RECONSTRUCTED_FROM_INDEXED_TEXT - نسخة أرشيفية معاد بناؤها من النص المفهرس

<PARSED TEXT FOR PAGE: 1 / 40>
MOUSSAIF UNIFIED AGI RESEARCH PROGRAM / 05 September 2026
نظرية التحسن العودي الحامل للبرهان وإغلاق البحث
العلمي
(Moussaif Proof-Carrying Recursive Improvement & Scientific Closure Theory (M-PCRI -
Dense formal research paper
Evidence-ranked Research Candidate Amiri Arabic RTL
زهير مساعف / Moussaif Zouhair
بنية افتراضية دقيقة قابلة للهدم والقياس. دمج داخلي لمخرجات MOUSSAIXT/HKOS-M/MUG/MUT مع قيود ونتائج المجتمع العلمي، من دون تحويل art prior إلى إثبات للمشروع.
1
<PARSED TEXT FOR PAGE: 2 / 40>
P14-E / M-PCRI / PAGE 002 / 05 September 2026
Recursive Carrying-Proof Moussaif - المشكلة المركزية وحدود الادعاء
(Improvement & Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
receipts effect ومسار rollback. في هذه الصفحة يختبر التركيز الخاص بالمشكلة المركزية وحدود الادعاء.
1. الكائن النظري والدقة الافتراضية
؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو المشكلة المركزية وحدود الادعاء، مع منع الخلط بين schema نظري ونتيجة تجريبية.
Promote(c) ⇔ ΔMetric>τ ∧ CriticalRegression=0 ∧ ProofValid ∧ AblationSurvives ∧ Replication≥k
المعادلة هنا object research formal وليست رقماً تجريبياً جاهزاً. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
APB 2026
Agent Planning Benchmark: 4,209 multimodal cases across 22 domains; planning,
broken tools, unsolvable tasks, refinement.
arXiv:2606.04874
AMA-Bench 2026
Long-horizon agent memory benchmark; highlights missing causality/objective
information and lossy similarity retrieval.
arXiv:2602.22769
AgentLAB 2026
Long-horizon attack benchmark covering intent hijacking, tool chaining, task
injection, objective drift and memory poisoning.
arXiv:2602.16901
RoadmapBench 2026
Long-horizon software-development benchmark over real version upgrades; exposes
multi-file, multi-target planning limits.
arXiv:2605.15846
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفياً محتملاً.
3. الربط الداخلي
MUT: theory unification, lineage, relation graph, research questions and experiment
candidates.
MUG: governance of placement, internalization, authority, evolution and evidence
status.

MOUSSAIXT: cross-environment execution shell binding models, tools, memory,
agents, verification and local/cloud runtimes.
MCD / QMCD / MCAI / HAKIM-MCD: difference-first, question-conditioned,
plural-path and falsification-oriented cognitive routing.
4. التجربة الحاسمة
candidate promotion؛ جمد model/tool/context، سجل raw trajectory، افصل process score عن outcome score، التجربة أعد ثم
تجربة صمم matched-budget حول tournament
مع حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرارٍ.

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عامًا.
5. شرط الهدم
إذا لم تقلل proof-carrying promotion regressions أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة
بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule،
إنشاء automatic rollback and scientific verdict
ledger. في محور المشكلة المركزية وحدود الادعاء، مع schema typed وid_trace وpointers evidence/source وreceipt rollback.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
2
<PARSED TEXT FOR PAGE: 3 / 40>
P14-E / M-PCRI / PAGE 003 / 05 September 2026
MUT/MUG/MOUSSAIXT - Moussaif Proof-Carrying Recursive إلى النسب
(Improvement & Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
MOUSSAIXT/MUG/MUT إلى بالنسب الخاص التركيز يختبر الصفحة هذه في. rollback ومسار receipts effect.
1. الكائن النظري والدقة الافتراضية
Paper-to-Kernel pipeline؛ promotion يقرر MUG؛ externalities/recourse يقيسان M-XEOS/CAROS تهاجم؛ Courts ؛
run يثبت M-RPROS؛ candidate يولد Evolution Lab
MOUSSAIXT/MUG/MUT إلى النسب هنا هو التركيز. البحث دورة يغلق، مع منع الخلط بين schema نظري ونتيجة تجريبية.
ΔState = State_after − State_before ; VerifiedSuccess ⇔ Postcondition(State_after)=true
object research formal هنا المعادلة وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو
scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
RoadmapBench 2026
Long-horizon software-development benchmark over real version upgrades; exposes
multi-file, multi-target planning limits.
arXiv:2605.15846
MM-ToolSandBox 2026
Stateful visual tool-calling across 500+ tools and 16 domains; visual precision
becomes a major bottleneck.

arXiv:2607.11818
World Model Agentic Mid-Training 2026
Capability-first world-model planning; distinguishes predictive capability from merely imitating foresight format.
arXiv:2606.27483
WorldEvolver 2026
Deployment-time world-model revision via episodic memory, semantic memory and selective foresight.
arXiv:2606.30639
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملاً.
3. الربط الداخلي
MCD / QMCD / MCAI / HAKIM-MCD: difference-first, question-conditioned, plural-path and falsification-oriented cognitive routing.
M-HKOS / KB-as-OS: knowledge as governed operational substrate, not passive archive.
HyperSkill -> HyperAbility -> Bounded HyperLoop -> Agent/Tool -> World Effect -< Self-Reconstruction Memory.
M-EAPOS / M-RPROS / M-XEOS / M-CAROS / M-MCBOS: action-state separation, reproducibility, externalities, accountability and multi-currency budgets.
4. التجربة الحاسمة
مع التجربة أعد ثم ،outcome score عن process score افصل ،raw trajectory سجل ،model/tool/context جمد ؛proof bundle
تجربة صمم matched-budget حول completeness
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كغضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرارٍ.

