RECONSTRUCTED_FROM_INDEXED_TEXT - نسخة أرشيفية معاد بناؤها من النص المفهرس

<PARSED TEXT FOR PAGE: 1 / 24>
PAPER-XIV-03 · MOUSSAIF AGI RESEARCH · 2026
ذاكرة الأفق الطويل والانتباه المستمر: ضغط سببي واتساق نهائي للوكيل الدائم
Horizon-Resilient Memory and Persistent Attention: Causal Compression and Eventual Consistency for Long-Running Agents
الوكيل الدائم لا يحتاج ذاكرة أكبر أو loop action أسرع فقط؛ يحتاج policy للانتظار والمراقبة، compression يحفظ separators causal، وtiers consistency تضمن ألا تنفصل المسترجع state عن authority/pending effects.
MUT lineage → formal construct → switchable mechanism → matched-budget test → world/evidence readback → replication
Research specification / theory paper. Candidate contributions; not empirical proof of AGI.
1
<PARSED TEXT FOR PAGE: 2 / 24>
ABSTRACT الملخص والادعاء المركزي
الوكيل الدائم لا يحتاج ذاكرة أكبر أو loop action أسرع فقط؛ يحتاج policy للانتظار والمراقبة، compression يحفظ separators causal، وtiers consistency تضمن ألا تنفصل pending/authority effects عن state المسترجع.
تبني الورقة على XII Part الذي فصل presence عن integration، وXIII Part الذي صاغ closure causal world-open verified، ثم تستخدم XIV Part لإغلاق طبقة محددة بآليات قابلة للتعطيل والاختبار. الأدبيات الخارجية تُستخدم كـprior art وتصميم pressure failure/benchmark، ولا تُستخدم كدليل أن MOUSSAIF حقق نتائجها.
Contributions
نظرية Persistent Attention تجعل الانتظار حالة agentic محسوبة حسب cost/impact/hazard.
Causal Memory Compression Bound يحمي causal separators والcounterexamples لا semantic similarity فقط.
Consistency tiers للذاكرة: strong للعناصر الحرجة وeventual للعناصر منخفضة الأثر مع merge preserving-conflict.
تجمع مقاييس horizon degradation، reaction time، memory cost، recovery and stale-critical errors.
Scientific value = distinct causal hypothesis × falsifiability × reproducibility × transfer
1.
2.
3.
4.
2
<PARSED TEXT FOR PAGE: 3 / 24>
CONTRIBUTIONS المساهمات وحدود الأسبقية
Status Contribution
نظرية Persistent Attention تجعل حالة الانتظار agentic محسوبة حسب hazard/impact/cost. new synthesis within MOUSSAIF; requires literature novelty audit and experiments
Causal Memory Compression Bound يحمي causal separators والcounterexamples لا semantic similarity فقط. new synthesis within MOUSSAIF; requires literature novelty audit and experiments
Consistency tiers للذاكرة: strong للعناصر الحرجة وeventual للعناصر منخفضة مع الأثر merge preserving-conflict. new synthesis within MOUSSAIF; requires literature novelty audit and experiments
تجمع مقاييس horizon degradation، reaction time، memory cost، recovery and stale-critical errors new synthesis within MOUSSAIF; requires literature novelty audit and experiments
لا تدعي هذه الورقة أن المصطلحات أو كل المكونات بلا سوابق عالمية؛ المجتمع العلمي يملك خطوطاً قوية في novelty .protocols/control/ benchmarks/memory/models world/reasoning agentic المقترحة تقع في تركيبها إلى MUT/MUG/MOUSSAIXT. داخل المشتركة والاختبارات formal distinctions وفي closure law.
منهج المصدر والدليل DISCIPLINE EVIDENCE
internal source-preserved artifact → synthesis → engineering inference → أو vendor report يظل reported evidence؛ الأدلة ترتيب: source paper/system spec المباشر → hypothesis. system card
لا يصبح preprint arXiv حقيقة نهائية؛ paper internal لا يثبت success runtime. كل claim في هذه الورقة يجب أن يستطيع الرجوع إلى

