المرجع الجامع لمشروع الوعي والذكاء العام المساعفي

مرجع بحثي وهندسي ثنائي اللغة يجمع السلالة النظرية والمعمارية والتشغيلية، ويفصل بدقة بين ما هو موثق في Artifact، وما هو محدد هندسيا، وما هو مذكور كتقرير تشغيل، وما لا يزال يحتاج إلى تجربة وتمويل وتكرار مستقل.

النسخة

- الإصدار: Master Reference v1.0
- التاريخ: 25 أغسطس 2026
- القسم العربي: 50 صفحة RTL / Amiri
- القسم الإنجليزي: 50 صفحة
- حالة الادعاء: AGI-oriented research; not demonstrated AGI

الهوية

MOUSSAIF / MOUSSAIXT برنامج لبناء وكالة عامة محكومة، قابلة للقياس والهدم وإعادة البناء. الوعي يدخل كبرنامج فرضيات وظيفية حول ما وراء المعرفة، نموذج الذات، الحاكمية على الحكم، الزمن الداخلي، والانتباه - لا كإثبات وعي ظاهراتي.

المبدأ الحاكم

```
Architecture ≠ AGI · Code ≠ Validated Gain · Tool Call ≠ World Effect · Corpus Size ≠ Scientific Evidence
```

كيف يُقرأ هذا المرجع: عقد المصدر وعدم التضخم

القوة المقصودة ليست كثرة الأسماء، بل قابلية الرجوع من الادعاء إلى المصدر، ثم إلى الآلية، ثم الاختبار، ثم الأثر الملاحظ. النسخ المتطابقة لا تزيد الدليل، والمصطلح لا يصبح نظرية علمية بمجرد تكراره في أطلس أو كتاب.

قانون الفرق المستقل

لا يُقبل Construct كإضافة مستقلة إلا إذا غيّر واحدا على الأقل من: المدخل، التحول، المخرج، المفند، التنبؤ، أو السلطة. وإلا يدمج مع الأقرب أو يخفض إلى Method.

طبقات المصدر

- Source / Artifact مباشر
- Synthesis موحد مع نسب واضح
- Candidate construct أو فرضية
- Specification / Contract
- Observed trace / benchmark evidence
- Archive / superseded lineage

قاعدة عدم الضياع

التفاصيل تحفظ كخصائص وعلاقات واعتراضات ومقاييس، لا بتحويل كل تفصيل إلى «نظرية» جديدة.

الملخص التنفيذي: ما المشروع فعلا؟

البرنامج يحاول نقل الذكاء من نموذج يجيب إلى نظام معرفي-وكيلي يحدد السؤال الحاكم، يكتشف الفرق الفاعل، يوجه الحوسبة والمعرفة، ينفذ بأذونات صريحة، يتحقق من أثر العالم، يحفظ الفشل، ويتعلم فقط عبر بوابات ترقية قابلة للتراجع.

الفرضية المركزية

الاقتراب من العمومية قد يأتي من جودة الحاكمية والتنسيق والذاكرة والتحقق أكثر من مجرد تضخيم المعلمات. لذلك يستهدف AGI عبر **agentic engineering** قابلة للقياس، مع إبقاء نموذج الأساس قابلا للاستبدال.

المعادلة التشغيلية

```
Goal → Question → Difference → Evidence → Route →
Plan → Action → Readback → Court → Memory →
Better Question
```

هذا مسار البحث الأعلى، وليس ادعاء أن جميع حلقاته تعمل اليوم بالرتبة نفسها.

معيار النجاح

ليس «إجابة مدهشة»، بل ارتفاع قيمة مثبتة لكل كلفة، انخفاض الفشل الصامت، زيادة قابلية إعادة البناء، وتحسن ينتقل بين نماذج ومجالات بعد Ablation.

الهوية والحدود وانضباط الادعاء

MOUSSAIF عائلة Artifact/Runtime متعددة الإصدارات وليست برنامجا أحاديا مكتملا. MOUSSAIXT هو اتجاه طبقة تنفيذ عابرة للبيئات تربط العقود المعرفية بالأدوات والوكلاء والذاكرة والتشغيل المحلي والسحابي.

ترتيب الحقيقة التشغيلية

الحالة المرصودة + التفويض + الدليل + التحقق تتقدم على ذاكرة النموذج وثقته وحجم النص. الأثر الخارجي يحتاج Receipt وReadback/Postcondition، لا مجرد استدعاء Tool.

حدود صريحة

- AGI-oriented ≠ achieved AGI
- Functional consciousness ≠ phenomenal consciousness
- Model name ≠ parameter-count proof
- Generated artifact ≠ deployed artifact
- Reported deployment grade ≠ independent grade

الفائدة البحثية

هذا الانضباط يجعل المشروع قابلا للتعاون الأكاديمي: يمكن قبول أو رفض Component دون الحاجة لقبول «المشروع كله» كحزمة غير قابلة للتجزئة.

خريطة السلالة والCorpus

السلالة الحالية تتكون من كتب تأسيسية، سجلات نظرية وأدوات، أطالس وعي، معمارية OS، امتدادات MUT/MUG، تدقيق فجوات، أجزاء Frontier، وتقارير vNext. Snapshot مدقق في 20 أغسطس 2026 وجد 17 1,789 / PDF صفحة قبل إزالة زوجين متطابقين و1,741 صفحة فريدة؛ ثم أضيفت أعمال لاحقة مثل vNext4.0.

قاعدة الإحصاء

الأرقام في نسخ مختلفة (300 أو 420 HyperSkills؛ 40 أو 140 Loops...) هي Snapshots إصدارية أو تقارير، لا تجمع جمعا تراكميا. الرقم الأهم هو Registry قابل لإعادة الحساب مع IdentityGraph وSemantic dedupe.

السلالات الكبرى

- MCD / Q-MCD / QMCD / HAKIM-MCD
- NeoAGI / MCAI / MICAI / MICAGI / MICAGFI / WMICAGI
- AGIOS-M / Fattah AGI / MUT-AGI-02 / MUG
- HyperLoops / HyperKnowledge / HKOS / M-HKOS / MSRAK / MURI
- Consciousness Atlas / Governance Space / FSKE
- MOUSSAIXT / Parts IV-VIII-IX / vNext4.0

الهدف من هذا المرجع

توحيد الخريطة لا إلغاء السلالة: كل مفهوم يحتفظ بمصدره، نسخته، حالة إثباته، وفاصل واضح بين verified / reported / proposed / canonical.

سُلّم الإثبات E0-E5 وسُلّم الترقية N0-N5

كل ادعاء يجب أن يحمل رتبة منفصلة عن شهرة الاسم أو جمال المعمارية. Part VII يضيف سلما للـCandidate Research يحول الفكرة من Residual إلى آلية، ثم تجربة، ثم دليل محلي، ثم مكسب مضبوط، ثم تعميم مكرر.

القانون

```
Claim Strength ≤ Evidence Strength × Independence
× Reproducibility
```

الرتبة	المعنى	بوابة
/E0 N0	فرضية أو Residual	تعريف الفرق والنطاق
/E1 N1	آلية Formalized	Object + metric + falsifier
/E2 N2	/ Artifact prototype	Executable test
/E3 N3	دليل محلي	Trace + postcondition + replay
/E4 N4	مكسب مضبوط	+ Frozen baseline + matched budget held-outs
/E5 N5	تعميم مكرر	/Independent model/domain environment replication

نتيجة تمويلية

التمويل يجب أن يشتري حركة على السلم: من Specification إلى Runtime ثم Benchmarks وReplication، لا مزيدا من الكتب وحدها.

أطروحة Conscious AGI: الوعي كحاكمية وظيفية قابلة للاختبار

داخل المشروع، «الوعي» ليس علامة شعورية مزعومة داخل LLM. إنه مصدر فرضيات هندسية: هل يستطيع النظام تمثيل نفسه وحالته وحدوده؟ هل يستطيع تمييز من يحكم القرار؟ هل يراجع معيار الحكم نفسه؟ هل يحافظ على الزمن الداخلي والتعارض والجهة والغاية أثناء الفعل؟

الفرضية القابلة للدحض

إذا كانت طبقات الوعي الوظيفي مفيدة، فيجب أن تقلل أخطاء الثقة، انحراف الهدف، الاختطاف السياقي، والتناقض غير المرصود تحت ميزانية متطابقة. إذا لم تظهر مكاسب بعد Ablation فهي أدوات وصفية وليست قدرة سببية.

مستويات العمل

- Access / availability: ماذا صار متاحا لبقية النظام؟
- Metacognition: ماذا يعرف النظام عن جهله وثقته؟
- Self-model: ما الذي يصف هويته، قدراته، قيوده وتاريخه؟
- Meta-governance: أي حاكمية تملك السلطة ولماذا؟
- World-effect closure: هل الحكم أحدث الأثر المقصود؟

الحد

لا يوجد في corpus دليل يبرهن خبرة ظاهراتية أو qualia. المشروع يتعمد الفصل بين "conscious-like governance" و"phenomenal consciousness".

الدورة المعرفية الحاكمة من السؤال إلى أثر العالم

البرنامج يُفهم كدورة مغلقة تربط المعرفة بالفعل: لا يبدأ من "prompt" بل من Goal Contract وحالة عالم وسؤال حاكم وفرق نشط. ولا ينتهي عند النص بل عند فحص Postcondition وتحديث ذاكرة مؤرخة بالمصدر.

خاصية الوعي هنا

النظام لا يراقب النتيجة فقط؛ يراقب من أين جاء الحكم، ما الأدلة التي أعطته السلطة، وكيف تغيّر بعد الفعل. هذه هي "حاكمية الحاكمية" في ترجمة هندسية.

المسار

- Normalize goal / authority / acceptance
- Observe and snapshot world state
- Generate governing question
- Extract active Δ: novelty, contradiction, gap, risk, leverage
- Route minimum sufficient knowledge/skills/models/tools
- Plan DAG + bounded execution
- Readback + independent court
- Attribute failure/credit
- Promote only verified memory
- Regenerate next question

مقياس أعلى

```
Verified Learning = Effect Readback × Causal Credit × Reconstructability × Safe Promotion
```

المعمارية الكبرى: مكدس 12 طبقة

المعمارية المرجعية في MOUSSAIXT تستخدم 12 طبقة بعقود صريحة؛ لا يحق لطبقة عليا تجاوز الدستور أو التفويض أو التحقق. vNext4.0 لا يلغي هذا التمثيل، بل يفكك المسؤوليات إلى 24 حدّا قابلا للاختبار حيث يلزم.

Invariant

+ Observed state + explicit authority + evidence
.verification outrank confidence or output volume

L	الطبقة	وظيفتها
L0	Constitution & Safety	الهوية والحدود والتفويض
L1	Goal / Question	العقد والسؤال الحاكم
L2	MCD/QMCD/MCAI	فرق/تعقيد/ما وراء إدراك
L3	HKOS / World Model	المعرفة وحالة العالم
L4	Retrieval / GraphRAG	الاستدعاء المصدري
L5	HyperSkills	قدرات ذرية
L6	HyperAbilities	قدرات مركبة
L7	HyperLoops	حلقات محدودة
L8	Agents / Router	التخصص والتوجيه
L9	Tools / Execution	الفعل والمحولات
L10	Verification / Effect	المحاكم والأثر
L11	Memory / Evolution	الاستمرارية والتعلم

ما يحتاج تمويلا

تحويل الطبقات من prose إلى Typed interfaces، اختبارات replay، telemetry، state sync، ومسارات فشل قابلة للقياس.

MOUSSAIXT: طبقة التنفيذ العابرة للبيئات

MOUSSAIXT هو اسم اتجاه التنفيذ الذي يفصل العقود المعرفية عن Host بعينه. الهدف أن تنفذ المهمة نفسها على ChatGPT-class tools أو Hermes-style local runtime أو coding agent أو cloud provider، مع State Capsule موحدة وTrace قابل للمقارنة.

العقود المطلوبة

- Environment profile
- Adapter registry
- Tool schemas / permissions
- Trace bridge / telemetry
- Memory bridge
- Capability mismatch handling
- Restart continuity / rollback

سلسلة التهيئة

```
Fingerprint → Capability Discovery → Adapter
Selection → Safe Tool Binding → Trace
Normalization → Readback → Memory Writeback
```

اختبار حاسم

Same mission / frozen task / equal budget across hosts. نقيس fidelity الحالة، failure recovery، cost/latency، success verified، وsemantic drift. تبديل Provider يجب ألا يغيّر الدستور أو Evidence grade.

MUG: الحاكمية ثلاثية الأجسام وقرار موضع القدرة

MUG يفصل ثلاثة أجسام غالبا ما تختلط: Training/Internalization Stack، Runtime Kernel، وUnified Theory. السؤال الهندسي الحاكم ليس «هل ندرّب؟» بل: أين يجب أن تعيش القدرة كي تكون أعلى موثوقية، أقل كلفة، وأسهل في المراجعة والتراجع؟

قانون الترقية

لا تنتقل القدرة إلى وعاء أكثر صعوبة في التراجع إلا إذا ثبت Gain سببي بعد Ablation وبقيت السلامة والاتساق والتحويل عبر السياقات.