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عامًا.
5. شرط الهدم
إذا لم تقلل promotion carrying-proof regressions أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، automatic rollback and scientific verdict إنشاء
MUT/MUG/MOUSSAIXT، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt. إلى النسب محور في ledger.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•

<PARSED TEXT FOR PAGE: 4 / 40>

قاموس الحالات والكائنات - Moussaif Proof-Carrying Recursive Improvement & (Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world- ولا self-improvement لا :الأطروحة

receipts effect ومسار rollback. في هذه الصفحة يختبر التركيز الخاص بقاموس الحالات والكائنات.
1. الكائن النظري والدقة الافتراضية
Paper-to-Kernel pipeline ؛ promotion يقرر MUG ؛ externalities/recourse يقيسان M-XEOS/CAROS ؛ Courts تهاجم
candidate يولد Evolution Lab ؛ run يثبت M-RPROS
يغلق دورة البحث. التركيز هنا هو قاموس الحالات والكائنات، مع منع الخلط بين schema نظري ونتيجة تجريبية.
(EvidenceStrength_eff = Σ w_i E_i − CorrelationPenalty(E_1…E_n
المعادلة هنا formal research object وليست رقماً تجريبياً جاهزاً. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
WorldEvolver 2026
Deployment-time world-model revision via episodic memory, semantic memory and selective foresight.
arXiv:2606.30639
Causal Memory Intervention 2026
Selects memories by causal usefulness under intervention rather than semantic relevance alone.
arXiv:2605.17641
Continual Learning Moves to Memory 2026
Shows stability-plasticity reappears at memory representation/retrieval level.
arXiv:2604.27003
LWM-Planner 2026
Fact-augmented lookahead planning with validated atomic facts and limited-horizon simulation.
AISTATS 2026 / PMLR 300
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفياً محتملاً.
3. الربط الداخلي
M-EAPOS / M-RPROS / M-XEOS / M-CAROS / M-MCBOS: action-state separation, reproducibility, externalities, accountability and multi-currency budgets.
Evidence discipline: AGI-oriented architecture is not evidence of achieved AGI; E3/E4/E5 require execution, controlled baselines and replication.
MUT: theory unification, lineage, relation graph, research questions and experiment candidates.
MUG: governance of placement, internalization, authority, evolution and evidence status.
4. التجربة الحاسمة
مع التجربة أعد ثم، outcome score عن process score افصل، raw trajectory سجل، model/tool/context جمد، correlated verifier
detection حول matched-budget تجربة صمم
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرار.

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عاماً.
5. شرط الهدم
إذا لم تقلل promotion carrying-proof regressions أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule،

إنشاء automatic rollback and scientific verdict
ledger. مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
.replication دون 5E لا ؛baseline frozen دون 4E لا ؛claim AGI لا .formalization theoretical Candidate :حالة هذا البناء
•
•
•
•
•
4
<PARSED TEXT FOR PAGE: 5 / 40>
P14-E / M-PCRI / PAGE 005 / 05 September 2026
الدنيا المسلمات - Moussaif Proof-Carrying Recursive Improvement &
(Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
الأطروحة: لا self-improvement ولا
. في هذه الصفحة يختبر التركيز الخاص بالمسلمات الدنيا. rollback ومسار receipts effect
1. الكائن النظري والدقة الافتراضية
؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو المسلمات الدنيا، مع منع الخلط بين schema نظري ونتيجة تجريبية.
RiskAdjustedUtility = ExpectedVerifiedUtility − λ_risk·Risk − λ_cost·Cost
المعادلة هنا object research formal وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو
scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
LWM-Planner 2026
Fact-augmented lookahead planning with validated atomic facts and limited-horizon
simulation.
AISTATS 2026 / PMLR 300
PaperBench 2025
Research replication benchmark: 20 ICML papers and 8,316 gradable outcomes;
stresses full research lifecycle competence.
OpenAI 2025
ToolSandbox 2024
Stateful conversational tool-use evaluation with dynamic milestone checks and state
dependencies.
arXiv:2408.04682
BrowseComp-Plus 2025
Fixed-corpus deep-research benchmark for retriever/citation/context-engineering
disentanglement.
arXiv:2508.06600
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها
«دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملاً.
3. الربط الداخلي
MUG: governance of placement, internalization, authority, evolution and evidence
status.
MOUSSAIXT: cross-environment execution shell binding models, tools, memory,
agents, verification and local/cloud runtimes.
MCD / QMCD / MCAI / HAKIM-MCD: difference-first, question-conditioned,
plural-path and falsification-oriented cognitive routing.
M-HKOS / KB-as-OS: knowledge as governed operational substrate, not passive
archive.
4. التجربة الحاسمة

