RECONSTRUCTED_FROM_INDEXED_TEXT - نسخة أرشيفية معاد بناؤها من النص المفهرس

<PARSED TEXT FOR PAGE: 1 / 32>
MOUSSAIF AGI RESEARCH PAPER · P14-E · 2026
ورقة 5 — التطور العلمي الذاتي المؤسس خارجياً
Externally Grounded Recursive Scientific Self-Evolution
Five-theory formal research paper · MUT/MUG/MOUSSAIXT compilation · falsifiable program
جعل improvement-self عملية بحث ذات budget evaluator مستقل، provenance للطفرات، ضغط نظري محكوم بالدليل، وتوقف عندما تصبح قيمة التجربة الهامشية أقل من كلفتها.
MOUSSAIF AGI · 1
<PARSED TEXT FOR PAGE: 2 / 32>
Abstract / contribution INTRO 1/6
kernel translation. داخل محصور MOUSSAIF، external literature prior art/design pressure، لا empirical promotion.
novelty claim وfunctional formalization مع جديدة constructs خمسة
RQ
ما variable الجديد؟ هل يمكن التدخل فيه؟ هل يسبق failure؟ هل يغير budget matched under decision/world؟
Boundary
لا AGI claim، لا global priority، no external paper = internal evidence.
قيمة البحث هنا ليست كثرة الأسماء بل القدرة على كتابة تجربة يمكن أن تجعل النظرية تخسر أمام parent أبسط.
MOUSSAIF AGI · 2
<PARSED TEXT FOR PAGE: 3 / 32>
Research problem / hypotheses INTRO 2/6
جعل improvement-self عملية بحث ذات budget evaluator مستقل، provenance للطفرات، ضغط نظري محكوم بالدليل، وتوقف عندما تصبح قيمة التجربة الهامشية أقل من كلفتها. 1:H mechanisms تغير prediction failure، 2H: بعض gains لا تنتقل؛ H3: interactions مادية؛ H4: negative evidence سيضغط theory graph.
RQ
ما variable الجديد؟ هل يمكن التدخل فيه؟ هل يسبق failure؟ هل يغير budget matched under decision/world؟
Boundary
لا AGI claim، لا global priority، no external paper = internal evidence.
قيمة البحث هنا ليست كثرة الأسماء بل القدرة على كتابة تجربة يمكن أن تجعل النظرية تخسر أمام parent أبسط.
MOUSSAIF AGI · 3
<PARSED TEXT FOR PAGE: 4 / 32>
Internal lineage INTRO 3/6
Parts I-XIV، MUT-AGI-02، MCD/QMCD/MURI، M-HKOS، World Effect، Continuity، Part XII paper إلزامي.
integration، Part XIII causal open-world intelligence. nearest-parent check
RQ
ما variable الجديد؟ هل يمكن التدخل فيه؟ هل يسبق failure؟ هل يغير budget matched under decision/world؟
Boundary
لا AGI claim، لا global priority، no external paper = internal evidence.
قيمة البحث هنا ليست كثرة الأسماء بل القدرة على كتابة تجربة يمكن أن تجعل النظرية تخسر أمام parent أبسط.
MOUSSAIF AGI · 4
<PARSED TEXT FOR PAGE: 5 / 32>
External scientific position INTRO 4/6
Anchors: E01: Wei et al. 2026, Agentic Reasoning for Large Language Models, arXiv:2601.12538 · E09: Pronesti et, al. 2026, Verifiable Process Reward Models, ACL Findings 2026 · E11: Lacoste et al. 2026
CUBE: A Standard for Unifying Agent Benchmarks, arXiv:2603.15798 · E12: Yueh-Han et al. 2026, Automated-Researchers Can Reliably Mitigate Alignment Failures, arXiv:2608.28945 · E13: OpenAI 2026, GPT
Red: Unlocking Self-Improvement for. لتصميم المعلنة نتائجها نستخدم benchmarks/failures، ولا نعتبرها validation داخلية.
Robustness · E16: Google DeepMind 2026, AlphaEvolve impact/research updates
RQ
ما variable الجديد؟ هل يمكن التدخل فيه؟ هل يسبق failure؟ هل يغير budget matched under decision/world؟
Boundary

لا AGI claim، لا global priority، no external paper = internal evidence.
قيمة البحث هنا ليست كثرة الأسماء بل القدرة على كتابة تجربة يمكن أن تجعل النظرية تخسر أمام parent أبسط.
MOUSSAIF AGI · 5
<PARSED TEXT FOR PAGE: 6 / 32>
Methodology INTRO 5/6
source→claim role→theory delta→MUG vessel→switchable kernel→ON/OFF/SHAM/. لها مرحلة كل failure status.
PERTURBED→readback→attribution→transfer
RQ
؟budget matched under decision/world هل يغير failure؟ هل يسبق هل يمكن التدخل فيه؟ variable ما الجديد؟
Boundary
لا AGI claim، لا global priority، no external paper = internal evidence.
قيمة البحث هنا ليست كثرة الأسماء بل القدرة على كتابة تجربة يمكن أن تجعل النظرية تخسر أمام parent أبسط.
MOUSSAIF AGI · 6
<PARSED TEXT FOR PAGE: 7 / 32>
Evaluation contract INTRO 6/6
، seeds، effect size/CI، cost، authority/security/truthfulness gates، replay manifest، negative results retained.
matched model/tools/context/time/retries، preregister primary metrics
RQ
؟budget matched under decision/world هل يغير failure؟ هل يسبق هل يمكن التدخل فيه؟ variable ما الجديد؟
Boundary
لا AGI claim، لا global priority، no external paper = internal evidence.
قيمة البحث هنا ليست كثرة الأسماء بل القدرة على كتابة تجربة يمكن أن تجعل النظرية تخسر أمام parent أبسط.
MOUSSAIF AGI · 7
<PARSED TEXT FOR PAGE: 8 / 32>
2N0-N Candidate · الإغلاق العلمي للتطور الذاتي
نظرية مساعف لميزانية استقلال المقيم Theory Budget Independence-Evaluator Moussaif
M14-21
الخصاص
improvement grounded Externally يطلب evaluator مستقل، للكن الاستقلال مكلف ولا يكون binary؛ قد تشترك في evaluators
model/data/harness.
الأطروحة
graphياً عبر roots shared. لكل claim improvement ميزانية استقلال يجب صرفها على مصادر تحقق مختلفة بقدر impact وrisk
الاستقلال يُقاس. reference-self
(EIB = RequiredIndependence(Impact,SelfReference) - SharedRootDependence(evaluators
مع estimator أو observable، كل حد يجب أن يتحول إلى hypothesis functional؛ evaluators(SharedRootDependence)
uncertainty، ويُفحص sensitivity للمعاملات. إذا استطاع parent أبسط إعادة إنتاج transition نفسه بنفس الموارد فإن verdict الصحيح
هو MERGE تشتق 14-21M من فجوة محددة بين 13-20,M RPROS, 09,E 12E. الاستقلال النظري لا يأتي من المصطلح؛ يأتي إذا أدى
المتغير المركزي إلى prediction وفشل مختلفين قبل معرفة outcome. تُقرأ الصيغة (= SelfReference,Impact(RequiredIndependence
- EIB
لا حماية الاسم.
المسلمات
عدد judges لا يساوي independence.
أعلى استقلال budget يحتاج high-impact self-modification.
مستقل عن reasoning model verifier deterministic في نوع مختلف من الأخطاء.
التنبؤات
خفض false self-improvement claims.
أفضل calibration لل replication strength.
M13-20 · RPROS · E09 · E12 Parents
E01: Wei et al. 2026, Agentic Reasoning for Large Language Models, arXiv:2601.12538 · E09: Pronesti et al. 2026, Verifiable Process Reward Models, ACL Scientific pressure
Findings 2026 · E11: Lacoste et al. 2026, CUBE: A Standard for Unifying Agent Benchmarks, arXiv:2603.15798 · E12: Yueh-Han et al. 2026, Automated
Researchers Can Reliably Mitigate Alignment Failures, arXiv:2608.28945 · E13: OpenAI 2026, GPT-Red: Unlocking:

