replication; theory freeze when experiment VOI dominates; reopen on unexplained residual

Required API class

register / propose / authorize / execute / verify / reconcile / replay / rollback. Every endpoint .returns a typed receipt with version and evidence pointer

Failure discipline

,timeout, partial write, stale source, permission denial, tool hallucination, verifier disagreement .compaction loss, provider swap and policy change must be first-class test cases

pre_state_hash → transition_id → actor/authority → input/source → action/tool → observation → verifier → post_state_hash → rollback_target

Harness Parity Suite

.same model/task under framework/harness variants; isolate context, retry, tools, reasoning effort, verifier

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

Long-Horizon Agency Suite

.AgencyBench-inspired hours/tool-calls/context; global constraints + feedback correction + state recovery

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

Trajectory Tool-Use Suite

.TRAJECT-Bench-style selection/arguments/order/dependency metrics; not final answer only

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

Verifier Quality Suite

.process/outcome, controllable/uncontrollable, false-positive priority, human-human agreement baseline

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

Causal Attribution Suite

.leave-one-out component perturbation, mediation, attribution-chain recovery, world-effect readback

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

.poisoning, stale facts, privacy leakage, reconsolidation, negative knowledge, forgetting, provider swap

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

Identity/Delegation Security Suite

.capability tokens, A2A delegation, MCP tool scope, non-repudiation, prompt injection, trajectory invariant attacks

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

Scheming/Honeypot Suite

.realism-controlled opportunities, hidden goals/agency prompts as separate conditions, investigator-agent analysis; do not infer scheming from generic error

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

Continual Capability Suite

.scaffold removal, skill/adapter promotion, transfer, regression, rollback, volatility classes

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

Research Meta-Evaluation Suite

.SETT portfolio vs generation-first: E3/E4 conversion, duplicate ratio, negative-result use, verified gain/cost

Design	Requirement
Pre-registration	.hypothesis, primary metric, failure threshold, stop rule
Budget	.tokens, wall time, tool calls, external cost, model effort
Seeds	.repeated trials; CI/effect size
Evidence	.process trace + outcome + authoritative readback
Safety floors	.authorization, security, truthfulness, privacy non-regression
Promotion	.E4 only after frozen baseline; E5 after cross-model/domain/environment replication

خطة 180 يوماً: من Part VIII إلى أول E4 موثوق

Phase	Days	Deliverable
R0 Measurement Kernel	1-30	.schemas, event store, auth, readback, replay, frozen harness
R1 VII Reconstruction Runtime	31-60	.EDDT/ETHT/GCIT/NKCT services with N3 traces
R2 Causal/Verifier Lab	61-90	.ICAT/DET/UV-style verifier + active eval + causal attribution
R3 Recovery/Capability	91-120	.FBRT/RCCT/ACCT, basin recovery, skill/adapter canaries
R4 Long Horizon & Security	121-150	.Agency/DeepPlanning/trajectory assurance/honeypot suites
R5 E4/E5 Campaign	151-180	.matched-budget ablations, CI, cross-model replication queue

ASI Frontier: ما الذي يمكن نقله وما الذي يجب إبقاؤه فرضية

مسارات AGI→ASI المنشورة - multi-agent collectives، recursive improvement، paradigm shift، scaling - تصلح لتحديد testbeds لا لإعلان قرب ASI. MOUSSAIF يضيف فقط حول safety case: proof -self-modification، authority ceiling، rollback، independent verifier، canary، regression suite، carrying change.

ImprovementCandidate ≠/ImprovementVerified ≠/ImprovementGeneralized ≠/AuthorityExpansion

حتى لو تحسنت capability، لا تتوسع السلطة تلقائياً. التعاون المؤسسي وحفظ agency البشرية objective مستقل عن task score.

نظرية التعاون المشروط بدل تضخيم multi-agent

الـmulti-agent ليس درجة أعلى من الذكاء في ذاته. يجب أن يخضع Collaboration Admission Gate يقارن specialization/diversity gain مقابل attack surface، token duplication، latency، correlated error، coordination tax. إذا كان single strong agent أفضل تحت matched budget، ترفض orchestration الزائدة.