مع التجربة أعد ثم ،outcome score عن process score افصل ،raw trajectory سجل ،model/tool/context جمد ؛continual learning
تجربة صمم matched-budget حول regression
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرارٍ

حيث تكون الثقة جزء error confidenceو Calibration
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
عندما يكون الادعاء عامًا. environment/domain/model عبر Transfer
5. شرط الهدم
إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CI أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict
ledger. في محور المسلمات الدنيا، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
.replication دون 5E لا ؛baseline frozen دون 4E لا ؛claim AGI لا .formalization theoretical Candidate :حالة هذا البناء
•
•
•
•
•
5
<PARSED TEXT FOR PAGE: 6 / 40>
الأولى - P14-E / M-PCRI / PAGE 006 / 05 September 2026 & Moussaif Proof-Carrying Recursive Improvement
الرياضية الصياغة
(Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
في هذه الصفحة يختبر التركيز الخاص بالصياغة الرياضية الأولى. rollback ومسار receipts effect
1. الكائن النظري والدقة الافتراضية
؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو الصياغة الرياضية الأولى، مع منع الخلط بين schema نظري ونتيجة تجريبية.
(Generalization = Conservation(Invariants | model,domain,tool,context change
المعادلة هنا formal research object وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
BrowseComp-Plus 2025
Fixed-corpus deep-research benchmark for retriever/citation/context-engineering
disentanglement.
arXiv:2508.06600
Agentic Misalignment 2025
Controlled simulations of harmful agentic actions under goal/replacement pressure;
motivates deployment-time governance.
Anthropic 2025
Automated Alignment Auditing 2025
Agents that discover hidden goals, build evaluations and surface concerning behavior.
Anthropic 2025
Bloom Auto-Evals 2025

Reproducible targeted behavioral evaluation generation and severity/frequency measurement.
Anthropic 2025
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملًا.
3. الربط الداخلي
M-HKOS / KB-as-OS: knowledge as governed operational substrate, not passive archive.
HyperSkill -> HyperAbility -> Bounded HyperLoop -> Agent/Tool -> World Effect -< Self-Reconstruction Memory.
M-EAPOS / M-RPROS / M-XEOS / M-CAROS / M-MCBOS: action-state separation, reproducibility, externalities, accountability and multi-currency budgets.
Evidence discipline: AGI-oriented architecture is not evidence of achieved AGI; E3/E4/E5 require execution, controlled baselines and replication.
4. التجربة الحاسمة
مع التجربة أعد ثم، outcome score عن process score افصل، raw trajectory سجل، model/tool/context جمد؛ research replication
تجربة صمم matched-budget حول pipeline
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرارٍ

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عامًا.
5. شرط الهدم
إذا لم تقلل promotion carrying-proof regressions أو improvements false مقارنةً بـbenchmark/CL baseline أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict
ledger. مع typed schema وtrace_id وsource/evidence pointers وrollback receipt، الأولى الرياضية الصياغة محور في
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
6
<PARSED TEXT FOR PAGE: 7 / 40>
P14-E / M-PCRI / PAGE 007 / 05 September 2026
identifiability - Moussaif Proof-Carrying Recursive وقابلية القياس
(Improvement & Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
الأطروحة: لا self-improvement ولا
identifiability. في هذه الصفحة يختبر التركيز الخاص بقابلية القياس ومسار rollback receipts effect
1. الكائن النظري والدقة الافتراضية
؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو قابلية القياس وidentifiability، مع منع الخلط بين schema نظري ونتيجة تجريبية.
Promote(c) ⇔ ΔMetric>τ ∧ CriticalRegression=0 ∧ ProofValid ∧ AblationSurvives ∧ Replication≥k
المعادلة هنا formal research object وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على operational definition، نطاق، وحدة أو

.runtime وطريقة قياس قبل استعماله في قرار ،scale
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
Bloom Auto-Evals 2025
Reproducible targeted behavioral evaluation generation and severity/frequency measurement.
Anthropic 2025
AI Scientist / Nature 2026
End-to-end scientific automation from idea through code, experiment, manuscript and review; shows research workflow can be agentized.
Nature 2026
Belief-Based World Models 2026
Exposes uncertainty about current state alongside simulations under partial observability.
arXiv:2609.00455
TASPO 2026
Action-level credit allocation grounded in verified trajectory outcomes; addresses process-supervision versus outcome-credit mismatch.
arXiv:2608.31077
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملًا.
3. الربط الداخلي
Evidence discipline: AGI-oriented architecture is not evidence of achieved AGI; E3/E4/E5 require execution, controlled baselines and replication.
MUT: theory unification, lineage, relation graph, research questions and experiment candidates.
MUG: governance of placement, internalization, authority, evolution and evidence status.
MOUSSAIXT: cross-environment execution shell binding models, tools, memory, agents, verification and local/cloud runtimes.
4. التجربة الحاسمة
التجربة أعد ثم ،outcome score عن process score افصل ،raw trajectory سجل ،model/tool/context جمد؛ E4 to E5 cross-model تجربة صمم matched-budget حول replication
مع حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرارٍ

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عامًا.
5. شرط الهدم
إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict
ledger. في القياس قابلية محور وidentifiability، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•