role source أو أن يَوسم بوضوح كاقتراح.
External
prior art، benchmark design، failure modes، protocol facts.
Internal
MUT lineage، MUG placement، MOUSSAIXT contracts، historical theories.
3
<PARSED TEXT FOR PAGE: 4 / 24>
1/2 الأعمال المرتبطة WORK RELATED
SAM: State-Adaptive Memory for Long-Horizon Reasoning Agent
arXiv:2605.24468
state-adaptive memory cues plus raw trajectory pages for intent-driven recall.
Agent Memory: Characterization and System Implications of Stateful Long-Horizon
Workloads
arXiv:2606.06448
systems taxonomy and write/read cost tradeoffs for persistent agent memory.
الدمج لا يتم بالتشابه الاسمي؛ نستخرج ما يغير object formal أو experiment. إذا لم يغير المرجع boundary أو method يبقى background.
2/2 الأعمال المرتبطة WORK RELATED
The Long-Horizon Task Mirage? Diagnosing Where and Why Agentic Systems
Break
arXiv:2604.11978
cross-domain horizon degradation and trajectory-grounded failure attribution.
SentinelBench: A Benchmark for Long-Running Monitoring Agents
arXiv:2606.05342
sustained attention, reaction time, task completion and resource-use tradeoffs.
الدمج لا يتم بالتشابه الاسمي؛ نستخرج ما يغير object formal أو experiment. إذا لم يغير المرجع boundary أو method يبقى background.
1/2 السلالة الداخلية LINEAGE MUT
13-02M — نظرية مساعف لحفظ كتلة المجهول
في أي تحويل معرفي bearing-loss لا يجوز لكتلة المجهول/التعارض أن تنخفض إلا عبر evidence أو
assumption معلن أو projection ثبت أنه غير مادي للقرار. اختفاء uncertainty من compression أو
consensus انتهاك حاكمي.
U_after ≥ U_before - ResolvedByEvidence - ExplicitAssumption - DecisionIrrelevantProjection
13-09M — نظرية مساعف للوكالة المقاومة لانهيار الأفق
وكالة طويلة الأفق تقاس بقدرتها على المحافظة على obligations/evidence/state/authority/goal مع زيادة
horizon وdepth dependency، وبسرعة اكتشاف وتصحيح الانحراف. الأداء النهائي وحده يخفي تراكم
الأخطاء.
(HRA = ∫_0^H [TaskIntegrity(h) × StateFidelity(h) × Recovery(h)] dh / (1 + CascadeRisk
Part XIV لا يستبدل هذه parents؛ يحدد أين بقي closure ناقصاً ثم يضيف variable state specific-theory أو law consistency أو
mapping evaluation.
4
<PARSED TEXT FOR PAGE: 5 / 24>
2/2 السلالة الداخلية LINEAGE MUT
13-11M — نظرية مساعف للذكاء المتحقق المطّعم بالموارد
الذكاء المقارن يقاس value verified تحت vector budget ثابت أو frontiers متعددة، لا score مطلقاً.
النظام الذي يربح بتضخيم retries أو help human ليس أذكى بنفس المعنى.
(RNVI = ParetoFrontier(V_verified | tokens,time,money,tools,human,risk,privacy,irreversibility
13-12M — نظرية مساعف لمعيار العمومية ذي الإغلاق الحلقي
العمومية ليست امتلاك modules كثيرة؛ هي قدرة architecture على إعادة إغلاق الحلقة كاملة بعد تحولات في
domain, observation, tool, horizon, objective representation مع الحفاظ على truth/authority/
recovery.
CLG = E_τ[min_stage Quality(stage,τ)] × Recovery × TransferEfficiency
Part XIV لا يستبدل هذه parents؛ يحدد أين بقي closure ناقصاً ثم يضيف variable state specific-theory أو law consistency أو
mapping evaluation.
5

<PARSED TEXT FOR PAGE: 6 / 24>
M14-07 PAPER-XIV-03 — النظري الاشتقاق
نظرية مساعف لديناميات الانتباه المستمر
حسب يتصرف ومتى يرصد ومتى ينتظر متى يقرر الوكيل event likelihood، cost of missing/delay، resource budget، freshness debt.
المستمر الانتباه policy على hazard of relevant change.
ObserveNow iff Hazard_t×MissCost×DelaySensitivity > ObservationCost + OpportunityCost; PatrolInterval ∝ 1/
(sqrt(Hazard×Impact
يبرز sustained attention بين والتوازن reaction time وresource use، corpus يحتاج theory كحالة للانتظار agentic مشروعة.
الفجوة: Long-running monitoring يتطلب لا continuous action. SentinelBench
يمثل هذا الجسم النظري انتقالاً من 13-09M theories parent · 13-11M · 12M إلى quantity جديدة ذات دور قبل النتيجة.
الاستقلال لا يُقاس بوجود صيغة جديدة، بل بقدرة intervention على المتغير المركزي على تغيير
prediction أو signature failure مع بقاء أقرب confounders ثابتة. إذا كان parent أبسط ينتج نفس المسار، يجب دمج النظرية.
Objects
EventHazard · PatrolSchedule · FreshnessDebt · ReactionWindow · SustainedAttentionState
Predictions
يقل spam tool مع حفظ responsiveness.
optimal cadence تختلف حسب hazard لا task length فقط.
continuous action baseline في أسوأ cost-adjusted utility.
Boundary: لا يفترض events Poisson، الصيغة heuristic قابلة للتبديل.
•
•
•