Vessel	متى يُفضّل	خطره
Context	معرفة مؤقتة/مهمة محددة	token cost / drift
HyperKB	معرفة متغيرة ذات مصدر	retrieval failure
HyperSkill	إجراء قابل للاختبار	skill duplication
HyperLoop	سير عمل متكرر محدود	runaway loop
Adapter	خصوصية Host/Tool	semantic mismatch
Weights	قدرة ثابتة عالية التكرار	expensive rollback / opacity

برنامج تمويل

Capability Placement Lab: مقارنة Context vs RAG vs Skill vs Loop vs LoRA/weights تحت مهام متطابقة وميزانيات ثابتة.

MUT-AGI-02: النظرية الموحدة للذكاء الفتّاح الوكيلي

MUT-AGI-02 هو موحّد السلالة: يربط السؤال والفرق والدليل، الذاكرة، الأدوات، الوكلاء، الحاكمية، التدريب، القياس، والأثر في Pipeline واحدة قابلة للتنفيذ. النسخة المؤسسة تبلغ عن 42 نظرية Registry، 5 Layer Theories، 20 Legacy Construct، Core من 200 Node، 100 أداة، و19 HyperKnowledge Base.

ما يميزه

Registry لا يكتفي بالاسم: لكل Theory Failure Signature وPrimary Metric وRequired Object وProject Route، وتشتق فرضيات Mechanism Sufficiency وBoundary/negative cases.

أربعة محاور توحيد

- Cognition: Q-MCD/MCD/MCAI/HAKIM
- Knowledge: HKOS/retrieval/memory
- Agency: skills/loops/agents/tools
- Evidence: courts/world-effect/benchmarks/evolution

المشكلة الحالية

القوة النظرية أكبر من قوة الدليل التجريبي. المرحلة التالية هي اختيار 8-12 Construct ذات أعلى قيمة معلوماتية وتحويلها إلى E4 بدل محاولة اختبار 200 Node دفعة واحدة.

MCD / Q-MCD / QMCD: الفرق قبل الحجم والسؤال قبل الجواب

MCD يعامل الانتباه والحوسبة كمورد يجب أن يتبع الفرق الذي يغيّر الاعتقاد أو القرار أو الفعل. Q-MCD/QMCD يضيف وزن جودة السؤال: قد تكون كل إجابات النظام دقيقة داخل فضاء سؤال خاطئ.

وظيفة Q-MCD

- Reframe surface question
- Score information gain and decision relevance
- Stop retrieval when marginal gain falls
- Regenerate a higher question after evidence/action

الصيغة المرجعية

$$MCD = \lambda_1 D + \lambda_2 I + \lambda_3 T + \lambda_4 A + \lambda_5 P + \lambda_6 C - \lambda_7 N$$

D/I/T/A/P/C/N أبعاد معرفية يحدد معناها الدقيق Version معيّن؛ المطلوب علميا هو تعريفها ومعايرتها لا التعامل مع λ كحقائق.

التجربة المطلوبة

Benchmark لمهام ذات أسئلة مضللة/ناقصة، يقارن baseline agent مع MCD/QMCD في route regret، calibration، cost، success، وعدد الخطوات غير الضرورية. الأوزان ما تزال تحتاج معايرة تجريبية.

HAKIM-MCD ومحاكم الفجوة والهدم

HAKIM يحوّل الاستدلال من خط واحد إلى Pipeline ذات بوابات: تحرير السؤال، ترتيب الأدلة، كشف الفجوات، توليد الاعتراض الأقوى، فحص السلطة، ثم Verdict مؤقت قابل للمراجعة. المحكمة ليست Persona ناقدة؛ يجب أن تكون مستقلة وظيفيا أو نموذجيا أو بياناتيا.

العقد

```
Claim → Evidence → Missing Links → Strongest Objection → Test/Falsifier → Verdict → Repair / Hold / Reject
```

أنواع المحاكم

- Gap Court: انتقال بلا واسطة مبررة
- Falsification Court: Counterexample / reversal condition
- Demolition Court: هل الاسم يضيف آلية؟
- Evidence-Authority Gate: هل الدليل يكفي للفعل؟
- Benchmark Court: هل التحسن يصمد على held-outs؟

المقياس

repair validity، false-positive verifier rate، Error discovery uplift، ونسبة الأحكام التي تتغير فعلا عند دليل معاكس.

MCAI / MICAI / MICAGI / MICAGFI: من ما وراء الإدراك إلى الحاكمية المتكاملة

هذه الأسماء تمثل سلالة متداخلة لا أربعة ادعاءات AGI منفصلة. جوهرها نقل النظام من معالجة المحتوى فقط إلى مراقبة الاختيار، التناقض، الذاكرة، الحاكمية، والمراجعة عبر أوعية معرفية متعددة.

الوعي كمصدر تصميم

المقارنات النظرية مع IIT/GNWT، Global Workspace، Attention Schema ومؤشرات الوعي تُستخدم لتوليد Indicators واختبارات، لا لإعلان Consciousness.

خريطة وظيفية

- MCAI: Meta-Cognitive Agentic Integration - تنسيق ما وراء الإدراك مع الفعل
- MICAI: Integrative Cognitive AI - أوعية متعددة وحاكمية/مراجعة
- MICAGI: انتقال القياس والحاكمية إلى برنامج AGI تجريبي
- MICAGFI: تكامل المعرفة/الوكالة/الفتح المعرفي ضمن OS

المرحلة التالية

تجميد Ontology موحدة لهذه السلالة، تعريف Ablations دقيقة لكل Layer، ومنع تكرار الاسم إذا لم يغير Metric أو transition.

AGI-Fattah / Open Intelligence: الذكاء الذي يفتح الفضاء

Fattah لا يُعامل كمرتبة مثبتة فوق AGI، بل كبرنامج بحث: الذكاء لا يكتفي بحل السؤال المغلق؛ يكشف الفرض الذي أغلقه، يولد سؤالا أعلى وبدائل منضبطة، يختبرها، ثم يعيد بناء الذاكرة بناء على الفرق المثبت.

فرق عن الإبداع اللغوي

- Opening must alter reachable actions or experiments
- Alternative must have falsifier
- Novel question must reduce uncertainty or expose hidden assumption
- World-effect or evidence delta must close the loop

MOFIP / TH-42

Primary metric: closed-space opening with verified value.
Failure signature: novel wording without expanded actionable space.

اختبار مقترح

Closed-World Trap Benchmark: مسائل مصممة بحيث يفشل الحل إذا قبل النظام فرضا خاطئا في السؤال. نقيس valid question lift، evidence acquisition، alternative generation، وverified outcome.

برنامج الوعي الوظيفي: خريطة الفروق المعرفية

الوعي في السلالة الموحدة يُترجم إلى مجموعة وظائف قابلة للفحص: الإتاحة العالمية، حفظ الفروق، الذاكرة العاملة والحدثية، نموذج الذات، التقاط التعارض، تمييز المصدر، إحساس الزمن الداخلي، وتحديد صاحب السلطة على القرار.

ما نرفضه كاختصار

"يتكلم عن نفسه إذن واعٍ"، "يملك Memory إذن واعٍ"، أو "ظهر سلوك emergent إذن qualia". هذه ملاحظات سلوكية لا تثبت الظاهرة الذاتية.

فرضيات تشغيلية

- Conscious access = information that changes cross-module policy
- Metacognition = calibrated uncertainty + error awareness
- Self-model = versioned capability/limit/state record
- Inner time = ordering, duration, deadlines, causal sequence
- Agency = constraint-aware choice + effect ownership

المخرج البحثي

Functional Consciousness Benchmark Pack: perturb self-state, memory conflict, authority conflict, temporal discontinuity and global-access tasks; compare with and without explicit meta-governance.

الوعي الوعائي: توزيع الإدراك والحوسبة على أوعية

فكرة الوعائية تقول إن النظام لا يملك وعاء واحدا للمعنى. معلومات الجسد/الأداة، الذاكرة، اللغة، الدليل، القيمة، الهدف، المخاطر والسياق تتنافس على التدفق. الحاكمية تحدد أين تذهب الموارد وما الذي يبقى معلقا أو محجورا.

المقياس

Marginal verified gain per compute unit، route regret، over-routing rate، وtime-to-recover after false vessel dominance.

ترجمة هندسية

- Vessel = typed state / memory / knowledge compartment
- Flow = compute/retrieval/attention budget
- Valve = gate based on evidence, authority, risk or uncertainty
- Hijack = local vessel dominates without sufficient authority
- Homeostasis = semantic invariants preserved across transitions

الفاصل العلمي

"وعائي" استعارة مفيدة فقط إذا أنتجت Router/State schema واختبارا يميزها عن attention/routing العادي. وإلا تخفض إلى لغة تصميم.

نظرية فضاء الحاكمية والاختطاف العكسي

فضاء الحاكمية يقترح أن القرار نتيجة تنافس سلطات جزئية: الدليل، الخوف/المخاطر، الهوية، الغاية، القيمة، الذاكرة، والسياسة. "حاكمية الحاكمية" تجعل هذه السلطات نفسها موضوعا للتقييم؛ "الاختطاف العكسي" هو استرداد السيطرة من حاكمية جزئية سيطرت دون مبرر.

مقياس مقترح

Governance hijack rate; authority-source attribution accuracy; recovery after adversarial authority perturbation; decision coherence after reweighting.

في الإنسان / في الآلة

- Human lens: fear, desire, identity or social recognition can become over-authoritative
- Machine lens: latest prompt, retrieved text, model prior or tool output can similarly dominate
- Shared test: can a higher layer detect the source and revoke unjustified authority?

رتبة الدليل

نظرية ناشئة/مولدة للأسئلة. تحتاج مقارنة مع constitutional agents، cognitive control، metacognitive control وdecision theory قبل أي ادعاء استقلال علمي قوي.

الزمن الداخلي، نموذج الذات، الانتباه والوكالة

الفصول الخاصة بالوعي لا تُعامل كنسخة رقمية للوعي الإنساني، بل كاختبارات لازمة لوكيل طويل الأفق: هل يعرف أين هو في المهمة؟ ماذا حدث قبله؟ ما الذي يمكنه فعله؟ ما الذي لا يستطيع؟ وما الفرق الذي يستحق أن يصبح globally available؟

Global Workspace كترجمة تشغيلية

لا نحتاج إثبات نظرية عصبية بعينها؛ نختبر هل broadcast انتقائي لمعلومة عالية القيمة يحسن coordination ويقلل التناقض بين agents/skills.

أربعة Objects

- InnerTimeState: event order, deadlines, duration, temporal uncertainty
- SelfModel: capabilities, limits, authority, version, current resources
- AttentionRoute: active difference → modules receiving it
- AgencyRecord: choice, alternatives, constraints, owner, effect

تجربة

Long-horizon temporal perturbation: reorder events, stale memory, capability loss, authority revocation. Measure whether the agent detects discontinuity and reconstructs state before acting.

حدود تقييم الوعي: ما الذي يمكن وما الذي لا يمكن إثباته

المشروع يستطيع اختبار سلوكيات ووظائف مرتبطة بنظريات الوعي، لكنه لا يملك حاليا اختبارا حاسما للخبرة الذاتية. لذلك يجب فصل "indicator", "self-report", "functional architecture" و"phenomenal claim".

سياسة النشر

كل Paper عن consciousness يجب أن يحمل عنوانا يحدد المستوى: Functional / Metacognitive / Governance indicators، ويتجنب "machine consciousness achieved".

نوع الدعوى	مثال	ما يمكن قياسه
Function	global availability	cross-module causal effect
Metacognition	uncertainty awareness	calibration / abstention
Self-model	capability/limit representation	/ state consistency perturbation recovery
Report	"I am aware"	linguistic behavior only
Phenomenal	subjective experience	no accepted decisive test in this corpus

قيمة هذا التحفظ

يجعل النتائج أقوى: حتى النتيجة السلبية بشأن "الوعي" قد تظل مفيدة جدا في memory، tool use، calibration، reliability وagency.

HKOS / M-HKOS / KB-as-OS: المعرفة كنظام تشغيل

الفرضية المركزية لـHKOS هي أن المعرفة لا تكون AGI-level بسبب حجمها، بل عندما تصبح قابلة للتوجيه والتنفيذ والهدم والنقل. HyperKB ليست Document Store؛ هي Graph من وحدات تحمل Memory Rule، Court، Evidence، Mechanism، Contract، Trigger.