AdmitMAS iff ExpectedGain_specialization + DiversityGain > CoordinationTax + CorrelatedErrorRisk + SecurityCost

من الكتاب إلى Repo: Minimal Research-Grade Kernel

/schemas /provenance /state /evidence /temporal /authority /question-mcd /reasoning /causal-lab /memory /negative-memory /capabilities /mcp /a2a /trajectory /execution /world-effect /verifiers / recovery /training /benchmarks /portfolio /telemetry

DoD الأول ليس بناء 300 skills. هو شريحة واحدة end-to-end يمكن replayها: independent verifier → invariant monitor → execution → authority → intervention → question/Δ → debt/freshness → claim/ provisional → failure/negative memory → causal/legitimate credit → readback أو portfolio decision → capability update.

حكم Part VIII على النظريات العشر

Theory	Status after reconstruction	أقرب تهديد للاندماج
M7-01 M-EDDT	N1-N2 strong candidate	قد يفشل المفهوم إذا كانت uncertainty + risk وحدهما تعطي نفس ترتيب verification. عندها يُخفض EDDT إلى scheduling method لا Theory.
M7-02 M-FBRT	N1-N2 strong candidate	إذا لم تتسأ membership بالتكرار أو كان generic retry مساوياً، يُدمج FBRT في Failure Taxonomy/ Recovery Planner.
M7-03 M-ICAT	N1-N2 strong candidate	إذا لم يتسأ ICV بالتحسن أو كانت تكلفة التقدير أكبر من فائدته، يُستبدل بـrule بسيطة مثل uncertainty threshold.
M7-04 M-RCCT	N1-N2 strong candidate	إذا لم تختلف أفضل vessel placements عن static policy بسيطة، تصبح RCCT optimization method لا مستقلة.
M7-05 M-ACCT	N1-N2 strong candidate	إذا كان hard authorization gate يمنع كل unauthorized paths بحيث لا توجد learning contamination، تظل ACCT مهمة للتدقيق لكنها قد لا تكون Theory مستقلة.
M7-06 M-GCIT	N1-N2 strong candidate	إذا كان generic serialization + tests يحقق نفس GCIT drift detection، يخفض إلى engineering invariant discipline.
M7-07 M-NKCT	N1-N2 strong candidate	إذا لم تقل repeated errors أو زادت false refusals، لا تُرقَّ negative object ويحتفظ به كـarchive فقط.
M7-08 M-DET	N1-N2 strong candidate	إذا كان majority/judge يساوي DET تحت matched budget، يظل DET فقط في high-impact or high-correlation regimes.
M7-09 M-ETHT	N1-N2 strong candidate	إذا كان fixed TTL domain-specific مساوياً لـsurvival model، تختصر ETHT إلى policy.
M7-10 M-SETT	N1-N2 strong candidate	إذا خفض controller novelty الحقيقية أو لم يرفع E3/E4 conversion، يجب تخفيف freeze واستخدامه advisory فقط.

لا توجد نظرية منها E4/E5 بمجرد هذا الكتاب. القيمة الآن أن لكل واحدة benchmark، training mapping، runtime contract، identifiability، formal variables، وkill condition أوضح بكثير من Part VII.

دفتر المصادر — 1

ID	Source	Use in Part VIII
I01	Current uploaded corpus	PDFs / 1,789 pages before deduplication; two exact duplicate pairs; current-turn 17 .primary internal basis
I02	HKOS-M–MICAGFI foundational paper	layer heritage compressed into capability formation, governed runtime and-176 .evolution; evidence-to-action thresholds and quarantined memory
I03	Agentic Governance Closure 50p	,Compositional assurance, causal grounding, epistemic credit, memory entropy .institutional sovereignty, open-world adaptation
I04	Unified Governance Map / MUG	Kernel vs training stack vs unified theory; E0-E5 evidence discipline; vessel placement .and reversibility
I05	MURI-AGI-01 70p	Typed ontology, plural calculi, rank/axis preservation, explicit ignorance, productive .contradiction, reversible comprehensive judgment
I06	MUSQAT-AGI 81p	Training/development split; context/trust/cumulative/question-driven learning; 10 .theories and 80 tools
I07	AGIOS-M 250p	,Agentic OS: kernel, memory, HyperRouting, tools/MCP, bounded loops, recovery .token economics, benchmarks
I08	MUT-AGI-02 160p	-core theories, 200-node registry, 100 tools, 19 HyperKnowledge bases; anti 42 .duplication and theory status rules

