RECONSTRUCTED_FROM_INDEXED_TEXT - نسخة أرشيفية معاد بناؤها من النص المفهرس

<PARSED TEXT FOR PAGE: 1 / 100>
MOUSSAIF UNIFIED AGI RESEARCH PROGRAM / 05 September 2026
MOUSSAIF UNIFIED AGI - PART XIV -
الإغلاق النظري والتطبيقي العلمي
Optimal Theoretical-Applied Closure Blueprint - MUT/MUG/M-HKOS/MOUSSAIXT +
Frontier Community Constraints
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 / 100>
PART XIV / PAGE 002 / 05 September 2026
ميثاق الإغلاق النظري-التطبيقي / لماذا لا يكفي المزيد من النظريات
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence، interface، metric، failure وقرار promote/ reject/merge. محور الصفحة: لماذا لا يكفي المزيد من النظريات.
1. الكائن النظري والدقة الافتراضية
يُربط لماذا لا يكفي المزيد من النظريات مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛ HKOS-M يحوله إلى object knowledge تشغيلي؛ Moussaif Unified Cognitive-State Transition Theory (M-UCST) هي الأقرب المرافقة الورقة. runtime contract كـ يطبقه.
MOUSSAIXT
SystemState_t=(World,Knowledge,Question,Difference,Belief,Actionability,Commitment,Memory,Governance,Budget)_t
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
أنشئ benchmark داخلي خاص بلماذا لا يكفي المزيد من النظريات مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من
أو PaperBench/ToolSandbox/Bench-AMA/APB
benchmark مناسب، مباشرة أرقامها تنقل ولا .نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.
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
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim
يضيق نطاقه domain/model-cross يفشل general
بدلاً من ترقيته بلاغيًا.
6. ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور لماذا لا يكفي المزيد من النظريات، ثم implementation reference صغيرة، ثم replay
failure، ثم ablation، ثم canary. لا توسع
corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
2
<PARSED TEXT FOR PAGE: 3 / 100>
PART XIV / PAGE 003 / 05 September 2026
ميثاق الإغلاق النظري-التطبيقي / تعريف الإغلاق النظري
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence
/promote وقرار failure، metric، interface
reject/merge. محور الصفحة: تعريف الإغلاق النظري.
1. الكائن النظري والدقة الافتراضية
يُربط تعريف الإغلاق النظري مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛
HKOS-M يحوله إلى object knowledge تشغيلي؛
MOUSSAIXT كـ runtime contract. هي الأقرب المرافقة الورقة (Moussaif Agentic Capability Closure Theory (M-ACCT.
Closure = TheoryIdentifiability × Executability × Verifiability × Reproducibility × Transfer
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
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 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
أنشئ benchmark داخلي خاص بتعريف الإغلاق النظري مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ benchmark أو ToolSandbox/Bench-AMA/APB
مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: mechanism unit، mission integrated، replication model-cross.
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 عاماً.
5. شرط الهدم
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim general يفشل domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغياً.
6. ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور تعريف الإغلاق النظري، ثم implementation reference صغيرة، ثم replay failure، ثم ablation، ثم canary. لا توسع corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•


<PARSED TEXT FOR PAGE: 4 / 100>

ميثاق الإغلاق النظري-التطبيقي / تعريف الإغلاق التطبيقي
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence، interface، metric، failure، وقرار promote/ reject/merge. محور الصفحة: تعريف الإغلاق التطبيقي.

1. الكائن النظري والدقة الافتراضية
يُ ربط تعريف الإغلاق التطبيقي مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛ HKOS-M يحوله إلى object knowledge تشغيلي؛
runtime. هي الأقرب المرافقة الورقة (Moussaif Question-Difference Geometry & Resource Rationality (M-QDGR.
MOUSSAIXT كـ يطبقه contract
Capability = Reachability × Reliability × Verification × Recovery × Transfer
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
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 تُدمج ورقة خارجية داخل MUT بوصفها «دليل صحة»؛ بل بوصفها مرجع آلية أو معيار اختبار أو نفيًا محتملاً.
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
أنشئ benchmark داخلي خاص بتعريف الإغلاق التطبيقي مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ benchmark أو ToolSandbox/Bench-AMA/APB
مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: mechanism unit، mission integrated، model-cross replication.
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
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim general يفشل domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغيًا.

6. ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور تعريف الإغلاق التطبيقي، ثم implementation reference صغيرة، ثم replay failure، ثم ablation، ثم canary. لا توسع corpus قبل أن
تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•


الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence، interface، metric، failure، وقرار promote/
reject/merge. محور الصفحة: تعريف الإغلاق العلمي.
1. الكائن النظري والدقة الافتراضية
يُربط تعريف الإغلاق العلمي مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛ HKOS-M يحوله إلى object knowledge تشغيلي؛ MOUSSAIXT
هي الأقرب المرافقة الورقة (Moussaif Self-Reconstruction, Causal Memory & Cross-Model Identity Theory (M-SRCI. يطبقه كـruntime contract.
Route* = argmax E[VerifiedUtility] subject to hard authority/privacy/safety constraints
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
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 نفيًا 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
أنشئ benchmark داخلي خاص بتعريف الإغلاق العلمي مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ benchmark أو ToolSandbox/Bench-AMA/APB
مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: mechanism unit، mission integrated، replication model-cross.
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
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim general يفشل domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغيً ا.
6. ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور تعريف الإغلاق العلمي، ثم implementation reference صغيرة، ثم replay failure، ثم ablation، ثم canary. لا توسع corpus قبل أن
تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
5
<PARSED TEXT FOR PAGE: 6 / 100>
PART XIV / PAGE 006 / 05 September 2026
الإغلاق ميثاق / Presence versus Integration versus Validation التطبيقي-النظري
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence، interface، metric، failure وقرار promote/
merge/reject. الصفحة محور: Presence versus Integration versus Validation.
1. الكائن النظري والدقة الافتراضية
بالمخطط مباشرة: MUT والنسب؛ المعنى يحفظ MUG يفرض placement/evidence/authority؛ M-HKOS إلى يحوله knowledge
ربط ي Presence versus Integration versus Validation
runtime. هي الأقرب المرافقة الورقة Moussaif Proof-Carrying Recursive Improvement & Scientific Closure Theory (M-
object؛ تشغيلي MOUSSAIXT كـ يطبقه contract
PCRI).
Promote ⇔ Δ>τ ∧ CriticalRegression=0 ∧ AblationSurvives ∧ Replication≥k
المعادلة هنا formal research object وليست رقم ًا تجريبي ًا جاهز ًا. 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 محتملًا.
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
APB/AMA-Bench/ToolSandbox/ من بضغط النتيجة اربط. held-out ومجموعة أبسط baseline مع Presence versus Integration versus Validation بـ خاص داخلي benchmark أنشئ
مستويات 3 نفذ. مباشرة أرقامها تنقل ولا ،مناسب benchmark أو: unit mechanism، integrated mission، cross-model replication.
PaperBench
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
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim يضيق نطاقه domain/model-cross يفشل general
بدلاً من ترقيته بلاغيًا.
6. ممر التطبيق
لمحور Presence versus Integration versus Validation، ثم reference implementation صغيرة، ثم failure replay، ثم ablation، ثم typed schema وtrace instrumentation بـ ابدأ
canary. لا توسع corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•


ميثاق الإغلاق النظري-التطبيقي / 5E0-E وإعادة ضبط الادعاءات
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence وقرار failure، metric، interface /promote reject/merge.
محور الصفحة: 5E0-E وإعادة ضبط الادعاءات.
1. الكائن النظري والدقة الافتراضية
يُربط 5E0-E وإعادة ضبط الادعاءات مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛

تشغيلي؛ object knowledge يحوله إلى HKOS-M
runtime contract. هي الأقرب المرافقة الورقة (Moussaif Unified Cognitive-State Transition Theory (M-UCST. كـ يطبقه
MOUSSAIXT
IdentityFidelity = InvariantConservation − CriticalInformationLoss
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
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
Agentic Misalignment 2025
Controlled simulations of harmful agentic actions under goal/replacement pressure;
motivates deployment-time governance.
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
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
أنشئ benchmark داخلي خاص بـ5E0-E وإعادة ضبط الادعاءات مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ToolSandbox/Bench-AMA/APB أو
benchmark مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.
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
أي claim يضيق نطاقه domain/model-cross يفشل general بدلاً من ترقيته بلاغيًا.
6. ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور 5E0-E وإعادة ضبط الادعاءات، ثم implementation reference صغيرة، ثم replay

لا توسع. canary ثم ،ablation ثم ،failure
تجريبية جديدة. bottleneck قبل أن تظهر corpus
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
7
<PARSED TEXT FOR PAGE: 8 / 100>
PART XIV / PAGE 008 / 05 September 2026
ميثاق الإغلاق النظري-التطبيقي / evidence versus Canonicality
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، evidence rank، interface، metric، failure وقرار promote/ merge/reject.
الصفحة محور: Canonicality versus evidence.
1. الكائن النظري والدقة الافتراضية
يُربط evidence versus Canonicality مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛ HKOS-M يحوله إلى object knowledge تشغيلي؛ MOUSSAIXT يطبقه كـruntime contract. هي الأقرب المرافقة الورقة (Moussaif Agentic Capability Closure Theory (M-ACCT.
Evidence_eff = Evidence_raw − DependencePenalty − LeakagePenalty
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
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 نفياً محتملاً.
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
أنشئ benchmark داخلي خاص بـevidence versus Canonicality مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ToolSandbox/Bench-AMA/APB أو

benchmark ،مناسب ولا تنقل أرقامها .مباشرة نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.
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
claim أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي general يفشل domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغيً ا.
6. ممر التطبيق
توسع لا canary. ثم ablation، ثم failure replay، ثم reference implementation صغيرة، ثم Canonicality versus evidence لمحور ابدأ typed schema وtrace instrumentation
corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•


ميثاق الإغلاق النظري-التطبيقي / الفجوة بين corpus وruntime
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence، interface، metric، failure، وقرار promote/ reject/merge.
محور الصفحة: الفجوة بين corpus وruntime.
1. الكائن النظري والدقة الافتراضية
يُ ربط الفجوة بين corpus وruntime مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛ HKOS-M يحوله إلى object knowledge تشغيلي؛
runtime. هي الأقرب المرافقة الورقة (Moussaif Question-Difference Geometry & Resource Rationality (M-QDGR. MOUSSAIXT كـ يطبقه contract
WorldEffect = ObservedPostState − VerifiedPreState
المعادلة هنا object research formal وليست رقمً ا تجريبيً ا 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
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
أنشئ benchmark داخلي خاص بالفجوة بين corpus وruntime مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ToolSandbox/Bench-AMA/APB أو
benchmark مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.
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 عاماً.
5. شرط الهدم
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim general يفشل domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغياً ا.
6. ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور الفجوة بين corpus وruntime، ثم implementation reference صغيرة، ثم replay failure، ثم ablation، ثم canary. لا توسع corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
9
<PARSED TEXT FOR PAGE: 10 / 100>
PART XIV / PAGE 010 / 05 September 2026
ميثاق الإغلاق النظري-التطبيقي / الفجوة بين runtime وbenchmark
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence، interface، metric، failure وقرار promote/reject/merge.
محور الصفحة: الفجوة بين runtime وbenchmark.
1. الكائن النظري والدقة الافتراضية
يُربط الفجوة بين runtime وbenchmark مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛ HKOS-M يحوله إلى object knowledge تشغيلي؛
هي الأقرب المرافقة الورقة (Moussaif Self-Reconstruction, Causal Memory & Cross-Model Identity Theory (M-SRCI.
MOUSSAIXT كـ يطبقه runtime contract.
SystemState_t=(World,Knowledge,Question,Difference,Belief,Actionability,Commitment,Memory,Governance,Budget)_t
المعادلة هنا formal object research وليست رقماً تجريبياً جاهزاً ا. كل متغير يجب أن يحصل على definition operational، 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
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
APB 2026
Agent Planning Benchmark: 4,209 multimodal cases across 22 domains; planning, broken tools, unsolvable tasks, refinement.
arXiv:2606.04874
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
أنشئ benchmark داخلي خاص بالفجوة بين runtime وbenchmark مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ToolSandbox/Bench-AMA/APB أو
benchmark مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.
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
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim general يفشل domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغياً ا.
.6 ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور الفجوة بين runtime وbenchmark، ثم implementation reference صغيرة، ثم replay failure، ثم ablation، ثم canary. لا توسع corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•