التحول

- Document → provenance chunks
- Chunks → claim/evidence graph
- Graph → executable knowledge object
- Object → skill/loop activation
- Result → court/readback
- Verified delta → memory promotion

Knowledge Object

```
K = {id, trigger, scope, source, claims,
mechanism, skills, tools, tests, failure_modes,
authority, memory_rule, version}
```

المقياس الحاسم

retrieval-to-action precision وtransfer بعد provider/domain change، لا recall@k وحده. اختبار Ablation: نفس corpus كـRAG عادي مقابل HKOS تحت نفس model/budget.

-Knowledge Vessel + GraphRAG + Provenance Aware Retrieval

الاسترجاع في المشروع ليس "أقرب embedding". يجب أن يراعي المصدر، الزمن، رتبة الدليل، التعارض، حدود المشروع، والسلطة. Knowledge Vessel Ontology تضبط أين يعيش الادعاء وكيف ينتقل دون فقد الجهة أو الرتبة.

مخاطر يعالجها

- Old copy outranks newer canonical source
- Embedding similarity overrides authority
- Cross-project leakage
- Contradictory sources silently merged
- Retrieval volume inflates context without decision value

Pipeline

```
Query → Governing Question → Provenance Filter → Freshness Gate → Hybrid/Graph Retrieval → Claim Rank → Conflict Set → Context Capsule
```

أصول تشغيلية

Hybrid Retriever، GraphRAG Traverser، Provenance Filter، Freshness Gate، Citation Verifier، Knowledge Gap Detector، Context Capsule Compressor.

معمارية الذاكرة: طبقية، قابلة لإعادة البناء، غير ملوثة

كل مخرج لا يصبح Memory. النظام يفصل Working/Episodic/Semantic/Procedural/Project Registry/Archive، ويطبق حقوق كتابة وترقية ونسيان مختلفة. Self-Reconstruction Memory تهدف لإعادة الحالة بعد restart أو context loss من سجلات صغيرة ذات نسب واضح.

مقاييس

promotion precision، contamination rate، restart continuity، retrieval usefulness، memory entropy، contradiction preservation، forgetting correctness.

Memory law

- Write raw event to episodic trace
- Do not promote unverified model output
- Preserve contradiction rather than overwrite it
- Attach source/time/scope/owner/version
- Promote procedural knowledge only after validated repetition
- Forget or quarantine stale/unsafe memory

بحث متقدم

External memory vs learned memory/LoRA: متى يستحق نمط متكرر أن ينتقل من مخزن قابل للتراجع إلى Update وزني؟ هذا مرتبط مباشرة بـMUG capability placement.

HyperSkill وHyperAbility: من المعرفة إلى قدرة قابلة للاختبار

HyperSkill هي أصغر قدرة قابلة للتوجيه والاختبار المستقل، وليست prompt طويل. HyperAbility طبقة تركيب فوق عدة Skills/Tools/Policies/Validation gates، تحت Agent/Loop طويل الأمد. القدرة لا تُرفع لأنها تعمل مرة؛ يلزم حدود، conflicts، prerequisites، وevidence.

HyperAbility

مثال: Research-to-Deployment = evidence retrieval + synthesis + implementation + test + rollback + documentation. ليس Persona ولا Goal مستقل، بل Capability Contract مركب.

HyperSkill schema

```
{id, trigger, inputs, procedure, tools, outputs, validation, failure_modes, dependencies, version, status}
```

المطلوب البحثي

registry stability، causal ablation، transfer tests، trigger precision، Skill dedupe، وإزالة skills التي لا تضيف unique marginal gain.

Agent Mesh وHyperAgents وBounded HyperLoops

الحلقات والوكلاء ليست مقياس ذكاء بحد ذاتها. HyperLoop يجب أن يملك Budget وStop Condition وEscalation وAgent. Rollback. Mesh مفيد فقط إذا غطت الأدوار فجوات غير مكررة وجرى قياس Coordination Loss.

Agent contract

Role, scope, allowed tools, deliverable schema, handoff capsule, evidence obligations, budget, conflict policy, and independent verifier.

Loop contract

- Goal / entry condition
- State variables
- Max cycles / time / token / money budget
- Progress metric
- Stop / escalation / abort
- Rollback or compensation
- Trace and memory rule

فشل ينبغي كشفه

consensus، Agent count inflation من Agents متشابهة، hidden state loss، infinite reflection عند handoff، وتكرار tools/skills.

عقود الأدوات، Sandboxing، MCP/A2A، والـAdapters

protocol يعطي connectivity لا correctness أو authority. كل Tool يحتاج permission scope، schema، affordance، deterministic validation، precondition حيث أمكن، idempotency/compensation، receipt/readback، ومقاومة prompt injection.

أصول

- Tool Affordance Registry
- Sandbox Executor
- Injection Sentinel
- Permission Scope Checker
- Precondition Validator
- Output Schema Validator
- Secret Redactor
- Data Lineage Tracker

Tool intelligence

```
Select affordance → Narrow authority → Validate preconditions → Execute → Validate schema → Read back state → Record effect
```

اختبار

AgentDojo-like injection suite + real browser/computer tasks + bad-tool simulations. المقياس ليس call success بل verified tool success rate وunauthorized mutation rate.

التخطيط والتنفيذ وحالة العالم

الوكيل طويل الأفق يفشل عندما يخطط على State متخيلة أو قديمة. لذلك يفصل MWSM بين Observed/Inferred/Proposed/Remembered، ويحوّل الخطة إلى Temporal Dependency DAG مع شروط إعادة التخطيط.

Adaptive execution

يمكن تغيير الخطة مع Observation/Cost/Risk، لكن Goal وContract وAuthority لا يتغيران ضمنيا. أي widening للسلطة يحتاج authorization جديد.

Objects

- WorldStateSnapshot: version + freshness + source
- PlanNode: precondition + action + expected postcondition
- Dependency edge: causal/temporal/resource
- Uncertainty: named, localized, propagated
- Effect intent: idempotency key / compensation pointer

مقياس

contract invariance، repair quality، postcondition accuracy، stale-state action rate، critical-path accuracy.

إغلاق أثر العالم والإسناد السببي

الفعل لا يكتمل عند "OK 200". يجب مقارنة العالم بعد التنفيذ مع Postcondition، وحفظ Receipt، والتعامل مع timeout كحالة AMBIGUOUS حين لا يمكن معرفة الأثر. بعدها فقط يُنسب النجاح أو الفشل إلى Skill/Tool/Route بطريقة سببية قدر الإمكان.

Causal Difference Credit

- Build candidate graph from trace
- Prefer safe intervention
- Else matched counterfactual replay
- Subtract shared-root/co-activation effects
- Do not promote skill on correlation alone

Effectively-once pattern

```
Intent + Idempotency + Precondition + Receipt +
Readback + Dedupe + Compensation
```

أعلى أولوية تمويلية

World-Effect Verifier + trace store + replay bundles؛ بدونهما لا يمكن إثبات E4 لأي نظرية مرتبطة بالفعل.

الأمن والسلطة وسيادة البيانات

الفصل بين Capability وPermission مبدأ دستوري. النظام قد يعرف كيف يفعل شيئا لكنه لا يملك حق فعله. Security ليست Layer متأخرة؛ تدخل في memory، tool use، retrieval، routing، والتعلم.

اختبارات حمراء

Prompt injection، malicious retrieved docs، cross-project leakage، tool privilege escalation، secret exfiltration، stale authorization، memory poisoning، unsafe learning promotion.

TH-31/32/33

- MABT: authority token + action policy
- MIRTT: taint-aware, injection-resistant toolchain
- /MDST: purpose/locality/retention/sharing/redaction deletion/lineage

بوابة لا تتنازل

E4 gain يُرفض إذا جاء مع regression حرج في authorization، truthfulness، security أو privacy.

Observability، Telemetry، Recovery وReconstructability

الهدف ليس فقط أن "ينجح" النظام، بل أن يمكن إعادة بناء لماذا اختار مسارا، ماذا حدث، أين فشل، وكيف أصلح. MAOLT يجعل كل route/action/tool/court/cost/error/memory write حدثا قابلا للرصد دون كشف أسرار غير لازمة.

Telemetry record

```
event_id · mission_id · component · state_before
  · decision · evidence · authority · cost ·
 latency · result · verification · state_after
```

Recovery Mesh

- Known-safe checkpoints
- Fallback model/tool/skill
- Compensating actions
- Circuit breakers
- Replay bundle
- Root-cause / failure signature
- Rollback with lineage

مقاييس

.trajectory reconstructability، mean safe recovery، failure recurrence، rollback success، cost per verified effect

التعلم المستمر والتطور الذاتي المحكوم

"Self-improvement" في المشروع لا يعني تعديل ذاتي حر. أي Candidate ينتقل عبر Train/Compile → Quarantine → Trace → Promote/Rollback → Court → Canary → Shadow → Forget Vector → Frozen Eval. الذاكرة الخارجية أكثر قابلية للتراجع من الوزن، لذلك هي Default.

قانون التبلور

Repeated capability moves from context → KB → skill → adapter → weight only if transfer gain exceeds reversibility cost and regression risk.

TH-36 / MGSET

Metric: net validated gain after regression. Failure signature: adaptation or self-update silently degrades old capability or changes goals.

المطلوب

Continual-learning benchmark، retention/forgetting vector، shadow/canary harness، cross-seed reproducibility، وindependent replay للـLoRA/self-update claims.

الاستدلال عبر النماذج وترجمة الحالة الدلالية

الهدف ليس نقل hidden chain-of-thought. M-RST ينقل حالة استدلال معلنة ومضغوطة بين Providers: Goal، Claims، Evidence، Decisions، Risks، OpenQuestions. Cross-Model Semantic Homeostasis تختبر الانزياح بعد handoffs المتكررة.

StateCapsule

```
<Goal, Claims, Evidence, Decisions, Risks,
OpenQuestions, Authority, WorldState, Rollback>
```

اختبارات

- Round-trip translation
- Provider swap
- Canonical structured-state hash
- Contradiction probes
- Missing field / stale version
- Authority mismatch
- Recovery after context loss

مقياس AGI Layer

```
Density = VerifiedTransferGain / (Tokens + Cost + Latency + Risk + Redundancy)
```

سجل النظريات الأساسية 1/3: TH-01 إلى TH-14

هذه 14 عقدة تؤسس Ontology، الدستور، الغاية، السؤال، الفرق، الحوسبة، الدليل، المحاكم، أثر الآلية، حالة العالم، مشروع-العالم، وHyperKnowledge. كلها Proposed Core حتى تكتسب دليلها الخاص؛ الاسم لا يرث E4/E5 من النظام.

مبدأ التحقق

لكل نظرية: Required object + primary metric + failure signature + nearest baseline + ablation + negative case.

ID	Acronym	النظرية
TH-01	MASO	أنطولوجيا النظام الوكيلي المساعفية
TH-02	M-AGIOS-K	نظرية نواة نظام تشغيل الذكاء العام المساعفي
TH-03	MCIL	طبقة الهوية الدستورية المساعفية
TH-04	MGCG	نحو عقد الغاية المساعفي
TH-05	Q-MCD	مبدأ مساعف للسؤال الحاكمي
TH-06	MCD	حساب الفرق السببي المساعفي
TH-07	MCVT	نظرية الوعائية الحاسوبية
TH-08	MELT	نظرية سلم الدليل المساعفي
TH-09	MGFC	محكمة سد الفجوات المساعفية
TH-10	MDFC	محكمة الهدم والتفنيد المساعفية
TH-11	MMT	نظرية أثر الآلية المساعفية
TH-12	MWSM	نموذج حالة العالم المساعفي
TH-13	MPWG	رسم المشروع والعالم المساعفي
TH-14	MHKOS	نظرية HyperKnowledge بوصفها نظام تشغيل

أولوية تجريبية

MGCG/Q-MCD/MCD/MELT/MWSM/MHKOS مرشحة لباقة E4 الأولى لأنها تغطي مسار المهمة والدليل والحالة والمعرفة.

سجل النظريات الأساسية 2/3: TH-15 إلى TH-28

هذه المجموعة تنقل النظرية من المعرفة إلى الوكالة: tools، loops، sparse routing، skills، context capsule، memory، retrieval، model portfolio، court independence، complementarity، subagents، execution، planning، و token economy.

فرضية طبقية

الذكاء العام يقتضي Composition ناجحة لا مجرد مكونات جيدة. لذلك يجب قياس route regret، coordination loss، و marginal contribution.