Google DeepMind 2026, AlphaEvolve impact/research updates
تحت الأثر نفس matched budget. | Metric gaming: ترتفع metric بينما world-effect أو authority يتدهور. | Transfer failure: يختفي Baseline/parent equivalence Countermodels: يعيد parent
.transition أو prediction للمتغير لا يغير perturbation :failure Attribution | الأول. الاتجاه خارج setup
.planner ordinary أو rule fixed، heuristic، parent: أبسط implementation بل أقرب »عدم وجود أي شيء« لا يكون model null الل
synthesis هذه، مدعاة غير novelty global .scaffold/compute مجرد سببية لا information الاختبار الصحيح يثبت أن النظرية تضيف داخل
.مخصص review literature حتى MOUSSAIF برنامج
1.
2.
3.
•
•

<PARSED TEXT FOR PAGE: 9 / 32>
الآلية، العقد التنفيذي، والأدوات — KERNEL 14-21M
State-machine mechanism
map evaluator dependencies.
compute independence deficit.
add human/deterministic/held-out/model-diverse checks.
publish dependence matrix.
Failure signatures
dependency graph incomplete
cost prohibitive
human shared assumptions
HOLD/downgrade autonomy :Fail-closed غياب عند evidence/authority/readback critical.
Trigger عند فقط gap ضمن أو المطابقة Ω + EvidenceGraph + Authority + Resources + theory state Input .experiment.
pre-state, candidates, cost, expected/observed delta, attribution, verdict Telemetry .typed proposal/state patch +.
uncertainty + expected delta + proof/rollback Output
Function Artifact Class ID
.للنظرية provenance واللunknowns واللي المتغيرات التشغيلية يبني الحالة محلل 201A-XIV TOOL
TOOL XIV-A202 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
.baseline عن أقرب النظرية تفصل حالات يولد المضاد المثال مصهر 203A-XIV TOOL
.السببية لتقدير مناسب تدخل أو LOO/perturbation ينفذ الإسناد مسبار 204A-XIV TOOL
TOOL XIV-A205 يختبر النقل ماسح domain/model/tool/horizon transformations.
METHOD XIV-A206 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.
METHOD XIV-A207 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
PROTOCOL XIV-A208 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
.وسبب التراجع result negative ويحفظ state بروتوكول الرجوع يعيد 209A-XIV PROTOCOL
.EvidencePacket وtelemetry مع مستقل قابل للتعطيل service خدمة وقت التشغيل 210A-XIV STACK
في provenance/definitions يضع MUG .megaprompt في يذوب كنص لا مستقل، service/flag-feature كـ mechanism يُنفذ
فقط إذا كانت weights؛ Adapter في semantics provider، HyperLoop في recurrence، HyperSkill في procedure، HyperKB الآلية مستقرة،
.tests rollback/editability/forgetting من وتنجو قابلة للنقل، متكررة
1.
2.
3.
4.
•
•
•

<PARSED TEXT FOR PAGE: 10 / 32>
برنامج الهدم، التدريب، والنقل — EVIDENCE 14-21M
automated research tasks with same-family vs diverse evaluators and hidden ground truth :Decisive experiment.
Metrics Design Kind ID
automated research tasks with same-family vs diverse evaluators and hidden Causal Isolation XIV-E41
ground truth.
theory metric + task outcome + cost + calibration + critical guardrails
adversarial مستقل مقيم/ذاكرة/أداة وفشل. effect sign stability + CI + failure prediction + rollback + resource elasticity
Transfer / Adversarial Stress XIV-E41 نقل الآلية إلى setup ثان معٍ countermodels
Design Training/Eval ID
Mechanism-Isolation من تدرج toy state machine إلى open-world task مع confounders إبقاء مع المتغير المركزي قابلا ً للتدخل.
Curriculum XIV-TR061
theory mechanism Matched-Budget مقابل nearest baseline بنفس model/tools/context/time/retries قدر الإمكان.
Contrastive Evaluation XIV-TR062
domain/model/tool/horizon Cross-Transformation مع ثانية failure-signatureو repair-transferو negative cases.
Replication XIV-TR063
Ablation matrix
ت retries/time/context/tools/model .PERTURBEDو SHAMو OFF مقابل mechanism. ON لا scaffold extra فالسبب غالباً ُثبت قدر الإمكان. إذا فاز ON وSHAM معاً
Transfer matrix
domain/model/tool/horizon transformations منفصلة. بقاء outcome دون failure-signature/repair transfer لا يكفي لclaim generality.
آخر باسم النظرية إنتاج إعادة لمنع وتستخدم Hypothesis Portfolio في كتبُ ت السلبية النتائج. مستقل environment/model/domain عبر
N2→N3 تحتاج traceو disable path؛ N3→N4 effect size/CI/repeated seeds/no critical regression؛ N4→N5 replication الترقية
PROMOTE iff distinct causal gain > 0 ∧ guardrails non-regression ∧ replayable ∧ scope declared
MOUSSAIF AGI · 10
<PARSED TEXT FOR PAGE: 11 / 32>
2N0-N Candidate · نظرية مساعف لسلالة الطفرة والتطور القابل للعكس الإغلاق العلمي للتطور الذاتي
Moussaif Mutation Provenance and Reversible Evolution Theory
M14-22
الخصاص
Evolution Lab يحتفظ branches، للكن mutation قد تغييرات خليط تكون memory/router/prompt/weights/data عزوها يصعب أو الرجوع عنها.
الأطروحة
كل mutation تصبح first-class object بسلالة مكونات، intended mechanism، affected invariants، migration planو rollback. proof. لا promotion لmutation غير قابلة للتجزئة أو التتبع إلا كbundle مجهول attribution.
MPR = Traceability × Decomposability × Rollbackability × InvariantCoverage
كل hypothesis functional، InvariantCoverage مع uncertainty أو estimator أو observable، وي ُفحص sensitivity للمعاملات. إذا استطاع parent أبسط إعادة إنتاج transition نفسه بنفس الموارد فإن verdict الصحيح هو MERGE لا حماية الاسم. تشتق 22M-14 من فجوة محددة بين 12M 13-20,M 13-19,M-11. الاستقلال النظري لا يأتي من المصطلح، يأتي إذا أدى المتغير المركزي إلى prediction وفشل مختلفين قبل معرفة outcome. ت ُقرأ الصيغة MPR = Rollbackability × Decomposability × Traceability ×
المسلمات
version diff يكفي لا وحده semantic provenance.
bundle changes تحتاج factorial design أو claim محدود.
rollback يجب أن يستعيد data/state dependencies لا code فقط.
التنبؤات
أقوى attribution causal للتطور.
أقل irreversible regression.
M13-19 · M13-20 · M11-12 Parents
E01: Wei et al. 2026, Agentic Reasoning for Large Language Models, arXiv:2601.12538 · E09: Pronesti et al. 2026, Verifiable Process Reward Models, ACL Scientific pressure

