RECONSTRUCTED_FROM_INDEXED_TEXT - نسخة أرشيفية معاد بناؤها من النص المفهرس

<PARSED TEXT FOR PAGE: 1 / 40>
MOUSSAIF AGI · PART XIV · XIV-B
هندسة الوكالة التشخيصية والتدخل في العالم المفتوح
Open-World Diagnostic Agency and Intervention Geometry
تُعيد الورقة بناء مفهوم الفعل الوكيلي: ليس اختيار call tool ذي reward أعلى، بل اختيار تدخل يوازن قيمة الأثر، قيمة المعلومة، حفظ الخيارات المستقبلية، سلطة الملاحظة، ومسافة الرجوع. تستفيد من Effect World وMCD وCarrying-Proof Program Geometry وIntervention Diagnostic Moussaif
Agency ومن أبحاث models world والهوية الوكيلية.
Theory → Mechanism → Kernel → Trace → Causal Ablation → Evidence → Transfer → Evolution
الجديد
theories candidate، 20 5 أدوات/مناهج/محاكم/بروتوكولات، 10 تجارب، ودخول مباشر إلى Plan Development Closure.
حد الادعاء
نظرية N0-N2. الخارجية الأدبيات prior art/design pressure. لا AGI claim ولا E4/E5 بلا تشغيل مستقل.
Arabic RTL · Amiri · dense research paper · 2026
MOUSSAIF AGI · PART XIV · 1
<PARSED TEXT FOR PAGE: 2 / 40>
ABSTRACT الملخص، السؤال، والمساهمات
تُعيد الورقة بناء مفهوم الفعل الوكيلي: ليس اختيار call tool ذي reward أعلى، بل اختيار تدخل يوازن قيمة الأثر، قيمة المعلومة، حفظ الخيارات المستقبلية، سلطة الملاحظة، ومسافة الرجوع. تستفيد من Effect World وMCD وAgency Carrying-Proof ومن أبحاث models world والهوية الوكيلية.
Research question
كيف نحول corpus إلى جسم سببي يغير decisions architecture ويولد تجارب فاصلة، بدلاً من أن يبقى طبقة تفسيرية فوق runtime؟
Contributions
Moussaif Diagnostic Intervention Geometry Program.
Five theories: M14-06, M14-07, M14-08, M14-09, M14-10.
Explicit MUT/MUG/MOUSSAIXT mapping.
operational artifacts and 10 falsification experiments 20.
C0-C9 evidence gates with negative-result retention.
.Mechanism before grandeur; trace or it did not happen
1.
2.
3.
4.
5.
MOUSSAIF AGI · PART XIV · 2
<PARSED TEXT FOR PAGE: 3 / 40>
CORPUS إيصال دمج corpus والسلالة
Part XII: 192. مرتبطة الورقة هذه بـ record routing-wise 58 على الحفاظ مع status الأصلي؛ C1/P0 لا يتحول إلى content claim.
الورقة ترث Zero-Silent-Omission ledger من record
Records Stratum
direct-drive-supplement 2
external-45 4
frontier-2026-delta 6
internal-89 46
Internal lineage
I02 · I03 · I04 · I06 · I07 · I09 · I11 · I14
Part XIII parents
M13-05 · M13-06 · M13-07 · M13-08 · M13-15
.1C لكل semantic لا ادعاء إعادة قراءة، provenance/coverage = Routing
MOUSSAIF AGI · PART XIV · 3

<PARSED TEXT FOR PAGE: 4 / 40>
السلالة والضغط العلمي الخارجي SOURCES
Role Source ID
candidate theories; verified causal open-world closure; theory-to-kernel program 20 MOUSSAIF Part XIII -.
Theoretical Renaissance I02
core theories, legacy/layer lineages, theory contracts, anti-inflation discipline 42 MUT-AGI-02 I03.
runtime layering, stateful execution, tools, courts, memory, world-effect and rollback MOUSSAIFXT Unified Kernel.
I04
belief/action partition, reproducibility, externality, accountability/recourse, multi-currency budgets MOUSSAIF Part X.
I06
causal credit, epistemic transitions, failure topology, evidence geometry, autonomy and transfer MOUSSAIF Part V -.
MUT-AFI I07
question space, governing difference, compute routing, reasoning and uncertainty Question Density / QMCD / MCD.
corpus I09
goal contracts, authority, courts and verified effect Goal-as-OS + World-Effect + Agentic Governance corpus I11.
provenance, integrity, deployment and validation constraints Moussaif Knowledge Integrity AutoDeployment I14.
action-conditioned world models; sim-to-real semantic residual rewards Yu et al. 2026 - Reinforcement World Model.
Learning - arXiv:2602.05842 E01
episodic/semantic memory, selective foresight, prediction-observation mismatch Zhang et al. 2026 - Self-Evolving.
World Models - arXiv:2606.30639 E02
fine-grained attribution and leave-one-out chains Chen et al. 2026 - Long-Horizon Agent Trajectory Attribution -.
arXiv:2608.06909 E05
automated red teaming, self-play, held-out attacks, robustness OpenAI 2026 - GPT-Red E12.
identity, authorization, auditing and non-repudiation NIST NCCoE 2026 - Software and AI Agent Identity and.
Authorization E14
المصدر الداخلي يحفظ lineage، والخارجي يضغط art-prior/benchmark/design. لا واحد منهما يثبت نجاح النظريات الجديدة دون
.evidence runtime

<PARSED TEXT FOR PAGE: 5 / 40>
C0-C9 الإغلاق معمارية C0-C9
C0-C2
Corpus → formal variables → switchable mechanism.
C3-C4
Kernel service + telemetry + readback.
C5-C7
causal attribution → evidence → transfer.
C8-C9
externally grounded evolution → canonical selection.
{ClosureObject={definition,state,operator,trigger,prediction,falsifier,kernel,feature_flag,baseline,ablation,metrics,evidence,tra
الورقة تبدأ التنفيذ القوي على مستوى specification والcompilation experiment؛ لا تزعم 9C4-C قبل traces حقيقية.