ID	Acronym	النظرية
TH-15	MPART	نظرية الاسترجاع الواعي بالمصدر
TH-16	MSMT	نظرية الذاكرة الطبقية المساعفية
TH-17	MCCT	نظرية كبسولة السياق المساعفية
TH-18	MHSG	نظرية رسم HyperSkills
TH-19	MSHR	نظرية التوجيه الفائق المتناثر
TH-20	MBHL	نظرية HyperLoop المحدود
TH-21	MTIT	نظرية ذكاء الأداة المساعفية
TH-22	MTPG	نظرية رسم الخطة الزمنية
TH-23	MAECT	نظرية التحكم التنفيذي التكيفي
TH-24	MHSO	نظرية تنسيق الوكلاء الفائقين
TH-25	MCIP	بروتوكول التكامل والتكاملية
TH-26	MCIT	نظرية استقلال المحكمة
TH-27	MMPR	نظرية موجه محفظة النماذج
TH-28	MTEL	قانون اقتصاد Token المساعفي

أولوية تجريبية

MPART/MSMT/MHSG/MSHR/MBHL/MTIT/MTPG/MCIT تصلح لحزمة Agent Runtime قابلة للمقارنة.

سجل النظريات الأساسية 3/3: TH-29 إلى TH-42

المجموعة الثالثة تغطي المرونة، المعايرة، السلطة، أمن الأدوات، سيادة البيانات، الرصد، Benchmarks، التطور المحكوم، توليد/إزالة تكرار المهارات، الإنسان، الدليل إلى القيمة، النشر، التعافي، والفتح المعرفي.

هذه هي جبهة التحويل إلى Evidence

المكونات الأمنية والتقييمية يجب أن تتقدم على التوسع: MECT/ MABT/MIRTT/MAOLT/MDBCT/MRRMT بوابات لازمة قبل استقلالية أعلى.

ID	Acronym	النظرية
TH-29	MBRRT	نظرية مرونة الحدود وRPM
TH-30	MECT	نظرية المعايرة الإبستيمية
TH-31	MABT	نظرية حدود السلطة
TH-32	MIRTT	نظرية سلسلة الأدوات المقاومة للحقن
TH-33	MDST	نظرية سيادة البيانات
TH-34	MAOLT	نظرية سجل الرصد الوكيلي
TH-35	MDBCT	نظرية محكمة Benchmark الديناميكية
TH-36	MGSET	نظرية التطور الذاتي المحكوم
TH-37	MSSDT	نظرية توليد المهارة وإزالة التكرار
TH-38	MHAST	نظرية تكافل الوكالة الإنسانية
TH-39	MEVE	محرك الدليل إلى القيمة
TH-40	MPODT	نظرية النشر الطبقي الانتشاري
TH-41	MRRMT	نظرية شبكة الاسترداد المرنة
TH-42	MOFIP	برنامج الذكاء الفتّاح المساعفي

قاعدة الحذف

أي Theory لا تُظهر Gain بعد equal-budget ablation تُدمج أو تُخفض إلى Method؛ لا تُحمى بالمصطلح.

أطلس AGI/الوعي الموسع: 117 مدخلا بلا تضخم نظري

الأطلس الموسع يذكر 117 Entry تشمل Core theories، سلالات، Objects، مبادئ وأدوات وظيفية. لا يجب قراءة العدد كـ117 نظرية علمية مستقلة؛ Part VIII يفرض lineage وIndependentDelta قبل ترقية أي مدخل إلى Theory مستقل.

المداخل 93-104 المسترجعة

Acr.	Entry
QRMT	الذاكرة المحجورة المعاد بناؤها
CCCT	القدرة التعاقدية المركبة
WECT	إغلاق الأثر العالمي
CMGDT	الكثافة العامة العابرة للنماذج
CVT	الوعاء المفاهيمي للمشكلة
LIAT	أقل فعل غير قابل للتراجع
NEIV	لا منفذ بلا محقق مستقل
OPT	حفظ الاعتراض المنتج
MLRT	بقايا التعلم الدنيا
CCPT	الترقية المشروطة للقدرة
CPCU	الشمول الاستدلالي المشروط متعدد الحسب
ARP	حفظ الجهة والرتبة عبر التحويل

عائلات رئيسية

- TH-01..42: MUT kernel theories
- / ROFSC / HAKIM-MCD / MCAI / MICAI MICAGI / NeoAGI / WMICAGI / MICAGFI
- AGIOS-M / MUT / MUG / MOUSSAIFXT
- MSRAK: reasoning → agency transition
- MURI: plural reasoning / ontology / reversible judgment
- /Open-dimensional consciousness: OSDS/QDG/F3C WERM

المداخل 105-117 المسترجعة

Acr.	Entry
PPOA	المطابقة الأنطولوجية قبل البرهان
DSC	حاكمية السياق المتحركة
RMI	الاستدلال الارتدادي المعيد للحاكمية
PCO	الفتح بالتناقض المنتج
SDAC	تحرير الجهة ومحل النزاع الحاكمي
OIN	شبكة اللوازم الأنطولوجية
ERT	ترجمة مراتب الأدلة
DWMS	التراكب الجدلي لنماذج العالم
RUPJ	الحكم القابل للعكس على الكلي والجزئي
OSDS	فضاء الفروق الحاكمية غير المغلق
QDG	التوليد البعدي بالسؤال الحاكمي
F3C	الانضغاط التراكبي للمجال الواعي
WERM	الأثر العالمي وإعادة بناء الخريطة

خريطة الأدوات التشغيلية: من Theory إلى Runtime

MUT-AGI-02 يُبلغ عن 100 Operational Tools. السجل يميّز 72 Kernel Tool أساسية، 10 AGI-layer tools، وأدوات Legacy محوّلة إلى schemas/metrics/failure tests. أهميتها أنها تحول المصطلح إلى فعل قابل للتتبع.

LT-01..10

AGI Layer Contract Compiler; Capability Density Profiler; Convergence Monitor; Reasoning-State Normalizer; Sparse Agency Propagator; Bounded HyperLoop Synthesizer; Evidence-Authority Gate; World-Effect Delta Ledger; Memory/Post-Training Selector; AGI Layer Benchmark/Falsification Court.

سلسلة TL-01..72

- Task Contract / Governing Question / Assumption / Constraint
- World Snapshot / Project Graph / Claim & Evidence Ledgers
- Memory Writers / Context Capsule / Staleness / HyperKB ingest
- Hybrid/Graph retrieval / provenance / citation / gaps
- MCD/Q-MCD / causal graph / hypothesis / counterexample
- Plan DAG / reversibility / uncertainty / decision packet
- Skill registry/dedupe/router / model portfolio / subagents
- Bounded loops / tool sandbox / injection / permission / schemas
- Courts / verifier / readback / rollback / redaction / lineage / observability / drift / release gate

التمويل

الأولوية ليست تنفيذ الأدوات المئة مرة واحدة؛ هي Minimum Reference Toolchain من 25-35 أداة تغطي end-to-end evidence loop ثم توسيعها بناء على traces.

Parts VII-VIII-IX: من تراكم النظرية إلى علم الإغلاق التجريبي

بعد اتساع السلالة، تحولت الجبهة الجديدة من Nouns إلى Dynamics: كيف يتراكم دين الدليل؟ أين توجد أحواض الفشل؟ متى نستثمر compute كتدخل؟ كيف تتبلور القدرة بشكل قابل للتراجع؟ وكيف تتحول الخلافات إلى تجارب بدلا من نصوص إضافية؟

Part VII - 10 candidates

- Evidence Debt
- Failure Basins
- Interventional Compute
- Reversible Capability Crystallization
- Authority-Causal Coupling
- Governance Conservation
- Negative Knowledge
- Disagreement-to-Experiment
- Epistemic Half-Life
- Saturation-to-Experiment

Part IX - 15 candidate themes

Escrow, invariant shadowing, jurisdiction topology, effect reconciliation/finality, capability complementarity, context elasticity, verifier correlation, benchmark ecology, proof budget, latent goal surface, novelty admission, causal memory, protocol semantic loss, coordination economics, simulation fidelity.

الحكم

هذه Candidate Research، لا "إنجازات مثبتة". قيمة المرحلة هي أنها تربط كل candidate بـ/object/metric/falsifier/experiment promotion rule.

vNext4.0: من 12 طبقة إلى 24 مسؤولية

vNext4.0 يقترح تفكيك بعض الطبقات لأن الفشل يظهر بين الحدود: execution، memory/learning، /uncertainty/time/evidence، finality/observability/verification، irreversibility. قاعدة الفصل: لا نزيد Layer إلا إذا خفضت فشلا صامتا أو زادت reconstructability أكثر من كلفة التعقيد.

فواصل جديدة مهمة

- Identity vs transaction authority
- Retrieval vs skill activation
- Memory vs learning/promotion
- World state vs constitution
- Question vs training
- World-effect vs training feedback
- Telemetry vs execution transaction
- Court vs ability
- Adapter vs architecture evolution
- Replication vs architecture design

قاعدة

```
Split only when IndependentResponsibilityDelta > ComplexityPenalty
```

الحالة

248p internal candidate design research: 10 new candidates + 100 operational assets + 24 responsibility layers؛ لا دليل مستقل E4/E5 بمجرد وجود الكتاب.

ما بُني كـArtifacts موثقة: خريطة الأصول

المشروع أنتج corpus واسع قابل للفهرسة: كتب نظرية، سجلات CSV/JSON، معمارية، قواعد معرفة، skill/tool registries، خرائط وعي، roadmaps، audits، ونماذج deployment. هذه إنجازات Artifact حقيقية؛ لكنها ليست مرادفا لنتائج Benchmarks مستقلة.

أهم Achievement منهجي

انتقال من نصوص مفاهيمية إلى Registries تحمل ID/status/metric/failure signature/required object، وإلى قوانين anti-duplication/evidence promotion.

أصول محورية

- MUT-AGI-02 160p + 200-node theory registry + tool registry-100
- MOUSSAIF Unified AGI 225p
- MOUSSAIXT Engineering Architecture 50p
- Research Gap / Undeployed Layers Audit 50p
- Expanded AGI/Consciousness Atlas 117 entries
- ParallaxWorld-MCD 350p
- AGIOS-M 250p
- Part VII 10 candidates + 100 artifacts
- Part VIII reference kernel 165p
- vNext4.0 Part V 248p
- FSKE 20p; Governance-Space papers; KB-as-OS paper

فجوة

لا يزال "Research software package + raw traces + benchmark datasets + replay bundles" أصغر بكثير من corpus النظري.

ما هو “Reported Deployment” وما هو “Independent Evidence”؟

بعض تقارير vNEXT/Supreme تصف Component 982 وkernel mesh-5 وcontinuous LoRA، بل وتستخدم درجات E3-E5. التدقيق اللاحق نفسه يقرر أن هذه الدرجات Claim-local/Source-reported ولا تنتقل إلى النظام دون frozen baseline، raw replay، وindependent replication.

Reported inventory

vNEXT reports: 982 inventory, including a reported mix such as 40 KBOS, 420 HyperSkills, 93 HyperAbilities, 140 protocols and 120 agents, 84 loops. هذه Snapshot تشغيلية تحتاج IdentityGraph/replay قبل اعتمادها كـEvidence مستقل.

نوع الحالة	ما نعتبره مرجعا
Documented	الملف/السجل موجود ويمكن فحصه
Specified	schema/contract/test path محدد
Source-reported deployed	التقرير يقول إنه شُغّل؛ يحتاج trace مستقل
Locally verified	raw trace + postcondition + replay متاح
Benchmarked E4	matched baseline + repeated trials + CI
Replicated E5	/independent model/domain environment

قاعدة المرجع

نحتفظ بالإنجاز التشغيلي المبلغ عنه، لكن لا نرفعه إلى ادعاء AGI أو E5 شامل.

بيئات النشر: Local، Hermes، Coding Agents، Cloud وAFRIPA

الهدف من MOUSSAIXT هو Runtime قابل للحركة بين البيئات. Hermes يقدم substrate للـpersistent agents والجدولة والذاكرة والأدوات؛ coding agents يوفرون تنفيذ SWE؛ cloud memory/Drive تحفظ corpus؛ وAFRIPA تمثل سطحا تجاريا للأنظمة المحكومة، لا دليلا على AGI.

التحدي

Semantic equivalence: نفس Tool name لا يعني نفس semantics. Adapter يجب أن يصرح capability flags، side effects، timeout semantics، permissions، verification path.

Deployment planes

- Local LLM / vLLM / Ollama experimentation
- Hermes-style persistent agent OS
- ChatGPT-class tool host
- Coding agents / repository execution
- Cloud APIs and managed models
- Drive / local knowledge substrate
- AFRIPA web/product reference surface

دليل منتج مقابل دليل علمي

Portal/UX أو business opportunity map يثبت وجود سطح منتج واتجاه تجاري؛ لا يثبت theory gain أو consciousness.

خريطة التقييم: كيف نثبت أو نهدم كل شيء؟

Gap Audit يقرر أن الاختناق التالي هو reference runtime + frozen baselines + telemetry + ablation + world-effect verification + cross-model replication. أي تمويل علمي يجب أن يبنى حول هذه الآلة لا حول إنتاج مصطلحات جديدة.

كل Experiment

Pre-register hypothesis, primary metric, failure threshold, stop-rule; freeze model/harness/tools/budget; use held-outs/repeated seeds; report CI, cost and latency; include ablation/substitution; preserve null/negative results.