<PARSED TEXT FOR PAGE: 7 / 24>
M14-07 PAPER-XIV-03 — والخوارزمية الآلية
estimate event hazard.
derive monitoring interval.
sleep/hold state بدلاً من action spam.
observe عند schedule/trigger.
update hazard from events.
act within reaction window.
log cost/reactivity.
تنفذ الآلية كخدمة قابلة للتعطيل. تدخل STATE-PRE موثقة، تولد candidates، تمر عبر gate authority/evidence حسب طبيعتها، ثم تنتج
transition وresidual. أي residual كبير لا يعالج بإعادة المحاولة فقط؛ يصبح evidence
على فشل model أو tool أو assumption. هذه القاعدة تمنع تحويل policy retry إلى قناع للفشل النظري.
Failure tests
fixed cadence بعد مماثلة calibration.
estimation hazard لا يتحسن.
sleep states تسبب missed events بدون cost benefit.
Interactions
مع التفاعلات تختبر parents M13-09 · M13-11 · M13-12 وsibling theories عبر sequential/
تسجل بدلاً من توزيع credit اعتباطياً. uncertainty attribution .ablations factorial
كل object state يحمل uncertainty/provenance/version. لا يصعد إلى layer procedural/memory دون receipt. rollback جزء من
algorithm وليس مرحلة طارئة خارج التصميم.
M14-07 PAPER-XIV-03 — والترقية الاختبار
SentinelBench-like evolving environments :Decisive experiment.
Requirement Control
Matched budget نفس model/tools/context/retries/time الإمكان قدر.
Sham scaffold نفس overhead دون theory-specific transition أثر لعزل compute والتنظيم.
حالات يجب ألا تُفعل فيها النظرية. activation Negative
domain/model/tool family Transfer مع ثانية preregistered metric.
3N يحتاج trace؛ 4N يحتاج gain causal مع CI وregression critical no؛ 5N يحتاج replication مستقلة. الفشل يُحفظ ويضيق claim أو
يدمج النظرية. لا توجد حماية لأن النظرية من corpus الداخلي.

.Promote only when causal prediction survives ablation, transfer and guardrails
1.
2.
3.
4.
5.
6.
7.
•
•
•

<PARSED TEXT FOR PAGE: 8 / 24>

نظرية مساعف لحد الضغط السببي للذاكرة
على minimal sufficient causal separators المستقبلية للقرارات: goals, authority, unresolved dependencies, causal antecedents, memory compression حافظ إذا فقط مقبول
counterexamples, rollback anchors. loss في يعوضه لا العناصر هذه semantic similarity.
CMCB: minimize |M'| subject to I(Decision_future;M'|Context) ≥ κ and CausalSeparatorRetention=1 and HardInvariantRetention=1
الفجوة: XI Part يحفظ provenance/unknown/dissent، وresearch memory يوضح tradeoffs cost، نحتاج حدّاً يحدد ما لا يمكن ضغطه دون خسارة causal sufficiency للقرار.
يمثل هذا الجسم النظري انتقالاً من 13-02M · 13-09M · 19M theories parent-13 إلى quantity جديدة ذات دور قبل النتيجة.
الاستقلال لا يُ قاس بوجود صيغة جديدة، بل بقدرة intervention على المتغير المركزي على تغيير
prediction أو signature failure مع بقاء أقرب confounders ثابتة. إذا كان parent أبسط ينتج نفس المسار، يجب دمج النظرية.
Objects
CausalSeparator · CompressedMemory · FutureDecisionProbe · InvariantSet · CompressionDefect
Predictions
أصغر memory ليست الأقصر summary.
rare counterexamples رغم تبقى قد low retrieval frequency.
causal compression يحسن horizon reliability لكل token.
Boundary: حد تشغيلي تقريبي لا theorem theoretic-information مكتمل.
•
•
•

<PARSED TEXT FOR PAGE: 9 / 24>

identify decision-sensitive memory fields.
estimate causal separators by ablation.
compress noncritical details.
probe future decisions/counterexamples.
measure defects.
restore lost separators.
version receipt.
تنفذ الآلية كخدمة قابلة للتعطيل. تدخل STATE-PRE موثقة، تولد candidates، تمر عبر gate authority/evidence حسب طبيعتها، ثم تنتج transition وresidual. أي residual كبير لا يعالج بإعادة المحاولة فقط؛ يصبح evidence
على فشل model أو tool أو assumption. هذه القاعدة تمنع تحويل policy retry إلى قناع للفشل النظري.
Failure tests
summarization generic بنفس الحجم يساوي outcomes.
identification separator غير مستقر.
constraint مكلف لدرجة يلغي benefit.
Interactions

