RECONSTRUCTED_FROM_INDEXED_TEXT - نسخة أرشيفية معاد بناؤها من النص المفهرس

<PARSED TEXT FOR PAGE: 1 / 24>
PAPER-XIV-01 · MOUSSAIF AGI RESEARCH · 2026
الذكاء الفعام ذو الإغلاق الحلقي: من الكفاية الرصدية إلى
نموذج الفعالم الموجّه للقرار
Closed-Loop General Intelligence: From Observational Sufficiency to Decision-Boundary World Models
ليست reasoning→answer، بل observation sufficiency→decision-relevant world model→diagnostic intervention→causal
المفتوح للذكاء الأساسية الحلقة أن نقترح
العمومية تبدأ عندما يعيد النظام إغلاق هذه الحلقة بعد تغير المجال والأداة. update verified→residual
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>
الملخص والادعاء المركزي ABSTRACT
observation sufficiency→decision-relevant world model→diagnostic intervention→causal residual→verified. تبدأ العمومية
ليست المفتوح للذكاء الأساسية الحلقة أن نقترح reasoning→answer، بل update
عندما يعيد النظام إغلاق هذه الحلقة بعد تغير المجال والأداة.
تبني الورقة على XII Part الذي فصل presence عن integration، وXIII Part الذي صاغ verified world-open causal closure، ثم
تستخدم XIV Part لإغلاق طبقة محددة بآليات قابلة للتعطيل والاختبار. الأدبيات الخارجية
ت ُستخدم كـart prior وتصميم pressure failure/benchmark، ولا ت ُستخدم كدليل أن MOSSAIF حقق نتائجها.
Contributions
تعريف Observational Sufficiency Frontier بدلاً من عدد مصادر أو confidence stop.
تعريف Boundary-Decision World Model الذي يتعلم dynamics ذات الصلة بالقرار لا observation next بالكامل.
تحويل residuals إلى graph سببي يولد تجارب وفروضاً بدلاً من patches منفصلة.
بروتوكول falsification يربط world-model accuracy بـplanning وworld-state readback.
Scientific value = distinct causal hypothesis × falsifiability × reproducibility × transfer
1.
2.
3.
4.
2

<PARSED TEXT FOR PAGE: 3 / 24>
المساهمات وحدود الأسبقية CONTRIBUTIONS
Status Contribution
new synthesis within MOUSSAIF; requires literature novelty audit and experiments .confidence stop أو مصادر عدد من ً
تعريف Observational Sufficiency Frontier بدلا
new synthesis within MOUSSAIF; requires literature novelty audit and experiments .بالكامل next observation لا بالقرار
تعريف Decision-Boundary World Model يتعلم الذي dynamics ذات الصلة
new synthesis within MOUSSAIF; requires literature novelty audit and experiments .منفصلة patches من ً بدلاً وفروضا
تحويل residuals إلى graph تجارب يولد سببي
new synthesis within MOUSSAIF; requires literature novelty audit and experiments .world-state readback وplanningبال
بروتوكول falsification يربط world-model accuracy
لا تدعي هذه الورقة أن المصطلحات أو كل المكونات بلا سوابق عالمية؛ المجتمع العلمي يملك خطوطاً قوية في /novelty .protocols/control
المقترحة تقع في تركيبها إلى benchmarks/memory/models world/reasoning agentic
closure law وفي formal distinctions داخل المشتركة والاختبارات MUT/MUG/MOSSAIXT.
منهج المصدر والدقيل 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 حقيقة نهائية؛ preprint arXiv لا يثبت success runtime. كل claim في هذه الورقة يجب أن يستطيع الرجوع إلى paper internal
أو أن ي ُوسم بوضوح كاقتراح. role source