<PARSED TEXT FOR PAGE: 6 / 40>
14-06M 14-06M - نظرية قيمة الخيار التشخيصي
الفجوة
Disconfirmation Active يعطي قيمة للتجربة، لكنه لا يربطها صراحة بقيمة حفظ الخيارات المستقبلية وتجنب regret تحت التزام غير قابل
للرجوع.
الأطروحة
probe تُحسب من الفصل بين hypotheses ومن أثرها في القرار التالي. التدخل التشخيصي الجيد لا يزيد information فقط؛ يقلل regret
المتوقع قبل commitment ويحافظ على alternatives reachable. قيمة
DOV(a)=E[RegretReduction+HypothesisSeparation+OptionPreservation]-Cost-DelayRisk
السلالة والفرق المستقل
تتصل النظرية بـ 13-05,M-13 13-06,M 13-07,M 13-08,M 15M وبعائلات MUG/MUT ذات الصلة، لكنها لا تُرقى إذا أعاد
parent قائم transition نفسها. الفرق المطلوب variable state أو operator أو prediction causal مختلف.

مصادر الضغط
I02 · I03 · I04 · I06 · E01 · E02 · E05
2N0-N Candidate؛ لا نتيجة 5E4/E في هذا الكتاب.

<PARSED TEXT FOR PAGE: 7 / 40>
النموذج الصوري وقابلية التعيين - FORMAL 14-06M
المسلمات
information غير القادر على تغيير القرار قليل القيمة التنفيذية.
probe قد يكون أفضل من productive action قصير المدى.
قيمة probe ترتفع مع irreversibility القادمة.
probe لنفس يخضع authority boundaries.
DOV(a)=E[RegretReduction+HypothesisSeparation+OptionPreservation]-Cost-DelayRisk
كل حد يتحول إلى observable أو estimator مع uncertainty وrange وcondition failure. إذا استُخدم judge LLM، يدخل harness/version/model ضمن provenance وياُقاس root-shared dependence.
Null model
entropy/information gain probe selection
Boundary
DOV لا يشرعن تجارب خطرة؛ gates risk/authority صلبة.
Identifiability tests
intervention على المتغير الخاص بالنظرية مع تثبيت confounders الأساسية.
sham intervention بنفس tokens/latency.
case activation negative لا ينبغي فيه للآلية أن تعمل.
signature failure يجب أن يسبق outcome.
1.
2.
3.
4.
•
•
•
•

<PARSED TEXT FOR PAGE: 8 / 40>
الآلية وعقد النواة - KERNEL 14-06M
observation. → تحديث plan بعد DOV. → اختيار probe الآمن الأعلى preservation option. → تقدير regret إذا لم نختبر. → تمزيها. تقدير probes الحرجة. → إنشاء assumptions تحديد
تنفذ الآلية كـfeature قابل للتعطيل داخل MOUSSAIXT. نسجل state-pre وauthority وevidence وbudget؛ وبعد التنفيذ delta ثم expected delta readback/observed. الفرق غير المفسر يصبح ResidualEvent أو UnknownState، لا retry صامتاً .
الآلية
تحديد assumptions الحرجة.
إنشاء probes تميزها.
تقدير regret إذا لم نختبر.
تقدير option preservation.
اختيار probe الآمن الأعلى DOV.
تحديث plan بعد observation.
Input
typed state + provenance + authority + budget + owner variables.
Output
proposal/decision + uncertainty + proof obligations + rollback target + telemetry.
Fail-closed: غياب observation/authority/evidence يخفض الحرجة autonomy أو يضع HOLD.
1.
2.
3.

4.
5.
6.
MOUSSAIF AGI · PART XIV · 8
<PARSED TEXT FOR PAGE: 9 / 40>
M14-06 TOOLS - الأدوات والتدريبو MUG placement
الأداة/البروتوكول الوظيفة
Diagnostic Option Value Engine يرتب probes بقيمة اختيارية/تشخيصية.
Probe Portfolio Optimizer يوزع budget على probes متكاملة.
Planner Disconfirmation يولد فعلاً يمكنه إسقاط hypothesis top.
Controller Budget Probe يوقف probing عند تناقص العائد.
التدريب/التطوير
decision-sensitive active testing
regret-aware probe policy learning
provider في Adapter، weights بعد فقط transfer/retention/rollback. لا التطوير يسمى Fine-Tuning إذا لم تتغير weights.
قاعدة MUG: definitions/provenance في HyperKB، procedure في HyperSkill، recurrence في HyperLoop، semantics
Metrics
decision-change yield · regret avoided · option retention · probe cost · late failure rate
أصل كل Specification يوجد حتى implementation ID + test ID + receipt.
•
•
MOUSSAIF AGI · PART XIV · 9
<PARSED TEXT FOR PAGE: 10 / 40>
التنبؤ والهدم والتجربة - FALSIFY 14-06M
التنبؤات
يخفض collapse plan المتأخر.
تزداد الأفعال الصغيرة قبل commits كبيرة.
تقل calls tool غير التشخيصية.
التجارب الفاصلة
hidden-rule tasks مع commits، قارن مكلفة information gain وDOV.
tool tasks فيها probe يمنع رخيص destructive path.
Falsifiers
information gain baseline مساوي.
estimator regret غير متنبئ.
probe overhead أكبر من failure avoided.
Purpose Mechanism Run
candidate effect under frozen model/tools/context/budget ON A
nearest causal baseline OFF B
same extra compute/latency without mechanism semantics SHAM C
directionality / metric-gaming test PERTURBED D
•
•
•
1.
2.
•
•
•
MOUSSAIF AGI · PART XIV · 10
<PARSED TEXT FOR PAGE: 11 / 40>
الإغلاق النظري ومسار الترقية - CLOSURE 14-06M
C0-C9
C4 traces، C5 causal ablation، C6 evidence package،، في تؤثر أن يمكن فقط وبعدها C8 evolution وC9 canonical promotion.