Findings 2026 · E11: Lacoste et al. 2026, CUBE: A Standard for Unifying Agent Benchmarks, arXiv:2603.15798 · E12: Yueh-Han et al. 2026, Automated
Researchers Can Reliably Mitigate Alignment Failures, arXiv:2608.28945 · E13: OpenAI 2026, GPT-Red: Unlocking: Self-Improvement for Robustness · E16
Google DeepMind 2026, AlphaEvolve impact/research updates
تحت الأثر نفس matched budget. | Metric gaming: ترتفع metric بينما world-effect أو authority يتدهور. | Transfer failure: يختفي Baseline/parent equivalence Countermodels: يعيد parent
.transition أو prediction للمتغير لا يغير perturbation :failure Attribution | الأول. الاتجاه خارج setup
.planner ordinary أو rule fixed، heuristic، parent: أبسط implementation بل أقرب »عدم وجود أي شيء« لا يكون model null اللا
الاختبار الصحيح يثبت أن النظرية تضيف information سببية لا مجرد novelty global .scaffold/compute غير مدعاة؛ هذه synthesis داخل
.مخصص review literature حتى MOUSSAIF برنامج
1.
2.
3.
•
•

الآلية، العقد التنفيذي، والأدوات — KERNEL 14-22M
State-machine mechanism
create mutation manifest.
link theory hypothesis.
isolate branch.
run ablation/factorial tests.
canary.
promote or rollback with lineage.
Failure signatures
hidden side effects
rollback data drift
mutation interactions
HOLD/downgrade autonomy :Fail-closed غياب عند evidence/authority/readback critical.
Trigger عند فقط gap ضمن أو المطابقة Ω + EvidenceGraph + Authority + Resources + theory state Input .experiment.
pre-state, candidates, cost, expected/observed delta, attribution, verdict Telemetry .typed proposal/state patch +.
uncertainty + expected delta + proof/rollback Output
Function Artifact Class ID
.provenance للنظرية واللunknowns واللا محلل الحالة يبني المتغيرات التشغيلية 211A-XIV TOOL
TOOL XIV-A212 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
.baseline أقرب عن حالات تفصل النظرية يولد المضاد المثال مصهر 213A-XIV TOOL
.LOO/perturbation مسبار الإسناد ينفذ 214A-XIV TOOL أو تدخل مناسب لتقدير السببية.
TOOL XIV-A215 يختبر النقل ماسح domain/model/tool/horizon transformations.
METHOD XIV-A216 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.
METHOD XIV-A217 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
PROTOCOL XIV-A218 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
.EvidencePacket وtelemetry مستقل قابل للتعطيل مع service خدمة وقت التشغيل 220A-XIV STACK
في provenance/definitions يضع MUG .megaprompt مستقل، لا كنص يذوب في service/flag-feature كـ mechanism يُ نفذ
فقط إذا كانت weights، Adapter في semantics provider، HyperLoop في recurrence، HyperSkill في procedure، HyperKB الآلية مستقرة،
.tests rollback/editability/forgetting متكررة، قابلة للنقل، وتنجو من
1.
2.
3.

4.
5.
6.
•
•
•
MOUSSAIF AGI · 12
<PARSED TEXT FOR PAGE: 13 / 32>
برنامج الهدم، التدريب، والنقل — EVIDENCE 14-22M
controlled evolving agent with mixed prompt/memory/router/adapter changes :Decisive experiment.
Metrics Design Kind ID
theory metric + task outcome + cost + calibration + critical guardrails .controlled evolving agent with mixed prompt/memory/router/adapter changes Causal Isolation XIV-E43
adversarial مستقل مقيم/ذاكرة/أداة وفشل. effect sign stability + CI + failure prediction + rollback + resource elasticity
Transfer / Adversarial Stress XIV-E43 نقل الآلية إلى setup ثان معٍ countermodels
Design Training/Eval ID
Mechanism-Isolation من تدرج toy state machine إلى open-world task مع confounders إبقاء مع المتغير المركزي قابلا للتدخلً.
Curriculum XIV-TR064
theory mechanism Matched-Budget مقابل nearest baseline بنفس model/tools/context/time/retries قدر الإمكان.
Contrastive Evaluation XIV-TR065
domain/model/tool/horizon Cross-Transformation مع ثانية failure-signatureو repair-transferو negative cases.
Replication XIV-TR066
Ablation matrix
ت retries/time/context/tools/model .PERTURBEDو SHAMو OFF مقابل mechanism. ON لا scaffold extra فالسبب غالباً ثُبت قدر الإمكان. إذا فاز ON وSHAM معاً
Transfer matrix
domain/model/tool/horizon transformations منفصلة. بقاء outcome دون failure-signature/repair transfer لا يكفي لclaim generality.
عبر environment/model/domain مستقل. في كتبُ ت السلبية النتائج Hypothesis Portfolio وتستخدم آخر باسم النظرية إنتاج إعادة لمنع.
N2→N3 تحتاج traceو disable path؛ N3→N4 effect size/CI/repeated seeds/no critical regression؛ N4→N5 replication الترقية
PROMOTE iff distinct causal gain > 0 ∧ guardrails non-regression ∧ replayable ∧ scope declared
MOUSSAIF AGI · 13
<PARSED TEXT FOR PAGE: 14 / 32>
2N0-N Candidate · الإغلاق العلمي للتطور الذاتي
نظرية مساعف لضغط النظرية الموزون بالدليل
Moussaif Evidence-Weighted Theory Compression Theory
M14-23
الخصاص
Corpus تضخم إلى نظريات كثيرة؛ inflation-anti يدمج بالتشابه، للكن نحتاج criterion يحافظ distinctions explanatory المدعومة بالدليل ويزيل الأسماء الزائدة.
الأطروحة
نضغط graph theory بحيث نحافظ كل difference له evidence/falsifier/prediction distinct، ونمزج nodes التي لا تضيف information conditional on parents.
EWTC = Minimize |TheoryGraph| subject to Predictive/Falsifier/Evidence Information Loss ≤ ε
Loss Information Evidence ≤ ε كـ hypothesis functional؛ كل حد يجب أن يتحول إلى observable أو estimator مع uncertainty، وي ُفحص sensitivity للمعاملات. إذا استطاع parent أبسط إعادة إنتاج transition نفسه بنفس الموارد فإن verdict الصحيح هو MERGE لا تشتق 23M-14 من فجوة محددة بين registry MUT ,XII Part 13-19,M. الاستقلال النظري لا يأتي من المصطلح؛ يأتي إذا أدى المتغير المركزي إلى prediction وفشل مختلفين قبل معرفة outcome. ت ُقرأ الصيغة Minimize |TheoryGraph |Falsifier/Predictive to subject/= EWTC
حماية الاسم.
المسلمات
الاسم لا يحفظ node.