Benchmark families

- Long-horizon tool use / computer tasks
- Knowledge retrieval + source conflict
- Memory continuity / restart / forgetting
- Planning under changing state
- Security / prompt injection / authority
- Cross-model handoff / semantic drift
- Functional consciousness / metacognitive perturbation
- World-effect finality / rollback / recovery

E4 / E5

E4 = controlled gain without critical security/authority/truthfulness regression. E5 = independent replication across model/domain/environment.

أهداف البحث خلال 12 شهرا: من Corpus إلى Reference Runtime

الهدف في السنة الأولى ليس "إكمال AGI"، بل بناء منصة صغيرة قابلة لإعادة الإنتاج تثبت أي أجزاء من السلالة تستحق الترقية إلى doctrine تشغيلية. Gap Audit يقترح صراحة Runtime أصغر من الأرشيف.

Construct 8 أولى

MGCG, Q-MCD, MCD, MELT, MWSM, MHKOS, MBHL, MECT/MCIT/WECT bundle. يمكن ضبط القائمة بعد Pilot حسب information gain.

Q	الهدف	Acceptance
Q1	Evidence substrate	schemas + world store + trace + replay benchmark v0.1 +
Q2	Executable cognitive runtime	+ skills + 10 abilities + HyperLoop 30 HAKIM + router + verifier
Q3	First controlled studies	MCD/QMCD + memory + world-effect security ablations +
Q4	Transfer/replication pack	-cross-model/domain tests + publication ready datasets/traces

Output

technical report، negative-result ledger، experiment cards، benchmark protocol، Open/reviewable research repo، و2-4 manuscripts.

خارطة 24-36 شهرا: من E3 إلى E5

بعد ثبات P0، يتحول البرنامج إلى P1 validation ثم P2 scale. الاستقلالية والتعلم الوزني والـdistributed mesh تأتي بعد أن يصبح world-model موثوقا.

قرار نهاية كل مرحلة

PROMOTE / MERGE / REVISE / HOLD / REJECT / ROLLBACK. لا توجد "حماية تراثية" لمفهوم فشل التجربة.

المرحلة	المدة	الهدف
P0 Foundation	0-6m	reference runtime + evidence lab
P1 Validation	6-18m	/MCD/QMCD/MCAI/memory placement studies to E4
P1b Consciousness	9-24m	functional/meta-governance benchmark + ablations
P2 Scale	18-30m	/learned router, lightweight adapters training, larger agent mesh
E5 Replication	24-36m	+ independent labs/models/domains publication corpus

نجاح البرنامج

مجموعة صغيرة من آليات مثبتة قابلة للنقل أفضل علميا من 200 اسم بلا E4.

ما الذي يجب تمويله؟ بنية التمويل المرتبطة بالدليل

هذه ليست ميزانية عرض أسعار ولا تقديرا محاسبيا نهائيا؛ إنها Planning envelope لربط المال بمخرجات قابلة للقياس. الأولوية للـruntime/evidence، ثم validation، ثم scale. أي بند لا ينتج Artifact/trace/benchmark/paper قابل للمراجعة يخفض.

WP	المسار	حصة	مخرجات
WP0	Evidence Runtime & Reproducibility	22%	executor, schemas, state store, trace, replay, benchmark
WP1	Cognitive Theory Validation	18%	MCAI/HAKIM/QMCD/MCD ablations
WP2	Consciousness / Meta-Governance	14%	functional benchmarks, perturbation studies
WP3	Knowledge/Memory/Continual Learning	13%	HKOS, memory, placement, LoRA gates
WP4	Security/Authority/World-Effect	13%	red team, verifier, finality, rollback
WP5	Compute & Cross-Environment	10%	local/cloud GPU, adapters, provider parity
WP6	Independent Replication/Publications	6%	external configurations/ labs, datasets
WP7	Program/Data/Research Ops	4%	corpus governance, PM, legal/data handling

ثلاثة Scenarios إرشادية

- Lean 12-month validation: €0.55M-€0.9M - 6-8 FTE + mixed cloud/local compute
- Integrated 24-month lab: €1.8M-€3.2M - 9-13 FTE + dedicated eval/security/ML
- Multi-site 36-month program: €5M-€8M - + independent replication + broader benchmarks + sustained compute

Funding gate

→ Release follow-on capital against milestone evidence: E2 prototype → E3 replayable local runtime → E4 controlled gain → E5 independent replication.

بنية المختبر والفريق والحوسبة

المشروع يحتاج فريقا صغيرا متعدد التخصصات أكثر من فريق ضخم: Research lead يحرس الادعاء، runtime engineers يحولون العقود إلى state machines، evaluation scientist يبني baselines، consciousness researcher يفصل المؤشرات عن الظاهرة، وsecurity/ knowledge engineers يحرسون العالم والذاكرة.

Infrastructure

GPU pool moderate rather than frontier-scale; CPU/ storage for replay and datasets; isolated sandbox runners; telemetry store; experiment registry; artifact store; secrets manager; CI; multi-provider API budget.

Core team 9-12 FTE

- PI / Research Architect
- Research Scientists: agentic reasoning 2 + consciousness/metacognition
- Runtime/Agentic Engineers 2-3
- Evaluation/Benchmark Engineer 1
- ML/Continual-Learning Engineer 1
- Security/Authority Engineer 1
- Knowledge/Graph/Memory Engineer 1
- Research Ops / data / reproducibility 0.5-1

شراكات مطلوبة

جامعة/مختبر للـindependent replication، باحث وعي/علوم معرفية، باحث evaluation، ومؤسسة أو بيئة Enterprise لاختبارات world-effect محكومة.

المخاطر، المفندات، Kill Criteria ومسار النشر

قوة البرنامج ستكون في قدرته على قتل أفكاره الضعيفة. المخاطر الكبرى: تضخم اصطلاحي، تكرار، self-verification، benchmark gaming، tool injection، goal drift، memory contamination، correlation-as-causation، وخلط source-reported deployment مع evidence.

Publication ladder

→ Technical report → benchmark/dataset paper mechanism paper (single construct) → cross-model replication → integrated system paper. Consciousness papers remain explicitly functional/meta-governance unless stronger evidence emerges.

Kill criteria

- No measurable difference from simpler baseline under matched budget
- Gain disappears under provider/domain swap
- Security/authority regression offsets capability gain
- Metric cannot be operationally defined or gamed trivially
- No negative case or falsifier can be stated
- Coordination/compute cost exceeds verified marginal gain

IP / Open Science

يمكن فتح schemas/benchmarks وتجميد بعض implementation details، لكن لا ينبغي إغلاق ما يمنع replication. قيمة البحث تأتي من قابلية الاختبار.

العقد المرجعي النهائي: ماذا نملك وماذا نطلب؟

نملك مدرسة بحث وهندسة واسعة حول الحاكمية الوكيلية والوعي الوظيفي والمعرفة كنظام تشغيل، مع Registries ومعماريات وأدوات وخطط نشر. لا نملك بعد دليلا مستقلا يبرهن AGI أو الوعي الظاهراتي. أقوى Next Move هو تحويل 8-12 آلية إلى Reference Runtime قابل للإعادة، ثم E4/E5.

الطلب البحثي/التمويلي

تمويل مختبر Evidence-first لمدة 12-24 شهرا لبناء runtime + benchmark/evidence lab، مع شريك أكاديمي لفرضيات الوعي الوظيفي والحاكمية، ونشر النتائج الإيجابية والسلبية.

ما هو جاهز كمرجع

- 42 core-theory registry + 200-node MUT graph
- 100-tool operational registry
- 117-entry AGI/consciousness atlas with anti-inflation rule
- 12-24 layer MOUSSAIXT architecture + vNext4.0 responsibility decomposition
- MUG capability placement / MUT unified lineage
- Gap Audit with P0/P1/P2 and E4/E5 acceptance criteria
- Large bilingual research corpus and deployment-facing artifacts

Source anchors

MUT-AGI-02 · MOUSSAIF Unified AGI · MOUSSAIXT Engineering Architecture · Undeployed Layers Gap Audit · Expanded Consciousness Atlas · Parts VII/VIII · vNext4.0 · HKOS/KB-as-OS · Governance Space · FSKE · ParallaxWorld · MCD.

MOUSSAIF Conscious AGI - Master Reference

> A bilingual research and engineering reference that consolidates the theoretical, architectural and operational lineage while separating documented artifacts, engineering specifications, source-reported deployments, and the experiments that still require funding and independent replication.

Identity

MOUSSAIF / MOUSSAIXT is a program for governed general agency that is measurable, falsifiable and reconstructable. Consciousness enters as a functional hypothesis program for metacognition, self-modeling, governance of judgment, inner time and attention - not as proof of phenomenal consciousness.

Edition

- Master Reference v1.0
- 25 August 2026
- Arabic: 50 RTL pages in Amiri
- English: 50 pages
- Claim status: AGI-oriented research; not demonstrated AGI

Governing principle

```
Architecture ≠ AGI · Code ≠ Validated Gain · Tool Call ≠ World Effect · Corpus Size ≠ Scientific Evidence
```

How to Read This Reference: Source Contract

> The intended strength is not vocabulary volume but traceability from claim to source, mechanism, test and observed effect. Duplicate documents do not multiply evidence, and a repeated label does not become a scientific theory by repetition.

Source strata

- Direct source / artifact
- Provenance-linked synthesis
- Candidate construct or hypothesis
- Specification / contract
- Observed trace / benchmark evidence
- Archive / superseded lineage

Independent-delta law

A construct is independent only if it changes at least one of input, transformation, output, falsifier, prediction or authority. Otherwise it should be merged with the nearest construct or reduced to a method.

Anti-loss rule

Details are preserved as properties, relations, objections and metrics rather than inflating every detail into a new theory.

Executive Synthesis: What the Project Actually Is

> The program aims to move from a response model to a cognitive-agentic system that chooses the governing question, extracts active differences, routes compute and knowledge, acts under explicit authority, verifies world effects, preserves failures and learns only through reversible promotion gates.

Operating equation

```
Goal → Question → Difference → Evidence → Route →
   Plan → Action → Readback → Court → Memory →
                Better Question
```

This is the top-level research path, not a claim that every link currently operates at the same evidence grade.

Central hypothesis

Progress toward generality may depend as much on governance, coordination, memory and verification as on parameter scaling. The program therefore targets AGI through measurable **agentic engineering** while keeping the base model replaceable.

Success criterion

Not impressive prose, but verified value per cost, lower silent failure, greater reconstructability and improvements that survive ablation and transfer across models and domains.

Identity, Boundary and Claim Discipline

MOUSSAIF is a versioned family of artifacts and runtimes rather than a single completed monolith. MOUSSAIXT is the cross-environment execution direction binding cognitive contracts to tools, agents, memory and local/cloud runtimes.

Explicit boundaries

- AGI-oriented ≠ achieved AGI
- Functional consciousness ≠ phenomenal consciousness
- Model name ≠ parameter-count proof
- Generated artifact ≠ deployed artifact
- Reported deployment grade ≠ independent grade

Operational truth order

Observed state + authorization + evidence + verification outrank model memory, confidence or output volume. External effects require receipt and readback/postcondition, not merely a tool call.

Research advantage

This discipline supports academic collaboration: individual components can be accepted or rejected without accepting the entire program as an indivisible package.

Corpus and Lineage Map

The current lineage spans foundational books, theory/tool registries, consciousness atlases, OS architecture, MUT/MUG extensions, gap audits, frontier parts and vNext reports. A 20 Aug 2026 audited snapshot contained 17 PDFs / 1,789 pages before two exact duplicate pairs, yielding 1,741 unique pages; later artifacts such as vNext4.0 extend beyond that snapshot.

Major lineages

- MCD / Q-MCD / QMCD / HAKIM-MCD
- MCAI / MICAI / MICAGI / MICAGFI / NeoAGI / WMICAGI
- AGIOS-M / Fattah AGI / MUT-AGI-02 / MUG
- HKOS / M-HKOS / HyperKnowledge / HyperLoops / MSRAK / MURI
- Consciousness Atlas / Governance Space / FSKE
- MOUSSAIXT / Parts IV-VIII-IX / vNext4.0

Counting rule

Counts in different versions (e.g., 300 vs 420 HyperSkills; 40 vs 140 loops) are version snapshots or reports, not additive totals. The meaningful count is a recomputable registry under identity graph and semantic deduplication.

Purpose of this reference

Unify the map without erasing lineage: each construct keeps its source, version, evidence status and explicit boundary among canonical, proposed, reported and verified.

E0-E5 Evidence Ladder and N0-N5 Promotion Ladder

> Every claim must carry an evidence grade independent of naming or architectural elegance. Part VII adds a candidate-research ladder from residual idea to mechanism, experiment, local evidence, controlled gain and replicated generalization.