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
Agentic Reasoning for Large Language Models
arXiv:2601.12538
foundational/self-evolving/collective agentic reasoning; open challenges include world models, long horizon, governance.
Self-Evolving World Models for LLM Agent Planning
arXiv:2606.30639
episodic/semantic memory and prediction-observation mismatch for test-time world-model revision.
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
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.
Long-Horizon Agent Trajectory Attribution
arXiv:2608.06909
fine-grained attribution chains and component-level leave-one-out perturbation.
الدمج لا يتم بالتشابه الاسمي؛ نستخرج ما يغير object formal أو experiment. إذا لم يغير المرجع boundary أو method يبقى background.
1/3 افسلالة الداخلية LINEAGE MUT
نظرية مساعف للعمل الإبستيمي المتحقق — 13-01M
العمل الإبستيمي هو مقدار التحول السببي القابل للتحقق في حالة المعرفة/القرار/العالم الناتج عن موارد معرفية.
الاستدلال الذي لا يغير calibration أو quality decision أو state-world verified لا يحسب عملاً إبستيمياً مهما طال.
VEW = (ΔCalibration + α·ΔDecision + β·ΔWorld + γ·DiagnosticGain) × Reproducibility × Transfer - Cost - Irreversibility
نظرية مساعف لحفظ كتلة المجهول — 13-02M
لا يجوز لكتلة المجهول/التعارض أن تنخفض إلا عبر evidence أو bearing-loss في أي تحويل معرفي
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
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>
2/3 افسلالة الداخلية LINEAGE MUT
نظرية مساعف للتحكم الراجع بالسؤال الحاكم — 13-03M
السؤال الحاكم input control يختار المعلومة التي تجعل النظام أكثر قدرة على التمييز بين سياسات أو نماذج عالم متنافسة. قيمة السؤال مقدار تغيير boundary decision/observability الذي يسببه، لا بلاغته.
(QValue(q)=E[DecisionBoundaryShift + HypothesisSeparation + UnknownReduction | q] / (Cost + Delay + Risk
نظرية مساعف تتعلم الفعالم من بقايا التنبؤ والملاحظة — 13-05M
العالم يُتعلم من residual بين الحالة المتوقعة والملاحظة بعد فعل معروف، مع فصل خطأ model world عن observation noise. البقايا unit واكتشاف تعلم unknown dynamics.
وعن tool execution
r_t = D_state(Ŵ_{t+1}, W_{t+1}); ΔH ∝ Attribution(r_t) × Recurrence × DecisionImpact
لا يستبدل هذه parents، يحدد أين بقي closure ناقصاً ثم يضيف variable state specific-theory أو law consistency أو mapping evaluation.
3/3 افسلالة الداخلية LINEAGE MUT