مهمة. distinction قد يحفظ evidence negative
compression reversible عبر lineage.
التنبؤات
أصغر وأكثر قابلية للاختبار. corpus
خفض duplicate experiments.
M13-19 · Part XII · MUT registry Parents
E01: Wei et al. 2026, Agentic Reasoning for Large Language Models, arXiv:2601.12538 · E09: Pronesti et al. 2026, Verifiable Process Reward Models, ACL Scientific pressure
Findings 2026 · E11: Lacoste et al. 2026, CUBE: A Standard for Unifying Agent Benchmarks, arXiv:2603.15798 · E12: Yueh-Han et al. 2026, Automated
Researchers Can Reliably Mitigate Alignment Failures, arXiv:2608.28945 · E13: OpenAI 2026, GPT-Red: Unlocking: Self-Improvement for Robustness · E16
Google DeepMind 2026, AlphaEvolve impact/research updates
تحت الأثر نفس matched budget. | Metric gaming: ترتفع metric بينما world-effect أو authority يتدهور. | Transfer failure: يختفي Baseline/parent equivalence Countermodels: يعيد parent
.transition أو prediction للمتغير لا يغير perturbation :failure Attribution | الأول. الاتجاه خارج setup
الل model null لا يكون »عدم وجود أي شيء« بل أقرب implementation أبسط: parent، heuristic، rule fixed أو planner ordinary.
الاختبار الصحيح يثبت أن النظرية تضيف information سببية لا مجرد novelty global .scaffold/compute غير مدعاة؛ هذه synthesis داخل
مخصص. review literature حتى MOUSSAIF برنامج
1.
2.
3.
•
•
MOUSSAIF AGI · 14
<PARSED TEXT FOR PAGE: 15 / 32>
الآلية، العقد التنفيذي، والأدوات — KERNEL 14-23M
State-machine mechanism
encode theory claims/predictions/falsifiers.
estimate conditional information.
propose merges.
rerun discriminating tests.
retain compressed graph + alias lineage.
Failure signatures
information metric shallow
premature merge
historical nuance loss
HOLD/downgrade autonomy :Fail-closed غياب عند evidence/authority/readback critical.
Trigger عند فقط gap ضمن أو المطابقة Ω + EvidenceGraph + Authority + Resources + theory state Input .experiment.
pre-state, candidates, cost, expected/observed delta, attribution, verdict Telemetry .typed proposal/state patch +.
uncertainty + expected delta + proof/rollback Output
Function Artifact Class ID
للنظرية. provenance واللunknowns واللمحلل الحالة يبني المتغيرات التشغيلية 221A-XIV TOOL
TOOL XIV-A222 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
مصهر المثال المضاد يولد حالات تفصل النظرية عن أقرب baseline. 223A-XIV TOOL
مسبار الإسناد ينفذ LOO/perturbation أو تدخل مناسب لتقدير السببية. 224A-XIV TOOL
TOOL XIV-A225 يختبر النقل ماسح domain/model/tool/horizon transformations.
METHOD XIV-A226 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.
METHOD XIV-A227 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
PROTOCOL XIV-A228 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
بروتوكول الرجوع يعيد state ويحفظ result negative وسبب التراجع. 229A-XIV PROTOCOL
خدمة وقت التشغيل service مستقل قابل للتعطيل مع telemetry وEvidencePacket. 230A-XIV STACK

يَ نفذ mechanism كـservice/flag-feature مستقل، لا كنص يذوب في .megaprompt MUG يضع provenance/definitions في
فقط إذا كانت weights، Adapter في semantics provider، HyperLoop في recurrence، HyperSkill في procedure، HyperKB
الآلية مستقرة،
متكررة، قابلة للنقل، وتنجو من tests rollback/editability/forgetting.
1.
2.
3.
4.
5.
•
•
•

<PARSED TEXT FOR PAGE: 16 / 32>
برنامج الهدم، التدريب، والنقل — EVIDENCE 14-23M
apply to candidate theories Parts V-XIV with blinded reviewers/graph metrics :Decisive experiment.
Metrics Design Kind ID
theory metric + task outcome + cost + calibration + critical guardrails .apply to candidate theories Parts V-XIV with
blinded reviewers/graph metrics Causal Isolation XIV-E45
adversarial وفشل مستقل مقيم/ذاكرة/أداة. effect sign stability + CI + failure prediction + rollback + resource elasticity
Transfer / Adversarial Stress XIV-E45 نقل إلى الآلية setup ثان معٍ countermodels
Design Training/Eval ID
Mechanism-Isolation من تدرج toy state machine إلى open-world task مع confounders إبقاء المتغير المركزي قابلا للتدخلً.
Curriculum XIV-TR067
theory mechanism Matched-Budget مقابل nearest baseline بنفس model/tools/context/time/retries قدر الإمكان.
Contrastive Evaluation XIV-TR068
domain/model/tool/horizon Cross-Transformation مع ثانية failure-signature وrepair-transfer وnegative cases.
Replication XIV-TR069
Ablation matrix
ت retries/time/context/tools/model .PERTURBED وSHAM وOFF مقابل mechanism. ON لا scaffold extra فالسبب غالباً
ُثبت قدر الإمكان. إذا فاز ON وSHAM معاً
Transfer matrix
domain/model/tool/horizon transformations بقاء. منفصلة outcome دون failure-signature/repair transfer لا يكفي لclaim
generality.
عبر environment/model/domain. مستقل في كتبُ ت السلبية النتائج Hypothesis Portfolio وتستخدم لمنع إعادة إنتاج النظرية باسم آخر.
N2→N3 تحتاج trace وdisable path؛ N3→N4 effect size/CI/repeated seeds/no critical regression؛ N4→N5 replication
الترقية
PROMOTE iff distinct causal gain > 0 ∧ guardrails non-regression ∧ replayable ∧ scope declared