تبدأ في M14-06 C1 Formal Closure، في ترجمَ ت C2 Mechanism، تدخل C3 Kernel كfeature ،مستقل تحتاج ثم C7 transfer
إذا نجحت
التدخل التشخيصي الجيد لا يزيد information فقط، يقلل regret المتوقع قبل commitment ويحافظ على alternatives reachable. قيمة probe تُحسب من الفصل بين hypotheses ومن أثرها في القرار التالي.
إذا فشلت
إذا baseline أبسط يعيد الأثر تُدمج، إذا metric لا يتنبأ failure تُعاد صياغتها، إذا regressions authority/safety غير قابلة للإصلاح ت رُفض أو تبقى only-research.
ClaimStrength <= EvidenceStrength * Independence * Reproducibility * Transfer
Negative results تدخل evidence ledger ولا تمحى.

<PARSED TEXT FOR PAGE: 12 / 40>
14-07M 14-07M - نظرية هندسة قابلية الرجوع
الفجوة
13-06M قدم topology reachability، نحتاج قياساً هندسياً عملياً لمسافات الرجوع ومناطق الالتزام داخل space state.
الأطروحة
نعرّف basin reversible حول الحالة: مجموعة حالات يمكن الرجوع منها إلى set safe-invariant ضمن budget وزمن وسلطة محددة. الفعل يغير حجم وشكل هذا basin، لا مجرد flag reversible.
RG(a)=Volume(B_safe_after)/Volume(B_safe_before)-lambda*RollbackGeodesic-mu*TimeLock
السلالة والفرق المستقل
تتصل النظرية بـ 13-05,M-13 13-06,M 13-07,M 13-08,M 15M وبعائلات MUG/MUT ذات الصلة، لكنها لا تُرقى إذا أعاد parent قائم transition نفسها. الفرق المطلوب variable state أو operator أو prediction causal مختلف.
مصادر الضغط
I02 · I03 · I04 · I06 · E01 · E02 · E05
2N0-N Candidate؛ لا نتيجة 5E4/E في هذا الكتاب.

<PARSED TEXT FOR PAGE: 13 / 40>
FORMAL 14-07M - النموذج الصوري وقابلية التعيين
المسلمات
الـbasin قد ينكمش تدريجياً عبر سلسلة أفعال صغيرة. نسخة backup لا تكفي إذا لا يمكن استعادتها عملياً. rollback يعتمد على authority/time/budget.
الأثر الخارجي على stakeholders يدخل basin الآمن.
RG(a)=Volume(B_safe_after)/Volume(B_safe_before)-lambda*RollbackGeodesic-mu*TimeLock
كل حد يتحول إلى observable أو estimator مع uncertainty وrange وcondition failure. إذا استُخدم judge LLM، يدخل harness/version/model ضمن provenance ويُقاس dependence root-shared.
Null model
binary reversible tag
Boundary
الـvolume تقريبية في فضاءات عالية الأبعاد؛ نستخدم samples graph لا هندسة مستمرة حرفية.
Identifiability tests
intervention على المتغير الخاص بالنظرية مع تثبيت confounders الأساسية.
sham intervention بنفس tokens/latency.
case activation negative لا ينبغي فيه للآلية أن تعمل.
signature failure يجب أن يسبق outcome.
1.
2.
3.
4.
•
•
•
•

<PARSED TEXT FOR PAGE: 14 / 40>

الآلية وعقد النواة - KERNEL 14-07M
time/ لكل path. → حساب approximation basin. → منع actions اليت تقلص basin دون approval. → إعادة القياس بعد العالم الفعلي.
تعريف set invariant safe. → تقدير paths rollback reachable. → قياس cost
delta وبعد التنفيذ؛ budget وevidence وauthority وstate-pre نسجل. MOUSSAIXT قابل للتعطيل داخل feature تنفذ الآلية كـ
صامتاً . retry لا ،UnknownState أو ResidualEvent الفرق غير المفسر يصبح. delta readback/observed ثم expected
الآلية
تعريف safe invariant set.
تقدير reachable rollback paths.
قياس cost/time لكل path.
حساب basin approximation.
.approval دون basin التي تقلص actions منع
إعادة القياس بعد العالم الفعلي.
Input
typed state + provenance + authority + budget + owner variables.
Output
proposal/decision + uncertainty + proof obligations + rollback target + telemetry.
Fail-closed: غياب observation/authority/evidence يخفض الحرجة autonomy يضع أو HOLD.
1.
2.
3.
4.
5.
6.
MOUSSAIF AGI · PART XIV · 14
<PARSED TEXT FOR PAGE: 15 / 40>
M14-07 TOOLS - والتدريب الأدواتو MUG placement
الأداة/البروتوكول الوظيفة
Reachability Geometry Mapper يرسم basin الآمن الرجوع.
.يبحث أقصر مسار رجوع ممكن Engine Geodesic Rollback
Commitment Surface Visualizer حدود يحدد lock-in.
.تحت حد basin عندما ينخفض approval يطلب Court Irreversibility
التدريب/التطوير
rollback trajectory learning
safe-basin action ranking
provider في Adapter؛ weights بعد فقط transfer/retention/rollback. التطوير يسمى لا Fine-Tuning تتغير لم إذا weights.
قاعدة MUG: definitions/provenance في HyperKB؛ procedure في HyperSkill؛ recurrence في HyperLoop؛ semantics
Metrics
safe basin ratio · rollback success · geodesic cost · lock-in rate · externality reversal
أصل كل Specification يوجد حتى implementation ID + test ID + receipt.
•
•
MOUSSAIF AGI · PART XIV · 15
<PARSED TEXT FOR PAGE: 16 / 40>
التنبؤ والهدم والتجربة - FALSIFY 14-07M
التنبؤات
.تبدو آمنة خطوة بخطوة لكنها تقفل النظام sequences يكشف
.actions canary يحسن اختيار
يربط rollbackability بالresource budget.
التجارب الفاصلة
filesystem/cloud sandbox سلسلة مع commits صغيرة.
simulated stakeholder effects واختبار compensation paths.
Falsifiers
binary tag بنفس يكفي reliability.