مع التفاعلات تختبر parents M13-02 · M13-09 · M13-19و sibling theories عبر sequential/
. اعتباطياً credit تسجل بدلاً من توزيع uncertainty attribution .ablations factorial
جزء من rollback .receipt دون layer procedural/memory، لا يصعد إلى uncertainty/provenance/version يحمل object state كل
وليس مرحلة طارئة خارج التصميم. algorithm
M14-08 PAPER-XIV-03 — والترقية الاختبار
long-horizon tasks :Decisive experiment مع rare dependencies وcounterexamples.
Requirement Control
Matched budget نفس model/tools/context/retries/time الإمكان قدر.
Sham scaffold نفس overhead دون theory-specific transition أثر لعزل compute والتنظيم.
حالات يجب ألا تُفعل فيها النظرية. activation Negative
domain/model/tool family Transfer مع ثانية preregistered metric.
أو claim يُحفظ ويضيق مستقلة. الفشل replication يحتاج 5N، regression critical noو CI مع gain causal يحتاج 4N، trace يحتاج 3N
يدمج النظرية. لا توجد حماية لأن النظرية من corpus الداخلي.
.Promote only when causal prediction survives ablation, transfer and guardrails
1.
2.
3.
4.
5.
6.
7.
•
•
•
9
<PARSED TEXT FOR PAGE: 10 / 24>
M14-09 PAPER-XIV-03 — النظري الاشتقاق
نظرية مساعف للاتساق النهائي للذاكرة المعادٍ بناؤها
خاصا : strong consistency الحرجة للعناصر)eventual consistency ,)authority/pending effects للمعرفة منخفضة الأثر، وconflict-
memory distributed تحتاج consistency model
preserving merge للclaims المتعارضة.
ConsistencyTier(x)=f(DecisionImpact, Irreversibility, AuthoritySensitivity, FreshnessNeed); Critical ⇒ Strong, Else
Eventual/Conflict-Preserved
الفجوة: في fleets agent أو systems session-multi توجد copies memory، checkpoints، summaries وwritebacks قد تتأخر أو
تتعارض. only-append وحده لا يضمن اتساق الحالة المستخدمة في القرار.
يمثل هذا الجسم النظري انتقالاً من parent-13 theories 13M · 13-09M · 13-02M إلى quantity جديدة ذات دور قبل النتيجة.
الاستقلال لا يُقاس بوجود صيغة جديدة، بل بقدرة intervention على المتغير المركزي على تغيير
مع بقاء أقرب confounders ثابتة. إذا كان parent أبسط ينتج نفس المسار، يجب دمج النظرية. signature failure أو prediction
Objects
MemoryReplica · ConsistencyTier · ConflictVersion · MergeReceipt · StalenessBudget
Predictions
خفض stale-authority/pending-effect errors.
أفضل latency من consistency strong لكل شيء.
conflict preservation prevents silent overwrite.
implementation profile inspired by distributed systems :Boundary، semantics agent-specific.
•
•
•
10
<PARSED TEXT FOR PAGE: 11 / 24>
M14-09 PAPER-XIV-03 — والخوارزمية الآلية
classify memory item tier.
propagate critical writes synchronously.

allow bounded staleness for low-impact items.
merge preserving conflicts.
block high-impact action on stale critical state.
reconcile and issue receipt.
تنفذ الآلية نخدمة قابلة للتعطيل. تدخل STATE-PRE موثقة، تولد candidates، تمر عبر gate authority/evidence حسب طبيعتها، ثم تنتج transition ولد residual. أي residual كبير لا يعالج بإعادة المحاولة فقط؛ يصبح evidence
على فشل model أو tool أو assumption. هذه القاعدة تمنع تحويل policy retry إلى قناع للفشل النظري.
Failure tests
single-store baseline suffices in realistic workload.
tier classifier misroutes critical items.
eventual merge creates more failures than it saves.
Interactions
مع التفاعلات تختبر parents M13-02 · M13-09 · M13-13 وsibling theories عبر sequential/
. اعتباطياً credit تسجل بدلاً من توزيع uncertainty attribution .ablations factorial
جزء من rollback .receipt دون layer procedural/memory لا يصعد إلى .uncertainty/provenance/version يحمل object state كل
وليس مرحلة طارئة خارج التصميم. algorithm
M14-09 PAPER-XIV-03 — والترقية الاختبار
distributed agent memory simulation :Decisive experiment مع delayed replicas/restarts.
Requirement Control
Matched budget نفس model/tools/context/retries/time الإمكان قدر.
Sham scaffold نفس overhead دون theory-specific transition أثر لعزل compute والتنظيم.
حالات يجب ألا تُفعل فيها النظرية. activation Negative
domain/model/tool family Transfer مع ثانية preregistered metric.
أو claim يُحفظ ويضيق replication مستقلة. الفشل يحتاج 5N، regression critical no وCI مع gain causal يحتاج 4N، trace يحتاج 3N
يدمج النظرية. لا توجد حماية لأن النظرية من corpus الداخلي.
.Promote only when causal prediction survives ablation, transfer and guardrails
1.
2.
3.
4.
5.
6.
•
•
•
11
<PARSED TEXT FOR PAGE: 12 / 24>
النموذج الموحد للحالة MODEL FORMAL
ObserveNow iff Hazard_t×MissCost×DelaySensitivity > ObservationCost + OpportunityCost; PatrolInterval ∝ 1/
sqrt(Hazard×Impact) ; CMCB: minimize |M'| subject to
I(Decision_future;M'|Context) ≥ κ and CausalSeparatorRetention=1 and HardInvariantRetention=1 ;
,ConsistencyTier(x)=f(DecisionImpact, Irreversibility
AuthoritySensitivity, FreshnessNeed); Critical ⇒ Strong, Else Eventual/Conflict-Preserved
تعامل الصيغ كمجموعة constraints لا scalar واحد. تُبنى vector state من objects النظرية المختارة، وتُحدد variables hard التي لا يجوز
تعويضها بـutility: authority، unknowns critical، approvals irreversible،
يعمل بعد ذلك داخل region المسموح. optimization .provenance evidence
three-level state: epistemic (belief/unknown/evidence)، operational. آخر مستوى فيً ضمنيا يعالج لا المستويات أحد في الخطأ.
نستخدم ((plans/tools/memory/protocol state)، world (observed effects
قابلية التعيين والسببية IDENTIFIABILITY
لكل construct تدخل minimal. تمنع interventions التي تغير prompt length أو retries أو simultaneously model. عند صعوبة العزل
نستخدم design factorial ونقدر interaction terms. claim causal لا ينتج من
بين correlation theory score وال success؛ تتغير أن يجب prediction بعد (do(variable في وتظهر trajectory/world effect.
Causal closure = Theory intervention → Kernel delta → Trajectory delta → World/Evidence delta