13-08M — نظرية مساعف للفعل اقتشخيصي ونقض الذات
الوكالة المتقدمة تخصص budget لأفعال ذات قيمة تشخيصية تستطيع إبطال hypothesis top أو plan top
مبكراً. disconfirmation ليس نقداً لغوياً بل فعل سببي يزيد قابلية التمييز بين نماذج العالم.
DValue(a)=E[PosteriorSeparation + FailureAvoidance + ModelRevision] - ActionCost - DelayCost
13-19M — نظرية مساعف للإغلاق اقسببي بين اقنظرية واقنواة
الإغلاق السببي يتحقق عندما يؤدي تدخل معرف في construct النظري إلى تغيير measurable في kernel
mechanism ثم trajectory ثم outcome evidence/world، مع attribution قابلة لإعادة التشغيل. غياب
أي حلقة يكسر claim التفسيري.
ClosureStrength = Causal(T→K) × Causal(K→τ) × Causal(τ→W) × EvidenceReproducibility
XIV Part 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>
M14-01 PAPER-XIV-01 — اقنظري الاشتقاق
نظرية مساعف لجبهة اقلكفاية الرصدية
لكل قرار/claim جبهة كفاية رصدية OSF تربط جودة القرار بحزمة الملاحظات واستقلالها وحداثتها وتغطيتها للمتغيرات السببية. التوقف عن جمع
البيانات مشروع فقط عندما تصبح المكاسب الهامشية في القرار/
التنبؤ أقل من التكلفة مع بقاء gates uncertainty hard.
OSF = ParetoFrontier(DecisionValidity, CausalCoverage, Independence, Freshness | ObservationCost); stop iff
ΔExpectedDecisionValue < Cost and HardUnknowns=0
اقفجوة: الحفاظ على المجهول واختيار السؤال وتعلم العالم من residuals لا يجيبان عن: متى نملك observation كافية فعلاً لاتخاذ قرار أو ترقية
claim، ومتى تكون المعلومات الإضافية ذات قيمة هامشية منخفضة؟
يمثل هذا الجسم النظري انتقالاً من 13-02M theories parent-13 · 13-03M · 05M 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
ObservationSet · CausalVariableCoverage · HardUnknown · MarginalObservationValue · StoppingReceipt
Predictions
يقلل over-browsing وpremature closure معاً.
الأسئلة عالية الأثر تحتاج sets observation أوسع أو أكثر استقلالاً.
عدد الملاحظات وحده predictor ضعيف للكفاية.
Boundary: جبهة كفاية لا تعني حقيقة مطلقة؛ هي scoped task/claim.
•
•
•
6
<PARSED TEXT FOR PAGE: 7 / 24>
M14-01 PAPER-XIV-01 — والخوارزمية الآقية
حدد decision-critical variables.
اربط كل observation بالمتغيرات التي يغطيها وبمصدرها.
قدر الاستقلال والحداثة.
احسب القيمة الهامشية لملاحظة جديدة.
توقف فقط إذا لم يبق HardUnknown وإذا انخفض value marginal.
أصدر sufficiency receipt أو HOLD.
تنفذ الآلية نكدمة 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
count evidence fixed يؤدي نفس القرارات بنفس التكلفة.
OSF يتنبأ لا unsupported certainty.
hard unknown labeling يحسن لا safety/accuracy.
Interactions
مع التفاعلات تختبر parents M13-02 · M13-03 · M13-05وsibling 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.
M14-01 PAPER-XIV-01 — واقترقية الاختبار
قارن مخفية؛ سببية fixed-k retrieval وconfidence stop وOSF على cost، calibration، decision validity، missed critical evidence.
Decisive experiment: بمتغيرات تصفح/بحث مهام
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 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.
•
•
•
7
<PARSED TEXT FOR PAGE: 8 / 24>
M14-02 PAPER-XIV-01 — اقنظري الاشتقاق
نظرية مساعف لنموذج اقعالم الموجّه لحدود اققرار
model world مفيد بقدر ما يتنبأ بالمتغيرات التي تفصل بين actions/policies المتنافسة. هدفه ليس reconstruction شامل للobservation بل تقليل uncertainty حول boundary decision تحت intervention.
DBWM-Loss = λ1·DecisionVariableError + λ2·CounterfactualTransitionError + λ3·Calibration - λ4·IrrelevantDetailFit
اقفجوة: نماذج العالم قد تحسن prediction state-next للكنها تنفق capacity على تفاصيل لا تغير القرار. أبحاث 2026 حول modeling world تضغط نحو supervision موجهة لحاجة oriented-decision/authored-agent policy.
يمثل هذا الجسم النظري انتقالاً من 13-03M theories parent · 13-05M · 08M 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
DecisionBoundary · PolicySet · DecisionVariable · CounterfactualTransition · SelectiveWorldState
Predictions
يمكن لنموذج أقل دقة نصياً أن يحسن planning أكثر.
selective world modeling يخفض compute.
فشل model يترك signature أوضح لأنه مربوط بvariables decision.
Boundary: لا يزعم model-world كامل؛ هو model وظيفي موجه للقرار.
•
•
•
8
<PARSED TEXT FOR PAGE: 9 / 24>
M14-02 PAPER-XIV-01 — والخوارزمية الآقية
استخرج السياسات المرشحة.
حدد variables world التي تغير ترتيبها.