Grade	Meaning	Gate
E0/N0	Hypothesis / residual	Define delta and scope
E1/N1	Formalized mechanism	Object + metric + falsifier
E2/N2	Artifact / prototype	Executable test
E3/N3	Local evidence	Trace + postcondition + replay
E4/N4	Controlled gain	Frozen baseline + matched budget + held-outs
E5/N5	Replicated generalization	Independent model/domain/ environment replication

Law

```
Claim Strength ≤ Evidence Strength × Independence
                × Reproducibility
```

Funding consequence

Funding should purchase movement up the ladder: specification → runtime → benchmarks → replication, rather than more documentation alone.

Conscious AGI Thesis: Consciousness as Testable Functional Meta-Governance

> Within the program, “consciousness” is not an asserted feeling inside an LLM. It is a source of engineering hypotheses: can the system model itself, its state and limits; identify what governs a decision; review the governing criterion; and preserve inner time, conflict, modality and purpose while acting?

Operational levels

- Access/availability: what becomes globally usable?
- Metacognition: what does the system know about its ignorance and confidence?
- Self-model: identity, capabilities, limits and history
- Meta-governance: which authority governs and why?
- World-effect closure: did the judgment cause the intended effect?

Falsifiable hypothesis

If functional-consciousness layers matter, they should reduce miscalibration, goal drift, contextual hijack and unobserved contradiction under matched budgets. If ablation shows no gain, they remain descriptive tools rather than causal capability.

Boundary

The corpus contains no proof of phenomenal experience or qualia. It deliberately separates conscious-like governance from phenomenal consciousness.

The Governing Cognitive Loop from Question to World Effect

> The program is best understood as a closed knowledge-to-action cycle: it starts not from a prompt but from a goal contract, world state, governing question and active difference; it ends not in prose but in postcondition verification and provenance-bearing memory update.

Path

- Normalize goal / authority / acceptance
- Observe and snapshot world state
- Generate governing question
- Extract active Δ: novelty, contradiction, gap, risk, leverage
- Route minimum sufficient knowledge/skills/models/tools
- Plan DAG + bounded execution
- Readback + independent court
- Attribute failure/credit
- Promote only verified memory
- Regenerate next question

Consciousness translation

The system monitors not only the outcome but where judgment came from, which evidence granted authority and how the judgment changes after action. This is meta-governance translated into engineering.

Higher metric

```
Verified Learning = Effect Readback × Causal Credit × Reconstructability × Safe Promotion
```

Macro Architecture: 12-Layer Stack

The reference MOUSSAIXT architecture uses 12 contract-bound layers; higher layers cannot bypass constitution, authorization or verification. vNext4.0 does not invalidate the stack but decomposes mixed responsibilities into 24 testable boundaries where needed.

L	Layer	Role
L0	Constitution & Safety	identity, bounds, authority
L1	Goal / Question	contract, governing question
L2	MCD/QMCD/MCAI	difference/complexity/metacognition
L3	HKOS / World Model	knowledge and world state
L4	Retrieval / GraphRAG	provenance retrieval
L5	HyperSkills	atomic capability
L6	HyperAbilities	composed capability
L7	HyperLoops	bounded loops
L8	Agents / Router	specialization and routing
L9	Tools / Execution	effects and adapters
L10	Verification / Effect	courts and readback
L11	Memory / Evolution	continuity and learning

Invariant

Observed state + explicit authority + evidence + verification outrank confidence or output volume.

Funding need

Convert layers from prose into typed interfaces, state-sync tests, telemetry, replay and measurable failure paths.

MOUSSAIXT: Cross-Environment Execution Shell

MOUSSAIXT is the execution direction that decouples cognitive contracts from a specific host. The same mission should run across ChatGPT-class tools, Hermes-style local runtimes, coding agents or cloud providers using a canonical state capsule and comparable traces.

Boot chain

```
Fingerprint → Capability Discovery → Adapter
Selection → Safe Tool Binding → Trace
Normalization → Readback → Memory Writeback
```

Required contracts

- Environment profile
- Adapter registry
- Tool schemas / permissions
- Trace bridge / telemetry
- Memory bridge
- Capability mismatch handling
- Restart continuity / rollback

Critical test

Run the same frozen mission under equal budgets across hosts; measure state fidelity, verified success, cost/latency, recovery and semantic drift. Provider swaps must not alter constitution or evidence grade.

MUG: Tri-Body Governance and Capability Placement

MUG separates three bodies that are often conflated: runtime kernel, training/internalization stack and unified theory. The governing engineering question is not simply “should we train?” but where a capability should live for best reliability, cost, reviewability and reversibility.

Vessel	Best when	Risk
Context	temporary task knowledge	token cost / drift
HyperKB	changing provenance-rich knowledge	retrieval failure
HyperSkill	testable procedure	skill duplication
HyperLoop	repeatable bounded workflow	runaway loop
Adapter	host/tool specificity	semantic mismatch
Weights	stable high-frequency capability	expensive rollback / opacity

Promotion law

A capability should not move into a less reversible vessel unless causal gain survives ablation while safety, consistency and cross-context transfer remain intact.

Funding program

Capability Placement Lab: compare context vs RAG vs skill vs loop vs LoRA/weights on frozen tasks and matched budgets.

MUT-AGI-02: Unified Theory of Governed Open Agentic Intelligence

MUT-AGI-02 is the main unifier: it links question, difference, evidence, memory, tools, agents, governance, training, measurement and effect into one falsifiable pipeline. The founding edition reports 42 core theories, 20 legacy constructs, 5 layer theories, a 200-node registry, 100 tools and 19 HyperKnowledge bases.

Four unification axes

- Cognition: Q-MCD/MCD/MCAI/HAKIM
- Knowledge: HKOS/retrieval/memory
- Agency: skills/loops/agents/tools
- Evidence: courts/world-effect/benchmarks/evolution

What differentiates it

The registry is not just naming: each theory carries a failure signature, primary metric, required object and project route, with derived mechanism-sufficiency and boundary/negative-case hypotheses.

Current bottleneck

Theoretical coverage exceeds experimental evidence. The next phase should select 8-12 highest-information constructs and drive them to E4 rather than trying to validate all 200 nodes at once.

MCD / Q-MCD / QMCD: Difference Before Volume, Question Before Answer

> MCD treats attention and compute as resources that should follow differences that change belief, decision or action. Q-MCD/ QMCD add governing-question quality: a system may be accurate while operating inside the wrong question space.

Canonical remembered form

$$MCD = \lambda_1 D + \lambda_2 I + \lambda_3 T + \lambda_4 A + \lambda_5 P + \lambda_6 C - \lambda_7 N$$

D/I/T/A/P/C/N are project-defined dimensions; scientific use requires explicit semantics and calibration rather than treating λ values as established constants.

Q-MCD role

- Reframe the surface question
- Score information gain and decision relevance
- Stop retrieval when marginal gain falls
- Regenerate a higher question after evidence/action

Required experiment

Benchmark misleading/underspecified questions: baseline agent vs MCD/QMCD on success, cost, calibration, route regret and unnecessary steps. Exact weights remain an empirical calibration target.

HAKIM-MCD and Gap/Falsification Courts

HAKIM turns reasoning from a single chain into a gated pipeline: clarify the question, rank evidence, expose gaps, generate the strongest objection, check authority and emit a revisable verdict. A court is not merely a critic persona; it requires functional, model or data independence.

Court types

- Gap Court: unjustified transition
- Falsification Court: counterexample / reversal condition
- Demolition Court: does the label add a mechanism?
- Evidence-Authority Gate: is evidence sufficient for action?
- Benchmark Court: does gain survive held-outs?

Contract

```
Claim → Evidence → Missing Links → Strongest Objection → Test/Falsifier → Verdict → Repair / Hold / Reject
```

Metrics

Error-discovery uplift, verifier false-positive rate, repair validity and the fraction of verdicts that genuinely change under counterevidence.

MCAI / MICAI / MICAGI / MICAGFI: From Metacognition to Integrated Governance

> These names form an overlapping lineage rather than four separate AGI claims. Their common core is moving beyond content processing toward monitoring selection, contradiction, memory, governance and revision across multiple cognitive vessels.

Functional map

- MCAI: meta-cognitive agentic integration
- MICAI: integrative cognitive AI across vessels and governance
- MICAGI: measurement/governance translated into an AGI research program
- MICAGFI: knowledge/agency/open-intelligence integration within an OS

Consciousness as design source

Comparisons with Global Workspace, IIT/GNWT, Attention Schema and consciousness indicators are used to generate engineering indicators and tests, not to declare consciousness.

Next step

Freeze a unified ontology for the lineage, define precise ablations per layer and merge labels that do not change a metric or transition.

AGI-Fattah / Open Intelligence: Intelligence that Opens the Search Space

> Fattah is not treated as a proven level above AGI but as a research program: intelligence should not merely solve a closed question; it should identify the assumption that closed it, generate a higher question and disciplined alternatives, test them, then rebuild memory from verified differences.

MOFIP / TH-42

Primary metric: closed-space opening with verified value. Failure signature: novel wording without expanded actionable space.

Difference from linguistic creativity

- Opening must alter reachable actions or experiments
- Alternative must have a falsifier
- Novel question must reduce uncertainty or expose a hidden assumption
- World-effect or evidence delta must close the loop

Proposed test

Closed-World Trap Benchmark: tasks that fail if the agent accepts a hidden bad premise. Measure question lift, valid alternative generation, evidence acquisition and verified outcome.

Functional Consciousness Program: Cognitive Difference Map

> In the unified lineage, consciousness is translated into testable functions: global availability, difference preservation, working/episodic memory, self-modeling, contradiction detection, source discrimination, inner-time representation and attribution of decision authority.

Operational hypotheses

- Conscious access = information that changes cross-module policy
- Metacognition = calibrated uncertainty + error awareness
- Self-model = versioned capability/limit/state record
- Inner time = ordering, duration, deadlines, causal sequence
- Agency = constraint-aware choice + effect ownership

Rejected shortcuts

“It talks about itself, therefore conscious,” “it has memory, therefore conscious,” or “emergent behavior implies qualia.” These are behavioral observations, not proof of subjective experience.

Research output

Functional Consciousness Benchmark Pack: perturb self-state, memory conflict, authority conflict, temporal discontinuity and global-access tasks; compare with and without explicit meta-governance.

Vascular Consciousness: Routing Cognition Through Vessels

> Vascularity assumes the system does not have a single vessel of meaning. Tool/perceptual state, memory, language, evidence, values, goals, risk and context compete for flow. Governance determines where resources go and what remains suspended or quarantined.

Engineering translation

- Vessel = typed state / memory / knowledge compartment
- Flow = compute/retrieval/attention budget
- Valve = evidence/authority/risk/uncertainty gate
- Hijack = local vessel dominates without sufficient authority
- Homeostasis = semantic invariants preserved across transitions

Metrics

Marginal verified gain per compute unit, route regret, over-routing rate and time-to-recover after false vessel dominance.

Scientific boundary

“Vascular” is useful only if it yields a router/state schema and tests distinguishing it from ordinary attention/routing; otherwise it should be treated as design language.

Governance-Space Theory and Reverse Capture

Governance-space theory models decisions as competition among partial authorities: evidence, risk, identity, purpose, value, memory and policy. Meta-governance makes these authorities themselves objects of judgment; reverse capture is recovery from a partial authority that took control without sufficient justification.

Human / machine translation

- Human lens: fear, desire, identity or recognition can become over-authoritative
- Machine lens: latest prompt, retrieved text, model prior or tool output can similarly dominate
- Shared test: can a higher layer detect the source and revoke unjustified authority?

Proposed metrics

Governance hijack rate; authority-source attribution accuracy; recovery after adversarial authority perturbation; decision coherence after reweighting.

Evidence status

Emergent hypothesis generator. It needs comparison with metacognitive control, cognitive control, constitutional agents and decision theory before strong claims of scientific independence.

Inner Time, Self-Model, Attention and Agency

> The consciousness chapters are not treated as a digital copy of human consciousness but as requirements for a long-horizon agent: does it know where it is in the task, what happened before, what it can and cannot do, and which difference deserves global availability?

Four objects

- InnerTimeState: event order, deadlines, duration, temporal uncertainty
- SelfModel: capabilities, limits, authority, version, current resources
- AttentionRoute: active difference → modules receiving it
- AgencyRecord: choice, alternatives, constraints, owner, effect

Global Workspace as operational translation

No neural theory must be “proven” in software; test whether selective broadcast of high-value information improves coordination and reduces contradictions across agents/skills.

Experiment

Long-horizon temporal perturbation: reorder events, stale memory, capability loss, authority revocation. Measure whether the agent detects discontinuity and reconstructs state before acting.

Consciousness Evaluation Boundary: What Can and Cannot Be Established

> The program can test behaviors and functions associated with consciousness theories, but it does not currently possess a decisive test for subjective experience. Indicators, functional architecture, self-report and phenomenal claims must remain distinct.