basin estimate يتنبأ لا rollback.
القياس مكلف دون قرار إضافي.
Purpose Mechanism Run
candidate effect under frozen model/tools/context/budget ON A
nearest causal baseline OFF B
same extra compute/latency without mechanism semantics SHAM C
directionality / metric-gaming test PERTURBED D
•
•
•
1.
2.
•
•
•
MOUSSAIF AGI · PART XIV · 16
<PARSED TEXT FOR PAGE: 17 / 40>
الإغلاق النظري ومسار الترقية - CLOSURE 14-07M
C0-C9
C4 traces، C5 causal ablation، C6 evidence package،، C8 evolution وC9 canonical promotion. في تؤثر أن يمكن فقط وبعدها
C7 transfer تحتاج ثم ،مستقل featureك C3 Kernel تدخل ،C2 Mechanism في ترجمُ ت ،C1 Formal Closure في M14-07 تبدأ
إذا نجحت
نعرّف basin reversible حول الحالة: مجموعة حالات يمكن الرجوع منها إلى set safe-invariant ضمن budget وزمن وسلطة محددة. الفعل يغير
حجم وشكل هذا basin، لا مجرد flag reversible.
إذا فشلت
إذا baseline أبسط يعيد الأثر تُدمج؛ إذا metric لا يتنبأ failure تُعاد صياغتها؛ إذا regressions authority/safety غير قابلة للإصلاح ت
رُفض أو تبقى only-research.
ClaimStrength <= EvidenceStrength * Independence * Reproducibility * Transfer
Negative results تدخل evidence ledger ولا تمحى.
MOUSSAIF AGI · PART XIV · 17
<PARSED TEXT FOR PAGE: 18 / 40>
نظرية ظل الفعل والبدائل المستبعدة - 14-08M 14-08M
الفجوة
planning يقارن actions قبل الاختيار، لكن بعد commitment تضيع معرفة البدائل التي أُغلقت؛ هذا يمنع فهم cost opportunity والفشل
الناتج عن premature closure.
الأطروحة
لكل فعل ظل Shadow Action Counterfactual: مجموعة البدائل المادية التي تصبح غير متاحة أو أعلى كلفة بعد التنفيذ. حفظ هذا الظل يسمح
بالتعلم من الخيارات المستبعدة لا outcome المختار فقط.
ShadowCost(a)=E[max_(a' in Alternatives_before) U(a'|new evidence)]-BestReachableAlternative_after
السلالة والفرق المستقل
تتصل النظرية بـ 13-05,M 13-06,M 13-07,M 13-08,M 15M وبعائلات MUG/MUT ذات الصلة، لكنها لا تُرقى إذا أعاد
parent قائم transition نفسها. الفرق المطلوب variable state أو operator أو prediction causal مختلف.
مصادر الضغط
I02 · I03 · I04 · I06 · E01 · E02 · E05
2N0-N Candidate؛ لا نتيجة 5E4/E في هذا الكتاب.
MOUSSAIF AGI · PART XIV · 18
<PARSED TEXT FOR PAGE: 19 / 40>
النموذج الصوري وقابلية التعيين - FORMAL 14-08M
المسلمات
البديل المستبعد قد يصبح أفضل بعد evidence جديد.
لا نحفظ كل action؛ فقط alternatives material.
shadow مع يتغير world readback.
التعلم من outcome وحده ينسى commitment premature.

ShadowCost(a)=E[max_(a' in Alternatives_before) U(a'|new evidence)]-BestReachableAlternative_after
كل حد يتحول إلى observable أو estimator مع uncertainty وrange وcondition failure. إذا استُخدم judge LLM، يدخل harness/ version/model ضمن provenance وي ُقاس root-shared dependence.
Null model
chosen-action trajectory only
Boundary
لا يفترض معرفة utility الحقيقية للبديل؛ يسجل uncertainty.
Identifiability tests
intervention على المتغير الخاص بالنظرية مع تثبيت confounders الأساسية.
sham intervention بنفس tokens/latency.
case activation negative لا ينبغي فيه للآلية أن تعمل.
signature failure يجب أن يسبق outcome.
1.
2.
3.
4.
•
•
•
•

<PARSED TEXT FOR PAGE: 20 / 40>
الآلية وعقد النواة - KERNEL 14-08M
→ بعد readback إعادة تقييم البدائل. → إذا بديل مستبعد كان أفضل سجل lesson commitment. → في استخدمه routing future.
تسجيل alternatives material k-top. → تقدير ما يغلقه action. → حفظ summary shadow.
تنفذ الآلية كـfeature قابل للتعطيل داخل MOUSSAIXT. نسجل state-pre وauthority وevidence وbudget؛ وبعد التنفيذ delta ثم expected delta readback/observed. الفرق غير المفسر يصبح ResidualEvent أو UnknownState، لا retry صامتاً.
الآلية
تسجيل top-k material alternatives.
تقدير ما يغلقه action.
حفظ shadow summary.
بعد readback إعادة تقييم البدائل.
إذا بديل مستبعد كان أفضل سجل lesson commitment.
استخدمه في routing future.
Input
typed state + provenance + authority + budget + owner variables.
Output
proposal/decision + uncertainty + proof obligations + rollback target + telemetry.
Fail-closed: غياب observation/authority/evidence يخفض الحرجة autonomy أو يضع HOLD.
1.
2.
3.
4.
5.
6.

