MOUSSAIF UNIFIED AGI · PART XIV MASTER · 2026

MOUSSAIF Unified AGI - Part XIV

Executable Theoretical Completion & Closure

الاستكمال النظري التنفيذي: من C0 Corpus Closure إلى C9 Canonical Closure.

18 نظرية 6 · M14 مسارات · 180 54 · 72 training recipes · asset specification تجربة حاكمة.

يتكون المتن المركزي من خمس أوراق بحثية مطولة مترابطة.

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

أطروحة Part XIV

بعد Parts السابقة التي وحّدت النظرية، المصادر، المعادلات، الذاكرة، العالم، الحاكمية والتطور، تتحول المشكلة إلى سؤال إغلاق: **ما الذي يجعل النظرية مكوّناً سببياً قابلاً للإبطال بدل أن تبقى وصفاً معمارياً؟** يقدم Part XIV جواباً بنيوياً: لا اسم جديد بلا فرق؛ لا فرق بلا operator؛ لا operator بلا feature flag؛ لا feature flag بلا trace؛ لا trace بلا world-state/evidence؛ لا evidence بلا attribution؛ لا attribution بلا baseline؛ لا generality بلا transfer؛ ولا self-evolution بلا evaluator/constitutional separation.

ΔTheory → ΔMechanism → ΔTrajectory → ΔWorld → ΔEvidence → ΔCanonical State

الجديد النظري هنا ليس ادعاء أن هذه السلسلة تحل AGI، بل تحويل مساحة واسعة من MOUSSAIF إلى برنامج يمكنه أن يفشل بطريقة مفيدة. إذا لم ينجُ construct من nearest baseline، يندمج أو يُخفض. إذا نجح محلياً فقط، يبقى envelope ضيقاً. إذا نُقل وأعيد، يزداد evidence grade.

خريطة المسارات الخمسة

Paper	المسار	النظريات
P14-A	الإغلاق الرصدي والسببي للعالم Observational & Causal World Closure	M14-01, M14-02, M14-03
P14-B	إغلاق المسار والالتزام والسلطة Trajectory, Commitment & Authority Closure	M14-04, M14-05, M14-06
P14-C	إغلاق الذاكرة والانتباه عبر الأفق Memory, Attention & Horizon Closure	M14-07, M14-08, M14-09
P14-D	إغلاق التنظيم متعدد الوكلاء والبروتوكولات Collective & Protocol Intelligence Closure	M14-10, M14-11, M14-12
P14-EF	إغلاق العمومية والتقييم والتطور العلمي الذاتي Generality, Evaluation & Scientific Self-Evolution Closure	M14-13, M14-14, M14-15, M14-16, M14-17, M14-18

المسارات ليست silos: P14-A يربط observation/world causality؛ P14-B يربط trajectory/commitment/authority؛ P14-C يحفظ state عبر الزمن؛ P14-D يضبط التنظيم والبروتوكولات؛ P14-E/F يقرر إن كان الأثر عاماً وقابلاً للتكاثر والتطور دون self-confirmation.

خريطة 18 نظرية M14

ID	النظرية	المسار	Formal core
M14-01	نظرية مساعف لجبهة الكفاية الرصدية Moussaif Observational Sufficiency Frontier Theory	P14-A	OSF = ParetoFrontier(V_d, C_cov, I_src, F | C_obs); stop iff ΔE[V_d] < C_obs and H_unknown=0
M14-02	نظرية مساعف لنموذج العالم الموجّه لحدود القرار Moussaif Decision-Boundary World Model Theory	P14-A	DBWM-L = λ1·Err_decision + λ2·Err_cf + λ3·Calib + λ4·Miss_critical - λ5·Fit_irrelevant
M14-03	نظرية مساعف لرسم السببية من البقايا Moussaif Residual Causality Graph Theory	P14-A	RCG(i→j)=R_rec·D_cond·I_cons·D_impact - P_conf - P_instability
M14-04	نظرية مساعف للمسار الحامل للبرهان Moussaif Proof-Carrying Trajectory Theory	P14-B	PCT = Coverage_critical · ChainConsistency · ReadbackValidity · AuthorityIntegrity · (1-OpaqueCriticalRate)
M14-05	نظرية مساعف لهندسة أفق الالتزام Moussaif Commitment-Horizon Geometry Theory	P14-B	CG_t = Δlog|Reachable(S_t)| + α·IrrevCost + β·DependencyLock + γ·AuthorityLock
M14-06	نظرية مساعف لحفظ السلطة عبر التفويض Moussaif Authority-Preserving Delegation Theory	P14-B	A_eff(n,t)=Grant_root ∩ Scope_task ∩ Cap_tool ∩ TimeWindow ∩ Purpose; Leak=ExecutedScope\A_eff
M14-07	نظرية مساعف لديناميات الانتباه المستمر Moussaif Persistent Attention Dynamics Theory	P14-C	A_i(t+1)=ρ_iA_i(t)+η·Surprise+ζ·Obligation+κ·Risk+χ·Dependency-ω·VerifiedClosure
M14-08	نظرية مساعف لحد الضغط السببي للذاكرة Moussaif Causal Memory Compression Bound Theory	P14-C	CCB = I(M_comp; D*, C*) / (Storage+RecallCost); admissible iff Regret_reconstruct≤ε and CausalRecall≥τ
M14-09	نظرية مساعف للاتساق النهائي للذاكرة المعاد بناؤها Moussaif Reconstructive Memory Eventual Consistency Theory	P14-C	ECons = 1 - d(S_rebuilt,S_auth); require ECons≥τ_risk before critical action; divergence budget B_d(t)
M14-10	نظرية مساعف لحمل المحاذاة التنظيمي Moussaif Organizational Alignment Load Theory	P14-D	OAL = RoleDivergence + IncentiveConflict + InfoAsymmetry + ResponsibilityDiffusion + ProtocolAmbiguity - GovernanceResolution
M14-11	نظرية مساعف لسيادة الدلالة فوق البروتوكول Moussaif Semantic Sovereignty over Protocol Theory	P14-D	SSP = Π_k Preserve(I_k | encode→transport→decode); SovereignHandoff iff min_k Preserve(I_k)≥τ_k
M14-12	نظرية مساعف لمعايرة الثقة في التفويض Moussaif Delegation Trust Calibration Theory	P14-D	DT(c)=Calib[P(success,compliance,recovery|c)]·IndependenceEvidence - Uncertainty - DriftPenalty
M14-13	نظرية مساعف لجبهة عمومية الإغلاق Moussaif Generality Closure Frontier Theory	P14-EF	GCF = ParetoFrontier(Quality, Calibration, Recovery, Cost | model,domain,tool,horizon,resource); GenScore=min_T normalized_effect(T)
M14-14	نظرية مساعف للتحويل بين المعيار والنشر Moussaif Benchmark-to-Deployment Morphism Theory	P14-EF	MorphLoss = Σ_k w_k·d(I_k^bench,I_k^deploy); ClaimTransfer ≤ exp(-MorphLoss)·EvidenceGrade
M14-15	نظرية مساعف لتفكيك القدرة المقاوم للهارنس Moussaif Harness-Resistant	P14-EF	Y = β0 + βM·Model + βH·Harness + βK·Mechanism + βMK + βHK + βMHK + ε; Credit_K = robust causal contrast

		Capability Decomposition Theory	
EID = w1·ModelOverlap+w2·DataOverlap+w3·AdaptiveCoupling+w4·Leakage+w5·MetricGaming - w6·ExternalGrounding	-P14 EF	نظرية مساعف لدَين استقلال المقيم Moussaif Evaluator Independence Debt Theory	M14-16
SRY = Σ_j Indep_j·Diversity_j·Reproduce_j / Σ_j Attempt_j; report with CI and failure taxonomy	-P14 EF	نظرية مساعف لعائد إعادة الإنتاج العلمي Moussaif Scientific Reproduction Yield Theory	M14-17
CMS = Gain_validated · InvariantPreservation · RollbackVerified · EvaluatorIndependence - CriticalRegressionRisk	-P14 EF	نظرية مساعف لثبات الطفرة الدستورية Moussaif Constitutional Mutation Stability Theory	M14-18

الأرقام لا تعني maturity: سجل الأصول

training recipes 72

isolation، adversarial boundaries، matched-budget causal evaluation، cross-model/domain replication.

assets 180

لكل نظرية: 4 runtime/replay + 2 protocols + 2 methods + tools assets.

papers 5

تحويل المسارات إلى أطروحات طويلة قابلة للعرض على مختبر أكاديمي.

core experiments 54

/minimal causal ablation، boundary/adversarial، transfer replication.

الحالة: theory/formal/experimental-design. هذه الأعداد coverage registry وليست benchmark result.

MCD/QMCD → Closure

MCD يبقى routing heuristic حتى المعايرة. Part XIV يضيف شرط أن routing gain يُنسب causalياً تحت matched budgets ويُختبر عبر models/horizons.

المعيار الحاكم: construct القديم لا يُستبدل لمجرد ظهور اسم M14؛ يحتفظ lineage، وتُضاف إليه only-the-delta. إذا كان M14 equivalent سلوكياً يُدمج بدلاً من إنشاء طبقة جديدة.

> Old construct + measurable new delta + falsifier + experiment = candidate extension; otherwise alias

HKOS/M-HKOS → Causal Memory

المعرفة كـOS تُغلق عبر proof/replay adapters، eventual consistency، versioned state، causal compression، لا عبر طول KB.

المعيار الحاكم: construct القديم لا يُستبدل لمجرد ظهور اسم M14؛ يحتفظ lineage، وتُضاف إليه only-the-delta. إذا كان M14 equivalent سلوكياً يُدمج بدلاً من إنشاء طبقة جديدة.

Old construct + measurable new delta + falsifier + experiment = candidate extension; otherwise alias

Governance Space/MURI → Authority

الحاكمية تنتقل من مفهوم أعلى إلى authority graphs وsemantic invariants وdelegation receipts وorganizational alignment load.

المعيار الحاكم: construct القديم لا يُستبدل لمجرد ظهور اسم M14؛ يحتفظ lineage، وتُضاف إليه only-the-delta. إذا كان M14 equivalent سلوكياً يُدمج بدلاً من إنشاء طبقة جديدة.

Old construct + measurable new delta + falsifier + experiment = candidate extension; otherwise alias

World Effect → Observational Sufficiency

world-effect loop يُقوّى بسؤال متى تكفي observation وبworld model موجه لحد القرار وresidual causal graph.