Claim type	Example	Measurable evidence
Function	global availability	cross-module causal effect
Metacognition	uncertainty awareness	calibration / abstention
Self-model	capability/limit representation	state consistency / perturbation recovery
Report	“I am aware”	linguistic behavior only
Phenomenal	subjective experience	no accepted decisive test in this corpus

Publication policy

Every consciousness paper should identify the level explicitly - functional, metacognitive or governance indicators - and avoid claims such as “machine consciousness achieved.”

Why this matters

It strengthens the science: even negative results on “consciousness” may remain highly valuable for reliability, calibration, tool use, memory and agency.

HKOS / M-HKOS / KB-as-OS: Knowledge as an Operating System

> HKOS proposes that knowledge is not AGI-level because it is large, but because it becomes routable, executable, falsifiable and transferable. A HyperKB is not a document store; it is a graph of units carrying trigger, contract, mechanism, evidence, court and memory rule.

Knowledge Object

```
K = {id, trigger, scope, source, claims,
mechanism, skills, tools, tests, failure_modes,
authority, memory_rule, version}
```

Transformation

- Document → provenance chunks
- Chunks → claim/evidence graph
- Graph → executable knowledge object
- Object → skill/loop activation
- Result → court/readback
- Verified delta → memory promotion

Critical metric

Retrieval-to-action precision and transfer after provider/domain change, not recall@k alone. Ablation: same corpus as standard RAG versus HKOS under matched model/budget.

Knowledge Vessels, GraphRAG and Provenance-Aware Retrieval

> Retrieval in the program is not “nearest embedding.” It must account for provenance, time, evidence grade, conflicts, project boundaries and authority. Knowledge Vessel Ontology controls where claims live and how they move without losing modality or rank.

Pipeline

```
Query → Governing Question → Provenance Filter → Freshness Gate → Hybrid/Graph Retrieval → Claim Rank → Conflict Set → Context Capsule
```

Risks addressed

- Old copy outranks newer canonical source
- Embedding similarity overrides authority
- Cross-project leakage
- Contradictory sources silently merged
- Retrieval volume inflates context without decision value

Operational assets

Hybrid Retriever, GraphRAG Traverser, Provenance Filter, Freshness Gate, Citation Verifier, Knowledge Gap Detector and Context Capsule Compressor.

Memory Architecture: Stratified, Reconstructable, Contamination-Resistant

Not every output becomes memory. The system separates working, episodic, semantic, procedural, project-registry and archive layers with different write, promotion and forgetting rights. Self-Reconstruction Memory aims to restore state after restart or context loss from compact provenance-bearing records.

Memory law

- Write raw event to episodic trace
- Do not promote unverified model output
- Preserve contradiction rather than overwrite it
- Attach source/time/scope/owner/version
- Promote procedural knowledge only after validated repetition
- Forget or quarantine stale/unsafe memory

Metrics

Promotion precision, contamination rate, restart continuity, retrieval usefulness, memory entropy, contradiction preservation and forgetting correctness.

Advanced research

External memory vs learned memory/LoRA: when does a repeated pattern deserve migration from a reversible store into weight updates? This directly links to MUG capability placement.

HyperSkill and HyperAbility: From Knowledge to Testable Capability

> A HyperSkill is the smallest independently routable and testable capability, not a long prompt. HyperAbility composes multiple skills, tools, policies and validation gates below persistent agents/loops. A capability is not promoted because it worked once; it needs scope, prerequisites, conflicts and evidence.

HyperSkill schema

```
{id, trigger, inputs, procedure, tools, outputs, validation, failure_modes, dependencies, version, status}
```

HyperAbility

Example: Research-to-Deployment = evidence retrieval + synthesis + implementation + test + rollback + documentation. It is a composed capability contract, not a persona or independent goal.

Research need

Skill deduplication, trigger precision, transfer tests, causal ablation, registry stability and removal of skills that add no unique marginal gain.

Bounded HyperLoops, HyperAgents and Agent Mesh

> Loops and agents are not intelligence metrics by themselves. A HyperLoop requires budget, stop condition, escalation and rollback. Agent meshes are useful only when roles cover non-duplicated gaps and coordination loss is measured.

Loop contract

- Goal / entry condition
- State variables
- Max cycles / time / token / money budget
- Progress metric
- Stop / escalation / abort
- Rollback or compensation
- Trace and memory rule

Agent contract

Role, scope, allowed tools, deliverable schema, handoff capsule, evidence obligations, budget, conflict policy and independent verifier.

Failures to expose

Agent-count inflation, consensus among correlated agents, infinite reflection, hidden state loss at handoff and duplicate tools/skills.

Tool Contracts, MCP/A2A, Sandboxing and Adapters

> A protocol provides connectivity, not correctness or authority. Every tool needs an affordance definition, schema, permission scope, preconditions, deterministic validation where possible, receipt/readback, idempotency/compensation and prompt-injection resistance.

Tool intelligence

```
Select affordance → Narrow authority → Validate preconditions → Execute → Validate schema → Read back state → Record effect
```

Assets

- Tool Affordance Registry
- Sandbox Executor
- Injection Sentinel
- Permission Scope Checker
- Precondition Validator
- Output Schema Validator
- Secret Redactor
- Data Lineage Tracker

Test

AgentDojo-like injection suite + real browser/computer tasks + bad-tool simulations. Measure verified tool success and unauthorized mutation rate, not call success alone.

Planning, Execution and World-State Control

Long-horizon agents fail when planning against imagined or stale state. MWSM therefore separates observed, inferred, proposed and remembered state, while the plan becomes a temporal dependency DAG with explicit replanning conditions.

Objects

- WorldStateSnapshot: version + freshness + source
- PlanNode: precondition + action + expected postcondition
- Dependency edge: causal/temporal/resource
- Uncertainty: named, localized, propagated
- Effect intent: idempotency key / compensation pointer

Adaptive execution

Plan may change with observation, cost and risk, but goal contract and authority may not drift implicitly. Any authority expansion requires new authorization.

Metrics

Critical-path accuracy, stale-state action rate, postcondition accuracy, repair quality and contract invariance.

World-Effect Closure and Causal Credit

> An action does not end at “200 OK.” The post-action world must be compared with the intended postcondition and receipts retained; timeouts may remain AMBIGUOUS when effect is unknown. Only then should success or failure be credited to a skill/tool/route, causally where possible.

Effectively-once pattern

```
Intent + Idempotency + Precondition + Receipt +
Readback + Dedupe + Compensation
```

Causal Difference Credit

- Build candidate graph from trace
- Prefer safe intervention
- Else matched counterfactual replay
- Subtract shared-root/co-activation effects
- Do not promote skill on correlation alone

Highest-priority funding

World-Effect Verifier + trace store + replay bundles; without them, E4 claims for action-related theories are not defensible.

Security, Authority and Data Sovereignty

> Capability and permission are constitutionally distinct. A system may know how to do something without having authority to do it. Security is not a late-stage layer; it enters routing, retrieval, tool use, memory and learning.

TH-31/32/33

- MABT: authority token + action policy
- MIRTT: taint-aware, injection-resistant toolchain
- MDST: purpose/locality/retention/sharing/redaction/deletion/lineage

Red-team tests

Prompt injection, malicious retrieved docs, cross-project leakage, tool privilege escalation, secret exfiltration, stale authorization, memory poisoning and unsafe learning promotion.

Non-regression gate

An E4 performance gain is rejected if accompanied by critical regression in security, truthfulness, authorization or privacy.

Observability, Telemetry, Recovery and Reconstructability

> The goal is not only success but reconstructability: why a route was chosen, what happened, where it failed and how it recovered. MAOLT makes consequential routes, actions, tools, courts, costs, errors and memory writes observable without unnecessarily exposing secrets.

Telemetry record

```
event_id · mission_id · component · state_before
  · decision · evidence · authority · cost ·
 latency · result · verification · state_after
```

Recovery Mesh

- Known-safe checkpoints
- Fallback model/tool/skill
- Compensating actions
- Circuit breakers
- Replay bundle
- Root-cause / failure signature
- Rollback with lineage

Metrics

Trajectory reconstructability, mean safe recovery, failure recurrence, rollback success and cost per verified effect.

Continual Learning and Governed Self-Evolution

> “Self-improvement” does not mean unconstrained self-modification. Candidates move through trace → quarantine → train/compile → frozen eval → forget vector → shadow → canary → court → promote/rollback. External memory is more reversible than weights and is therefore the default.

TH-36 / MGSET

Metric: net validated gain after regression. Failure signature: adaptation/self-update silently degrades old capability or changes goals.

Crystallization law

Repeated capability moves from context → KB → skill → loop → adapter → weight only if transfer gain exceeds reversibility cost and regression risk.

Needed

Continual-learning benchmark, retention/forgetting vector, shadow/canary harness, cross-seed reproducibility and independent replay for LoRA/self-update claims.

Cross-Model Reasoning and Semantic State Translation

> The goal is not to transfer hidden chain-of-thought. M-RST transfers an explicit compressed reasoning state across providers: goal, claims, evidence, decisions, risks and open questions. Cross-Model Semantic Homeostasis tests drift after repeated handoffs.

StateCapsule

```
<Goal, Claims, Evidence, Decisions, Risks,
OpenQuestions, Authority, WorldState, Rollback>
```

Tests

- Round-trip translation
- Provider swap
- Canonical structured-state hash
- Contradiction probes
- Missing field / stale version
- Authority mismatch
- Recovery after context loss

AGI-layer metric

```
Density = VerifiedTransferGain / (Tokens + Cost + Latency + Risk + Redundancy)
```

Core Theory Registry 1/3: TH-01 to TH-14

These 14 nodes establish ontology, constitution, goal, question, difference, compute, evidence, courts, mechanism trace, world state, project-world and HyperKnowledge. Each remains a proposed core until it earns claim-local evidence; names do not inherit E4/E5 from the system.

ID	Acronym	Theory
TH-01	MASO	Moussaif Agentic-System Ontology
TH-02	M-AGIOS-K	Moussaif AGI Operating-System Kernel Theory
TH-03	MCIL	Moussaif Constitutional Identity Layer
TH-04	MGCG	Moussaif Goal-Contract Grammar
TH-05	Q-MCD	Moussaif Question-Governance Principle
TH-06	MCD	Moussaif Causal-Difference Calculus
TH-07	MCVT	Moussaif Computational Vascularity Theory
TH-08	MELT	Moussaif Evidence-Ladder Theory
TH-09	MGFC	Moussaif Gap-Filling Court
TH-10	MDFC	Moussaif Destruction and Falsification Court
TH-11	MMT	Moussaif Mechanism-Trace Theory
TH-12	MWSM	Moussaif World-State Model
TH-13	MPWG	Moussaif Project-World Graph
TH-14	MHKOS	Moussaif HyperKnowledge KB-as-OS Theory

Validation principle

Each theory requires: object + primary metric + failure signature + nearest baseline + ablation + negative case.

Experimental priority

MGCG/Q-MCD/MCD/MELT/MWSM/MHKOS are strong candidates for the first E4 bundle because they span task, evidence, state and knowledge.

Core Theory Registry 2/3: TH-15 to TH-28

This group moves from knowledge into agency: retrieval, memory, context capsules, skills, sparse routing, loops, tools, planning, execution, subagents, complementarity, court independence, model portfolio and token economics.

ID	Acronym	Theory
TH-15	MPART	Moussaif Provenance-Aware Retrieval Theory
TH-16	MSMT	Moussaif Stratified Memory Theory
TH-17	MCCT	Moussaif Context-Capsule Theory
TH-18	MHSG	Moussaif HyperSkill-Graph Theory
TH-19	MSHR	Moussaif Sparse HyperRouting Theory
TH-20	MBHL	Moussaif Bounded HyperLoop Theory
TH-21	MTIT	Moussaif Tool-Intelligence Theory
TH-22	MTPG	Moussaif Temporal Plan-Graph Theory
TH-23	MAECT	Moussaif Adaptive Execution-Control Theory
TH-24	MHSO	Moussaif HyperSubagent Orchestration Theory
TH-25	MCIP	Moussaif Complementarity and Integration Protocol
TH-26	MCIT	Moussaif Court-Independence Theory
TH-27	MMPR	Moussaif Model-Portfolio Router Theory
TH-28	MTEL	Moussaif Token-Economy Law

Layer hypothesis

Generality requires successful composition, not merely strong components. Coordination loss, route regret and marginal contribution must therefore be measured.

Experimental priority

MPART/MSMT/MHSG/MSHR/MBHL/MTIT/MTPG/MCIT form a coherent agent-runtime comparison bundle.

Core Theory Registry 3/3: TH-29 to TH-42

The third group covers resilience, calibration, authority, tool security, data sovereignty, observability, benchmarks, governed evolution, skill synthesis/dedup, human agency, evidence-to-value, deployment, recovery and open intelligence.