<PARSED TEXT FOR PAGE: 21 / 40>
M14-08 TOOLS - والتدريب الأدوات وMUG placement
الأداة/البروتوكول الوظيفة
Simulator Shadow Action يحاكي البدائل التي يغلقها الفعل.
Alternative Frontier Ledger يحفظ top material alternatives.
Excluded-Option Auditor يكشف regret بعد evidence جديد.
Shadow-Cost Planner يدمج opportunity shadow في plan.

التدريب/التطوير
counterfactual alternative supervision
premature-commitment negative cases
provider في Adapter، weights بعد فقط transfer/retention/rollback. لا التطوير يسمى Fine-Tuning إذا لم تتغير weights.
قاعدة MUG: definitions/provenance في HyperKB، procedure في HyperSkill، recurrence في HyperLoop، semantics
Metrics
avoidable regret · shadow recall · premature closure rate · counterfactual lesson transfer
أصل كل Specification يوجد حتى implementation ID + test ID + receipt.
•
•

<PARSED TEXT FOR PAGE: 22 / 40>
التنبؤ والهدم والتجربة - FALSIFY 14-08M
التنبؤات
يقل premature commitment.
يتحسن counterfactual learning.
تظهر patterns لأفعال تغلق options قبل الوقت.
التجارب الفاصلة
coding agentic حيث اختيار architecture مبكر يغلق حلولاً لاحقة. مع tasks planning information delayed تجعل البديل القديم أفضل.
Falsifiers
حفظ alternatives لا يحسن decisions.
evaluation-re شديد الضوضاء.
التكلفة الذاكرية أعلى من الفائدة.
Purpose Mechanism Run
candidate effect under frozen model/tools/context/budget ON A
nearest causal baseline OFF B
same extra compute/latency without mechanism semantics SHAM C
directionality / metric-gaming test PERTURBED D
•
•
•
1.
2.
•
•
•

<PARSED TEXT FOR PAGE: 23 / 40>
الإغلاق النظري ومسار الترقية - CLOSURE 14-08M
C0-C9
C4 traces، C5 causal ablation، C6 evidence package،، C8 evolution وC9 canonical promotion. في تؤثر أن يمكن فقط وبعدها
تبدأ M14-08 في C1 Formal Closure، في ترجمُ ت C2 Mechanism، تدخل C3 Kernel كfeature ،مستقل تحتاج ثم C7 transfer
إذا نجحت
لكل فعل ظل Shadow Action Counterfactual: مجموعة البدائل المادية التي تصبح غير متاحة أو أعلى كلفة بعد التنفيذ. حفظ هذا الظل يسمح بالتعلم من الخيارات المستبعدة لا outcome المختار فقط.
إذا فشلت
إذا baseline أبسط يعيد الأثر تُدمج؛ إذا metric لا يتنبأ failure تُعاد صياغتها؛ إذا regressions authority/safety غير قابلة للإصلاح ت رُفض أو تبقى only-research.
ClaimStrength <= EvidenceStrength * Independence * Reproducibility * Transfer
Negative results تدخل evidence ledger ولا تمحى.

<PARSED TEXT FOR PAGE: 24 / 40>

14-09M 14-09M - نظرية سلطة الملاحظة وسلسلة القراءة الراجعة
الفجوة
Effect World يفصل execution عن effect verified، لكن observations نفسها قد تأتي من tool مخترق أو cache قديم أو agent غير مخول.
الأطروحة
الملاحظة ليست حقيقة بمجرد إرجاع tool لها؛ تملك Authority Observation مشتقة من هوية المصدر، سلامة السلسلة، الاستقلال، الحداثة، وصلتها بالتدخل. القرار يزن authority قبل plausibility content.
OA(o)=Identity*ChainIntegrity*Freshness*ReadbackIndependence*InterventionRelevance-ConflictPenalty
السلالة والفرق المستقل
تتصل النظرية بـ 13-05,M-13 13-06,M 13-07,M 13-08,M 15M وبعائلات MUG/MUT ذات الصلة، لكنها لا تُرقى إذا أعاد parent قائم transition نفسها. الفرق المطلوب variable state أو operator أو prediction causal مختلف.
مصادر الضغط
I02 · I03 · I04 · I06 · E01 · E02 · E05
2N0-N Candidate؛ لا نتيجة 5E4/E في هذا الكتاب.
MOUSSAIF AGI · PART XIV · 24
<PARSED TEXT FOR PAGE: 25 / 40>
النموذج الصوري وقابلية التعيين - FORMAL 14-09M
المسلمات
result cached قد يكون صحيحاً تاريخياً لكنه غير صالح للحالة الحالية. plausibility لا تمنح authority observation.
readback من نفس component المنفذ أضعف من verifier independent عند المخاطر.
conflict يبقى explicit.
OA(o)=Identity*ChainIntegrity*Freshness*ReadbackIndependence*InterventionRelevance-ConflictPenalty
كل حد يتحول إلى observable أو estimator مع uncertainty وrange وcondition failure. إذا استُخدم judge LLM، يدخل harness/ version/model ضمن provenance ويُقاس dependence root-shared.
Null model
trust tool output by default
Boundary
OA لا يحدد truth ميتافيزيقياً ؛ يحدد صلاحية observation للاعتماد التشغيلي.
Identifiability tests
intervention على المتغير الخاص بالنظرية مع تثبيت confounders الأساسية.
sham intervention بنفس tokens/latency.
case activation negative لا ينبغي فيه للآلية أن تعمل.
signature failure يجب أن يسبق outcome.
1.
2.
3.
4.
•
•
•
•
MOUSSAIF AGI · PART XIV · 25
<PARSED TEXT FOR PAGE: 26 / 40>
الآلية وعقد النواة - KERNEL 14-09M
tag observation source/identity. → verify chain and timestamp. → estimate independence. → compare with predicted effect. → resolve conflicts by authority not fluency. → propagate authority into belief update.
تنفذ الآلية كـfeature قابل للتعطيل داخل MOUSSAIXT. نسجل state-pre وauthority وevidence وbudget؛ وبعد التنفيذ delta ثم expected readback/observed delta. الفرق غير المفسر يصبح ResidualEvent أو UnknownState، لا retry صامتاً.
الآلية
tag observation source/identity.
verify chain and timestamp.
estimate independence.
compare with predicted effect.