.
.
7
<PARSED TEXT FOR PAGE: 8 / 40>

theory promotion بلا proof-carrying candidate، يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
receipts effect ومسار rollback. في هذه الصفحة يختبر التركيز الخاص بالحدود الأنطولوجية.
1. الكائن النظري والدقة الافتراضية
Paper-to-Kernel pipeline؛ MUG يقرر promotion، externalities/recourse؛ M-XEOS/CAROS يقيسان Courts ؛ تهاجم
M-RPROS يثبت run، Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو الحدود الأنطولوجية، مع منع الخلط بين schema نظري ونتيجة تجريبية.
ΔState = State_after − State_before ; VerifiedSuccess ⇔ Postcondition(State_after)=true
المعادلة هنا object research formal وليست رقماً تجريبياً جاهزاً. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو
scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
TASPO 2026
Action-level credit allocation grounded in verified trajectory outcomes; addresses
process-supervision versus outcome-credit mismatch.
arXiv:2608.31077
APB 2026
Agent Planning Benchmark: 4,209 multimodal cases across 22 domains; planning,
broken tools, unsolvable tasks, refinement.
arXiv:2606.04874
AMA-Bench 2026
Long-horizon agent memory benchmark; highlights missing causality/objective
information and lossy similarity retrieval.
arXiv:2602.22769
AgentLAB 2026
Long-horizon attack benchmark covering intent hijacking, tool chaining, task
injection, objective drift and memory poisoning.
arXiv:2602.16901
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها
«دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفياً محتملاً.
3. الربط الداخلي
MOUSSAIXT: cross-environment execution shell binding models, tools, memory,
agents, verification and local/cloud runtimes.
MCD / QMCD / MCAI / HAKIM-MCD: difference-first, question-conditioned,
plural-path and falsification-oriented cognitive routing.
M-HKOS / KB-as-OS: knowledge as governed operational substrate, not passive
archive.
HyperSkill -> HyperAbility -> Bounded HyperLoop -> Agent/Tool -> World Effect -<
Self-Reconstruction Memory.
4. التجربة الحاسمة
candidate promotion؛ جمد model/tool/context، سجل raw trajectory، افصل process score عن outcome score، ثم أعد التجربة
تجربة صمم matched-budget حول tournament
مع حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرار.

ً

حيث تكون الثقة جزء error confidenceو Calibration
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
عندما يكون الادعاء عامًا. environment/domain/model عبر Transfer
5. شرط الهدم
إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict
ledger. في typed schema وtrace_id وsource/evidence pointers وrollback receipt، مع الأنطولوجية الحدود محور
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
8
<PARSED TEXT FOR PAGE: 9 / 40>
P14-E / M-PCRI / PAGE 009 / 05 September 2026
MCD - Moussaif Proof-Carrying Recursive Improvement & :بـ العلاقة
(Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate، يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
receipts effect ومسار rollback. في هذه الصفحة يختبر التركيز الخاص بالعلاقة بـMCD.
1. الكائن النظري والدقة الافتراضية
Paper-to-Kernel pipeline؛ MUG يقرر promotion؛ M-XEOS/CAROS يقيسان externalities/recourse؛ Courts تهاجم؛
candidate يولد Evolution Lab؛ run يثبت M-RPROS
يغلق دورة البحث. التركيز هنا هو العلاقة بـMCD، مع منع الخلط بين schema نظري ونتيجة تجريبية.
(EvidenceStrength_eff = Σ w_i E_i − CorrelationPenalty(E_1…E_n
المعادلة هنا formal research object وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
AgentLAB 2026
Long-horizon attack benchmark covering intent hijacking, tool chaining, task
injection, objective drift and memory poisoning.
arXiv:2602.16901
RoadmapBench 2026
Long-horizon software-development benchmark over real version upgrades; exposes
multi-file, multi-target planning limits.
arXiv:2605.15846
MM-ToolSandBox 2026
Stateful visual tool-calling across 500+ tools and 16 domains; visual precision
becomes a major bottleneck.
arXiv:2607.11818
World Model Agentic Mid-Training 2026
Capability-first world-model planning; distinguishes predictive capability from merely
imitating foresight format.
arXiv:2606.27483
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملاً .

3. الربط الداخلي
HyperSkill -> HyperAbility -> Bounded HyperLoop -> Agent/Tool -> World Effect -<
Self-Reconstruction Memory.
M-EAPOS / M-RPROS / M-XEOS / M-CAROS / M-MCBOS: action-state
separation, reproducibility, externalities, accountability and multi-currency budgets.
Evidence discipline: AGI-oriented architecture is not evidence of achieved AGI; E3/
E4/E5 require execution, controlled baselines and replication.
MUT: theory unification, lineage, relation graph, research questions and experiment
candidates.
4. التجربة الحاسمة
مع التجربة أعد ثم ،outcome score عن process score افصل ،raw trajectory سجل ،model/tool/context جمد ؛proof bundle
تجربة صمم matched-budget حول completeness
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرارٍ.