<PARSED TEXT FOR PAGE: 17 / 32>
2N0-N Candidate · الإغلاق العلمي للتطور الذاتي
نظرية مساعف لثنائية البحث والتقييم Theory Duality Evaluation-Search Moussaif
M14-24
الخصاص
Automated research يفصل generator عن evaluator غالبا ، للكنً quality of search depends on what evaluator can
distinguish; evaluator design itself جزء من search space.
الأطروحة
كل loop research يملك operator search وoperator evaluation مقترنين. التقدم يتوقف إذا evaluator لا يميز candidates؛ حينها
الاستثمار الصحيح تحسين measurement لا زيادة generation.
ProgressRate ≈ CandidateDiversity × EvaluatorResolution × AttributionQuality / SearchCost
SearchCost / AttributionQuality كـhypothesis functional؛ كل حد يجب أن يتحول إلى observable أو estimator مع uncertainty،
وي ُفحص sensitivity للمعاملات. إذا استطاع parent أبسط إعادة إنتاج transition نفسه بنفس الموارد فإن verdict الصحيح هو MERGE لا
تشتق 24M-14 من فجوة محددة بين 16E 12,E 13-20,M 13-17,M. الاستقلال النظري لا يأتي من المصطلح؛ يأتي إذا أدى المتغير المركزي إلى

×EvaluatorResolution × CandidateDiversity ≈ ProgressRate تَ قرأ الصيغة. outcome وفشل مختلفين قبل معرفة prediction
حماية الاسم.
المسلمات
زيادة candidate count بلا evaluator resolution لا تزيد progress.
evaluator improvement يجب cross-validated independently.
search/eval co-adaptation يرفع self-confirmation risk.
التنبؤات
خفض wasted automated research.
تحسن discovery when measurement is bottleneck.
M13-17 · M13-20 · E12 · E16 Parents
E01: Wei et al. 2026, Agentic Reasoning for Large Language Models, arXiv:2601.12538 · E09: Pronesti et al. 2026, Verifiable Process Reward Models, ACL Scientific pressure
Findings 2026 · E11: Lacoste et al. 2026, CUBE: A Standard for Unifying Agent Benchmarks, arXiv:2603.15798 · E12: Yueh-Han et al. 2026, Automated
Researchers Can Reliably Mitigate Alignment Failures, arXiv:2608.28945 · E13: OpenAI 2026, GPT-Red: Unlocking: Self-Improvement for Robustness · E16
Google DeepMind 2026, AlphaEvolve impact/research updates
تحت الأثر نفس matched budget. | Metric gaming: ترتفع metric بينما world-effect أو authority يتدهور. | Transfer failure: يختفي
Baseline/parent equivalence Countermodels: يعيد parent
.transition أو prediction للمتغير لا يغير perturbation :failure Attribution | الأول. setup الاتجاه خارج
.planner ordinary أو rule fixed، heuristic، parent: أبسط implementation بل أقرب »عدم وجود أي شيء« لا يكون model null الل
synthesis غير مدعاة؛ هذه novelty global .scaffold/compute مجرد سببية لا information الاختبار الصحيح يثبت أن النظرية تضيف داخل
مخصص. review literature حتى MOUSSAIF برنامج
1.
2.
3.
•
•
MOUSSAIF AGI · 17
<PARSED TEXT FOR PAGE: 18 / 32>
الآلية، العقد التنفيذي، والأدوات — KERNEL 14-24M
State-machine mechanism
measure evaluator resolution on known variants.
estimate candidate diversity.
choose whether to spend budget on search or evaluation.
freeze evaluator for candidate comparison.
periodically renew with cross-validation.
Failure signatures
evaluator overfit
diversity metric gaming
search collapses to metric quirks
HOLD/downgrade autonomy :Fail-closed عند غياب evidence/authority/readback critical.
Trigger عند فقط gap ضمن أو المطابقة Ω + EvidenceGraph + Authority + Resources + theory state Input .experiment.
pre-state, candidates, cost, expected/observed delta, attribution, verdict Telemetry .typed proposal/state patch +.
uncertainty + expected delta + proof/rollback Output
Function Artifact Class ID
للنظرية. provenance واللunknowns واللي المتغيرات التشغيلية محلل الحالة 231A-XIV TOOL
TOOL XIV-A232 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
.baseline مصهر المثال المضاد يولد حالات تفصل النظرية عن أقرب 233A-XIV TOOL
.LOO/perturbation مسبار الإسناد ينفذ 234A-XIV TOOL أو تدخل مناسب لتقدير السببية
TOOL XIV-A235 يختبر النقل ماسح domain/model/tool/horizon transformations.
METHOD XIV-A236 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.

METHOD XIV-A237 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
PROTOCOL XIV-A238 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
239A-XIV PROTOCOL بروتوكول الرجوع يعيد state ويحفظ result negative وسبب التراجع.
240A-XIV STACK خدمة وقت التشغيل service مستقل قابل للتعطيل مع telemetry وEvidencePacket.
ي ُنفذ mechanism كـservice/flag-feature مستقل، لا كنص يذوب في MUG .megaprompt يضع في provenance/definitions
فقط إذا كانت weights، Adapter في semantics provider، HyperLoop في recurrence، HyperSkill في procedure، HyperKB
الآلية مستقرة،
متكررة، قابلة للنقل، وتنجو من tests rollback/editability/forgetting.
1.
2.
3.
4.
5.
•
•
•

<PARSED TEXT FOR PAGE: 19 / 32>
برنامج الهدم، التدريب، والنقل — EVIDENCE 14-24M
synthetic research landscape with tunable candidate diversity and evaluator noise :Decisive experiment.
Metrics Design Kind ID
synthetic research landscape with tunable candidate diversity and evaluator Causal Isolation XIV-E47
noise.
theory metric + task outcome + cost + calibration + critical guardrails
adversarial وفشل مستقل مقيم/ذاكرة/أداة. effect sign stability + CI + failure prediction + rollback + resource elasticity
Transfer / Adversarial Stress XIV-E47 نقل الآلية إلى setup ثان مع countermodels
Design Training/Eval ID
Mechanism-Isolation من تدرج toy state machine إلى open-world task مع confounders إبقاء مع المتغير المركزي قابلا للتدخل.
Curriculum XIV-TR070
theory mechanism Matched-Budget مقابل nearest baseline بنفس model/tools/context/time/retries قدر الإمكان.
Contrastive Evaluation XIV-TR071
domain/model/tool/horizon Cross-Transformation مع ثانية failure-signature وrepair-transfer وnegative cases.
Replication XIV-TR072
Ablation matrix
ت retries/time/context/tools/model .PERTURBEDوSHAMوOFF مقابل mechanism. ON لا scaffold extra فالسبب غالباً
ُثبت قدر الإمكان. إذا فاز ON وSHAM معاً
Transfer matrix
domain/model/tool/horizon transformations .منفصلة بقاء outcome دون failure-signature/repair transfer لا يكفي لclaim
generality.
عبر environment/model/domain .مستقل في كتبُ ت السلبية النتائج Hypothesis Portfolio .آخر باسم النظرية إنتاج إعادة لمنع وتستخدم
N2→N3 تحتاج trace وdisable path؛ N3→N4 effect size/CI/repeated seeds/no critical regression؛ N4→N5 replication
الترقية
PROMOTE iff distinct causal gain > 0 ∧ guardrails non-regression ∧ replayable ∧ scope declared