.variables موجهة لهذه evidence transition اجمع
درب/حدّث نموذجاً انتقائياً .
اختبر counterfactual action outcomes.
استخدم foresight فقط إذا معاير.
تنفذ الآلية نخدمة قابلة للتعطيل. تدخل STATE-PRE موثقة، تولد candidates، تمر عبر gate authority/evidence حسب طبيعتها، ثم تنتج residual وtransition. 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
.planning بنفس الحجم يتفوق في model observation-next
.trajectories لا تستقر عبر variables decision
.selectivity يحذف dynamics حرجة غير متوقعة
Interactions
مع التفاعلات تختبر parents M13-03 · M13-05 · M13-08وsibling theories عبر sequential/
. تسجل بدلاً من توزيع credit اعتباطياً 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
M14-02 PAPER-XIV-01 — واقترقية الاختبار
ALFWorld/ScienceWorld/tool simulators؛ قارن next-observation، generic state prediction، DBWM نفس تحت compute.
: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.
•
•
•
9
<PARSED TEXT FOR PAGE: 10 / 24>
M14-03 PAPER-XIV-01 — اقنظري الاشتقاق
البقايا ليست errors مستقلة؛ تبني Graph Causality Residual حيث العقد أنماط انحراف، والحواف hypotheses حول causes المشتركة. graph يوجه السؤال التالي والتجربة التالية ويمنع إصلاح الأعراض نظرية مساعف لرسم اقسببية من اقبقايا
منفصلة.
RCG-edge(i→j)=Recurrence×ConditionalDependence×InterventionConsistency×DecisionImpact - ConfoundingPenalty
اقفجوة: residual observation-Prediction يُستخدم للتحديث، للكن البقايا المتكررة عبر contexts قد تكشف structure causal hidden. نحتاج طريقة تربط clusters من residuals إلى mechanisms hypothesized وتجارب
فاصلة.
يمثل هذا الجسم النظري انتقالاً من parent-13 theories 13-05M · 13-19M · 17M 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
ResidualCluster · CandidateCause · CausalEdge · InterventionTest · RepairFamily
Predictions
يقل patch proliferation.

تنشأ اختبارات أعلى gain information.
يمكن التنبؤ بفشل جديد قبل وقوعه من neighborhood graph.
graph :Boundary حتى فرضي intervention، correlation edges لا تصبح causation.
•
•
•

<PARSED TEXT FOR PAGE: 11 / 24>

cluster typed residuals.
بال اربطها actions/context/tool states.
اقترح أسباباً مشتركة.
نفذ interventions صغيرة.
حدّث strengths edge.
اجمع repairs المتشابهة في family.
اكتب causal residue بعد replication.
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
clustering لا يزيد failure prediction.
edges causal غير مستقرة عبر environment.
simple frequency rules تحقق نفس repair efficiency.
Interactions
مع التفاعلات تختبر parents M13-05 · M13-19 · M13-17 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.

Decisive: بيئات agentic مع متعددة failure causes قارن متنكرة، مشتركة per-failure patches وRCG-guided interventions.
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.
أو claim يُحفظ ويضيق مستقلة. الفشل replication 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.
7.
•
•
•

<PARSED TEXT FOR PAGE: 12 / 24>
النموذج الموحد للحالة MODEL FORMAL
OSF = ParetoFrontier(DecisionValidity, CausalCoverage, Independence, Freshness | ObservationCost); stop iff