حيث تكون الثقة جزء error confidence وCalibration
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
عندما يكون الادعاء عامًا. environment/domain/model عبر Transfer
5. شرط الهدم
إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CL، تصبح إضافة
بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule،
إنشاء automatic rollback and scientific verdict
ledger. بـMCD، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
9
<PARSED TEXT FOR PAGE: 10 / 40>
P14-E / M-PCRI / PAGE 010 / 05 September 2026
QMCD - Moussaif Proof-Carrying Recursive Improvement & بـالعلاقة
(Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
receipts effect ومسار rollback. في هذه الصفحة يختبر التركيز الخاص بالعلاقة بـQMCD.
1. الكائن النظري والدقة الافتراضية
Paper-to-Kernel pipeline ؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو العلاقة بـQMCD، مع منع الخلط بين schema نظري ونتيجة تجريبية.
RiskAdjustedUtility = ExpectedVerifiedUtility − λ_risk·Risk − λ_cost·Cost
المعادلة هنا object research formal وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو
scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
World Model Agentic Mid-Training 2026
Capability-first world-model planning; distinguishes predictive capability from merely

imitating foresight format.
arXiv:2606.27483
WorldEvolver 2026
Deployment-time world-model revision via episodic memory, semantic memory and selective foresight.
arXiv:2606.30639
Causal Memory Intervention 2026
Selects memories by causal usefulness under intervention rather than semantic relevance alone.
arXiv:2605.17641
Continual Learning Moves to Memory 2026
Shows stability-plasticity reappears at memory representation/retrieval level.
arXiv:2604.27003
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفياً محتملاً.

3. الربط الداخلي

MUT: theory unification, lineage, relation graph, research questions and experiment candidates.

MUG: governance of placement, internalization, authority, evolution and evidence status.

MOUSSAIXT: cross-environment execution shell binding models, tools, memory, agents, verification and local/cloud runtimes.

MCD / QMCD / MCAI / HAKIM-MCD: difference-first, question-conditioned, plural-path and falsification-oriented cognitive routing.

4. التجربة الحاسمة

مع التجربة أعد ثم، outcome score عن process score افصل، raw trajectory سجل، model/tool/context جمد؛ correlated verifier
تجربة صمم matched-budget حول detection
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرار.

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عاماً.

5. شرط الهدم

إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.

6. ممر التطبيق

immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict
ledger. في QMCD، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.

•
•
•
•
•

(Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
receipts effect ومسار rollback. في هذه الصفحة يختبر التركيز الخاص بالعلاقة بـMCAL.
1. الكائن النظري والدقة الافتراضية
Paper-to-Kernel pipeline ؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو العلاقة بـMCAL، مع منع الخلط بين schema نظري ونتيجة تجريبية.
(Generalization = Conservation(Invariants | model,domain,tool,context change
المعادلة هنا object research formal وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو
scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
Continual Learning Moves to Memory 2026
Shows stability-plasticity reappears at memory representation/retrieval level.
arXiv:2604.27003
LWM-Planner 2026
Fact-augmented lookahead planning with validated atomic facts and limited-horizon
simulation.
AISTATS 2026 / PMLR 300
PaperBench 2025
Research replication benchmark: 20 ICML papers and 8,316 gradable outcomes;
stresses full research lifecycle competence.
OpenAI 2025
ToolSandbox 2024
Stateful conversational tool-use evaluation with dynamic milestone checks and state
dependencies.
arXiv:2408.04682
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها
«دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملاً.
3. الربط الداخلي
MCD / QMCD / MCAI / HAKIM-MCD: difference-first, question-conditioned,
plural-path and falsification-oriented cognitive routing.
M-HKOS / KB-as-OS: knowledge as governed operational substrate, not passive
archive.
HyperSkill -> HyperAbility -> Bounded HyperLoop -> Agent/Tool -> World Effect -<
Self-Reconstruction Memory.
M-EAPOS / M-RPROS / M-XEOS / M-CAROS / M-MCBOS: action-state
separation, reproducibility, externalities, accountability and multi-currency budgets.
4. التجربة الحاسمة
continual learning؛ جمد model/tool/context، سجل raw trajectory، افصل process score عن outcome score، مع التجربة أعد ثم
regression حول matched-budget تجربة صمم
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرار.

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عامًا.
5. شرط الهدم
إذا لم تقلل promotion carrying-proof regressions أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة

بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict
ledger. في MCAIبـ العلاقة محور مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
.replication دون 5E لا ؛baseline frozen دون 4E لا ؛claim AGI لا .formalization theoretical Candidate :حالة هذا البناء
•
•
•
•
•
11
<PARSED TEXT FOR PAGE: 12 / 40>
P14-E / M-PCRI / PAGE 012 / 05 September 2026
HAKIM-MCD - Moussaif Proof-Carrying Recursive Improvement بـ العلاقة
& (Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate، يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
الأطروحة: لا self-improvement ولا
.MCD-HAKIM بـ العلاقة الخاص التركيز يختبر الصفحة هذه في .rollback ومسار receipts effect
1. الكائن النظري والدقة الافتراضية
؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse، MUG يقرر promotion، Paper-to-Kernel pipeline
M-RPROS يثبت run، Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو العلاقة بـMCD-HAKIM، مع منع الخلط بين schema نظري ونتيجة تجريبية.
Promote(c) ⇔ ΔMetric>τ ∧ CriticalRegression=0 ∧ ProofValid ∧ AblationSurvives ∧ Replication≥k
المعادلة هنا formal research object وليست رقماً تجريبياً جاهزاً. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
ToolSandbox 2024
Stateful conversational tool-use evaluation with dynamic milestone checks and state dependencies.
arXiv:2408.04682
BrowseComp-Plus 2025
Fixed-corpus deep-research benchmark for retriever/citation/context-engineering disentanglement.
arXiv:2508.06600
Agentic Misalignment 2025
Controlled simulations of harmful agentic actions under goal/replacement pressure; motivates deployment-time governance.
Anthropic 2025
Automated Alignment Auditing 2025
Agents that discover hidden goals, build evaluations and surface concerning behavior.
Anthropic 2025
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفياً محتملاً.
3. الربط الداخلي
M-EAPOS / M-RPROS / M-XEOS / M-CAROS / M-MCBOS: action-state separation, reproducibility, externalities, accountability and multi-currency budgets.
Evidence discipline: AGI-oriented architecture is not evidence of achieved AGI; E3/E4/E5 require execution, controlled baselines and replication.
MUT: theory unification, lineage, relation graph, research questions and experiment candidates.
MUG: governance of placement, internalization, authority, evolution and evidence