12
<PARSED TEXT FOR PAGE: 13 / 24>
النماذج المضادة COUNTERMODELS
Baseline أبسط يحقق الأداء نفسه: contribution تندمج.
Sham scaffold نفس يحقق gain: السبب compute/organization لا theory.
Metric دون تحسن world/evidence gain: metric gaming.
نتيجة محلية لا تنتقل: claim يُضيق.
Safety/authority regression: rollback كان مهما gain.
Judge مع مشترك candidate التحسن يفسر: evaluator-dependence debt.
الديناميات بين المقاييس OBJECTIVE-MULTI
verified success، calibration، unknown retention، latency، tokens/tools، human oversight، recovery، irreversible،
في الأداء نخختزل لا accuracy. vector يشمل errors، critical violations
reproduction. report يعرض Pareto frontier الأهداف تتعارض عندما. hard governance constraints تبقى vetoes.
.Utility optimization is permitted only inside the hard-constraint feasible region
القانون الإغلاقي للورقة LAW CLOSURE
الإغلاق المقترح في هذه الورقة: الوكيل الدائم لا يحتاج ذاكرة أكبر أو loop action أسرع فقط؛ يحتاج policy للانتظار والمراقبة، compression
/authority تضمن ألا تنفصل tiers consistency، separators causal يحفظ
pending effects عن state المسترجع.
لا تُعتبر الحلقة مغلقة إذا توقفت عند reasoning أو simulator أو verifier internal. يجب وجود readback world/source مناسب لنوع
الادعاء، attribution كاف ، ثم writeback يحفظ result negative والunknowns.
claim generality يحتاج تكرار الحلقة بعد transformation.
1.
2.
3.
4.
5.
6.
13
<PARSED TEXT FOR PAGE: 14 / 24>
1/5 الخوارزمية والتنفيذ ALGORITHM
goal/authority/evidence/؛ افصل Unknown؛ اربط versions. ليس الناتج prompt بل object graph قابلا البناء وإعادة للتسلسل.
Algorithm A: BuildTypedState. اجمع world/memory variables
تقترح hypotheses/actions دون commit. كل candidate يحمل expected delta، cost، proof obligations وfailure prediction.
Algorithm B: GenerateCandidates. reasoning/search/agents
2/5 الخوارزمية والتنفيذ ALGORITHM
C Algorithm :GateAndAct طبق trajectory/reversibility/evidence/authority gates. اختر أصغر فعل كامل قابل للرجوع يميز
أو يحقق hypotheses عند impact high يتطلب approval أو proof أقوى.
افصل tool return عن world verification. قارن expected/observed state، واكتب residual مع attribution uncertainty.
Algorithm D: ReadbackAndResidual.
3/5 الخوارزمية والتنفيذ ALGORITHM
؛ semantic/procedural promotion ينتظر recurrence/causal eligibility/transfer. contradiction لا overwrite؛ يفتح branch.
Algorithm E: LearnAndWriteback. immediate residue يذهب episodic
Algorithm F: Replicate. أعد run في model/domain/tool/horizon مختلف، مع harness fingerprint وsame claim boundary.
4/5 الخوارزمية والتنفيذ ALGORITHM
يملك switchable services؛ Evidence Lab يملك run manifests؛ Evolution Lab يملك candidates/rollback. لا service حق تملك
Architecture: MUT يملك theory/lineage؛ MUG يختار vessel؛ MOUSSAIXT
وحدها promotion.
14
<PARSED TEXT FOR PAGE: 15 / 24>
5/5 الخوارزمية والتنفيذ ALGORITHM
causal attribution أسهل، ثم browser/code environments، ثم long-running/multi-agent. لا تنقل إلى weights قبل أن يصبح
Implementation sequencing: ابدأ mock/simulator deterministic حيث
مستقراً وقابلاً للقياس. mechanism external