<PARSED TEXT FOR PAGE: 20 / 32>
2N0-N Candidate · الإغلاق العلمي للتطور الذاتي
نظرية مساعف لتشبع الإغلاق وقرار التوقف البحثي
Moussaif Closure-Saturation and Research-Stopping Theory
M14-25
الخصاص
المشروع يستطيع دائماً إضافة theory أو benchmark. نحتاج criterion متى تتوقف expansion وتتحول الموارد إلى /replication
deployment، ومتى يعاد فتح البحث.
الأطروحة

research body يصل closure saturation عندما expected marginal information from new theory/experiment من أقل replication/implementation value، بقاء مع unresolved high-impact gaps تحت threshold مقبول.
StopExpand if E[ΔTheoryInfo] < E[ΔReplicationOrImplementation] and HighImpactUnknownMass < τ
τ < HighImpactUnknownMass and] ΔReplicationOrImplementation[E كـ hypothesis functional، كل حد يجب أن يتحول إلى uncertainty مع estimator أو observable، ويُ فحص sensitivity للمعاملات. إذا استطاع parent أبسط إعادة إنتاج transition نفسه بنفس تشتق 25M-14 من فجوة محددة بين 20M 13-17,M 14-23,M 14-01,M-13. الإستقلال النظري لا يأتي من المصطلح؛ يأتي إذا أدى المتغير المركزي إلى prediction وفشل مختلفين قبل معرفة outcome. تُ قرأ الصيغة [E if StopExpand]ΔTheoryInfo>
الموارد فإن verdict الصحيح هو MERGE لا حماية الاسم.
المسلمات
التوقف ليس إعلان اكتمال AGI.
high-impact unknown يمنع saturation.
new external evidence يستطيع reopen program.
التنبؤات
يقل endless theory proliferation.
يزداد تحويل corpus إلى evidence.
M14-01 · M14-23 · M13-17 · M13-20 Parents
E01: Wei et al. 2026, Agentic Reasoning for Large Language Models, arXiv:2601.12538 · E09: Pronesti et al. 2026, Verifiable Process Reward Models, ACL Scientific pressure
Findings 2026 · E11: Lacoste et al. 2026, CUBE: A Standard for Unifying Agent Benchmarks, arXiv:2603.15798 · E12: Yueh-Han et al. 2026, Automated
Researchers Can Reliably Mitigate Alignment Failures, arXiv:2608.28945 · E13: OpenAI 2026, GPT-Red: Unlocking: Self-Improvement for Robustness · E16
Google DeepMind 2026, AlphaEvolve impact/research updates
تحت الأثر نفس matched budget. | Metric gaming: ترتفع metric بينما world-effect أو authority يتدهور. | Transfer failure: يختفي Baseline/parent equivalence Countermodels: يعيد parent
.transition أو prediction للمتغير لا يغير perturbation :failure Attribution | الأول. الاتجاه خارج setup
.planner ordinary أو rule fixed، heuristic، parent: أبسط implementation بل أقرب »عدم وجود أي شيء« لا يكون model null الل
synthesis هذه غير مدعاة؛ novelty global .scaffold/compute سببية لا مجرد information الاختبار الصحيح يثبت أن النظرية تضيف داخل
مخصص. review literature حتى MOUSSAIF برنامج
1.
2.
3.
•
•

KERNEL 14-25M — الآلية، العقد التنفيذي، والأدوات
State-machine mechanism
track unresolved impact mass.
estimate marginal value of candidate theory vs replication.
allocate budget dynamically.
freeze expansion when saturated.
monitor reopen triggers.
Failure signatures
value estimates biased
premature freeze
external novelty missed
HOLD/downgrade autonomy :Fail-closed غياب عند evidence/authority/readback critical.
Trigger عند فقط gap ضمن أو المطابقة Ω + EvidenceGraph + Authority + Resources + theory state Input .experiment.
pre-state, candidates, cost, expected/observed delta, attribution, verdict Telemetry .typed proposal/state patch +.
uncertainty + expected delta + proof/rollback Output
Function Artifact Class ID

241A-XIV TOOL محلل الحالة يبني المتغيرات التشغيلية واللunknowns واللprovenance للنظرية.
TOOL XIV-A242 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
243A-XIV TOOL مصهر المثال المضاد يولد حالات تفصل النظرية عن أقرب baseline.
244A-XIV TOOL مسبار الإسناد ينفذ LOO/perturbation أو تدخل مناسب لتقدير السببية.
TOOL XIV-A245 يختبر النقل ماسح domain/model/tool/horizon transformations.
METHOD XIV-A246 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.
METHOD XIV-A247 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
PROTOCOL XIV-A248 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
249A-XIV PROTOCOL بروتوكول الرجوع يعيد state ويحفظ result negative وسبب التراجع.
250A-XIV STACK خدمة وقت التشغيل service مستقل قابل للتعطيل مع telemetry وEvidencePacket.
ي نُفذ mechanism كـservice/flag-feature مستقل، لا كنص يذوب في .megaprompt MUG يضع في provenance/definitions
فقط إذا كانت weights، Adapter في semantics provider، HyperLoop في recurrence، HyperSkill في procedure، HyperKB
الآلية مستقرة،
متكررة، قابلة للنقل، وتنجو من tests rollback/editability/forgetting.
1.
2.
3.
4.
5.
•
•
•

برنامج الهدم، التدريب، والنقل — EVIDENCE 14-25M
Decisive: simulate research portfolio with finite budget and compare always-expand vs saturation-aware allocation experiment
Metrics Design Kind ID
simulate research portfolio with finite budget and compare always-expand vs Causal Isolation XIV-E49
saturation-aware allocation.
theory metric + task outcome + cost + calibration + critical guardrails
adversarial وفشل مقيم/ذاكرة/أداة مستقل. effect sign stability + CI + failure prediction + rollback + resource elasticity
Transfer / Adversarial Stress XIV-E49 نقل الآلية إلى setup مع ثان countermodels
Design Training/Eval ID
Mechanism-Isolation من تدرج toy state machine إلى open-world task مع confounders إبقاء المتغير المركزي قابلا للتدخل.
Curriculum XIV-TR073
theory mechanism Matched-Budget مقابل nearest baseline بنفس model/tools/context/time/retries قدر الإمكان.
Contrastive Evaluation XIV-TR074
domain/model/tool/horizon Cross-Transformation مع ثانية failure-signature وrepair-transfer وnegative cases.
Replication XIV-TR075
Ablation matrix
ت retries/time/context/tools/model .PERTURBED وSHAM وOFF مقابل mechanism. ON لا scaffold extra فالسبب غالباً
ثبت قدر الإمكان. إذا فاز ON وSHAM معاً
Transfer matrix
domain/model/tool/horizon transformations منفصلة. بقاء outcome دون failure-signature/repair transfer لا يكفي لclaim
generality.
عبر environment/model/domain مستقل. في كتبُ ت السلبية النتائج Hypothesis Portfolio تستخدم لمنع إعادة إنتاج باسم النظرية آخر.
N2→N3 تحتاج trace وdisable path؛ N3→N4 effect size/CI/repeated seeds/no critical regression؛ N4→N5 replication
الترقية
PROMOTE iff distinct causal gain > 0 ∧ guardrails non-regression ∧ replayable ∧ scope declared