.
.
.
10
<PARSED TEXT FOR PAGE: 11 / 100>
PART XIV / PAGE 011 / 05 September 2026
ميثاق الإغلاق النظري-التطبيقي / الفجوة بين benchmark وgeneralization
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence، interface، metric، failure وقرار promote/ reject/merge.
محور الصفحة: الفجوة بين benchmark وgeneralization.
1. الكائن النظري والدقة الافتراضية
يُربط الفجوة بين benchmark وgeneralization مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/ placement؛ HKOS-M يحوله إلى object knowledge
هي الأقرب المرافقة الورقة (Moussaif Proof-Carrying Recursive Improvement & Scientific Closure Theory (M-PCRI. تشغيلي؛ MOUSSAIXT يطبقه كـ runtime contract.
Closure = TheoryIdentifiability × Executability × Verifiability × Reproducibility × Transfer
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
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
أنشئ benchmark داخلي خاص بالفجوة بين benchmark وgeneralization مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من /ToolSandbox/Bench-AMA/APB
أو benchmark مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.
PaperBench
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
claim أي مجمد يدمج أو يخفض. أي baseline تحت risk/cost أو localization failure أو calibration أو outcome لا يغير construct أي
يضيق نطاقه domain/model-cross يفشل general
بدلاً من ترقيته بلاغيً ا.
6. ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور الفجوة بين benchmark وgeneralization، ثم implementation reference صغيرة،
ثم replay failure، ثم ablation، ثم canary. لا
توسع corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
11
<PARSED TEXT FOR PAGE: 12 / 100>
PART XIV / PAGE 012 / 05 September 2026
ضغط المجتمع العلمي 2024-2026 / planning horizon-long
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence
/promote وقرار failure، metric، interface
merge/reject. الصفحة محور: long-horizon planning.
1. الكائن النظري والدقة الافتراضية
يُ ربط planning horizon-long مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛ HKOS-
M يحوله إلى object knowledge تشغيلي؛
هي الأقرب المرافقة الورقة. runtime contract كـ يطبقه (Moussaif Unified Cognitive-State Transition Theory (M-UCST.
MOUSSAIXT
Capability = Reachability × Reliability × Verification × Recovery × Transfer
المعادلة هنا formal research object وليست رقمً ا تجريبيً ا 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.

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

أنشئ benchmark داخلي خاص بـplanning horizon-long مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ ToolSandbox/Bench-AMA/APB أو

benchmark مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.

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

أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim general يفشل domain/model-cross يضيق نطاقه

بدلاً من ترقيته بلاغيًا.

6. ممر التطبيق

قبل corpus توسع لا. canary ثم، ablation ثم، failure replay ثم، reference implementation صغيرة ثم، long-horizon planning لمحور trace instrumentation وtyped schema بـ ابدأ

أن تظهر bottleneck تجريبية جديدة.