; ΔExpectedDecisionValue < Cost and HardUnknowns=0
DBWM-Loss = λ1·DecisionVariableError + λ2·CounterfactualTransitionError + λ3·Calibration - λ4·IrrelevantDetailFit ;
RCG⊛edge(i→j)=Recurrence×ConditionalDependence×InterventionConsistency×DecisionImpact - ConfoundingPenalty
تعامل الصيغ كمجموعة constraints لا scalar واحد. تُبنى vector state من objects النظرية المختارة، وتُحدد variables hard التي لا يجوز تعويضها ب: utility، authority :unknowns critical، approvals irreversible،
optimization .provenance evidence يعمل بعد ذلك داخل region المسموح.
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 اقتعيين واقسببية IDENTIFIABILITY
لكل construct تدخل minimal. نمنع interventions التي تغير length prompt أو retries أو simultaneously model. عند صعوبة العزل نستخدم design factorial ونقدر interaction terms. claim causal لا ينتج من
correlation بين theory score والsuccess؛ بعد prediction تتغير أن يجب (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 تتحسن 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، errors، critical violations يشمل accuracy. vector في الأداء نخختزل لا
reproduction. report يعرض Pareto frontier الأهداف تتعارض عندما. hard governance constraints تبقى vetoes.
.Utility optimization is permitted only inside the hard-constraint feasible region
اقتانون الإغلاقي ڤلورقة LAW CLOSURE
الإغلاق المقترح في هذه الورقة: نقترح أن الحلقة الأساسية للذكاء المفتوح ليست answer→reasoning، بل causal→intervention
diagnostic→model world relevant-decision→sufficiency observation
العمومية تبدأ عندما يعيد النظام إغلاق هذه الحلقة بعد تغير المجال والأداة. update verified→residual
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.
يحتاج تكرار الحلقة بعد claim generality transformation.
1.
2.
3.
4.
5.
6.
13
<PARSED TEXT FOR PAGE: 14 / 24>
الخوارزمية واقتنفيذ ALGORITHM 1/5
goal/authority/evidence/؛ افصل Unknown؛ اربط versions. ليس الناتج بل prompt object graph قابلا للتسلسل وإعادة البناء.
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
الخوارزمية واقتنفيذ ALGORITHM 2/5
Algorithm C :GateAndAct طبق gates trajectory/reversibility/evidence/authority. اختر أصغر فعل كامل قابل للرجوع يميز hypotheses أو يحقق utility. 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 واقتنفيذ 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 واقتنفيذ 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 واقتنفيذ ALGORITHM
causal attribution أسهل، ثم browser/code environments، ثم long-running/multi-agent. يصبح أن قبل weights إلى تنقل لا
Implementation sequencing: ابدأ mock/simulator deterministic حيث
مستقراً وقابلاً للقياس. mechanism external
1/4 اقتجربة EXPERIMENT
Partially observable web/research tasks مع hidden critical variables.
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 اقتجربة EXPERIMENT
ALFWorld/ScienceWorld-like transition tasks مع same compute world models.
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 اقتجربة EXPERIMENT
Failure-cause environments لاختبار residual causal graph prediction.
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 اقتجربة EXPERIMENT
Cross-domain transfer من browsing إلى tool/simulator decision loops.
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 اقتياس واقفشل
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 افعلمية 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 افتنفيذ
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 وافتدريب
عندما Adapter .use repeated بعد HyperLoop/HyperSkill انقل إلى editability. للحفاظ على خارجياً mechanism/knowledge ابدأ
transfer فقط عندما يثبت weights .specific-protocol/provider يكون السلوك
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
بل observation sufficiency→decision-relevant world model→diagnostic intervention→causal residual→verified
ليست المفتوح للذكاء الأساسية الحلقة أن نقترح reasoning→answer،
update. العمومية تبدأ عندما يعيد النظام إغلاق هذه الحلقة بعد تغير المجال والأداة.
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
مستقلة، يمكن أن تصبح جزءاً أقوى من MUG/MUT مع evidence transformations.
18
<PARSED TEXT FOR PAGE: 19 / 24>
المراجع ومصادر افسلالة REFERENCES
Role Reference External source
foundational/self-evolving/collective agentic reasoning; open challenges include arXiv:2601.12538 Agentic Reasoning for Large Language Models
world models, long horizon, governance.
episodic/semantic memory and prediction-observation mismatch for test-time arXiv:2606.30639 Self-Evolving World Models for LLM Agent Planning
world-model revision.