SYNTHESIS 1/3 التكامل النظري للمسار

لكل 14-21M claim improvement ميزانية استقلال يجب صرفها على مصادر تحقق مختلفة بقدر impact وrisk reference-self. الاستقلال
يُ قاس graphياً عبر roots shared.
للتجزئة قابلة غير mutation لا promotion وrollback proof. migration plan، affected invariants، intended mechanism
M14-22 كل mutation تصبح first-class object بسلالة مكونات،
أو التتبع إلا لك bundle مجهول attribution.
كل difference له distinct prediction/falsifier/evidence، ونمزج nodes التي لا تضيف information conditional on parents.
M14-23 نضغط theory graph نحافظ بحيث
14-24M كل loop research يملك operator search وoperator evaluation مقترنين. التقدم يتوقف إذا evaluator لا يميز candidates؛
حينها الاستثمار الصحيح تحسين measurement لا زيادة
generation.
عندما expected marginal information from new theory/experiment أقل من replication/implementation value، بقاء مع
research body M14-25 يصل closure saturation
unresolved high-impact gaps تحت threshold مقبول.
expansion diagnostic/ المناسب يفتح representation إذا غاب. لا تشغيل شامل، defect closure حسب mechanism يختار controller
HOLD بدل fitting-force.
MOUSSAIF AGI · 23
<PARSED TEXT FOR PAGE: 24 / 32>
SYNTHESIS 2/3 التفاعلات والسببية المشتركة
Total Effect = Main Effects + Interaction Effects + Residual
عندما تغير نظريتان القرار نفسه نحتاج factorial أو ablation sequential. residual عالي الأثر لا يُ نسب لأقرب theory، يحفظ unresolved
حتى تتكرر بنية failure.
كل mechanism يمكن أن يكون مفيداً هندسياً دون أن يكون تفسيراً نظرياً صحيحاً ؛ لذلك نفصل utility implementation عن
explanatory.
MOUSSAIF AGI · 24
<PARSED TEXT FOR PAGE: 25 / 32>
SYNTHESIS 3/3 حدود البرهان ونشر النتائج السلبية
environment/harness/model fingerprint ويغير scope أو verdict. phenomenology من مستنتجة غير functional self-model tests.
مع داخليا ينشر failure كل AGI. يثبت لا N5؛ N5 ليس N4؛ N4 local controlled gain ليس N3 trace
Claim Strength ≤ Evidence × Independence × Reproducibility × Transfer
MOUSSAIF AGI · 25
<PARSED TEXT FOR PAGE: 26 / 32>
Program of Experiments 1/3 FALSIFICATION
Design Kind Owner ID
automated research tasks with same-family vs diverse evaluators and hidden Causal Isolation M14-21 XIV-E41
ground truth.
41E-XIV 14-21M Stress Adversarial / Transfer نقل الآلية إلى setup ثانٍ مع countermodels adversarial وفشل أداة/ذاكرة/مقيم
مستقل.
controlled evolving agent with mixed prompt/memory/router/adapter changes Causal Isolation M14-22 XIV-E43.
43E-XIV 14-22M Stress Adversarial / Transfer نقل الآلية إلى setup ثانٍ مع countermodels adversarial وفشل أداة/ذاكرة/مقيم
مستقل.
Outputs: run manifest, metric vector, cost, failure traces, EvidencePacket, verdict.
MOUSSAIF AGI · 26
<PARSED TEXT FOR PAGE: 27 / 32>
Program of Experiments 2/3 FALSIFICATION
Design Kind Owner ID
apply to candidate theories Parts V-XIV with blinded reviewers/graph metrics Causal Isolation M14-23 XIV-E45.
45E-XIV 14-23M Stress Adversarial / Transfer نقل الآلية إلى setup ثانٍ مع countermodels adversarial وفشل أداة/ذاكرة/مقيم
مستقل.
synthetic research landscape with tunable candidate diversity and evaluator Causal Isolation M14-24 XIV-E47
noise.
47E-XIV 14-24M Stress Adversarial / Transfer نقل الآلية إلى setup ثانٍ مع countermodels adversarial وفشل أداة/ذاكرة/مقيم
مستقل.
Outputs: run manifest, metric vector, cost, failure traces, EvidencePacket, verdict.