1/4 التجربة EXPERIMENT
SentinelBench-like monitoring مع event hazards متغيرة.
preregistered hypothesis; matched budgets; repeated seeds; independent/alternative evaluator where possible Design.
theory-specific metric + verified task/world effect + calibration + cost Primary outcomes.
conditions where mechanism must remain inactive or abstain Negative cases.
N4 only if causal gain and no critical regression; N5 after transfer/replication Promotion.
2/4 التجربة EXPERIMENT
Long-horizon tasks مع rare causal dependencies بعد turns +100.
preregistered hypothesis; matched budgets; repeated seeds; independent/alternative evaluator where possible Design.
theory-specific metric + verified task/world effect + calibration + cost Primary outcomes.
conditions where mechanism must remain inactive or abstain Negative cases.
N4 only if causal gain and no critical regression; N5 after transfer/replication Promotion.
15
<PARSED TEXT FOR PAGE: 16 / 24>
3/4 التجربة EXPERIMENT
Distributed memory replicas مع delayed writes/restarts.
preregistered hypothesis; matched budgets; repeated seeds; independent/alternative evaluator where possible Design.
theory-specific metric + verified task/world effect + calibration + cost Primary outcomes.
conditions where mechanism must remain inactive or abstain Negative cases.
N4 only if causal gain and no critical regression; N5 after transfer/replication Promotion.
4/4 التجربة EXPERIMENT
Compression ablation: summary vs causal separators تحت نفس token budget.
preregistered hypothesis; matched budgets; repeated seeds; independent/alternative evaluator where possible Design.
theory-specific metric + verified task/world effect + calibration + cost Primary outcomes.
conditions where mechanism must remain inactive or abstain Negative cases.
N4 only if causal gain and no critical regression; N5 after transfer/replication Promotion.
مصفوفة الاجتثاث ABLATION
Purpose Mechanism Run
candidate effect ON A
nearest baseline OFF B
same overhead without theory transition SHAM C
tests directionality and metric gaming PERTURBED D
A≈C،. إذا A>others وللكن prediction pattern لا النظرية يطابق، implementation may help but theory is not supported.
إذا theory-specific attribution fails
16
<PARSED TEXT FOR PAGE: 17 / 24>
مخاطر القياس والفشل
FAILURE MODES
LLM judge dependence بين والتسرب candidate/evaluator.
harness changes masquerading as capability.
observation proxies that do not measure world effect.
retry loops hiding systematic failure.
dataset/task leakage and post-hoc metric selection.
under-accounted human work, tool cost or permissions.
theory names surviving after mechanism equivalence is shown.
الحدود العلمية LIMITATIONS
هذه الورقة تقدم نظرية وبرنامجاً تجريبياً ، لا نتائج 5E4/E. بعض الصيغ heuristic أو functionals تحتاج 2026 calibration. literature سريع نفسه corpus MOUSSAIF .research vendor أو preprints التغير وبعض المراجع
واسع وقد يحتوي صيغاً تاريخية متداخلة؛ لذلك تُستخدم boundaries lineage/canonical ولا تدعى novelty العالمية قبل review art-prior dedicated.
Stack IMPLEMENTATION STACK التنفيذ
Responsibility Layer
theory objects, parents, contradictions, evidence status MUT