المعيار الحاكم: construct القديم لا يُستبدل لمجرد ظهور اسم M14؛ يحتفظ lineage، وتُضاف إليه only-the-delta. إذا كان M14 equivalent سلوكياً يُدمج بدلاً من إنشاء طبقة جديدة.

Old construct + measurable new delta + falsifier + experiment = candidate extension; otherwise alias

Parts XII/XIII → Reproduction

تكامل الأوراق والنظريات الجديدة يُغلق بالـconstitutional mutation stability، reproduction yield، evaluator debt، harness decomposition.

المعيار الحاكم: construct القديم لا يُستبدل لمجرد ظهور اسم M14؛ يحتفظ lineage، وتُضاف إليه only-the-delta. إذا كان M14 equivalent سلوكياً يُدمج بدلاً من إنشاء طبقة جديدة.

Old construct + measurable new delta + falsifier + experiment = candidate extension; otherwise alias

مصفوفة المجتمع العلمي: ما الذي يضغط على Part XIV؟

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

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

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

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

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

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

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

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

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

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

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

مصفوفة المجتمع العلمي - الاستكمال

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

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

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

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

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

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

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

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

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

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

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

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

الورقة الأولى: الإغلاق الرصدي والسببي للعالم

& Observational Sufficiency, Decision-Boundary World Models Residual Causality

P14-A - الإغلاق الرصدي والسببي للعالم

MOUSSAIF Unified AGI Part XIV - Executable Theoretical Completion ضمن

Moussaif Zouhair / زهير مساعف · MUT × MUG × MOUSSAIXT

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

الملخص والأطروحة

تعالج هذه الورقة فجوة محددة في برنامج MOUSSAIF: إغلاق الفجوة بين المعرفة الداخلية والعالم عبر world models، sufficiency of observation موجهة للقرار، ورسوم سببية للبقايا التي تربط prediction بالفعل والreadback. لا تقدم مساهمة الورقة على أنها «إثبات AGI»، بل كحزمة فرضيات ذات state variables وآليات قابلة للتعطيل وأدوات وتجارب ومفندات. معيار القيمة هو ما إذا أضافت النظرية فرقاً سببياً فوق baseline بسيط وتحمّل ذلك الفرق النقل بين شروط مختلفة.

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

> **المساهمة:** نظرية مساعف لجبهة الكفاية الرصدية؛ نظرية مساعف لنموذج العالم الموجّه لحدود القرار؛ نظرية مساعف لرسم السببية من البقايا. كل مساهمة مرتبطة بمعادلة falsifier، feature flag، design-level، تجربة transfer envelope، matched-budget، وأداة runtime.

سؤال الورقة: ما الحد الأدنى من البنية الإضافية التي تجعل الإغلاق الرصدي والسببي للعالم قابلاً للقياس والتفنيد والربط بأثر العالم، دون تحويل density النظرية إلى claim غير مدعوم؟

العقد الإبستيمي ومنع تضخم الادعاء

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

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

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

مصفوفة الضغط من حالة الفن 2026

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

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

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

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

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

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

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

خريطة الاندماج الداخلي: corpus → canonical lineage

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

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

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

M14-01 - الفجوة والأطروحة والنسب

Moussaif Observational Sufficiency Frontier Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M13-02 Unknown-Mass Conservation**، **M13-03 Governing-Question Search**، **M13-05 Residual World Learning**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

DeepPlanning 2026، AutoResearchBench 2026، PRInTS 2026. هذه الأعمال تُعامل كـevidence problem وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-01 - الصياغة الرسمية والobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
V_d	صلاحية القرار تحت إعادة الفحص
C_cov	تغطية المتغيرات السببية الحاكمة
I_src	استقلال أدلة الرصد
F	الحداثة/الصلاحية الزمنية
C_obs	كلفة الملاحظة
H_unknown	مجهول صلب يمكنه قلب القرار

افتراضات الصياغة

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

قاعدة formal closure

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

M14-01 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-01 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. استخرج decision-critical variables من السؤال الحاكم.
2. حوّل كل observation إلى تغطية typed للمتغيرات مع source/provenance.
3. قدّر dependence بين المصادر بدل عدّها كأدلة مستقلة.
4. قدّر marginal decision value للملاحظة لاحقة.
5. امنع stop عند وجود HardUnknown عالي الأثر.
6. أصدر SufficiencyReceipt أو HOLD مع missing-evidence set.

عقد النواة

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

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

M14-01 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- تنخفض حالتا over-browsing وpremature closure معاً.
- المهام الأعلى irreversible impact تتطلب جبهة أوسع أو مصادر أكثر استقلالاً.
- عدد المصادر الخام أقل قدرة على التنبؤ بالصواب من CausalCoverage×Independence.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-01 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A001	Observational Sufficiency Frontier State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A002	Observational Sufficiency Frontier Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A003	Observational Sufficiency Frontier Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A004	Observational Sufficiency Frontier Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A005	Observational Sufficiency Frontier Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A006	Observational Sufficiency Frontier Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-02 - الفجوة والأطروحة والنسب

Moussaif Decision-Boundary World Model Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M13-03 Governing-Question Search**؛ **M13-05 Residual World Learning**؛ **M13-08 Closed-Loop Generality**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

Self-Evolving World Models 2026؛ Agent-Authored World Modeling 2026؛ DeepPlanning 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-02 - الصياغة الرسمية والـobservables

$$DBWM\text{-}L = \lambda1 \cdot Err_decision + \lambda2 \cdot Err_cf + \lambda3 \cdot Calib + \lambda4 \cdot Miss_critical - \lambda5 \cdot Fit_irrelevant$$

الرمز/المتغير	المعنى التشغيلي المقترح
Err_decision	خطأ المتغيرات التي تغير ترتيب السياسات
Err_cf	خطأ الانتقال المضاد للواقع
Calib	خطأ المعايرة
Miss_critical	ديناميكا حرجة حُذفت بالانتقائية
Fit_irrelevant	قدرة صرفت على تفاصيل غير حاكمة

افتراضات الصياغة

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

قاعدة formal closure

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

M14-02 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-02 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. استخرج PolicySet قبل بناء النموذج.
2. حدد variables التي يمكنها قلب ترتيب السياسات.
3. اجمع transitions موجهة لهذه variables.
4. حافظ على unknown channels بدل ملئها تخمينياً.
5. نفذ counterfactual rollouts فقط حيث تتغير قيمة القرار.
6. اقرن كل تحديث residual بسبب/سياق/أثر.

عقد النواة

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

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

M14-02 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- نموذج أقل دقة نصياً قد يتفوق في planning إذا كان أكثر دقة عند boundary.
- الانتقائية تخفض compute دون فقد قرار عندما تكون boundary variables صحيحة.
- failure signatures تصبح أضيق لأن كل متغير له policy consequence.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-02 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A011	Decision Boundary World Model State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A012	Decision Boundary World Model Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A013	Decision Boundary World Model Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A014	Decision Boundary World Model Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A015	Decision Boundary World Model Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A016	Decision Boundary World Model Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-03 - الفجوة والأطروحة والنسب

Moussaif Residual Causality Graph Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M13-05 Residual World Learning**؛ **M13-17 Failure Topology**؛ **M13-19 Causal Repair**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

Trajectory Attribution 2026؛ Fission-GRPO 2026؛ Tool-Reflection-Bench 2026. هذه الأعمال تُعامل كـevidence problem وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-03 - الصياغة الرسمية والobservables

$$RCG(i \to j) = R_rec \cdot D_cond \cdot I_cons \cdot D_impact - P_conf - P_instability$$

الرمز/المتغير	المعنى التشغيلي المقترح
R_rec	تكرار النمط
D_cond	اعتماد مشروط بعد ضبط السياق
I_cons	ثبات تحت interventions
D_impact	أثر على قرار/فعل
P_conf	عقوبة confounding
P_instability	عدم الاستقرار عبر البيئات

افتراضات الصياغة

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

قاعدة formal closure

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

M14-03 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-03 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. cluster residuals حسب النوع والمرحلة والسياق.
2. اربطها بـactions/tools/memory states.
3. ولّد candidate causes بأقل تعقيد تفسيري.
4. اختر interventions صغيرة عالية information gain.
5. حدّث edge posterior بعد التجربة.
6. اجمع repairs المتشابهة في RepairFamily.
7. اكتب causal residue فقط بعد شروط الدليل.

عقد النواة

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

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

M14-03 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- يخفض patch proliferation.
- يتحسن اختيار التجربة التالية مقارنة بإصلاح كل خطأ منفرد.
- يمكن التنبؤ بفشل جديد من neighborhood قبل وقوعه.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-03 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
XIV-A021	Residual Causality Graph State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A022	Residual Causality Graph Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A023	Residual Causality Graph Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A024	Residual Causality Graph Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A025	Residual Causality Graph Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
XIV-A026	Residual Causality Graph Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

الاقتران 1: كيف تتحول النظريات إلى نظام لا قائمة

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

- M14-01: نظرية مساعف لجبهة الكفاية الرصدية - output حاكم: صلاحية القرار تحت إعادة الفحص
- M14-02: نظرية مساعف لنموذج العالم الموجّه لحدود القرار - output حاكم: خطأ المتغيرات التي تغير ترتيب السياسات
- M14-03: نظرية مساعف لرسم السببية من البقايا - output حاكم: تكرار النمط

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

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

الاقتران 2: Graph نظري إلى Graph تنفيذي

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

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

Invariants المشتركة

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

مختبر الضغط 1: العدسة السببية: من الارتباط إلى intervention

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

سجل الأدوات والأصول التنفيذية - 1/2

ID	Owner	Asset	Class	Purpose
-XIV A001	M14-01	Observational Sufficiency Frontier State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A002	M14-01	Observational Sufficiency Frontier Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A003	M14-01	Observational Sufficiency Frontier Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A004	M14-01	Observational Sufficiency Frontier Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A005	M14-01	Observational Sufficiency Frontier Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A006	M14-01	Observational Sufficiency Frontier Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A007	M14-01	Observational Sufficiency Frontier Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A008	M14-01	Observational Sufficiency Frontier Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A009	M14-01	Observational Sufficiency Frontier Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A010	M14-01	Observational Sufficiency Frontier Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A011	M14-02	Decision Boundary World Model State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A012	M14-02	Decision Boundary World Model Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A013	M14-02	Decision Boundary World Model Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A014	M14-02	Decision Boundary World Model Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A015	M14-02	Decision Boundary World Model Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.