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

•

•

•

•

•





ضغط المجتمع العلمي 2024-2026 / use tool stateful

الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence، interface، metric، failure وقرار promote/ merge/reject. الصفحة محور: stateful tool use.

1. الكائن النظري والدقة الافتراضية

MOUSSAIXT تشغيلي، knowledge object إلى يحوله M-HKOS، placement/evidence/authority يفرض MUG والنسب؛ المعنى يحفظ ربطًا ي stateful tool use مباشرة بالمخطط: MUT

هي الأقرب المرافقة الورقة. (Moussaif Agentic Capability Closure Theory (M-ACCT. يطبقه كـ runtime contract.

Route* = argmax E[VerifiedUtility] subject to hard authority/privacy/safety constraints

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
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 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
benchmark أو APB/AMA-Bench/ToolSandbox/PaperBench من بضغط النتيجة اربط held-out. ومجموعة أبسط baseline مع stateful tool use
benchmark خاص داخلي بـ أنشئ
مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: mechanism unit، mission integrated، model-cross replication.
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
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim يفشل general domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغياً.
6. ممر التطبيق
أن قبل corpus توسع لا canary. ثم ablation، ثم failure replay، صغيرة reference implementation، ثم stateful tool use لمحور ابدأ بـ typed schema وtrace instrumentation
تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•

.
13
<PARSED TEXT FOR PAGE: 14 / 100>
PART XIV / PAGE 014 / 05 September 2026
causality memory / 2024-2026 ضغط المجتمع العلمي
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence
/promote وقرار failure، metric، interface
merge/reject. الصفحة محور: memory causality.
1. الكائن النظري والدقة الافتراضية
يُ ربط causality memory مباشرةً بالمخطط: MUT يحفظ المعنى والنسب؛ MUG يفرض authority/evidence/placement؛ HKOS-M
MOUSSAIXT تشغيلي؛ object knowledge يحوله إلى
runtime. هي الأقرب المرافقة الورقة (Moussaif Question-Difference Geometry & Resource Rationality (M-QDGR.
contract كـ يطبقه
Promote ⇔ Δ>τ ∧ CriticalRegression=0 ∧ AblationSurvives ∧ Replication≥k
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 نفيًا 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
أنشئ benchmark داخلي خاص بـcausality memory مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/
benchmark أو ToolSandbox/Bench-AMA/APB
مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: mechanism unit، mission integrated، model-cross replication.
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 عاماً.
5. شرط الهدم
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim general يفشل domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغياً.
6. ممر التطبيق
ابدأ بـschema typed وinstrumentation trace لمحور causality memory، ثم implementation reference صغيرة، ثم replay failure، ثم ablation، ثم canary. لا توسع corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
14
<PARSED TEXT FOR PAGE: 15 / 100>
PART XIV / PAGE 015 / 05 September 2026
ضغط المجتمع العلمي continual learning in memory / 2026-2024
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence /promote وقرار failure، metric، interface
merge/reject. الصفحة محور: continual learning in memory.
1. الكائن النظري والدقة الافتراضية
تشغيلي؛ knowledge object إلى يحوله M-HKOS؛ placement/evidence/authority يفرض MUG والنسب؛ المعنى يحفظ MUT: بالمخطط
مباشرة continual learning in memory ربطُ ي
هي الأقرب المرافقة الورقة (Moussaif Self-Reconstruction, Causal Memory & Cross-Model Identity Theory (M-SRCI.
MOUSSAIXT كـ يطبقه runtime contract.
IdentityFidelity = InvariantConservation − CriticalInformationLoss
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
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 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
continual learning مع baseline ومجموعة أبسط held-out. من بضغط النتيجة اربط APB/AMA-Bench/ToolSandbox/PaperBench أو
أنشئ benchmark خاص داخلي بـ in memory
benchmark مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.
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
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim
يضيق نطاقه domain/model-cross يفشل general
بدلاً من ترقيته بلاغيًا.
6. ممر التطبيق
توسع لا. canary ثم، ablation ثم، failure replay ثم، صغيرة reference implementation ثم، continual learning in memory لمحور
ابدأ بـ typed schema وtrace instrumentation
corpus قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: Candidate formalization theoretical. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•
15
<PARSED TEXT FOR PAGE: 16 / 100>
PART XIV / PAGE 016 / 05 September 2026
foresight model-world / 2024-2026 ضغط المجتمع العلمي
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، evidence rank،
/promote وقرار failure، metric، interface
merge/reject. الصفحة محور: world-model foresight.
1. الكائن النظري والدقة الافتراضية
HKOS- ؛authority/evidence/placement يفرض MUG ؛يحفظ المعنى والنسب MUT :مباشرةً بالمخطط foresight model-world يُربط
تشغيلي؛ object knowledge إلى M يحوله
هي الأقرب المرافقة الورقة (Moussaif Proof-Carrying Recursive Improvement & Scientific Closure Theory (M-PCRI.
MOUSSAIXT كـ يطبقه runtime contract.
Evidence_eff = Evidence_raw − DependencePenalty − LeakagePenalty
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
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 نفياً محتملاً.
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
أنشئ benchmark داخلي خاص بـworld-model foresight مع baseline أبسط ومجموعة out-held. اربط النتيجة بضغط من PaperBench/ أو ToolSandbox/Bench-AMA/APB
benchmark مناسب، ولا تنقل أرقامها مباشرة. نفذ 3 مستويات: unit mechanism، integrated mission، cross-model replication.
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
أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline مجمد يدمج أو يخفض. أي claim general يفشل domain/model-cross يضيق نطاقه
بدلاً من ترقيته بلاغياً.
6. ممر التطبيق
قبل توسع corpus. canary، ثم ablation، ثم failure replay، ثم reference implementation صغيرة، ثم world-model foresight.
ابدأ بـtyped schema وtrace instrumentation لمحور
أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: Candidate theoretical formalization. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•