status.
4. التجربة الحاسمة
research replication، جمد model/tool/context، سجل raw trajectory، افصل process score عن outcome score، مع التجربة أعد ثم
تجربة صمم matched-budget حول pipeline
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرارٍ

حيث تكون الثقة جزء error confidence وCalibration
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
عندما يكون الادعاء عامًا. environment/domain/model عبر Transfer
5. شرط الهدم
أبسط، تصبح إضافة baseline benchmark/CL مقارنةً بـ improvements false أو regressions promotion carrying-proof إذا لم تقلل
بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule،
إنشاء automatic rollback and scientific verdict
ledger. في HAKIM-MCD، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
12
<PARSED TEXT FOR PAGE: 13 / 40>
P14-E / M-PCRI / PAGE 013 / 05 September 2026
M-HKOS - Moussaif Proof-Carrying Recursive Improvement & بالعلاقة
(Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
receipts effect ومسار rollback. في هذه الصفحة يختبر التركيز الخاص بالعلاقة بـHKOS-M.
1. الكائن النظري والدقة الافتراضية
M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline تهاجم؛ Courts ؛
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو العلاقة بـHKOS-M، مع منع الخلط بين schema نظري ونتيجة تجريبية.
ΔState = State_after − State_before ; VerifiedSuccess ⇔ Postcondition(State_after)=true
object research formal هنا المعادلة وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو
scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
Automated Alignment Auditing 2025
Agents that discover hidden goals, build evaluations and surface concerning behavior.
Anthropic 2025
Bloom Auto-Evals 2025
Reproducible targeted behavioral evaluation generation and severity/frequency
measurement.
Anthropic 2025
AI Scientist / Nature 2026
End-to-end scientific automation from idea through code, experiment, manuscript and
review; shows research workflow can be agentized.

Nature 2026
Belief-Based World Models 2026
Exposes uncertainty about current state alongside simulations under partial observability.
arXiv:2609.00455
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفياً محتملاً.
3. الربط الداخلي
MUG: governance of placement, internalization, authority, evolution and evidence status.
MOUSSAIXT: cross-environment execution shell binding models, tools, memory, agents, verification and local/cloud runtimes.
MCD / QMCD / MCAI / HAKIM-MCD: difference-first, question-conditioned, plural-path and falsification-oriented cognitive routing.
M-HKOS / KB-as-OS: knowledge as governed operational substrate, not passive archive.
4. التجربة الحاسمة
التجربة أعد ثم outcome score عن process score افصل، raw trajectory سجل، model/tool/context جمد؛ E4 to E5 cross-model replication حول matched-budget تجربة صمم
مع حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرار.

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عاماً.
5. شرط الهدم
إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
إنشاء automatic rollback and scientific verdict ledger. في العلاقة محور بـM-HKOS، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•

العلاقة بـHyperSkills - Moussaif Proof-Carrying Recursive Improvement & (Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-الأطروحة: لا self-improvement ولا
HyperSkills. في هذه الصفحة يختبر التركيز الخاص بالعلاقة بـrollback ومسار receipts effect.
1. الكائن النظري والدقة الافتراضية
؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو العلاقة بـHyperSkills، مع منع الخلط بين schema نظري ونتيجة تجريبية.

(EvidenceStrength_eff = Σ w_i E_i − CorrelationPenalty(E_1…E_n
المعادلة هنا object research formal وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
Belief-Based World Models 2026
Exposes uncertainty about current state alongside simulations under partial observability.
arXiv:2609.00455
TASPO 2026
Action-level credit allocation grounded in verified trajectory outcomes; addresses process-supervision versus outcome-credit mismatch.
arXiv:2608.31077
APB 2026
Agent Planning Benchmark: 4,209 multimodal cases across 22 domains; planning, broken tools, unsolvable tasks, refinement.
arXiv:2606.04874
AMA-Bench 2026
Long-horizon agent memory benchmark; highlights missing causality/objective information and lossy similarity retrieval.
arXiv:2602.22769
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملًا.
3. الربط الداخلي
M-HKOS / KB-as-OS: knowledge as governed operational substrate, not passive archive.
HyperSkill -> HyperAbility -> Bounded HyperLoop -> Agent/Tool -> World Effect -< Self-Reconstruction Memory.
M-EAPOS / M-RPROS / M-XEOS / M-CAROS / M-MCBOS: action-state separation, reproducibility, externalities, accountability and multi-currency budgets.
Evidence discipline: AGI-oriented architecture is not evidence of achieved AGI; E3/E4/E5 require execution, controlled baselines and replication.
4. التجربة الحاسمة
التجربة أعد ثم، outcome score عن process score افصل، raw trajectory سجل، model/tool/context جمد؛ candidate promotion
تجربة صمم matched-budget حول tournament
مع حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرار.