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

سجل الأدوات والأصول التنفيذية - 2/2

ID	Owner	Asset	Class	Purpose
-XIV A016	M14-02	Decision Boundary World Model Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A017	M14-02	Decision Boundary World Model Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A018	M14-02	Decision Boundary World Model Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A019	M14-02	Decision Boundary World Model Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A020	M14-02	Decision Boundary World Model Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A021	M14-03	Residual Causality Graph State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A022	M14-03	Residual Causality Graph Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A023	M14-03	Residual Causality Graph Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A024	M14-03	Residual Causality Graph Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A025	M14-03	Residual Causality Graph Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A026	M14-03	Residual Causality Graph Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A027	M14-03	Residual Causality Graph Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A028	M14-03	Residual Causality Graph Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A029	M14-03	Residual Causality Graph Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A030	M14-03	Residual Causality Graph Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.

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

برنامج التدريب: من التعريف إلى قدرة قابلة للنقل

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

- M14-01 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-01 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-01 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-01 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-02 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-02 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-02 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-02 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-03 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-03 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-03 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-03 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

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

مصفوفة التجارب - 1/1

ID	Owner	Kind	Design	Metrics
-XIV E01	M14-01	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E02	M14-01	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E03	M14-01	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E04	M14-02	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E05	M14-02	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E06	M14-02	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E07	M14-03	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E08	M14-03	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E09	M14-03	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier

ترجمة الورقة إلى MOUSSAIXT Reference Kernel

```
service registry: M14-01,M14-02,M14-03
feature_flags: independent
state: versioned typed objects
authority: L0/L1 gates
telemetry: trigger,state_before,decision,tool,observation,state_after,residual
verifier: independent where possible
rollback: mandatory for reversible mutations
writeback: evidence-bearing delta only
```

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

Acceptance at kernel level

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

القيود والمشكلات المفتوحة

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

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

المراجع الخارجية الأساسية

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

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

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

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

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

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

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

المراسي الداخلية والسلالة

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

الورقة الثانية: المسار الحامل للبرهان وهندسة الالتزام والسلطة

& Proof-Carrying Trajectories, Commitment Geometry Authority-Preserving Delegation

P14-B - إغلاق المسار والالتزام والسلطة

ضمن MOUSSAIF Unified AGI Part XIV - Executable Theoretical Completion

Moussaif Zouhair / زهير مساعف · MUT × MUG × MOUSSAIXT

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

الملخص والأطروحة

تعالج هذه الورقة فجوة محددة في برنامج MOUSSAIF: نقل proof-carrying من action مفرد إلى trajectory كاملة، تمثيل هندسة الالتزام، وحفظ السلطة أثناء delegation عبر tools/agents/protocols. لا تقدم مساهمة الورقة على أنها «إثبات AGI»، بل كحزمة فرضيات ذات state variables وآليات قابلة للتعطيل وأدوات وتجارب ومفندات. معيار القيمة هو ما إذا أضافت النظرية فرقاً سببياً فوق baseline بسيط وتحمّل ذلك الفرق النقل بين شروط مختلفة.

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

> **المساهمة:** نظرية مساعف للمسار الحامل للبرهان؛ نظرية مساعف لهندسة أفق الالتزام؛ نظرية مساعف لحفظ السلطة عبر التفويض. كل مساهمة مرتبطة بمعادلة design-level، feature flag، falsifier، تجربة transfer envelope، matched-budget، وأداة runtime.

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

العقد الإبستيمي ومنع تضخم الادعاء

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

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

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

مصفوفة الضغط من حالة الفن 2026

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

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

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

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

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

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

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

خريطة الاندماج الداخلي: corpus → canonical lineage

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

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

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

M14-04 - الفجوة والأطروحة والنسب

Moussaif Proof-Carrying Trajectory Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M13-07 Proof-Carrying Agency**، **M13-09 World-Effect Receipts**، **M13-19 Causal Attribution**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

Long-Horizon Trajectory Attribution 2026، Plan-RewardBench 2026، ToolPRMBench 2026. هذه الأعمال تُعامل كـevidence وproblem وmechanism وpriors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-04 - الصياغة الرسمية والobservables

PCT = Coverage_critical · ChainConsistency · ReadbackValidity · AuthorityIntegrity · (1-OpaqueCriticalRate)

الرمز/المتغير	المعنى التشغيلي المقترح
Coverage_critical	نسبة الخطوات الحرجة ذات شاهد قابل للفحص
ChainConsistency	اتساق الاعتماد بين الخطوات
ReadbackValidity	صحة قراءة الحالة بعد الأفعال
AuthorityIntegrity	عدم تجاوز السلطة
OpaqueCriticalRate	خطوات حاسمة بلا تفسير/trace كاف

افتراضات الصياغة

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

قاعدة formal closure

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

M14-04 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-04 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. حوّل Mission DAG إلى proof obligations.
2. اربط كل action بـprecondition وauthority receipt.
3. اربط كل observation بـsource snapshot.
4. اجمع dependency chain لا final score فقط.
5. أغلق postconditions عبر readback مستقل.
6. حافظ على rollback checkpoints.
7. أصدر TrajectoryProofPacket قابل لإعادة التشغيل.

عقد النواة

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

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

M14-04 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- يخفض false-success في المهام ذات الأدوات.
- تتحسن attribution للخطأ البعيد.
- يمكن للمدقق رفض trajectory ناجحة outcome-wise إذا كسر invariant حرجاً.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-04 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
XIV-A031	Proof Carrying Trajectory State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A032	Proof Carrying Trajectory Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A033	Proof Carrying Trajectory Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A034	Proof Carrying Trajectory Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A035	Proof Carrying Trajectory Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
XIV-A036	Proof Carrying Trajectory Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-05 - الفجوة والأطروحة والنسب

Moussaif Commitment-Horizon Geometry Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M13-07 Proof-Carrying Agency**؛ **M13-12 Reversible Autonomy**؛ **Goal-as-OS**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

DeepPlanning 2026؛ CostBench 2026؛ AgentGym2 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-05 - الصياغة الرسمية والـobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
Reachable	مجموعة الحالات الممكنة تحت الميزانية والسلطة
IrrevCost	كلفة الاسترداد/التعويض
DependencyLock	اعتماديات تلتصق بالفعل
AuthorityLock	تفويضات يصبح الرجوع عنها صعباً

افتراضات الصياغة

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

قاعدة formal closure

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

M14-05 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-05 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. قدر reachable alternatives تقريبياً لا exhaustively.
2. صنف actions إلى probe/prepare/commit/irreversible.
3. اختر probe إذا كانت information value مرتفعة وcommitment منخفضاً.
4. أنشئ checkpoint قبل commit.
5. ارفع approval عند crossing threshold.
6. حدّث commitment debt بعد كل side effect.

عقد النواة

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

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

M14-05 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- تقل irreversible errors مع تكلفة متوسطة محدودة.
- يزداد جمع المعلومات قبل القفز في المهام عالية الغموض.
- الوكيل يختار خططاً تختلف عندما تتغير reversibility حتى مع نفس utility الاسمية.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-05 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A041	Commitment Horizon Geometry State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A042	Commitment Horizon Geometry Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A043	Commitment Horizon Geometry Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A044	Commitment Horizon Geometry Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A045	Commitment Horizon Geometry Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A046	Commitment Horizon Geometry Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-06 - الفجوة والأطروحة والنسب

Moussaif Authority-Preserving Delegation Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **MUG governance**؛ **MOUSSAIXT authority gates**؛ **M13-12 Reversible Autonomy**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

NIST Agent Identity/Authorization 2026؛ MCP authorization 2026؛ Anthropic containment 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-06 - الصياغة الرسمية والobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
Grant_root	تفويض الأصل
Scope_task	حد المهمة
Cap_tool	قدرة الأداة
TimeWindow	صلاحية زمنية
Purpose	قيد الغرض
Leak	أثر خارج التفويض

افتراضات الصياغة

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

قاعدة formal closure

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

M14-06 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-06 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. سجل authority graph قبل التخطيط.
2. احسب effective scope عند كل delegation.
3. لا تورث credentials بلا explicit edge.
4. تحقق من identity/target عند action time.
5. راجع authority بعد context change.
6. أوقف chain عند ambiguity.
7. اكتب revocation وresidual state.

عقد النواة

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

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

M14-06 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- تخفض privilege escalation الصامتة.
- الرفض يحدث قبل mutation لا بعده.
- تظهر اختلافات بين صحة task وصحة authority وتصبح قابلة للقياس.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-06 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A051	Authority Preserving Delegation State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A052	Authority Preserving Delegation Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A053	Authority Preserving Delegation Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A054	Authority Preserving Delegation Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A055	Authority Preserving Delegation Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A056	Authority Preserving Delegation Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

الاقتران 1: كيف تتحول النظريات إلى نظام لا قائمة

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

- M14-04: نظرية مساعف للمسار الحامل للبرهان - output حاكم: نسبة الخطوات الحرجة ذات شاهد قابل للفحص
- M14-05: نظرية مساعف لهندسة أفق الالتزام - output حاكم: مجموعة الحالات الممكنة تحت الميزانية والسلطة
- M14-06: نظرية مساعف لحفظ السلطة عبر التفويض - output حاكم: تفويض الأصل

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

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

الاقتران 2: Graph نظري إلى Graph تنفيذي

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

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

Invariants المشتركة

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

مختبر الضغط 1: العدسة السببية: من الارتباط إلى intervention

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

سجل الأدوات والأصول التنفيذية - 1/2

ID	Owner	Asset	Class	Purpose
XIV-A031	M14-04	Proof Carrying Trajectory State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A032	M14-04	Proof Carrying Trajectory Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A033	M14-04	Proof Carrying Trajectory Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A034	M14-04	Proof Carrying Trajectory Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A035	M14-04	Proof Carrying Trajectory Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
XIV-A036	M14-04	Proof Carrying Trajectory Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
XIV-A037	M14-04	Proof Carrying Trajectory Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
XIV-A038	M14-04	Proof Carrying Trajectory Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
XIV-A039	M14-04	Proof Carrying Trajectory Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
XIV-A040	M14-04	Proof Carrying Trajectory Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
XIV-A041	M14-05	Commitment Horizon Geometry State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A042	M14-05	Commitment Horizon Geometry Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A043	M14-05	Commitment Horizon Geometry Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A044	M14-05	Commitment Horizon Geometry Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A045	M14-05	Commitment Horizon Geometry Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.