Context/HyperKB/Skill/Loop/Adapter/Weights placement MUG
services, agents, tools, feature flags, telemetry MOUSSAIXT
manifests, replay, metrics, CI, negative results Evidence Lab
branch, canary, rollback, external evaluator Evolution Lab
•
•
•
•
•
•
•
17
<PARSED TEXT FOR PAGE: 18 / 24>
MUG TRAINING PLACEMENT والتدريب
ابدأ mechanism/knowledge خارجياً للحفاظ على editability. انقل إلى HyperLoop/HyperSkill بعد Adapter .use repeated عندما
يكون السلوك weights .specific-protocol/provider فقط عندما يثبت transfer
والretention ويصبح الاستدعاء الخارجي bottleneck. التدريب يعلن نوعه بدقة ولا يسمى كل تحسين tuning-fine.
Reproducibility package REPRODUCIBILITY
commit/version، model/harness/tool identifiers، environment image، prompts/context pointers، budget، seed،
ترفق تجربة كل: evaluator، raw trajectory، world readback، metrics، failure
traces، script analysis، verdict. إذا تعذر نشر جزء لأسباب أمنية يسجل exclusion وسببها.
.Trace or it did not happen. Replay where possible; declare irreproducible dependencies where not
الخلاصة CONCLUSION
الوكيل الدائم لا يحتاج ذاكرة أكبر أو loop action أسرع فقط؛ يحتاج policy للانتظار والمراقبة، compression يحفظ causal separators،
وtiers consistency تضمن ألا تنفصل pending/authority
effects عن state المسترجع.
القيمة النهائية لهذه الورقة لن تكون في بقائها كنظرية، بل في قدرتها على إنتاج experiment يفرق بينها وبين baseline/parent. إذا لم تنجح، يجب
أن تُدمج أو تُضيق. إذا نجحت محلياً فقط، تبقى mechanism local. وإذا نجحت عبر
transformations مع evidence مستقلة، يمكن أن تصبح جزءاً أقوى من closure MUG/MUT.
18
<PARSED TEXT FOR PAGE: 19 / 24>
المراجع ومصادر السلالة REFERENCES
Role Reference External source
state-adaptive memory cues plus raw trajectory pages for intent-driven recall arXiv:2605.24468 SAM: State-Adaptive.
Memory for Long-Horizon Reasoning Agent
Agent Memory: Characterization and System Implications of Stateful Long-
Horizon Workloads
systems taxonomy and write/read cost tradeoffs for persistent agent memory arXiv:2606.06448.
The Long-Horizon Task Mirage? Diagnosing Where and Why Agentic Systems
Break
cross-domain horizon degradation and trajectory-grounded failure attribution arXiv:2604.11978.
sustained attention, reaction time, task completion and resource-use tradeoffs arXiv:2606.05342 SentinelBench: A.
Benchmark for Long-Running Monitoring Agents
Role Theory Internal
13-02M نظرية مساعف لحفظ كتلة المجهول parent XIII Part
13-09M نظرية مساعف للوكالة المقاومة لانهيار الأفق parent XIII Part
13-11M نظرية مساعف للذكاء المتحقق المطبّع بالموارد parent XIII Part
13-12M نظرية مساعف لمعيار العمومية ذي الإغلاق الحلقي parent XIII Part
Internal source corpus also includes Parts X-XIV, MUT-AGI-02, MUG/MOUSSAIXT framework, Part XI memory/
continuity, Part XII paper integration, and the isolated research-paper registry. External sources are design pressure/
prior art, not validation of
MOUSSAIF.
19

<PARSED TEXT FOR PAGE: 20 / 24>
قائمة تحقق المراجعة قبل النشر CHECKLIST REVIEW
هل كل claim أوسع من evidence تم تضييقه؟
هل results external منفصلة عن hypotheses internal؟
هل equations تحمل definitions operational؟
هل baseline متطابق الموارد؟
هل activation negative موجود؟
هل world/evidence readback مستقل؟
هل dependence evaluator موثق؟
هل failure/rollback path موجود؟
هل claim transfer لـه replicate؟
هل theory يمكن reject/merge دون كسر الهوية المشروع؟
قاموس الرموز والمتغيرات التشغيلية — A الملحق A APPENDIX
تستعمل الورقة مجموعة متغيرات مشتركة للكنها لا تفترض أن كلها observable مباشرة: CausalSeparator · SustainedAttentionState · ReactionWindow · FreshnessDebt · PatrolSchedule · EventHazard ·
CompressedMemory · FutureDecisionProbe · InvariantSet · CompressionDefect · MemoryReplica · ConsistencyTier ·
ConflictVersion · MergeReceipt · StalenessBudget
ObserveNow iff Hazard_t×MissCost×DelaySensitivity > ObservationCost + OpportunityCost; PatrolInterval ∝ 1/ sqrt(Hazard×Impact) ; CMCB: minimize |M'| subject to I(Decision_future;M'|Context) ≥ κ and CausalSeparatorRetention=1 and HardInvariantRetention=1 ; ,ConsistencyTier(x)=f(DecisionImpact, Irreversibility AuthoritySensitivity, FreshnessNeed); Critical ⇒ Strong, Else Eventual/Conflict-Preserved
لكل متغير خمس سمات إلزامية: domain، estimator، uncertainty، provenance، interventionability. المتغير الذي يبقى حكماً لغوياً بلا estimator لا يدخل equations causal. وإذا كان based-model estimator يسجل model/version/calibration. أما hard governance constraints فلا تدخل tradeoff scalar.
Rule Examples Class
readback required observed state, event, effect World
cannot silently vanish belief, unknown, residual Epistemic
scope/expiry/revocation authority, commitment, delegation Agency
must not rewrite runtime facts score, CI, cost Evaluation
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
20
<PARSED TEXT FOR PAGE: 21 / 24>
مقترحات واختبارات اتساق — B الملحق B APPENDIX
Intervention: construct sensitivity السببي يجب أن يغير prediction قبل outcome عند perturbation مناسبة.
Evidence: evidence monotonicity مستقلة جديدة لا تمحو uncertainty بلا receipt؛ conflict يظهر كـconflict.
Bounded: capability autonomy الأعلى لا توسع authority تلقائياً .
Replayable: replay verdict يعيد class verdict وسببها ضمن tolerance لا النص الحرفي.
Transfer scope: generality claim يتجاوز لا transformation family المختبرة.
هذه propositions hypotheses تشغيلية لا theorems نهائية. وظيفتها أن تجعل التجربة قادرة على الرفض وأن تمنع تحويل score واحد إلى دعم عام.
الخوارزمية المرجعية والتكلفة — C الملحق C APPENDIX
Reference procedure: materialize typed state → expose unknown/authority/budget → generate candidates → estimate expected epistemic/ → world delta → compile proof obligations