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عامًا.
5. شرط الهدم
إذا لم تقلل promotion carrying-proof أو regressions false improvements مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، automatic rollback and scientific verdict إنشاء
ledger. في HyperSkills، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt. بـالعلاقة محور
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
.

.
.
.
.
14
<PARSED TEXT FOR PAGE: 15 / 40>
P14-E / M-PCRI / PAGE 015 / 05 September 2026
العلاقة بـHyperAbilities - Moussaif Proof-Carrying Recursive Improvement
& (Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
ولا self-improvement لا :الأطروحة
.HyperAbilities بـ يختبر التركيز الخاص بالعلاقة هذه الصفحة في .rollback ومسار receipts effect
1. الكائن النظري والدقة الافتراضية
؛ Courts تهاجم، M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو العلاقة بـHyperAbilities، مع منع الخلط بين schema نظري ونتيجة تجريبية.
RiskAdjustedUtility = ExpectedVerifiedUtility − λ_risk·Risk − λ_cost·Cost
المعادلة هنا object research formal وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو
scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
AMA-Bench 2026
Long-horizon agent memory benchmark; highlights missing causality/objective
information and lossy similarity retrieval.
arXiv:2602.22769
AgentLAB 2026
Long-horizon attack benchmark covering intent hijacking, tool chaining, task
injection, objective drift and memory poisoning.
arXiv:2602.16901
RoadmapBench 2026
Long-horizon software-development benchmark over real version upgrades; exposes
multi-file, multi-target planning limits.
arXiv:2605.15846
MM-ToolSandBox 2026
Stateful visual tool-calling across 500+ tools and 16 domains; visual precision
becomes a major bottleneck.
arXiv:2607.11818
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها
«دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملاً.
3. الربط الداخلي
Evidence discipline: AGI-oriented architecture is not evidence of achieved AGI; E3/
E4/E5 require execution, controlled baselines and replication.
MUT: theory unification, lineage, relation graph, research questions and experiment
candidates.
MUG: governance of placement, internalization, authority, evolution and evidence
status.
MOUSSAIXT: cross-environment execution shell binding models, tools, memory,
agents, verification and local/cloud runtimes.
4. التجربة الحاسمة
proof bundle؛ جمد model/tool/context، سجل raw trajectory، افصل process score عن outcome score، ثم أعد التجربة مع
completeness حول matched-budget تجربة صمم
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى

Task/world-effect success تحت budget مجمد.
ا من القرارٍ.

حيث تكون الثقة جزء error confidence وCalibration
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
عندما يكون الادعاء عامًا. environment/domain/model عبر Transfer
5. شرط الهدم
إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict
ledger. العلاقة محور في بـHyperAbilities، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
15
<PARSED TEXT FOR PAGE: 16 / 40>
P14-E / M-PCRI / PAGE 016 / 05 September 2026
Bounded HyperLoops - Moussaif Proof-Carrying Recursive بـالعلاقة
(Improvement & Scientific Closure Theory (M-PCRI
theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
الأطروحة: لا self-improvement ولا
HyperLoops Bounded بـالعلاقة الخاص التركيز يختبر الصفحة هذه في. rollback ومسار receipts effect.
1. الكائن النظري والدقة الافتراضية
؛ Courts تهاجم؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline
M-RPROS يثبت run؛ Evolution Lab يولد candidate
يغلق دورة البحث. التركيز هنا هو العلاقة بـHyperLoops Bounded، مع منع الخلط بين schema نظري ونتيجة تجريبية.
(Generalization = Conservation(Invariants | model,domain,tool,context change
المعادلة هنا formal research object وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو scale، وطريقة قياس قبل استعماله في قرار runtime.
الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.
2. ضغط المجتمع العلمي
MM-ToolSandBox 2026
Stateful visual tool-calling across 500+ tools and 16 domains; visual precision becomes a major bottleneck.
arXiv:2607.11818
World Model Agentic Mid-Training 2026
Capability-first world-model planning; distinguishes predictive capability from merely imitating foresight format.
arXiv:2606.27483
WorldEvolver 2026
Deployment-time world-model revision via episodic memory, semantic memory and selective foresight.
arXiv:2606.30639
Causal Memory Intervention 2026
Selects memories by causal usefulness under intervention rather than semantic relevance alone.
arXiv:2605.17641

الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تَدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملًا.

3. الربط الداخلي

MOUSSAIXT: cross-environment execution shell binding models, tools, memory, agents, verification and local/cloud runtimes.

MCD / QMCD / MCAI / HAKIM-MCD: difference-first, question-conditioned, plural-path and falsification-oriented cognitive routing.