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

سجل الأدوات والأصول التنفيذية - 2/2

ID	Owner	Asset	Class	Purpose
-XIV A046	M14-05	Commitment Horizon Geometry Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A047	M14-05	Commitment Horizon Geometry Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A048	M14-05	Commitment Horizon Geometry Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A049	M14-05	Commitment Horizon Geometry Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A050	M14-05	Commitment Horizon Geometry Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A051	M14-06	Authority Preserving Delegation State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A052	M14-06	Authority Preserving Delegation Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A053	M14-06	Authority Preserving Delegation Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A054	M14-06	Authority Preserving Delegation Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A055	M14-06	Authority Preserving Delegation Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A056	M14-06	Authority Preserving Delegation Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A057	M14-06	Authority Preserving Delegation Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A058	M14-06	Authority Preserving Delegation Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A059	M14-06	Authority Preserving Delegation Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A060	M14-06	Authority Preserving Delegation Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.

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

برنامج التدريب: من التعريف إلى قدرة قابلة للنقل

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

- M14-04 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-04 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-04 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-04 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-05 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-05 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-05 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-05 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-06 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-06 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-06 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-06 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

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

مصفوفة التجارب - 1/1

ID	Owner	Kind	Design	Metrics
-XIV E10	M14-04	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E11	M14-04	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E12	M14-04	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E13	M14-05	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E14	M14-05	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E15	M14-05	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E16	M14-06	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E17	M14-06	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E18	M14-06	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier

ترجمة الورقة إلى MOUSSAIXT Reference Kernel

```
service registry: M14-04,M14-05,M14-06
feature_flags: independent
state: versioned typed objects
authority: L0/L1 gates
telemetry: trigger,state_before,decision,tool,observation,state_after,residual
verifier: independent where possible
rollback: mandatory for reversible mutations
writeback: evidence-bearing delta only
```

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

Acceptance at kernel level

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

القيود والمشكلات المفتوحة

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

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

المراجع الخارجية الأساسية

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

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

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

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

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

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

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

المراسي الداخلية والسلالة

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

الورقة الثالثة: الذاكرة السببية والانتباه والاستمرار عبر الأفق

& Causal Memory Compression, Persistent Attention Reconstructive Consistency

P14-C - إغلاق الذاكرة والانتباه عبر الأفق

MOUSSAIF Unified AGI Part XIV - Executable Theoretical Completion ضمن

Moussaif Zouhair / زهير مساعف · MUT × MUG × MOUSSAIXT

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

الملخص والأطروحة

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

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

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

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

العقد الإبستيمي ومنع تضخم الادعاء

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

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

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

مصفوفة الضغط من حالة الفن 2026

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

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

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

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

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

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

خريطة الاندماج الداخلي: corpus → canonical lineage

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

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

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

M14-07 - الفجوة والأطروحة والنسب

Moussaif Persistent Attention Dynamics Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

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

ضغط المجتمع العلمي

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

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

M14-07 - الصياغة الرسمية والـobservables

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

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

افتراضات الصياغة

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

قاعدة formal closure

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

M14-07 - الميكانيزم والعقد التشغيلي

State transition

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

الخطوات

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

عقد النواة

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

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

M14-07 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

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

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-07 - التجربة، الأدوات والتدريب

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

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

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-08 - الفجوة والأطروحة والنسب

Moussaif Causal Memory Compression Bound Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

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

ضغط المجتمع العلمي

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

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

M14-08 - الصياغة الرسمية والـobservables

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

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

افتراضات الصياغة

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

قاعدة formal closure

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

M14-08 - الميكانيزم والعقد التشغيلي

State transition

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

الخطوات

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

عقد النواة

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

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

M14-08 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

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

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-08 - التجربة، الأدوات والتدريب

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

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

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-09 - الفجوة والأطروحة والنسب

Moussaif Reconstructive Memory Eventual Consistency Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

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

ضغط المجتمع العلمي

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

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

M14-09 - الصياغة الرسمية والobservables

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

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

افتراضات الصياغة

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

قاعدة formal closure

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

M14-09 - الميكانيزم والعقد التشغيلي

State transition

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

الخطوات

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

عقد النواة

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

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

M14-09 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

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

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-09 - التجربة، الأدوات والتدريب

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

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

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

الاقتران 1: كيف تتحول النظريات إلى نظام لا قائمة

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

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

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

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

الاقتران 2: Graph نظري إلى Graph تنفيذي

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

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

Invariants المشتركة

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

مختبر الضغط 1: العدسة السببية: من الارتباط إلى intervention

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

سجل الأدوات والأصول التنفيذية - 1/2

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

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

سجل الأدوات والأصول التنفيذية - 2/2

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

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

برنامج التدريب: من التعريف إلى قدرة قابلة للنقل

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

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

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

مصفوفة التجارب - 1/1

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

ترجمة الورقة إلى MOUSSAIXT Reference Kernel

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

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

Acceptance at kernel level

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

القيود والمشكلات المفتوحة

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

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

المراجع الخارجية الأساسية

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

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

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

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

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

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

المراسي الداخلية والسلالة

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

الورقة الرابعة: الذكاء التنظيمي والبروتوكولي متعدد الوكلاء

Collective Intelligence, Semantic Sovereignty & Governed Delegation

P14-D - إغلاق التنظيم متعدد الوكلاء والبروتوكولات

ضمن MOUSSAIF Unified AGI Part XIV - Executable Theoretical Completion

Moussaif Zouhair / زهير مساعف · MUT × MUG × MOUSSAIXT

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

الملخص والأطروحة

تعالج هذه الورقة فجوة محددة في برنامج MOUSSAIF: فصل interoperability عن sovereignty، قياس alignment load التنظيمي، ومعايرة الثقة والتفويض في منظمات وكلائية غير متجانسة. لا تُقدم مساهمة الورقة على أنها «إثبات AGI»، بل كحزمة فرضيات ذات state variables وآليات قابلة للتعطيل وأدوات وتجارب ومفندات. معيار القيمة هو ما إذا أضافت النظرية فرقاً سببياً فوق baseline بسيط وتحمّل ذلك الفرق النقل بين شروط مختلفة.

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

> **المساهمة:** نظرية مساعف لحمل المحاذاة التنظيمي؛ نظرية مساعف لسيادة الدلالة فوق البروتوكول؛ نظرية مساعف لمعايرة الثقة في التفويض. كل مساهمة مرتبطة بمعادلة falsifier، feature flag، design-level، تجربة transfer envelope، matched-budget، وأداة runtime.

سؤال الورقة: ما الحد الأدنى من البنية الإضافية التي تجعل إغلاق التنظيم متعدد الوكلاء والبروتوكولات قابلاً للقياس والتفنيد والربط بأثر العالم، دون تحويل density النظرية إلى claim غير مدعوم؟

العقد الإبستيمي ومنع تضخم الادعاء

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

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

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

مصفوفة الضغط من حالة الفن 2026

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

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

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

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

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

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

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

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

خريطة الاندماج الداخلي: corpus → canonical lineage

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

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

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

M14-10 - الفجوة والأطروحة والنسب

Moussaif Organizational Alignment Load Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

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

ضغط المجتمع العلمي

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

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

M14-10 - الصياغة الرسمية والobservables

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

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

افتراضات الصياغة

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

قاعدة formal closure

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

M14-10 - الميكانيزم والعقد التشغيلي

State transition

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

الخطوات

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

عقد النواة

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

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

M14-10 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

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

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-10 - التجربة، الأدوات والتدريب

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

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

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-11 - الفجوة والأطروحة والنسب

Moussaif Semantic Sovereignty over Protocol Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **MURI plural reasoning**؛ **MUG governance**؛ **MOUSSAIXT adapters**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

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

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

M14-11 - الصياغة الرسمية والـobservables

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

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

افتراضات الصياغة

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

قاعدة formal closure

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

M14-11 - الميكانيزم والعقد التشغيلي

State transition

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

الخطوات

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

عقد النواة

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

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

M14-11 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

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

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-11 - التجربة، الأدوات والتدريب

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

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

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-12 - الفجوة والأطروحة والنسب

Moussaif Delegation Trust Calibration Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M-AFI Delegation Trust Dynamics**؛ **Authority-Preserving Delegation**؛ **Agent Mesh**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

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

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

M14-12 - الصياغة الرسمية والـobservables

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

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

افتراضات الصياغة

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

قاعدة formal closure

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

M14-12 - الميكانيزم والعقد التشغيلي

State transition

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

الخطوات

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

عقد النواة

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

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

M14-12 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

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

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-12 - التجربة، الأدوات والتدريب

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

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

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

الاقتران 1: كيف تتحول النظريات إلى نظام لا قائمة

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

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

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

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

الاقتران 2: Graph نظري إلى Graph تنفيذي

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

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

Invariants المشتركة

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

مختبر الضغط 1: العدسة السببية: من الارتباط إلى intervention

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

سجل الأدوات والأصول التنفيذية - 1/2

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

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

سجل الأدوات والأصول التنفيذية - 2/2

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

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

برنامج التدريب: من التعريف إلى قدرة قابلة للنقل

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

- M14-10 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-10 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-10 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-10 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-11 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-11 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-11 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-11 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-12 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-12 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-12 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-12 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

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

مصفوفة التجارب - 1/1

ID	Owner	Kind	Design	Metrics
-XIV E28	M14-10	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	primary theory metric + task outcome + calibration + cost + critical regressions
-XIV E29	M14-10	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E30	M14-10	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E31	M14-11	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	primary theory metric + task outcome + calibration + cost + critical regressions
-XIV E32	M14-11	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E33	M14-11	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E34	M14-12	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	primary theory metric + task outcome + calibration + cost + critical regressions
-XIV E35	M14-12	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E36	M14-12	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier

ترجمة الورقة إلى MOUSSAIXT Reference Kernel

```
service registry: M14-10,M14-11,M14-12
feature_flags: independent
state: versioned typed objects
authority: L0/L1 gates
telemetry: trigger,state_before,decision,tool,observation,state_after,residual
verifier: independent where possible
rollback: mandatory for reversible mutations
writeback: evidence-bearing delta only
```

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

Acceptance at kernel level

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

القيود والمشكلات المفتوحة

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

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

المراجع الخارجية الأساسية

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

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

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

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

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

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

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

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

المراسي الداخلية والسلالة

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

الورقة الخامسة: العمومية والتقييم والتطور العلمي الذاتي

Generality Frontiers, Harness Decomposition, Evaluator Independence & Scientific Self-Evolution

P14-EF - إغلاق العمومية والتقييم والتطور العلمي الذاتي

ضمن MOUSSAIF Unified AGI Part XIV - Executable Theoretical Completion

Moussaif Zouhair / زهير مساعف · MUT × MUG × MOUSSAIXT

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

الملخص والأطروحة

تعالج هذه الورقة فجوة محددة في برنامج MOUSSAIF: تعريف frontier للعمومية، morphism بين benchmark وdeployment، فصل model capability عن harness، ثم ضبط self-evolution باستقلال المقيم وإنتاجية إعادة الإنتاج وثبات الدستور. لا تُقدم مساهمة الورقة على أنها «إثبات AGI»، بل كحزمة فرضيات ذات state variables وآليات قابلة للتعطيل وأدوات وتجارب ومفندات. معيار القيمة هو ما إذا أضافت النظرية فرقاً سببياً فوق baseline بسيط وتحمّل ذلك الفرق النقل بين شروط مختلفة.