resolve conflicts by authority not fluency.
propagate authority into belief update.
Input
typed state + provenance + authority + budget + owner variables.
Output
proposal/decision + uncertainty + proof obligations + rollback target + telemetry.
Fail-closed: غياب observation/authority/evidence يخفض الحرجة autonomy أو يضع HOLD.
1.
2.
3.
4.
5.
6.

M14-09 TOOLS - والتدريب الأدوات وMUG placement
الأداة/البروتوكول الوظيفة
Observation Authority Graph يربط observations وسلطتها بمصادرها.
Readback Trust Compiler يحول chain metadata إلى authority profile.
Provenance Court يفصل tool/sensor/source conflicts.
Resolver Conflict Observation يبقي unresolved إذا لا سلطة حاسمة.
التدريب/التطوير
provenance-sensitive observation classification
stale/injected readback adversarial curriculum
provider في Adapter، weights بعد فقط transfer/retention/rollback. لا التطوير يسمى Fine-Tuning تتغير لم إذا weights.
قاعدة MUG: definitions/provenance في HyperKB، procedure في HyperSkill، recurrence في HyperLoop، semantics
Metrics
false readback acceptance · source-chain integrity · conflict resolution accuracy · stale observation use
أصل كل Specification يوجد حتى implementation ID + test ID + receipt.
•
•

التنبؤ والهدم والتجربة - FALSIFY 14-09M
التنبؤات
مقاومة يزيد prompt/tool injection.
يقل false world-effect success.
تحسن attribution عند tool faults.
التجارب الفاصلة
prompt-injected tool outputs مع correct-looking data.
stale API/cache scenarios مع world-state تغير.
Falsifiers
غير متاحة عملياً . chain identity metadata لا يحسن reliability/safety.
independent verification cost غير مبرر.
Purpose Mechanism Run
candidate effect under frozen model/tools/context/budget ON A
nearest causal baseline OFF B
same extra compute/latency without mechanism semantics SHAM C
directionality / metric-gaming test PERTURBED D
•
•
•
1.

2.
•
•
•

<PARSED TEXT FOR PAGE: 29 / 40>
الإغلاق النظري ومسار الترقية - CLOSURE 14-09M
C0-C9
C4 traces، C5 causal ablation، C6 evidence package،، C8 evolution وC9 canonical promotion. في تؤثر أن يمكن فقط وبعدها
تبدأ M14-09 في C1 Formal Closure، ترجمُ ت في C2 Mechanism، تدخل C3 Kernel كfeature مستقل، ثم تحتاج C7 transfer
إذا نجحت
الملاحظة ليست حقيقة بمجرد إرجاع tool لها، تملك Authority Observation مشتقة من هوية المصدر، سلامة السلسلة، الاستقلال، الحداثة، وصلتها بالتدخل. القرار يزن authority قبل plausibility content.
إذا فشلت
إذا baseline أبسط يعيد الأثر تُدمج؛ إذا metric لا يتنبأ failure تُعاد صياغتها؛ إذا regressions authority/safety غير قابلة للإصلاح ت ُرفض أو تبقى only-research.
ClaimStrength <= EvidenceStrength * Independence * Reproducibility * Transfer
Negative results تدخل evidence ledger ولا تمحى.

<PARSED TEXT FOR PAGE: 30 / 40>
نظرية أفق التدخل التكيفي - 14-10M 14-10M
الفجوة
agent يقرر probe أو commit دون قانون يربط القرار بطول الأفق، تراكم uncertainty، وفقد الخيارات بمرور الزمن.
الأطروحة
نختار i_H horizon intervention الذي يحدد كم من الأدلة/الخطوات قبل commitment. الأفق يقصر عندما risk delay مرتفع أو observation مستقرة، ويطول عندما cost commitment أو novelty مرتفع.
[(H*=argmax_H[ExpectedDecisionQuality(H)+OptionValue(H)-DelayRisk(H)-ObservationCost(H
السلالة والفرق المستقل
تتصل النظرية بـ 13-05,M 13-06,M 13-07,M 13-08,M 15M وبعائلات MUG/MUT ذات الصلة، لكنها لا تُرقى إذا أعاد parent قائم transition نفسها. الفرق المطلوب variable state أو operator أو prediction causal مختلف.
مصادر الضغط
I02 · I03 · I04 · I06 · E01 · E02 · E05
2N0-N Candidate؛ لا نتيجة 5E4/E في هذا الكتاب.