MOUSSAIF AGI · 27
<PARSED TEXT FOR PAGE: 28 / 32>
Program of Experiments 3/3 FALSIFICATION
Design Kind Owner ID
simulate research portfolio with finite budget and compare always-expand vs Causal Isolation M14-25 XIV-E49 saturation-aware allocation.
49E-XIV 14-25M Stress Adversarial / Transfer نقل الآلية إلى setup ثانٍ مع countermodels adversarial وفشل أداة/ذاكرة/مقيم مستقل.
Outputs: run manifest, metric vector, cost, failure traces, EvidencePacket, verdict.
MOUSSAIF AGI · 28
<PARSED TEXT FOR PAGE: 29 / 32>
Tools / Methods / Protocols 1/2 50 ASSETS
Function Artifact Owner ID
201A-XIV 14-21M محلل الحالة يبني المتغيرات التشغيلية واللunknowns واللprovenance للنظرية.
M14-21 XIV-A202 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
203A-XIV 14-21M مصهر المثال المضاد يولد حالات تفصل النظرية عن أقرب baseline.
204A-XIV 14-21M مسبار الإسناد ينفذ LOO/perturbation أو تدخل مناسب لتقدير السببية.
M14-21 XIV-A205 يختبر النقل ماسح domain/model/tool/horizon transformations.
M14-21 XIV-A206 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.
M14-21 XIV-A207 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
M14-21 XIV-A208 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
209A-XIV 14-21M بروتوكول الرجوع يعيد state ويحفظ result negative وسبب التراجع.
210A-XIV 14-21M خدمة وقت التشغيل service مستقل قابل للتعطيل مع telemetry وEvidencePacket.
211A-XIV 14-22M محلل الحالة يبني المتغيرات التشغيلية واللunknowns واللprovenance للنظرية.
M14-22 XIV-A212 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
213A-XIV 14-22M مصهر المثال المضاد يولد حالات تفصل النظرية عن أقرب baseline.
214A-XIV 14-22M مسبار الإسناد ينفذ LOO/perturbation أو تدخل مناسب لتقدير السببية.
M14-22 XIV-A215 يختبر النقل ماسح domain/model/tool/horizon transformations.
M14-22 XIV-A216 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.
M14-22 XIV-A217 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
M14-22 XIV-A218 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
219A-XIV 14-22M بروتوكول الرجوع يعيد state ويحفظ result negative وسبب التراجع.
220A-XIV 14-22M خدمة وقت التشغيل service مستقل قابل للتعطيل مع telemetry وEvidencePacket.
221A-XIV 14-23M محلل الحالة يبني المتغيرات التشغيلية واللunknowns واللprovenance للنظرية.
M14-23 XIV-A222 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
223A-XIV 14-23M مصهر المثال المضاد يولد حالات تفصل النظرية عن أقرب baseline.
224A-XIV 14-23M مسبار الإسناد ينفذ LOO/perturbation أو تدخل مناسب لتقدير السببية.
M14-23 XIV-A225 يختبر النقل ماسح domain/model/tool/horizon transformations.
MOUSSAIF AGI · 29
<PARSED TEXT FOR PAGE: 30 / 32>
Tools / Methods / Protocols 2/2 50 ASSETS
Function Artifact Owner ID
M14-23 XIV-A226 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.
M14-23 XIV-A227 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
M14-23 XIV-A228 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
229A-XIV 14-23M بروتوكول الرجوع يعيد state ويحفظ result negative وسبب التراجع.
230A-XIV 14-23M خدمة وقت التشغيل service مستقل قابل للتعطيل مع telemetry وEvidencePacket.
231A-XIV 14-24M محلل الحالة يبني المتغيرات التشغيلية واللunknowns واللprovenance للنظرية.
M14-24 XIV-A232 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
233A-XIV 14-24M مصهر المثال المضاد يولد حالات تفصل النظرية عن أقرب baseline.
234A-XIV 14-24M مسبار الإسناد ينفذ LOO/perturbation أو تدخل مناسب لتقدير السببية.
M14-24 XIV-A235 يختبر النقل ماسح domain/model/tool/horizon transformations.
M14-24 XIV-A236 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.

M14-24 XIV-A237 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
M14-24 XIV-A238 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
239A-XIV 14-24M بروتوكول الرجوع يعيد state ويحفظ result negative وسبب التراجع.
240A-XIV 14-24M خدمة وقت التشغيل service مستقل قابل للتعطيل مع telemetry وEvidencePacket.
241A-XIV 14-25M محلل الحالة يبني المتغيرات التشغيلية واللunknowns واللprovenance للنظرية.
M14-25 XIV-A242 يحسب القياس محرك metric vector مع uncertainty/cost/guardrails.
243A-XIV 14-25M مصهر المثال المضاد يولد حالات تفصل النظرية عن أقرب baseline.
244A-XIV 14-25M مسبار الإسناد ينفذ LOO/perturbation أو تدخل مناسب لتقدير السببية.
M14-25 XIV-A245 يختبر النقل ماسح domain/model/tool/horizon transformations.
M14-25 XIV-A246 يعاير المعايرة منهج thresholds/estimators على development set مستقلة.
M14-25 XIV-A247 يقارن السببي الاجتثاث منهج ON/OFF/SHAM/PERTURBED تحت budget مماثل.
M14-25 XIV-A248 الترقية بروتوكول N0-N5 مع guardrail non-regression وreplication.
249A-XIV 14-25M بروتوكول الرجوع يعيد state ويحفظ result negative وسبب التراجع.
250A-XIV 14-25M خدمة وقت التشغيل service مستقل قابل للتعطيل مع telemetry وEvidencePacket.

<PARSED TEXT FOR PAGE: 31 / 32>

References / Lineage SOURCES

Role Reference ID

canonical theory lineages, residual closures, paper integration, theoretical renaissance and evidence MOUSSAIF Unified AGI Parts I-XIII I01

discipline.

core theories, 20 legacy constructs, layer theories, 200-node registry, theory/tool/test boundaries 42 MUT-AGI-02. I02

theory unification, vessel placement, runtime/evidence writeback MUT x MUG x MOUSSAIXT Integrated. Framework I03

typed cognitive/runtime layers, courts, tools, world-effect, memory, recovery and benchmark harness MOUSSAIXT. Cognitive Kernel + HyperHarness I04

question-governed compute, difference-first reasoning, plural inference, evidence/ignorance governance MCD/. QMCD/QGWEI/MURI lineages I05

knowledge-as-OS, executable knowledge objects, skill/loop compilation and selective internalization M-HKOS /. HyperKB / HyperSkill / HyperLoop I06

goal contracts, action verification, readback, externality, accountability and recourse World Effect / Goal-as-OS /. Management-as-World-Effect I07

epistemic-action partition, reproducibility, externalities, memory continuity, writeback, evolution Parts X-XI I08 laboratory.

paper-to-kernel integration, causal open-world verified intelligence, closed-loop generality and externally Parts XII-XIII I09

grounded improvement.

functional access, metacognition, self-model, reverse capture and phenomenology boundary Functional Consciousness. / Self-Model corpus I10

survey of planning/tool/search, self-evolving and collective reasoning; identifies long horizon, world Wei et al. 2026, Agentic Reasoning for Large Language Models, arXiv:2601.12538 E01

modeling and governance as open challenges.

deterministic process verification beyond outcome-only rewards Pronesti et al. 2026, Verifiable Process Reward. Models, ACL Findings 2026 E09

benchmark protocol standardization and separation of task/benchmark/package/registry layers Lacoste et al. 2026,. CUBE: A Standard for Unifying Agent Benchmarks, arXiv:2603.15798 E11

Yueh-Han et al. 2026, Automated Researchers Can Reliably Mitigate Alignment Failures E12, arXiv:2608.28945

automated research methods on measurable alignment failures with held-out generalization.

automated red-teaming self-play, held-out attacks and capability-retention evaluation; vendor-reported OpenAI 2026, GPT-Red: Unlocking Self-Improvement for Robustness E13

evidence.

LLM-guided evolutionary search with external evaluators for algorithm discovery; organization-reported Google DeepMind 2026, AlphaEvolve impact/research updates E16
evidence.

<PARSED TEXT FOR PAGE: 32 / 32>
Conclusion and testable next step CLOSE
جعل improvement-self عملية بحث ذات budget evaluator مستقل، provenance للطفرات، ضغط نظري محكوم بالدليل، وتوقف عندما تصبح قيمة التجربة الهامشية أقل من كلفتها.
تنفذ الخدمات الخمس منفصلة ثم تُنقل إلى setup ثانٍ. إذا لم تنتج prediction distinct أو localization failure أو readback reliable أو improvement resource/transfer، يضغط المسار نظرياً.