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

المساهمة: نظرية مساعف لجبهة عمومية الإغلاق؛ نظرية مساعف للتحويل بين المعيار والنشر؛ نظرية مساعف لتفكيك القدرة المقاوم للهارنس؛ نظرية مساعف لدَين استقلال المقيم؛ نظرية مساعف لعائد إعادة الإنتاج العلمي؛ نظرية مساعف لثبات الطفرة الدستورية. كل مساهمة مرتبطة بمعادلة falsifier، feature flag، design-level، تجربة matched-budget، transfer envelope، وأداة runtime.

سؤال الورقة: ما الحد الأدنى من البنية الإضافية التي تجعل إغلاق العمومية والتقييم والتطور العلمي الذاتي قابلاً للقياس والتفنيد والربط بأثر العالم، دون تحويل density النظرية إلى claim غير مدعوم؟

العقد الإبستيمي ومنع تضخم الادعاء

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

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

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

مصفوفة الضغط من حالة الفن 2026

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

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

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

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

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

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

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

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

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

خريطة الاندماج الداخلي: corpus → canonical lineage

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

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

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

M14-13 - الفجوة والأطروحة والنسب

Moussaif Generality Closure Frontier Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M13-08 Closed-Loop Generality**؛ **MUT transfer program**؛ **MUG evidence promotion**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

CUBE 2026، AgencyBench 2026، AgentGym2 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-13 - الصياغة الرسمية والobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
TransformSet	تحويلات نموذج/مجال/أداة/أفق/موارد
normalized_effect	فرق سببي عن baseline
Recovery	استرداد بعد perturbation

افتراضات الصياغة

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

قاعدة formal closure

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

M14-13 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-13 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. عرّف transformations قبل الاختبار.
2. طابق budgets قدر الإمكان.
3. احسب effect relative لا raw score فقط.
4. سجل sign stability وfailure boundaries.
5. افصل adaptation المسموح من retraining.
6. ارسم resource frontier.

عقد النواة

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

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

M14-13 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- mechanism الحقيقي يحافظ على sign في subset واسع لا جميع الحالات.
- قد تقل الجودة الخام بينما تتحسن calibration/recovery.
- عمومية مزعومة ستكشفها transformations صغيرة في harness.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-13 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
XIV-A121	Generality Closure Frontier State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A122	Generality Closure Frontier Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A123	Generality Closure Frontier Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A124	Generality Closure Frontier Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A125	Generality Closure Frontier Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
XIV-A126	Generality Closure Frontier Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-14 - الفجوة والأطروحة والنسب

Moussaif Benchmark-to-Deployment Morphism Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M13-08 Generality**؛ **World-Effect Closure**؛ **MUG evaluation**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو evidence أو recovery.

ضغط المجتمع العلمي

CUBE 2026؛ DeepPlanning 2026؛ AgencyBench 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-14 - الصياغة الرسمية والobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
I_k	invariant وظيفي
d	اختلاف semantics/measurement
ClaimTransfer	سقف قوة النقل

افتراضات الصياغة

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

قاعدة formal closure

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

M14-14 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-14 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. صف benchmark contract.
2. صف deployment contract.
3. ابن invariant crosswalk.
4. اختبر mismatches بمهمات bridge.
5. لا تنقل claim خارج preserved dimensions.
6. حدث morphism عند تغير harness أو product.

عقد النواة

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

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

M14-14 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- تفسر بعض leaderboard gains التي لا تتحول إلى production value.
- bridge tasks تقلل surprise عند deployment.
- claims تصبح أضيق لكنها أكثر قابلية للدفاع.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-14 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A131	Benchmark to Deployment Morphism State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A132	Benchmark to Deployment Morphism Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A133	Benchmark to Deployment Morphism Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A134	Benchmark to Deployment Morphism Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A135	Benchmark to Deployment Morphism Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A136	Benchmark to Deployment Morphism Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-15 - الفجوة والأطروحة والنسب

Moussaif Harness-Resistant Capability Decomposition Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **MUT anti-inflation rule**؛ **M13 Theory-to-Kernel Closure**؛ **MCD routing**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

AgencyBench 2026؛ Trajectory Attribution 2026؛ Plan-RewardBench 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-15 - الصياغة الرسمية والـobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
Model	backbone
Harness	prompt/router/tools/retries
Mechanism	المكوّن النظري
Credit_K	فرق سببي مقاوم للتبديل

افتراضات الصياغة

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

قاعدة formal closure

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

M14-15 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-15 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. ثبت task/data.
2. أنشئ minimal baseline.
3. بدّل mechanism feature flag.
4. طابق token/tool/time budget.
5. نفذ factorial subsets مع seeds.
6. حلل interactions لا main effect فقط.
7. انشر null results.

عقد النواة

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

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

M14-15 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- تختفي بعض "النظريات" عند إزالة harness advantage.
- mechanisms الحقيقية تظهر sign ثابتاً نسبياً.
- interaction effect قد يكون الإسهام الحقيقي بدلاً من component منفرد.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-15 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A141	Harness Resistant Capability Decomposition State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A142	Harness Resistant Capability Decomposition Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A143	Harness Resistant Capability Decomposition Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A144	Harness Resistant Capability Decomposition Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A145	Harness Resistant Capability Decomposition Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A146	Harness Resistant Capability Decomposition Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر /domain/tool/model horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-16 - الفجوة والأطروحة والنسب

Moussaif Evaluator Independence Debt Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **MUG**، **Falsification Court**، **evidence governance**، **M13 External Grounding**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

Recursive Self-Improvement survey 2026، Plan-RewardBench 2026، GPT-Red 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-16 - الصياغة الرسمية والobservables

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

الرمز/المتغير	المعنى التشغيلي المقترح
ModelOverlap	تشابه النموذج/العائلة
DataOverlap	تداخل بيانات
AdaptiveCoupling	تكيّف المقيم والنظام معاً
Leakage	انكشاف test
MetricGaming	تحسين على المقاس
ExternalGrounding	دليل مستقل/صوري/عالمي

افتراضات الصياغة

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

قاعدة formal closure

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

M14-16 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-16 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. سجل evaluator lineage.
2. احسب overlap indicators.
3. استخدم multiple evaluator classes.
4. أدخل hidden renewal.
5. افصل attacker/critic عن deployed system حيث يلزم.
6. لا تسمح للنظام بتعديل acceptance gate منفرداً.

عقد النواة

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

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

M14-16 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- self-improvement gains تتكمش عند independent evaluator إذا كان الدين مرتفعاً.
- formal/world-state evidence أقل عرضة للدين من intrinsic self-judge.
- التجارب الطويلة تحتاج renewal لأن agent يتعلم rubric.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-16 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
XIV-A151	Evaluator Independence Debt State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A152	Evaluator Independence Debt Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A153	Evaluator Independence Debt Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A154	Evaluator Independence Debt Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A155	Evaluator Independence Debt Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
XIV-A156	Evaluator Independence Debt Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-17 - الفجوة والأطروحة والنسب

Moussaif Scientific Reproduction Yield Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **MUG publication governance**؛ **Evidence Closure**؛ **M13 External Grounding**. معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

CUBE 2026؛ Trajectory Attribution 2026؛ Recursive Self-Improvement 2026. هذه الأعمال تُعامل كـevidence problem وpriors mechanism وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-17 - الصياغة الرسمية والobservables

SRY = Σ_j Indep_j·Diversity_j·Reproduce_j / Σ_j Attempt_j; report with CI and failure taxonomy

الرمز/المتغير	المعنى التشغيلي المقترح
Indep_j	استقلال الفريق/الهارنس/البيانات
Diversity_j	تنوع شروط الإعادة
Reproduce_j	حفظ sign/criterion لا التطابق الحرفي

افتراضات الصياغة

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

قاعدة formal closure

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

M14-17 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-17 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. أنشئ replay package.
2. ثبت hashes/configs.
3. سجل attempts قبل معرفة النتائج.
4. فرق exact reproduction عن conceptual replication.
5. حافظ على negative results.
6. حدث theory envelope من failures.

عقد النواة

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

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

M14-17 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- theories الأقوى تملك yield قابل للحفاظ خارج harness الأصلي.
- إعادة الإنتاج تكشف hidden dependencies أكثر من إضافة benchmarks مشابهة.
- نشر null results يخفض theory inflation.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-17 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
XIV-A161	Scientific Reproduction Yield State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
XIV-A162	Scientific Reproduction Yield Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
XIV-A163	Scientific Reproduction Yield Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
XIV-A164	Scientific Reproduction Yield Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
XIV-A165	Scientific Reproduction Yield Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
XIV-A166	Scientific Reproduction Yield Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

M14-18 - الفجوة والأطروحة والنسب

Moussaif Constitutional Mutation Stability Theory

الفجوة الدقيقة

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

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

لماذا ليست alias؟

لا تُرقّى هذه العقدة لأنها تملك اسماً جديداً، بل لأنها تقترح متغيراً أو operator أو prediction غير ممثل مباشرة في أقرب الآباء. أقرب السلالة: **M13 MUG constitution؛ Recursive Improvement؛ Governance Space.** معيار novelty هو أن حذف العقدة أو استبدالها بأقرب parent يجب أن ينتج فرقاً قابلاً للقياس في route أو decision أو recovery أو evidence.

ضغط المجتمع العلمي

Recursive Self-Improvement 2026؛ GPT-Red 2026؛ DeepMind AI Control 2026؛ Anthropic trust/containment 2026. هذه الأعمال تُعامل كـproblem evidence وmechanism priors وbaselines، لا كتصديق على النظرية الداخلية.