M-HKOS / KB-as-OS: knowledge as governed operational substrate, not passive archive.

HyperSkill -> HyperAbility -> Bounded HyperLoop -> Agent/Tool -> World Effect -< Self-Reconstruction Memory.

4. التجربة الحاسمة

مع التجربة أعد ثم، outcome score عن process score افصل، raw trajectory سجل، model/tool/context جمد؛ correlated verifier

تجربة صمم matched-budget حول detection

حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.

مقاييس الحد الأدنى

Task/world-effect success تحت budget مجمد.

ا من القرارٍ

Calibration وerror confidence حيث تكون الثقة جزء

Recovery / rollback / failure localization.

Cost, latency, tool calls, context load.

Transfer عبر environment/domain/model عندما يكون الادعاء عامًا.

5. شرط الهدم

إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.

6. ممر التطبيق

immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict

ledger. العلاقة محور في Bounded HyperLoops، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.

حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.

•

•

•

•

•

Agent Mesh - Moussaif Proof-Carrying Recursive Improvement & العلاقة بـ
(Scientific Closure Theory (M-PCRI

theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-

الأطروحة: لا self-improvement ولا

Mesh Agent بـ العلاقة بـ الخاص المركز يختبر الصفحة هذه في. rollback ومسار receipts effect.

1. الكائن النظري والدقة الافتراضية

تهاجم؛ Courts ؛ M-XEOS/CAROS يقيسان externalities/recourse؛ MUG يقرر promotion؛ Paper-to-Kernel pipeline M-RPROS يثبت run؛ Evolution Lab يولد candidate

يغلق دورة البحث. التركيز هنا هو العلاقة بـMesh Agent، مع منع الخلط بين schema نظري ونتيجة تجريبية.

Promote(c) ⇔ ΔMetric>τ ∧ CriticalRegression=0 ∧ ProofValid ∧ AblationSurvives ∧ Replication≥k

المعادلة هنا formal research object وليست رقمًا تجريبيًا جاهزًا. كل متغير يجب أن يحصل على definition operational، نطاق، وحدة أو scale، وطريقة قياس قبل استعماله في قرار runtime.

الغاية هي منع انتقال المشروع من غنى اصطلاحي إلى دقة زائفة.

2. ضغط المجتمع العلمي

Causal Memory Intervention 2026
Selects memories by causal usefulness under intervention rather than semantic relevance alone.
arXiv:2605.17641
Continual Learning Moves to Memory 2026
Shows stability-plasticity reappears at memory representation/retrieval level.
arXiv:2604.27003
LWM-Planner 2026
Fact-augmented lookahead planning with validated atomic facts and limited-horizon simulation.
AISTATS 2026 / PMLR 300
PaperBench 2025
Research replication benchmark: 20 ICML papers and 8,316 gradable outcomes; stresses full research lifecycle competence.
OpenAI 2025
الاستفادة العلمية تتم عبر تحويل نتيجة المجتمع إلى constraint أو baseline أو mode failure. لا تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملًا.
3. الربط الداخلي
HyperSkill -> HyperAbility -> Bounded HyperLoop -> Agent/Tool -> World Effect -< Self-Reconstruction Memory.
M-EAPOS / M-RPROS / M-XEOS / M-CAROS / M-MCBOS: action-state separation, reproducibility, externalities, accountability and multi-currency budgets.
Evidence discipline: AGI-oriented architecture is not evidence of achieved AGI; E3/E4/E5 require execution, controlled baselines and replication.
MUT: theory unification, lineage, relation graph, research questions and experiment candidates.
4. التجربة الحاسمة
مع التجربة أعد ثم، outcome score عن process score افصل، raw trajectory سجل، model/tool/context جمد؛ continual learning regression حول matched-budget تجربة صمم
حذف المكوّن المستهدف. استخدم cases failure كبيانات أساسية لا كضوضاء.
مقاييس الحد الأدنى
Task/world-effect success تحت budget مجمد.
ا من القرارٍ

Calibration وerror confidence حيث تكون الثقة جزء
Recovery / rollback / failure localization.
Cost, latency, tool calls, context load.
Transfer عبر environment/domain/model عندما يكون الادعاء عامًا.
5. شرط الهدم
إذا لم تقلل regressions promotion carrying-proof أو improvements false مقارنةً بـbaseline benchmark/CL أبسط، تصبح إضافة بيروقراطية غير مبررة.
6. ممر التطبيق
immutable CandidateProofBundle، frozen harness fingerprint، ablation matrix، canary registry، replication capsule، إنشاء automatic rollback and scientific verdict
ledger. في محور العلاقة بـAgent Mesh، مع typed schema وtrace_id وsource/evidence pointers وrollback receipt.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•

<PARSED TEXT FOR PAGE: 18 / 40>

العلاقة بـM-EAPOS - Moussaif Proof-Carrying Recursive Improvement &
(Scientific Closure Theory (M-PCRI

theory promotion بلا proof-carrying candidate؛ يحمل أن يجب تغيير كل baseline، raw run، causal ablation، security/world-
الأطروحة: لا self-improvement ولا
receipts effect ومسار rollback. في هذه الصفحة يختبر التركيز الخاص بالعلاقة بـEAPOS-M.

1. الكائن النظري والدقة الافتراضية