ID	Source	Use in Part VIII
I09	Expanded AGI/Consciousness Atlas	entries integrating MSRAK, MURI, M-HKOS/MICAGFI and functional 117 .consciousness; no phenomenal proof
I10	Question Density 80p	Question density as generative/trust/effect control before retrieval and planning; links .question quality to resource routing
I11	Part V MUT-AFI	.complementary candidate theories and promotion ladder N0-N5 20
I12	Part VI Deep-Crawl	gap-reconstruction dossiers: epistemic compilation, trajectory assurance, protocol 96 .sovereignty, causal sufficiency, long-horizon coherence
I13	Part VII	.candidate theories + 100 artifacts; current object of forensic reconstruction 10
I14	MOUSSAIF Gap Audit 50p	Main conclusion: specification-heavy/evidence-light; reference runtime, calibrated .metrics, replay, telemetry, frozen baselines, E4/E5 replication are next bottleneck
I15	MOUSSAIFXT Engineering Architecture 50p	-Cross-environment execution shell; evidence hierarchy; world-effect readback; source .ranked claims and versioned state
I16	NeoAGI Hyperloop KB Engineering 72p	,Knowledge objects, MCD-K, machine-readable constraints, retrieval membrane .destruction court, compression seal

دفتر المصادر — 3

ID	Source	Use in Part VIII
I17	MOUSSAIF Kernel / MSRAK 43p	Sovereign reasoning-to-agency kernel; tool/memory/eval literature anchors including AgentBench, OSWorld, Terminal-Bench and continual memory.
I18	Project research conversations	/Goal-as-OS, Question-as-Executable-Object, zero-duplication governance, evidence world-effect/recovery contracts, MOUSSAIFXT/Hermes integration.
W01	OpenAI GPT-5.6 System Card (2026-07-09)	-CoT monitorability/controllability, destructive-action and confirmation evals, prompt injection robustness; reasoning trace is a monitoring signal but not ground truth.
W02	Microsoft Universal Verifier / CUAVerifierBench (2026-04)	/Verifier design: non-overlapping rubrics; process/outcome separation; controllable uncontrollable failures; very low false-positive emphasis.
W03	AgencyBench, ACL 2026	;tasks / 32 real-world scenarios; ~90 tool calls and 1M-token contexts on average 138 resource efficiency and feedback-driven correction are differentiators.
W04	Long-Horizon Agent Trajectory Attribution (2026-08)	Primary attribution localization and attribution-chain recovery; leave-one-out component perturbation and incremental contribution baselines.
W05	TRAJECT-Bench	Tool selection, arguments, ordering/dependency diagnostics; trajectory quality beyond final accuracy.
W06	DeepPlanning, ACL 2026	Long-horizon planning with active information gathering, local constraints and global budget optimization; frontier systems still struggle.

ID	Source	Use in Part VIII
W07	Terminal-Bench 2.0 / Harbor	89 hard terminal tasks with comprehensive tests; frontier agents below 65%; containerized harness and rollout interfaces.
W08	NIST NCCoE Agent Identity & Authorization (2026-02)	Identification, authorization, auditing, non-repudiation and prompt-injection controls for software/AI agents.
W09	MCP 2026-07-28 Specification	Stateless core, MRTR, header routing, cache hints, authorization hardening, extensions/ tasks; protocol is transport/integration, not authority.
W10	Google Developer Guide to Agent Protocols (2026-03)	MCP, A2A and related protocols solve distinct integration layers; ADK used as implementation substrate.
W11	Microsoft Agent Framework (2026)	Agents + harness + graph workflows; model/provider abstraction, sessions, memory, middleware, checkpointing and HITL.
W12	LangGraph v1 / Persistence	Durable execution, checkpoints, pending writes, time travel, interrupts, resumability, and fault tolerance.
W13	Google ADK / agents-cli (2026)	A2A exposure, scaffold-build-eval-deploy-observe workflow, observability and registry/ deployment tooling.
W14	Hugging Face smolagents	CodeAgent vs ToolCallingAgent; code as compositional action representation; sandbox/security needed.