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

M14-18 - الصياغة الرسمية والobservables

CMS = Gain_validated · InvariantPreservation · RollbackVerified · EvaluatorIndependence - CriticalRegressionRisk

الرمز/المتغير	المعنى التشغيلي المقترح
Gain_validated	مكسب متحقق causalياً
InvariantPreservation	حفظ invariants
RollbackVerified	نجاح الاسترجاع
EvaluatorIndependence	استقلال التقييم
CriticalRegressionRisk	مخاطر فشل حرج

افتراضات الصياغة

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

قاعدة formal closure

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

M14-18 - الميكانيزم والعقد التشغيلي

State transition

```
trigger → validate inputs → capture state S_t → activate M14-18 → emit predicted_delta → execute bounded action/probe → independent readback → court → memory residue
```

الخطوات

1. افصل constitution root عن mutable branch.
2. حوّل mutation إلى typed diff.
3. شغّل pre/post invariant tests.
4. اختبر adversarial/negative activation.
5. استخدم evaluator مستقل.
6. نفذ rollback drill.
7. PROMOTE فقط بعد replication، وإلا HOLD/REVISE.

عقد النواة

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

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

M14-18 - التنبؤات، المفندات ومحكمة الهدم

مفندات مباشرة

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

تنبؤات فارقة

- self-evolution الأكثر أماناً سيكون أبطأ لكنه أكثر قابلية لإعادة الإنتاج.
- mutations التي تغيّر evaluator والدستور معاً ستظهر optimism أعلى.
- rollback drills تكشف hidden state غير مرجوع.

أربعة خصوم نظريين

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

قرار المحكمة

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

M14-18 - التجربة، الأدوات والتدريب

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

ID	الأداة/الأصل	النوع	الدور
-XIV A171	Constitutional Mutation Stability State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A172	Constitutional Mutation Stability Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A173	Constitutional Mutation Stability Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A174	Constitutional Mutation Stability Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A175	Constitutional Mutation Stability Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A176	Constitutional Mutation Stability Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.

وصفات التدريب/التقييم

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

الحد الأدنى للتجارب

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

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

الاقتران 1: كيف تتحول النظريات إلى نظام لا قائمة

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

- M14-13: نظرية مساعف لجبهة عمومية الإغلاق - output حاكم: تحويلات نموذج/مجال/أداة/أفق/موارد
- M14-14: نظرية مساعف للتحويل بين المعيار والنشر - output حاكم: invariant وظيفي
- M14-15: نظرية مساعف لتفكيك القدرة المقاوم للهارنس - output حاكم: backbone
- M14-16: نظرية مساعف لدَين استقلال المقيم - output حاكم: تشابه النموذج/العائلة
- M14-17: نظرية مساعف لعائد إعادة الإنتاج العلمي - output حاكم: استقلال الفريق/الهارنس/البيانات
- M14-18: نظرية مساعف لثبات الطفرة الدستورية - output حاكم: مكسب متحقق causalياً

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

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

الاقتران 2: Graph نظري إلى Graph تنفيذي

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

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

Invariants المشتركة

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

مختبر الضغط 1: العدسة السببية: من الارتباط إلى intervention

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

سجل الأدوات والأصول التنفيذية - 1/2

ID	Owner	Asset	Class	Purpose
-XIV A121	M14-13	Generality Closure Frontier State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A122	M14-13	Generality Closure Frontier Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A123	M14-13	Generality Closure Frontier Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A124	M14-13	Generality Closure Frontier Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A125	M14-13	Generality Closure Frontier Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A126	M14-13	Generality Closure Frontier Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر /domain/tool/model horizon.
-XIV A127	M14-13	Generality Closure Frontier Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A128	M14-13	Generality Closure Frontier Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A129	M14-13	Generality Closure Frontier Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A130	M14-13	Generality Closure Frontier Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A131	M14-14	Benchmark to Deployment Morphism State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A132	M14-14	Benchmark to Deployment Morphism Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A133	M14-14	Benchmark to Deployment Morphism Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A134	M14-14	Benchmark to Deployment Morphism Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A135	M14-14	Benchmark to Deployment Morphism Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A136	M14-14	Benchmark to Deployment Morphism Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر /domain/tool/model horizon.
-XIV A137	M14-14	Benchmark to Deployment Morphism Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A138	M14-14	Benchmark to Deployment Morphism Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A139	M14-14	Benchmark to Deployment Morphism Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A140	M14-14	Benchmark to Deployment Morphism Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A141	M14-15	Harness Resistant Capability Decomposition State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A142	M14-15	Harness Resistant Capability Decomposition Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A143	M14-15	Harness Resistant Capability Decomposition Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A144	M14-15	Harness Resistant Capability Decomposition Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
	M14-15		METHOD	

-XIV A145		Harness Resistant Capability Decomposition Matched-Budget Method		يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A146	M14-15	Harness Resistant Capability Decomposition Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر /domain/tool/model horizon.
-XIV A147	M14-15	Harness Resistant Capability Decomposition Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A148	M14-15	Harness Resistant Capability Decomposition Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A149	M14-15	Harness Resistant Capability Decomposition Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A150	M14-15	Harness Resistant Capability Decomposition Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.

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

سجل الأدوات والأصول التنفيذية - 2/2

ID	Owner	Asset	Class	Purpose
-XIV A151	M14-16	Evaluator Independence Debt State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A152	M14-16	Evaluator Independence Debt Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A153	M14-16	Evaluator Independence Debt Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A154	M14-16	Evaluator Independence Debt Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A155	M14-16	Evaluator Independence Debt Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A156	M14-16	Evaluator Independence Debt Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A157	M14-16	Evaluator Independence Debt Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A158	M14-16	Evaluator Independence Debt Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A159	M14-16	Evaluator Independence Debt Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A160	M14-16	Evaluator Independence Debt Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A161	M14-17	Scientific Reproduction Yield State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A162	M14-17	Scientific Reproduction Yield Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A163	M14-17	Scientific Reproduction Yield Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A164	M14-17	Scientific Reproduction Yield Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
-XIV A165	M14-17	Scientific Reproduction Yield Matched-Budget Method	METHOD	يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A166	M14-17	Scientific Reproduction Yield Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A167	M14-17	Scientific Reproduction Yield Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A168	M14-17	Scientific Reproduction Yield Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A169	M14-17	Scientific Reproduction Yield Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A170	M14-17	Scientific Reproduction Yield Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.
-XIV A171	M14-18	Constitutional Mutation Stability State Profiler	TOOL	يستخرج متغيرات النظرية والمجهول والاعتماد على المصادر قبل التفعيل.
-XIV A172	M14-18	Constitutional Mutation Stability Counterfactual Probe	TOOL	يولد intervention يفصل causal mechanism عن أقرب baseline.
-XIV A173	M14-18	Constitutional Mutation Stability Failure Topology Mapper	TOOL	يجمع failure signatures ويربط recurrence بالسياق والسبب المحتمل.
-XIV A174	M14-18	Constitutional Mutation Stability Calibration Engine	TOOL	يعاير thresholds/uncertainty/cost على held-out data.
	M14-18	Constitutional Mutation Stability Matched-Budget Method	METHOD	

-XIV A175				يطابق model/tools/context/retries/time ويمنع attribution الزائف.
-XIV A176	M14-18	Constitutional Mutation Stability Cross-Environment Transfer	METHOD	يفحص حفظ causal role عبر domain/tool/model/horizon.
-XIV A177	M14-18	Constitutional Mutation Stability Evidence Promotion Protocol	PROTOCOL	ينظم N0-N5/E0-E5 ويمنع claim expansion قبل replication.
-XIV A178	M14-18	Constitutional Mutation Stability Negative-Result Protocol	PROTOCOL	يحفظ الفشل والحدود ويمنع إعادة تسمية الفرضية بعد عدم نجاحها.
-XIV A179	M14-18	Constitutional Mutation Stability Runtime Service	STACK	خدمة feature-flagged مستقلة لها state/telemetry/rollback.
-XIV A180	M14-18	Constitutional Mutation Stability Proof/Replay Adapter	STACK	يحوّل trajectory إلى manifest + evidence packet + replay receipt.

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

برنامج التدريب: من التعريف إلى قدرة قابلة للنقل

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

- M14-13 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-13 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-13 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-13 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-14 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-14 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-14 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-14 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-15 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-15 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-15 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-15 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

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

برنامج التدريب - الاستكمال والانتقال

- M14-16 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-16 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-16 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-16 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-17 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-17 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-17 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-17 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.
- M14-18 / Mechanism Isolation Curriculum: حالات بسيطة تعزل المتغير المركزي قبل open-world confounders.
- M14-18 / Adversarial Boundary Curriculum: أمثلة داخل/خارج الحد وnegative activation لمنع overgeneralization.
- M14-18 / Matched-Budget Causal Evaluation: تدريب/تقييم يقارن تحت model/tools/context/retries/time متقاربة مع sham scaffold.
- M14-18 / Cross-Model/Domain Replication: إعادة التدريب/الاختبار في model/domain/tool family مختلفة مع stable-sign criterion.

الانتقال لا يعني استنساخ الأوزان أو prompts فقط. نختبر transfer of causal role: هل يبقى سبب التحسن نفسه عند تغيير model/domain/tool؟ إذا تغير، يُسجل كـinteraction أو domain-specific variant.