العلمي المجتمع ضغط partial observability and belief / 2026-2024
الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence
/promote وقرار failure، metric، interface
merge/reject. الصفحة محور: partial observability and belief.
1. الكائن النظري والدقة الافتراضية
تشغيلي؛ knowledge object إلى يحوله M-HKOS، placement/evidence/authority يفرض MUG والنسب؛ المعنى يحفظ MUT: بالمخطط
مباشرة partial observability and belief ربطُ ي
runtime contract. هي الأقرب المرافقة الورقة (Moussaif Unified Cognitive-State Transition Theory (M-UCST. كـ يطبقه
MOUSSAIXT
WorldEffect = ObservedPostState − VerifiedPreState
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
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 نفيًا 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
partial مع baseline ومجموعة أبسط held-out. من بضغط النتيجة اربط APB/AMA-Bench/ToolSandbox/PaperBench أو
observability and belief بـ خاص داخلي benchmark أنشئ
benchmark مناسب، مباشرة أرقامها تنقل ولا. مستويات 3 نفذ: unit mechanism، integrated mission، cross-model replication.
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

أي claim مجمد يدمج أو يخفض. أي construct لا يغير outcome أو calibration أو localization failure أو risk/cost تحت baseline يفشل general يضيق نطاقه domain/model-cross
بدلاً من ترقيته بلاغيًا.
6. ممر التطبيق
لمحور partial observability and belief، ثم reference implementation صغيرة، ثم failure replay، ثم ablation، ثم canary. لا توسع ابدأ بـtyped schema وtrace instrumentation
قبل أن تظهر bottleneck تجريبية جديدة.
حالة هذا البناء: formalization theoretical Candidate. لا claim AGI؛ لا 4E دون baseline frozen؛ لا 5E دون replication.
•
•
•
•
•


الأطروحة: الإغلاق الأمثل لا يعني أن كل سؤال أجيب؛ بل أن كل claim مهم يملك حالة، آلية، mode، experiment، rank evidence
/promote وقرار failure، metric، interface