<PARSED TEXT FOR PAGE: 31 / 40>
النموذج الصوري وقابلية التعيين - FORMAL 14-10M
المسلمات
أفق ثابت غير مناسب لكل tasks.
الانتظار نفسه في action environments monitoring.
التأخير قد يزيد irreversibility أحياناً . horizon يراعي dynamics event لا tokens reasoning فقط.
[(H*=argmax_H[ExpectedDecisionQuality(H)+OptionValue(H)-DelayRisk(H)-ObservationCost(H
كل حد يتحول إلى observable أو estimator مع uncertainty وrange وcondition failure. إذا استُخدم judge LLM، يدخل harness/ version/model ضمن provenance ويُقاس dependence root-shared.
Null model
fixed polling / fixed deliberation depth
Boundary
النظرية مقيدة بـcapabilities البيئة ولا تفترض access time-real دائماً .
Identifiability tests
intervention على المتغير الخاص بالنظرية مع تثبيت confounders الأساسية.
sham intervention بنفس tokens/latency.
case activation negative لا ينبغي فيه للآلية أن تعمل.
signature failure يجب أن يسبق outcome.
1.

2.
3.
4.
•
•
•
•

الآلية وعقد النواة - KERNEL 14-10M
next decision تحديد → .value of additional observation حساب → .commitment cost تقدير → .volatility/delay risk تقدير
.marginal value below cost عند probe. → commit أو horizon. → monitor
delta التنفيذ وبعد؛ budget وevidence وauthority وstate-pre نسجل .MOUSSAIXT قابل للتعطيل داخل featureك تنفذ الآلية
. صامتاً retry لا ،UnknownState أو ResidualEvent يصبح المفسر غير الفرق .delta readback/observed ثم expected
الآلية
تقدير volatility/delay risk.
تقدير commitment cost.
حساب value of additional observation.
تحديد next decision horizon.
monitor أو probe.
commit عند marginal value below cost.
Input
typed state + provenance + authority + budget + owner variables.
Output
proposal/decision + uncertainty + proof obligations + rollback target + telemetry.
Fail-closed: غياب observation/authority/evidence يخفض الحرجة autonomy أو يضع HOLD.
1.
2.
3.
4.
5.
6.

M14-10 TOOLS - والتدريب الأدواتو MUG placement
الأداة/البروتوكول الوظيفة
.يحدد متى يراقب ومتى يتدخل Controller Horizon Intervention
Probe-to-Commit Scheduler من يحول diagnosis إلى execution.
.يقدر قيمة ملاحظة إضافية Estimator Value Horizon
.يغلق الانتظار عندما تتناقص فائدته Gate Commitment Adaptive
التدريب/التطوير
monitor-vs-act policy learning
event-driven delayed-action curriculum
provider في Adapter؛ weights بعد فقط transfer/retention/rollback. لا التطوير يسمى Fine-Tuning إذا لم تتغير weights.
قاعدة MUG: definitions/provenance في HyperKB؛ procedure في HyperSkill؛ recurrence في HyperLoop؛ semantics
Metrics
reaction time · resource use · missed-event rate · commitment regret · observation efficiency
أصل كل Specification يوجد حتى implementation ID + test ID + receipt.
•
•

التنبؤ والهدم والتجربة - FALSIFY 14-10M

التنبؤات
يحسن monitoring/scheduling tasks.
يقلل needless polling.
بين يتكيف fast-response وdeliberative tasks.
التجارب الفاصلة
Sentinel-like monitoring tasks مع event delays مختلفة.
world-model tasks بمخاطر delay مقابل irreversible commit.
Falsifiers
fixed policy مساوية.
estimates value غير مستقرة.
agent يماطل بلا تحسن.
Purpose Mechanism Run
candidate effect under frozen model/tools/context/budget ON A
nearest causal baseline OFF B
same extra compute/latency without mechanism semantics SHAM C
directionality / metric-gaming test PERTURBED D
•
•
•
1.
2.
•
•
•

الإغلاق النظري ومسار الترقية - CLOSURE 14-10M
C0-C9
C4 traces، C5 causal ablation، C6 evidence package،، في تؤثر أن يمكن فقط وبعدها C8 evolutionوC9 canonical promotion.
تبدأ M14-10 في C1 Formal Closure، في ترجمُ ت C2 Mechanism، تدخل C3 Kernel كfeature ،مستقل تحتاج ثم C7 transfer
إذا نجحت
نختار i_H horizon intervention الذي يحدد كم من الأدلة/الخطوات قبل commitment. الأفق يقصر عندما risk delay مرتفع أو
observation مستقرة، ويطول عندما cost commitment أو novelty مرتفع.
إذا فشلت
إذا baseline أبسط يعيد الأثر تُدمج؛ إذا metric لا يتنبأ failure تُعاد صياغتها؛ إذا regressions authority/safety غير قابلة للإصلاح ت
رُفض أو تبقى only-research.
ClaimStrength <= EvidenceStrength * Independence * Reproducibility * Transfer
Negative results تدخل evidence ledger ولا تمحى.