مصفوفة التجارب - 1/2

ID	Owner	Kind	Design	Metrics
-XIV E37	M14-13	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E38	M14-13	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E39	M14-13	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E40	M14-14	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E41	M14-14	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E42	M14-14	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E43	M14-15	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E44	M14-15	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E45	M14-15	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E46	M14-16	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E47	M14-16	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E48	M14-16	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier

مصفوفة التجارب - 2/2

ID	Owner	Kind	Design	Metrics
-XIV E49	M14-17	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E50	M14-17	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E51	M14-17	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier
-XIV E52	M14-18	Minimal Causal Ablation	switchable implementation مقابل nearest parent/baseline تحت matched budget.	+ primary theory metric + task outcome + calibration + cost critical regressions
-XIV E53	M14-18	Boundary/Adversarial Test	اختبار failure modes والحدود بحالات مصممة لكسر التعميم.	false activation + missed failure + rollback + evidence integrity
-XIV E54	M14-18	Transfer/Replication	نقل الآلية إلى model/domain/tool/horizon مختلف مع preregistered metric.	effect sign stability + CI + failure-signature transfer + resource frontier

ترجمة الورقة إلى MOUSSAIXT Reference Kernel

```
service registry: M14-13,M14-14,M14-15,M14-16,M14-17,M14-18
feature_flags: independent
state: versioned typed objects
authority: L0/L1 gates
telemetry: trigger,state_before,decision,tool,observation,state_after,residual
verifier: independent where possible
rollback: mandatory for reversible mutations
writeback: evidence-bearing delta only
```

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

Acceptance at kernel level

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

القيود والمشكلات المفتوحة

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

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

المراجع الخارجية الأساسية

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

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

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

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

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

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

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

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

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

المراسي الداخلية والسلالة

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

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

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

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

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

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

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

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

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

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

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

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

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

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

الملحق A - سجل الأصول 1-30

Name	Class	Owner	ID
Observational Sufficiency Frontier State Profiler	TOOL	M14-01	XIV-A001
Observational Sufficiency Frontier Counterfactual Probe	TOOL	M14-01	XIV-A002
Observational Sufficiency Frontier Failure Topology Mapper	TOOL	M14-01	XIV-A003
Observational Sufficiency Frontier Calibration Engine	TOOL	M14-01	XIV-A004
Observational Sufficiency Frontier Matched-Budget Method	METHOD	M14-01	XIV-A005
Observational Sufficiency Frontier Cross-Environment Transfer	METHOD	M14-01	XIV-A006
Observational Sufficiency Frontier Evidence Promotion Protocol	PROTOCOL	M14-01	XIV-A007
Observational Sufficiency Frontier Negative-Result Protocol	PROTOCOL	M14-01	XIV-A008
Observational Sufficiency Frontier Runtime Service	STACK	M14-01	XIV-A009
Observational Sufficiency Frontier Proof/Replay Adapter	STACK	M14-01	XIV-A010
Decision Boundary World Model State Profiler	TOOL	M14-02	XIV-A011
Decision Boundary World Model Counterfactual Probe	TOOL	M14-02	XIV-A012
Decision Boundary World Model Failure Topology Mapper	TOOL	M14-02	XIV-A013
Decision Boundary World Model Calibration Engine	TOOL	M14-02	XIV-A014
Decision Boundary World Model Matched-Budget Method	METHOD	M14-02	XIV-A015
Decision Boundary World Model Cross-Environment Transfer	METHOD	M14-02	XIV-A016
Decision Boundary World Model Evidence Promotion Protocol	PROTOCOL	M14-02	XIV-A017
Decision Boundary World Model Negative-Result Protocol	PROTOCOL	M14-02	XIV-A018
Decision Boundary World Model Runtime Service	STACK	M14-02	XIV-A019
Decision Boundary World Model Proof/Replay Adapter	STACK	M14-02	XIV-A020
Residual Causality Graph State Profiler	TOOL	M14-03	XIV-A021
Residual Causality Graph Counterfactual Probe	TOOL	M14-03	XIV-A022
Residual Causality Graph Failure Topology Mapper	TOOL	M14-03	XIV-A023
Residual Causality Graph Calibration Engine	TOOL	M14-03	XIV-A024
Residual Causality Graph Matched-Budget Method	METHOD	M14-03	XIV-A025
Residual Causality Graph Cross-Environment Transfer	METHOD	M14-03	XIV-A026
Residual Causality Graph Evidence Promotion Protocol	PROTOCOL	M14-03	XIV-A027
Residual Causality Graph Negative-Result Protocol	PROTOCOL	M14-03	XIV-A028
Residual Causality Graph Runtime Service	STACK	M14-03	XIV-A029
Residual Causality Graph Proof/Replay Adapter	STACK	M14-03	XIV-A030

الملحق A - سجل الأصول 31-60

Name	Class	Owner	ID
Proof Carrying Trajectory State Profiler	TOOL	M14-04	XIV-A031
Proof Carrying Trajectory Counterfactual Probe	TOOL	M14-04	XIV-A032
Proof Carrying Trajectory Failure Topology Mapper	TOOL	M14-04	XIV-A033
Proof Carrying Trajectory Calibration Engine	TOOL	M14-04	XIV-A034
Proof Carrying Trajectory Matched-Budget Method	METHOD	M14-04	XIV-A035
Proof Carrying Trajectory Cross-Environment Transfer	METHOD	M14-04	XIV-A036
Proof Carrying Trajectory Evidence Promotion Protocol	PROTOCOL	M14-04	XIV-A037
Proof Carrying Trajectory Negative-Result Protocol	PROTOCOL	M14-04	XIV-A038
Proof Carrying Trajectory Runtime Service	STACK	M14-04	XIV-A039
Proof Carrying Trajectory Proof/Replay Adapter	STACK	M14-04	XIV-A040
Commitment Horizon Geometry State Profiler	TOOL	M14-05	XIV-A041
Commitment Horizon Geometry Counterfactual Probe	TOOL	M14-05	XIV-A042
Commitment Horizon Geometry Failure Topology Mapper	TOOL	M14-05	XIV-A043
Commitment Horizon Geometry Calibration Engine	TOOL	M14-05	XIV-A044
Commitment Horizon Geometry Matched-Budget Method	METHOD	M14-05	XIV-A045
Commitment Horizon Geometry Cross-Environment Transfer	METHOD	M14-05	XIV-A046
Commitment Horizon Geometry Evidence Promotion Protocol	PROTOCOL	M14-05	XIV-A047
Commitment Horizon Geometry Negative-Result Protocol	PROTOCOL	M14-05	XIV-A048
Commitment Horizon Geometry Runtime Service	STACK	M14-05	XIV-A049
Commitment Horizon Geometry Proof/Replay Adapter	STACK	M14-05	XIV-A050
Authority Preserving Delegation State Profiler	TOOL	M14-06	XIV-A051
Authority Preserving Delegation Counterfactual Probe	TOOL	M14-06	XIV-A052
Authority Preserving Delegation Failure Topology Mapper	TOOL	M14-06	XIV-A053
Authority Preserving Delegation Calibration Engine	TOOL	M14-06	XIV-A054
Authority Preserving Delegation Matched-Budget Method	METHOD	M14-06	XIV-A055
Authority Preserving Delegation Cross-Environment Transfer	METHOD	M14-06	XIV-A056
Authority Preserving Delegation Evidence Promotion Protocol	PROTOCOL	M14-06	XIV-A057
Authority Preserving Delegation Negative-Result Protocol	PROTOCOL	M14-06	XIV-A058
Authority Preserving Delegation Runtime Service	STACK	M14-06	XIV-A059
Authority Preserving Delegation Proof/Replay Adapter	STACK	M14-06	XIV-A060

الملحق A - سجل الأصول 61-90

Name	Class	Owner	ID
Persistent Attention Dynamics State Profiler	TOOL	M14-07	XIV-A061
Persistent Attention Dynamics Counterfactual Probe	TOOL	M14-07	XIV-A062
Persistent Attention Dynamics Failure Topology Mapper	TOOL	M14-07	XIV-A063
Persistent Attention Dynamics Calibration Engine	TOOL	M14-07	XIV-A064
Persistent Attention Dynamics Matched-Budget Method	METHOD	M14-07	XIV-A065
Persistent Attention Dynamics Cross-Environment Transfer	METHOD	M14-07	XIV-A066
Persistent Attention Dynamics Evidence Promotion Protocol	PROTOCOL	M14-07	XIV-A067
Persistent Attention Dynamics Negative-Result Protocol	PROTOCOL	M14-07	XIV-A068
Persistent Attention Dynamics Runtime Service	STACK	M14-07	XIV-A069
Persistent Attention Dynamics Proof/Replay Adapter	STACK	M14-07	XIV-A070
Causal Memory Compression Bound State Profiler	TOOL	M14-08	XIV-A071
Causal Memory Compression Bound Counterfactual Probe	TOOL	M14-08	XIV-A072
Causal Memory Compression Bound Failure Topology Mapper	TOOL	M14-08	XIV-A073
Causal Memory Compression Bound Calibration Engine	TOOL	M14-08	XIV-A074
Causal Memory Compression Bound Matched-Budget Method	METHOD	M14-08	XIV-A075
Causal Memory Compression Bound Cross-Environment Transfer	METHOD	M14-08	XIV-A076
Causal Memory Compression Bound Evidence Promotion Protocol	PROTOCOL	M14-08	XIV-A077
Causal Memory Compression Bound Negative-Result Protocol	PROTOCOL	M14-08	XIV-A078
Causal Memory Compression Bound Runtime Service	STACK	M14-08	XIV-A079
Causal Memory Compression Bound Proof/Replay Adapter	STACK	M14-08	XIV-A080
Reconstructive Memory Eventual Consistency State Profiler	TOOL	M14-09	XIV-A081
Reconstructive Memory Eventual Consistency Counterfactual Probe	TOOL	M14-09	XIV-A082
Reconstructive Memory Eventual Consistency Failure Topology Mapper	TOOL	M14-09	XIV-A083
Reconstructive Memory Eventual Consistency Calibration Engine	TOOL	M14-09	XIV-A084
Reconstructive Memory Eventual Consistency Matched-Budget Method	METHOD	M14-09	XIV-A085
Reconstructive Memory Eventual Consistency Cross-Environment Transfer	METHOD	M14-09	XIV-A086
Reconstructive Memory Eventual Consistency Evidence Promotion Protocol	PROTOCOL	M14-09	XIV-A087
Reconstructive Memory Eventual Consistency Negative-Result Protocol	PROTOCOL	M14-09	XIV-A088
Reconstructive Memory Eventual Consistency Runtime Service	STACK	M14-09	XIV-A089
Reconstructive Memory Eventual Consistency Proof/Replay Adapter	STACK	M14-09	XIV-A090