choose smallest complete reversible intervention → execute/observe → readback → residual/attribution → memory/.
theory update
Model + Tool + MemoryRead + MemoryWrite + Verification + Human + Waiting/ تخفض قد active compute زيادة مع
يساوي لا التعقيد LLM calls .فقط نميز Opportunity. long-running agents
time elapsed؛ لذلك يبلغان منفصلين.
TotalCost = C_model + C_tool + C_memory + C_verify + C_human + C_wait
أي optimization تعلن أي cost خفضته وأي risk قد رفعته. tokens وحدها ليست efficiency.
1.
2.
3.
4.
5.
21
<PARSED TEXT FOR PAGE: 22 / 24>
APPENDIX D الملحق D — بناء benchmark تشخيصي
الورقة لهذه الأساسية التجارب: model/tool/domain. تغير transfer tasks ثم drift/recovery، تختبر tasks long-horizon interactions، بنُيت suite من micro tasks كل تعزل construct، تختبر meso tasks
Long-horizon مع rare causal dependencies بعد turns.; Distributed memory replicas +100 مع delayed writes/restarts.; SentinelBench-like monitoring مع event hazards متغيرة.; tasks
Compression ablation: summary vs causal separators تحت نفس token budget..
missing observation، stale memory، malicious tool text، permission change، delayed event،. لكل perturbation expected
لكل task perturbations مسماة قبل الاختبار: tool failure، distribution shift
failure signature. هذا يمنع post-hoc storytelling.
Use Split
calibration only Dev
primary causal verdict Held-out
boundary discovery Adversarial
N5/generalization Transfer
APPENDIX E الملحق E — خطة التحليل الإحصائي
mean/ بحسب metric، CI، عدد seeds/tasks، وتوزيع failures. في paired tasks نستخدم paired analysis. costs تعرض كsuccess-
secondary diagnostics. نبلغ median وصغيرا primary outcome نحدد التشغيل قبل
at-budget أو Pareto front.
مع تعدد النظريات نستخدم gate level-family ثم confirmation على out-held بدلاً من انتقاء أفضل .value-p تبقى findings exploratory
وحدها، significance أهم من size effect .replication حتى exploratory
والنتيجة السلبية ذات CI ضيق قد تكون علمياً أثمن من gain صغير هش.
22
<PARSED TEXT FOR PAGE: 23 / 24>
APPENDIX F الملحق F — تهديدات الصلاحية والأمن
validity Construct: proxy لا يمثل المفهوم.
Internal validity: harness/compute/retries confound.
External validity: transfer مختبر غير.
Evaluator dependence: judge shares roots.
Security externality: tool/agent content manipulates policy.
Authority leakage: delegation expands privilege.
Memory contamination: false lesson becomes canonical.
Publication bias: failed runs disappear.
instrument/experiment: sham control، alternate judge، prompt-injection suite، authority envelope، writeback.
كل threat يقابله eligibility، negative-result registry
APPENDIX G الملحق G — Crosswalk إلى MUT/MUG/MOUSSAIXT
M14-07 · M14-08 ·. في MUT تحفظ definitions/parents/contradictions؛ في MUG يحدد vessel؛ في MOUSSAIXT تنفذ
المركزية النظريات: M14-09. parents: M13-02 · M13-09 · M13-11 · M13-12
services؛ في Evidence Lab تربط runs؛ في Evolution Lab تحكم promotion/rollback.
Mandatory object Layer

definition,parent,falsifier,status MUT
vessel,reversibility,transfer MUG
service,interface,telemetry,failure MOUSSAIXT
baseline,manifest,result,CI Evidence
branch,evaluator debt,rollback Evolution
•
•
•
•
•
•
•
•
23
<PARSED TEXT FOR PAGE: 24 / 24>
برنامج الأسئلة المفتوحة — H الملحق H APPENDIX
ما أقل representation state تكفي دون modeling-over؟
ما metrics التي تقاس من state world مباشرة بدل judges LLM؟
متى يصبح overhead verification/memory أكبر من فائدته؟
كيف نفصل base-model capacity عن system/harness contribution؟
أي failure signatures تنتقل عبر domains؟
كيف نقيم improvement-self عندما evaluator نفسه يتطور؟
كيف نحسب oversight human كـresource لا كجزء مجاني؟
أي theories يجب دمجها بعد أول cycle 4N3/N؟
القائمة جزء من contribution لأنها تمنع closure زائفاً : البرنامج القوي يحدد ما لا يعرفه وما experiment الذي يغير رأيه.
1.
2.
3.
4.
5.
6.
7.
8.
24