Beyond Next-Observation Prediction: Agent-Authored World Modeling for
Sequential Decision Making
decision-oriented world-model targets derived from policy information needs arXiv:2606.25421.
fine-grained attribution chains and component-level leave-one-out perturbation arXiv:2608.06909 Long-Horizon.
Agent Trajectory Attribution
Role Theory Internal
parent XIII Part نظرية مساعف للعمل الإبستيمي المتحقق 13-01M
parent XIII Part نظرية مساعف لحفظ كتلة المجهول 13-02M
parent XIII Part نظرية مساعف للتحكم الراجع بالسؤال الحاكم 13-03M
parent XIII Part نظرية مساعف لتعلم العالم من بقايا التنبؤ والملاحظة 13-05M
parent XIII Part نظرية مساعف للفعل التشخيصي ونقض الذات 13-08M
parent XIII Part نظرية مساعف للإغلاق السببي بين النظرية والنواة 13-19M
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 افنشر 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 افتشغيلية — A الملحق A APPENDIX
تستعمل الورقة مجموعة متغيرات مشتركة للكنها لا تفترض أن كلها observable مباشرة: StoppingReceipt · MarginalObservationValue · ·HardUnknown · CausalVariableCoverage · ObservationSet
DecisionBoundary · PolicySet · DecisionVariable · CounterfactualTransition · SelectiveWorldState · ResidualCluster ·. CandidateCause · CausalEdge · InterventionTest · RepairFamily
OSF = ParetoFrontier(DecisionValidity, CausalCoverage, Independence, Freshness | ObservationCost); stop iff ; ΔExpectedDecisionValue < Cost and HardUnknowns=0
DBWM-Loss = λ1·DecisionVariableError + λ2·CounterfactualTransitionError + λ3·Calibration - λ4·IrrelevantDetailFit ;
RCG⊕edge(i→j)=Recurrence×ConditionalDependence×InterventionConsistency×DecisionImpact - ConfoundingPenalty
لكل متغير خمس سمات إلزامية: domain، estimator، uncertainty، provenance، interventionability. المتغير الذي يبقى حكماً لغوياً بلا estimator لا يدخل equations causal. وإذا كان estimator يسجل based-model
model/version/calibration. أما hard governance constraints تدخل فلا 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
.perturbation عند outcome قبل prediction يجب أن يغير السببي sensitivity Intervention: construct
.conflict يظهر كـ conflict ؛receipt بلا uncertainty مستقلة جديدة لا تمحو monotonicity Evidence: evidence
. تلقائياً authority الأعلى لا توسع autonomy Bounded: capability
.tolerance لا النص الحرفي. وسببها ضمن class verdict يعيد verdict Replayable: replay
generality claim :Transfer scope يتجاوز لا transformation family المختبرة.
هذه hypotheses propositions تشغيلية لا theorems نهائية. وظيفتها أن تجعل التجربة قادرة على الرفض وأن تمنع تحويل score واحد إلى دعم عام.
الخوارزمية المرجعية واقتكلفة — C الملحق C APPENDIX
materialize typed state → expose unknown/authority/budget → generate candidates → estimate expected epistemic/ → world delta → compile proof obligations :Reference procedure
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
.efficiency وحدها ليست tokens . قد رفعته risk خفضته وأي cost تعلن أي optimization أي
1.
2.
3.
4.
5.
21
<PARSED TEXT FOR PAGE: 22 / 24>
APPENDIX D الملحق D — بناء benchmark تشخيصي
interactions، long- تختبر drift/recovery، ثم transfer tasks تغير model/tool/domain. الورقة لهذه الأساسية التجارب: Partially
بنُى ت suite من micro tasks كل تعزل construct، meso tasks تختبر horizon tasks
hidden critical variables.; مع same compute world models.; Failure-cause environments لاختبار residual causal
observable web/research tasks مع ALFWorld/ScienceWorld-like transition tasks
graph prediction.; Cross-domain transfer من browsing إلى tool/simulator decision loops..
missing observation، stale memory، malicious tool text، permission change، delayed event،. لكل perturbation expected
لكل task perturbations قبل مسماة الاختبار: tool failure، distribution shift
failure signature. يمنع هذا post-hoc storytelling.
Use Split
calibration only Dev
primary causal verdict Held-out
boundary discovery Adversarial
N5/generalization Transfer
خطة اقتحليل الإحصائي — E الملحق E APPENDIX
mean/ بحسب metric، CI، عدد seeds/tasks، وتوزيع failures. في paired tasks نستخدم paired analysis. costs تعرض كـsuccess-
نحدد التشغيل قبل primary outcome وصغيرا ًsecondary diagnostics. نبلغ median
at-budget أو Pareto front.
تبقى findings exploratory .value-p بدلاً من انتقاء أفضل out-held على confirmation ثم gate level-family مع تعدد النظريات نستخدم
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 اقصلاحية والأمن
Construct validity: proxy لا يمثل المفهوم.
Internal validity: harness/compute/retries confound.
External validity: transfer غير مختبر.
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
M14-01 · M14-02 · M14-03. parents:. في MUT تحفظ definitions/parents/contradictions؛ في MUG يحدد vessel؛
المركزية النظريات: M13-01 · M13-02 · M13-03 · M13-05 · M13-08 · M13-19
في MOUSSAIXT تنفذ 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>
APPENDIX H الملحق H — برنامج الأسئلة المفتوحة
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 لـresource لا تجزء 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