الملحق A - سجل الأصول 91-120

Name	Class	Owner	ID
Organizational Alignment Load State Profiler	TOOL	M14-10	XIV-A091
Organizational Alignment Load Counterfactual Probe	TOOL	M14-10	XIV-A092
Organizational Alignment Load Failure Topology Mapper	TOOL	M14-10	XIV-A093
Organizational Alignment Load Calibration Engine	TOOL	M14-10	XIV-A094
Organizational Alignment Load Matched-Budget Method	METHOD	M14-10	XIV-A095
Organizational Alignment Load Cross-Environment Transfer	METHOD	M14-10	XIV-A096
Organizational Alignment Load Evidence Promotion Protocol	PROTOCOL	M14-10	XIV-A097
Organizational Alignment Load Negative-Result Protocol	PROTOCOL	M14-10	XIV-A098
Organizational Alignment Load Runtime Service	STACK	M14-10	XIV-A099
Organizational Alignment Load Proof/Replay Adapter	STACK	M14-10	XIV-A100
Semantic Sovereignty over Protocol State Profiler	TOOL	M14-11	XIV-A101
Semantic Sovereignty over Protocol Counterfactual Probe	TOOL	M14-11	XIV-A102
Semantic Sovereignty over Protocol Failure Topology Mapper	TOOL	M14-11	XIV-A103
Semantic Sovereignty over Protocol Calibration Engine	TOOL	M14-11	XIV-A104
Semantic Sovereignty over Protocol Matched-Budget Method	METHOD	M14-11	XIV-A105
Semantic Sovereignty over Protocol Cross-Environment Transfer	METHOD	M14-11	XIV-A106
Semantic Sovereignty over Protocol Evidence Promotion Protocol	PROTOCOL	M14-11	XIV-A107
Semantic Sovereignty over Protocol Negative-Result Protocol	PROTOCOL	M14-11	XIV-A108
Semantic Sovereignty over Protocol Runtime Service	STACK	M14-11	XIV-A109
Semantic Sovereignty over Protocol Proof/Replay Adapter	STACK	M14-11	XIV-A110
Delegation Trust Calibration State Profiler	TOOL	M14-12	XIV-A111
Delegation Trust Calibration Counterfactual Probe	TOOL	M14-12	XIV-A112
Delegation Trust Calibration Failure Topology Mapper	TOOL	M14-12	XIV-A113
Delegation Trust Calibration Calibration Engine	TOOL	M14-12	XIV-A114
Delegation Trust Calibration Matched-Budget Method	METHOD	M14-12	XIV-A115
Delegation Trust Calibration Cross-Environment Transfer	METHOD	M14-12	XIV-A116
Delegation Trust Calibration Evidence Promotion Protocol	PROTOCOL	M14-12	XIV-A117
Delegation Trust Calibration Negative-Result Protocol	PROTOCOL	M14-12	XIV-A118
Delegation Trust Calibration Runtime Service	STACK	M14-12	XIV-A119
Delegation Trust Calibration Proof/Replay Adapter	STACK	M14-12	XIV-A120

الملحق A - سجل الأصول 121-150

Name	Class	Owner	ID
Generality Closure Frontier State Profiler	TOOL	M14-13	XIV-A121
Generality Closure Frontier Counterfactual Probe	TOOL	M14-13	XIV-A122
Generality Closure Frontier Failure Topology Mapper	TOOL	M14-13	XIV-A123
Generality Closure Frontier Calibration Engine	TOOL	M14-13	XIV-A124
Generality Closure Frontier Matched-Budget Method	METHOD	M14-13	XIV-A125
Generality Closure Frontier Cross-Environment Transfer	METHOD	M14-13	XIV-A126
Generality Closure Frontier Evidence Promotion Protocol	PROTOCOL	M14-13	XIV-A127
Generality Closure Frontier Negative-Result Protocol	PROTOCOL	M14-13	XIV-A128
Generality Closure Frontier Runtime Service	STACK	M14-13	XIV-A129
Generality Closure Frontier Proof/Replay Adapter	STACK	M14-13	XIV-A130
Benchmark to Deployment Morphism State Profiler	TOOL	M14-14	XIV-A131
Benchmark to Deployment Morphism Counterfactual Probe	TOOL	M14-14	XIV-A132
Benchmark to Deployment Morphism Failure Topology Mapper	TOOL	M14-14	XIV-A133
Benchmark to Deployment Morphism Calibration Engine	TOOL	M14-14	XIV-A134
Benchmark to Deployment Morphism Matched-Budget Method	METHOD	M14-14	XIV-A135
Benchmark to Deployment Morphism Cross-Environment Transfer	METHOD	M14-14	XIV-A136
Benchmark to Deployment Morphism Evidence Promotion Protocol	PROTOCOL	M14-14	XIV-A137
Benchmark to Deployment Morphism Negative-Result Protocol	PROTOCOL	M14-14	XIV-A138
Benchmark to Deployment Morphism Runtime Service	STACK	M14-14	XIV-A139
Benchmark to Deployment Morphism Proof/Replay Adapter	STACK	M14-14	XIV-A140
Harness Resistant Capability Decomposition State Profiler	TOOL	M14-15	XIV-A141
Harness Resistant Capability Decomposition Counterfactual Probe	TOOL	M14-15	XIV-A142
Harness Resistant Capability Decomposition Failure Topology Mapper	TOOL	M14-15	XIV-A143
Harness Resistant Capability Decomposition Calibration Engine	TOOL	M14-15	XIV-A144
Harness Resistant Capability Decomposition Matched-Budget Method	METHOD	M14-15	XIV-A145
Harness Resistant Capability Decomposition Cross-Environment Transfer	METHOD	M14-15	XIV-A146
Harness Resistant Capability Decomposition Evidence Promotion Protocol	PROTOCOL	M14-15	XIV-A147
Harness Resistant Capability Decomposition Negative-Result Protocol	PROTOCOL	M14-15	XIV-A148
Harness Resistant Capability Decomposition Runtime Service	STACK	M14-15	XIV-A149
Harness Resistant Capability Decomposition Proof/Replay Adapter	STACK	M14-15	XIV-A150

الملحق A - سجل الأصول 151-180

Name	Class	Owner	ID
Evaluator Independence Debt State Profiler	TOOL	M14-16	XIV-A151
Evaluator Independence Debt Counterfactual Probe	TOOL	M14-16	XIV-A152
Evaluator Independence Debt Failure Topology Mapper	TOOL	M14-16	XIV-A153
Evaluator Independence Debt Calibration Engine	TOOL	M14-16	XIV-A154
Evaluator Independence Debt Matched-Budget Method	METHOD	M14-16	XIV-A155
Evaluator Independence Debt Cross-Environment Transfer	METHOD	M14-16	XIV-A156
Evaluator Independence Debt Evidence Promotion Protocol	PROTOCOL	M14-16	XIV-A157
Evaluator Independence Debt Negative-Result Protocol	PROTOCOL	M14-16	XIV-A158
Evaluator Independence Debt Runtime Service	STACK	M14-16	XIV-A159
Evaluator Independence Debt Proof/Replay Adapter	STACK	M14-16	XIV-A160
Scientific Reproduction Yield State Profiler	TOOL	M14-17	XIV-A161
Scientific Reproduction Yield Counterfactual Probe	TOOL	M14-17	XIV-A162
Scientific Reproduction Yield Failure Topology Mapper	TOOL	M14-17	XIV-A163
Scientific Reproduction Yield Calibration Engine	TOOL	M14-17	XIV-A164
Scientific Reproduction Yield Matched-Budget Method	METHOD	M14-17	XIV-A165
Scientific Reproduction Yield Cross-Environment Transfer	METHOD	M14-17	XIV-A166
Scientific Reproduction Yield Evidence Promotion Protocol	PROTOCOL	M14-17	XIV-A167
Scientific Reproduction Yield Negative-Result Protocol	PROTOCOL	M14-17	XIV-A168
Scientific Reproduction Yield Runtime Service	STACK	M14-17	XIV-A169
Scientific Reproduction Yield Proof/Replay Adapter	STACK	M14-17	XIV-A170
Constitutional Mutation Stability State Profiler	TOOL	M14-18	XIV-A171
Constitutional Mutation Stability Counterfactual Probe	TOOL	M14-18	XIV-A172
Constitutional Mutation Stability Failure Topology Mapper	TOOL	M14-18	XIV-A173
Constitutional Mutation Stability Calibration Engine	TOOL	M14-18	XIV-A174
Constitutional Mutation Stability Matched-Budget Method	METHOD	M14-18	XIV-A175
Constitutional Mutation Stability Cross-Environment Transfer	METHOD	M14-18	XIV-A176
Constitutional Mutation Stability Evidence Promotion Protocol	PROTOCOL	M14-18	XIV-A177
Constitutional Mutation Stability Negative-Result Protocol	PROTOCOL	M14-18	XIV-A178
Constitutional Mutation Stability Runtime Service	STACK	M14-18	XIV-A179
Constitutional Mutation Stability Proof/Replay Adapter	STACK	M14-18	XIV-A180

الملحق B - سجل التجارب 1-18

ID	Owner	Kind	Metrics
XIV-E01	M14-01	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E02	M14-01	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E03	M14-01	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E04	M14-02	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E05	M14-02	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E06	M14-02	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E07	M14-03	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E08	M14-03	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E09	M14-03	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E10	M14-04	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E11	M14-04	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E12	M14-04	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E13	M14-05	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E14	M14-05	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E15	M14-05	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E16	M14-06	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E17	M14-06	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E18	M14-06	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier

الملحق B - سجل التجارب 19-36

ID	Owner	Kind	Metrics
XIV-E19	M14-07	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E20	M14-07	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E21	M14-07	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E22	M14-08	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E23	M14-08	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E24	M14-08	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E25	M14-09	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E26	M14-09	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E27	M14-09	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E28	M14-10	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E29	M14-10	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E30	M14-10	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E31	M14-11	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E32	M14-11	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E33	M14-11	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E34	M14-12	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E35	M14-12	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E36	M14-12	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier

الملحق B - سجل التجارب 37-54

ID	Owner	Kind	Metrics
XIV-E37	M14-13	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E38	M14-13	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E39	M14-13	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E40	M14-14	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E41	M14-14	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E42	M14-14	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E43	M14-15	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E44	M14-15	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E45	M14-15	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E46	M14-16	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E47	M14-16	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E48	M14-16	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E49	M14-17	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E50	M14-17	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E51	M14-17	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier
XIV-E52	M14-18	Minimal Causal Ablation	primary theory metric + task outcome + calibration + cost + critical regressions
XIV-E53	M14-18	Boundary/Adversarial Test	false activation + missed failure + rollback + evidence integrity
XIV-E54	M14-18	Transfer/Replication	effect sign stability + CI + failure-signature transfer + resource frontier

الدستور الختامي لـ Part XIV

- لا نظرية بلا فرق سببي محتمل عن أقرب baseline.
- لا فرق بلا observable/estimator أو unresolved flag.
- لا آلية بلا feature flag وtrace.
- لا فعل بلا authority وworld-state verification.
- لا نجاح بلا raw evidence وnegative runs.
- لا عمومية بلا transfer envelope.
- لا self-improvement بلا evaluator independence وrollback.
- لا canonical promotion بلا MUG verdict وlineage writeback.

Difference → Question → Mechanism → Action → Readback → Attribution → Evidence → Transfer → Reproduction → Canonical Revision

بهذا يُغلق المسار النظري في معنى محدد: ليس لأن كل الأسئلة حُلّت، بل لأن كل نظرية رئيسية أصبحت تعرف ما الذي يجب أن يحدث كي تكتسب أو تفقد القوة العلمية.