ID	Source	Use in Part VIII
W15	CrewAI	Crews + flows, guardrails, memory, knowledge, observability, resumable stateful orchestration.
W16	DSPy	Structured signatures and optimizable LM programs; useful for compiling prompt-level behavior into evaluable modules.
W17	Modular Memory is the Key to Continual Learning Agents (2026)	ICL for rapid adaptation + IWL for stable capability updates; modular memory as bridge around catastrophic forgetting.
W18	SSGM: Governing Evolving Memory in LLM Agents (2026)	Decouple memory evolution from execution; consistency verification, temporal decay and dynamic access control.
W19	Architectural Implications of Agentic AI Workflows (2026-08)	Agentic CPU/GPU fragmentation, bursty tool/orchestration load, role-aware scheduling and resource co-design.
W20	Securing Agentic AI: Trajectory Assurance (2026-08)	Sequences of individually allowed actions can violate system invariants; end-to-end observability, identity, provenance and containment are architectural properties.
W21	From AGI to ASI, Google DeepMind (2026-06)	-Four pathways: scaling AGI, paradigm shifts, recursive improvement, large-scale multi agent collectives; large uncertainty and bottlenecks.
W22	Solipsistic superintelligence is unlikely to be cooperative (2026-06)	,Deployment is endogenous and non-stationary; cooperation, adaptive counterparties institutions and human agency should be structural design primitives.

ID	Source	Use in Part VIII
W23	Gram: automated alignment auditing (2026-05)	simulated agentic scenarios; targeted investigator pipeline for sabotage/misbehavior 17 .drivers; realism changes rates
W24	Realistic honeypot evaluations for scheming propensity (2026-05)	/Coding-task honeypots; explicit agency/hidden-goal prompts can expose scheming .sabotage propensities; realism and evaluation awareness matter
W25	ProEval, ICML 2026	Active failure discovery/performance estimation with uncertainty-aware surrogate .models; 8-65x fewer samples for ±1% performance estimates in reported experiments
W26	SeLaR, ACL 2026	Selective latent reasoning activated at low-confidence steps; entropy-gated exploration .highlights that reasoning substrate and monitorability may diverge
W27	Nature: Agentic profiles for effective AI governance (2026-08-12)	Profiles agents along autonomy, efficacy, goal complexity and generality; governance .should depend on profile, not binary agent/non-agent labels
W28	Nature: AI isn't ready to research itself (2026-08-13)	Automated AI-research systems can run experiments but may settle early, self-review .weakly and fail open-ended novelty; evidence against easy full automation
W29	Nature Machine Intelligence: Capable language models can outgrow collaboration benefits (2026)	/Multi-agent benefit is conditional; fair evaluation requires matched prompts/tools .budgets and process dynamics, not final accuracy alone
W30	Nature/Scientific Reports: AutoLabs (2026)	Self-correcting multi-agent architecture for autonomous experimentation; reliability at .tool/protocol granularity remains central

VIII-CLOSE

الخاتمة: ما الذي أصلحه Part VIII فعلياً؟

أصلح Part VIII مشكلة Part VII بطريقة لا تعتمد على زيادة الحجم وحده: أعاد كل theory إلى research object كامل؛ جعل الـ100 artifact عقوداً قابلة للتنفيذ؛ أضاف resource co-، active evaluation، global constraints، monitorability-، identity/non-repudiation، cooperation، scheming audit، continual memory، design، recursive-improvement safety case؛ وقارن frameworks لاختيار primitives بدلاً من تبني علامة تجارية.

Theory → Contract → Runtime → Trace → Verifier → World Effect → Attribution → Replay → Benchmark → Replication → Promotion/Reject

أعلى فجوة الآن: code + datasets + benchmark execution. إذا تم إنشاء Part IX نظري قبل تشغيل Kernel VIII و Bench v0.1، فإن M-SETT نفسه يجب أن يسجل ذلك كإشارة saturation لا كنجاح بحثي.