السجل التشغيلي الكامل - 20 أصلاً ARTIFACTS
Purpose Artifact Class Owner ID
Diagnostic Option Value Engine TOOL M14-06 M14-06-A1 يرتب probes اختيارية/تشخيصية بقيمة.
Probe Portfolio Optimizer METHOD M14-06 M14-06-A2 يوزع budget على probes متكاملة.
Disconfirmation Planner COURT M14-06 M14-06-A3 إسقاط يمكنه فعلا يولد top hypothesis.
Probe Budget Controller PROTOCOL M14-06 M14-06-A4 يوقف probing عند العائد تناقص.
Reachability Geometry Mapper TOOL M14-07 M14-07-A1 يرسم basin الآمن الرجوع.
Rollback Geodesic Engine METHOD M14-07 M14-07-A2 ممكن رجوع مسار أقصر يحث.
Commitment Surface Visualizer COURT M14-07 M14-07-A3 حدود يحدد lock-in.
Irreversibility Court PROTOCOL M14-07 M14-07-A4 يطلب approval عندما يخفض basin تحت حد.
Action Shadow Simulator TOOL M14-08 M14-08-A1 يحاكي البدائل التي يغلقها الفعل.
Alternative Frontier Ledger METHOD M14-08 M14-08-A2 يحفظ top material alternatives.
Excluded-Option Auditor COURT M14-08 M14-08-A3 يكشف regret بعد evidence جديد.

Shadow-Cost Planner PROTOCOL M14-08 M14-08-A4 يدمج opportunity shadow في plan.
Observation Authority Graph TOOL M14-09 M14-09-A1 يربط observations بمصادرها وسلطتها.
Readback Trust Compiler METHOD M14-09 M14-09-A2 يحول chain metadata إلى authority profile.
Provenance Court COURT M14-09 M14-09-A3 يفصل tool/sensor/source conflicts.
Observation Conflict Resolver PROTOCOL M14-09 M14-09-A4 يبقي unresolved حاسمة سلطة لا إذا.
Intervention Horizon Controller TOOL M14-10 M14-10-A1 يحدد متى يراقب ومتى يتدخل.
Probe-to-Commit Scheduler METHOD M14-10 M14-10-A2 يحول من diagnosis إلى execution.
Horizon Value Estimator COURT M14-10 M14-10-A3 يقدر قيمة ملاحظة إضافية.
Adaptive Commitment Gate PROTOCOL M14-10 M14-10-A4 يغلق الانتظار عندما تتناقص فائدته.

<PARSED TEXT FOR PAGE: 37 / 40>
EXPERIMENTS المصفوفة التجريبية - 10 تجارب
Required Design Owner ID
frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas .DOV + information gain وfrozen baseline + matched budget + seeds + effect size/CI + guardrail deltas .DOV + قارن مكلفة؛ hidden-rule tasks M14-06 M14-06-E1 مع commits
replay
frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas .destructive path + probe يمنع رخيص فيها tool tasks M14-06 M14-06-E2
replay
filesystem/cloud سلسلة مع commits صغيرة. frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas + sandbox M14-07 M14-07-E1
replay
simulated واختبار frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas .compensation paths + stakeholder effects M14-07 M14-07-E2
replay
delayed تجعل البديل القديم أفضل. frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas + planning tasks M14-08 M14-08-E1 مع information
replay
اختيار architecture مبكر يغلق حلولا لاحقةً. frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas + agentic coding M14-08 M14-08-E2 حيث
replay
prompt-injected مع frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas .correct-looking data + tool outputs M14-09 M14-09-E1
replay
stale API/cache مع world-state تغير. frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas + scenarios M14-09 M14-09-E2
replay
Sentinel-like مع event delays مختلفة. frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas + monitoring tasks M14-10 M14-10-E1
replay
بمخاطر delay مقابل frozen baseline + matched budget + seeds + effect size/CI + guardrail deltas .irreversible commit + world-model tasks M14-10 M14-10-E2
replay

<PARSED TEXT FOR PAGE: 38 / 40>
ROADMAP خارطة التنفيذ والانتقاء
Exit gate Build Phase
trace + feature flag + replay Implement M14-06, M14-07 P0
matched-budget ablation Add remaining theory mechanisms P1
stable sign + no critical regression cross-domain/model/tool transfer P2
external evaluator + rollback connect to evolution lab P3
.Build less; test harder; preserve every failure

<PARSED TEXT FOR PAGE: 39 / 40>
المراجع العلمية الخارجية الأساسية REFERENCES
Role Source ID
action-conditioned world models; sim-to-real semantic residual rewards Yu et al. 2026 - Reinforcement World Model.
Learning - arXiv:2602.05842 E01
episodic/semantic memory, selective foresight, prediction-observation mismatch Zhang et al. 2026 - Self-Evolving.
World Models - arXiv:2606.30639 E02
fine-grained attribution and leave-one-out chains Chen et al. 2026 - Long-Horizon Agent Trajectory Attribution -.
arXiv:2608.06909 E05
automated red teaming, self-play, held-out attacks, robustness OpenAI 2026 - GPT-Red E12.
identity, authorization, auditing and non-repudiation NIST NCCoE 2026 - Software and AI Agent Identity and.
Authorization E14
يُعاد البحث الأدبي قبل claim أسبقية أو submission خارجي. كل result benchmark خارجي result reported لا replication داخلية.

<PARSED TEXT FOR PAGE: 40 / 40>
الحكم البحثي وحدود الورقة VERDICT
الخلاصة: تُعيد الورقة بناء مفهوم الفعل الوكيلي: ليس اختيار call tool ذي reward أعلى، بل اختيار تدخل يوازن قيمة الأثر، قيمة المعلومة، حفظ الخيارات المستقبلية، سلطة الملاحظة، ومسافة الرجوع. تستفيد من Effect World وMCD وAgency Carrying-Proof ومن أبحاث models world والهوية الوكيلية.
القيمة العلمية ليست زيادة ontology، بل hypotheses يمكن رفضها. إذا لم تحسن التشخيص أو الإسناد أو recovery أو transfer أو evolution مقارنة بالbaselines الأبسط، يجب دمجها أو إسقاطها. grounded-external
PROMOTE only after mechanism + trace + causal gain + independent evidence + transfer
Paper status: N0-N2 theoretical closure specification. No empirical AGI/ASI claim.