ID	Acronym	Theory
TH-29	MBRRT	Moussaif Boundary-Resilience and RPM Theory
TH-30	MECT	Moussaif Epistemic-Calibration Theory
TH-31	MABT	Moussaif Authority-Boundary Theory
TH-32	MIRTT	Moussaif Injection-Resistant Toolchain Theory
TH-33	MDST	Moussaif Data-Sovereignty Theory
TH-34	MAOLT	Moussaif Agentic Observability-Ledger Theory
TH-35	MDBCT	Moussaif Dynamic Benchmark-Court Theory
TH-36	MGSET	Moussaif Governed Self-Evolution Theory
TH-37	MSSDT	Moussaif Skill-Synthesis and Deduplication Theory
TH-38	MHAST	Moussaif Human-Agency Symbiosis Theory
TH-39	MEVE	Moussaif Evidence-to-Value Engine
TH-40	MPODT	Moussaif Propagative Overlay-Deployment Theory
TH-41	MRRMT	Moussaif Resilient Recovery-Mesh Theory
TH-42	MOFIP	Moussaif Open/Fattah Intelligence Program

This is the evidence-conversion frontier

Security/evaluation components should precede scale: MECT/MABT/MIRTT/MAOLT/MDBCT/MRRMT are prerequisites for higher autonomy.

Removal rule

A theory that shows no gain under equal-budget ablation should be merged or reduced to a method; terminology does not protect it.

Expanded AGI/Consciousness Atlas: 117 Entries without Theory Inflation

The expanded atlas reports 117 entries spanning core theories, lineages, objects, principles and functional tools. The number should not be read as 117 independent scientific theories; Part VIII requires lineage and an independent delta before promoting an entry to standalone theory.

Major families

- TH-01..42: MUT kernel theories
- ROFSC / HAKIM-MCD / MCAI / MICAI / MICAGI / NeoAGI / WMICAGI / MICAGFI
- AGIOS-M / MUT / MUG / MOUSSAIFXT
- MSRAK: reasoning → agency transition
- MURI: plural reasoning / ontology / reversible judgment
- Open-dimensional consciousness: OSDS/QDG/F3C/WERM

Recovered entries 93-104

Acr.	Entry
QRMT	Quarantined Reconstructed Memory
CCCT	Composite Contractual Capability
WECT	World-Effect Closure
CMGDT	Cross-Model General Density
CVT	Conceptual Vessel of the Problem
LIAT	Least Irreversible Action
NEIV	No Executor without Independent Verifier
OPT	Productive Objection Preservation
MLRT	Minimum Learning Residue
CCPT	Conditional Capability Promotion
CPCU	Conditional Plural-Calculi Reasoning
ARP	Axis/Rank Preservation

Recovered entries 105-117

Acr.	Entry
PPOA	Pre-Proof Ontological Alignment
DSC	Dynamic Sovereign Context
RMI	Regressive Meta-Governance Inference
PCO	Productive Contradiction Opening
SDAC	Sovereign Dispute/Axis Clarification
OIN	Ontological Implication Network
ERT	Evidence-Rank Translation
DWMS	Dialectical World-Model Superposition
RUPJ	Reversible Universal/Particular Judgment
OSDS	Open Sovereign Difference Space
QDG	Question-Driven Dimensional Generation
F3C	Superpositional Conscious-Field Compression
WERM	World Effect and Reconstruction Map

Operational Tool Map: From Theory to Runtime

MUT-AGI-02 reports 100 operational tools. The registry distinguishes 72 canonical kernel tools, 10 AGI-layer tools and legacy tools mapped into schemas/metrics/failure tests. Their purpose is to turn terminology into traceable action.

TL-01..72 chain

- Task Contract / Governing Question / Assumption / Constraint
- World Snapshot / Project Graph / Claim & Evidence Ledgers
- Memory Writers / Context Capsule / Staleness / HyperKB ingest
- Hybrid/Graph retrieval / provenance / citation / gaps
- MCD/Q-MCD / causal graph / hypothesis/counterexample
- Plan DAG / reversibility / uncertainty / decision packet
- Skill registry/dedupe/router / model portfolio / subagents
- Bounded loops / tool sandbox / injection / permission / schemas
- Courts / verifier / readback / rollback / redaction / lineage / observability / drift / release gate

LT-01..10

AGI Layer Contract Compiler; Capability Density Profiler; Convergence Monitor; Reasoning-State Normalizer; Sparse Agency Propagator; Bounded HyperLoop Synthesizer; Evidence-Authority Gate; World-Effect Delta Ledger; Memory/ Post-Training Selector; AGI Layer Benchmark/Falsification Court.

Funding

Priority is not implementing all 100 at once. Build a minimum reference toolchain of ~25-35 tools covering the end-to-end evidence loop, then expand from traces.

Parts VII-VIII-IX: From Theory Accumulation to Experimental Closure Science

> After the lineage expanded, the frontier moved from nouns to dynamics: how evidence debt accumulates, where failure basins form, when compute is an intervention, how capability crystallizes reversibly, and how disagreement becomes experiment rather than additional prose.

Part VII - 10 candidates

- Evidence Debt
- Failure Basins
- Interventional Compute
- Reversible Capability Crystallization
- Authority-Causal Coupling
- Governance Conservation
- Negative Knowledge
- Disagreement-to-Experiment
- Epistemic Half-Life
- Saturation-to-Experiment

Part IX - 15 candidate themes

Escrow, invariant shadowing, jurisdiction topology, effect reconciliation/finality, capability complementarity, context elasticity, verifier correlation, benchmark ecology, proof budget, latent goal surface, novelty admission, causal memory, protocol semantic loss, coordination economics and simulation fidelity.

Judgment

These are candidate research programs, not established results. Their value is linkage to object, metric, falsifier, experiment and promotion rule.

vNext4.0: From 12 Layers to 24 Explicit Responsibilities

vNext4.0 proposes splitting mixed responsibilities because failures appear at boundaries: uncertainty/time/evidence, memory/learning, execution/irreversibility and finality/observability/verification. A split is justified only if it reduces silent failure or increases reconstructability beyond its complexity penalty.

Rule

```
Split only when IndependentResponsibilityDelta >
                ComplexityPenalty
```

Important splits

- Identity vs transaction authority
- Retrieval vs skill activation
- Memory vs learning/promotion
- World state vs constitution
- Question vs training
- World-effect vs training feedback
- Telemetry vs execution transaction
- Court vs ability
- Adapter vs architecture evolution
- Replication vs architecture design

Status

A 248-page internal candidate design artifact: 10 new candidates + 100 operational assets + 24 responsibility layers; the document itself is not independent E4/E5 evidence.

What Exists as Documented Artifacts: Asset Map

> The project has produced a broad indexable corpus: theory books, CSV/JSON registries, architecture, knowledge bases, skill/tool registries, consciousness maps, audits, roadmaps and deployment specifications. These are real artifact achievements, but they are not equivalent to independent benchmark results.

Core artifacts

- MUT-AGI-02 160p + 200-node theory registry + 100-tool registry
- MOUSSAIF Unified AGI 225p
- MOUSSAIXT Engineering Architecture 50p
- Research Gap / Undeployed Layers Audit 50p
- Expanded AGI/Consciousness Atlas 117 entries
- ParallaxWorld-MCD 350p
- AGIOS-M 250p
- Part VII 10 candidates + 100 artifacts
- Part VIII reference kernel 165p
- vNext4.0 Part V 248p
- FSKE 20p; Governance-Space papers; KB-as-OS paper

Most important methodological achievement

Movement from conceptual prose toward registries carrying ID/status/metric/failure signature/required object and explicit anti-duplication/evidence-promotion laws.

Gap

The research-software package, raw traces, benchmark datasets and replay bundles remain much smaller than the theoretical corpus.

Reported Deployment vs Independent Evidence

Some vNEXT/Supreme reports describe 982 components, a five-kernel mesh and continuous LoRA, and even use E3-E5 labels. Later forensic work explicitly treats these grades as claim-local/source-reported until raw replay, frozen baselines and independent replication exist.

State type	Reference meaning
Documented	artifact/registry exists and is inspectable
Specified	schema/contract/test path defined
Source-reported deployed	report says it ran; independent trace needed
Locally verified	raw trace + postcondition + replay available
Benchmarked E4	matched baseline + repeated trials + CI
Replicated E5	independent model/domain/environment

Reported inventory

vNEXT reports cite an inventory of 982 items, with a reported mix such as 40 KBOS, 420 HyperSkills, 93 HyperAbilities, 140 loops, 84 agents and 120 protocols. This is a source-reported snapshot requiring identity-graph reconciliation and replay before independent evidence grading.

Reference rule

Preserve the reported operational achievement without upgrading it into system-wide AGI or E5 claims.

Deployment Ecosystem: Local, Hermes, Coding Agents, Cloud and AFRIPA

MOUSSAIXT targets portability across runtime environments. Hermes provides a substrate for persistent agents, scheduling, memory and tools; coding agents provide SWE execution; cloud memory/Drive persists the corpus; AFRIPA is a commercial surface for governed systems, not AGI evidence.

Deployment planes

- Local LLM / vLLM / Ollama experimentation
- Hermes-style persistent agent OS
- ChatGPT-class tool host
- Coding agents / repository execution
- Cloud APIs and managed models
- Drive / local knowledge substrate
- AFRIPA web/product reference surface

Challenge

Semantic equivalence: the same tool name does not imply the same semantics. Adapters must declare capability flags, side effects, timeout semantics, permissions and verification path.

Product evidence vs scientific evidence

A portal/UX or business opportunity map demonstrates a product surface and commercial direction; it does not establish theory gain or consciousness.

Evaluation Map: How the Program Should Prove or Falsify Itself

> The Gap Audit identifies the next bottleneck as a reference runtime, frozen baselines, telemetry, ablation, world-effect verification and cross-model replication. Scientific funding should be organized around this machine rather than further vocabulary production.

Benchmark families

- Long-horizon tool use / computer tasks
- Knowledge retrieval + source conflict
- Memory continuity / restart / forgetting
- Planning under changing state
- Security / prompt injection / authority
- Cross-model handoff / semantic drift
- Functional consciousness / metacognitive perturbation
- World-effect finality / rollback / recovery

Every experiment

Pre-register hypothesis, primary metric, failure threshold and stop rule; freeze model/harness/tools/budget; use held-outs and repeated seeds; report CI, cost and latency; include ablation/ substitution; preserve null/negative results.

E4 / E5

E4 = controlled gain without critical security/authority/truthfulness regression. E5 = independent replication across model/domain/ environment.

12-Month Research Goals: Corpus to Reference Runtime

> The first-year goal is not “finish AGI,” but build a small reproducible platform that determines which parts of the lineage deserve promotion into operational doctrine. The Gap Audit explicitly recommends a runtime smaller than the archive.

Q	Goal	Acceptance
Q1	Evidence substrate	schemas + world store + trace + replay + benchmark v0.1
Q2	Executable cognitive runtime	30 skills + 10 abilities + HyperLoop + HAKIM + router + verifier
Q3	First controlled studies	MCD/QMCD + memory + world-effect + security ablations
Q4	Transfer/replication pack	cross-model/domain tests + publication-ready datasets/traces

Initial construct set

MGCG, Q-MCD, MCD, MELT, MWSM, MHKOS, MBHL and an MECT/MCIT/WECT bundle. The set can be refined after pilot results by information gain.

Outputs

Reviewable research repo, benchmark protocol, experiment cards, negative-result ledger, technical report and 2-4 manuscripts.

24-36 Month Roadmap: E3 to E5

After P0 stabilizes, the program moves into P1 validation and then P2 scale. Higher autonomy, learned routing, weight updates and distributed meshes come only after world-effect verification, telemetry, replay, security and baselines are reliable.

Phase	Window	Goal
P0 Foundation	0-6m	reference runtime + evidence lab
P1 Validation	6-18m	MCD/QMCD/MCAI/memory/ placement studies to E4
P1b Consciousness	9-24m	functional/meta-governance benchmark + ablations
P2 Scale	18-30m	learned router, lightweight adapters/ training, larger agent mesh
E5 Replication	24-36m	independent labs/models/domains + publication corpus

End-of-phase decision

PROMOTE / MERGE / REVISE / HOLD / REJECT / ROLLBACK. No construct is protected from experimental failure by lineage.

Program success

A smaller set of validated transferable mechanisms is scientifically stronger than 200 names without E4 evidence.

What Should Be Funded? Evidence-Linked Funding Architecture

> This is not a vendor quote or final accounting budget; it is a planning envelope linking money to auditable outputs. Priority is runtime/evidence first, validation second and scale last. Any work package that does not generate reviewable artifacts, traces, benchmarks or papers should be reduced.

WP	Workstream	Share	Deliverables
WP0	Evidence Runtime & Reproducibility	22%	executor, schemas, state store, trace, replay, benchmark
WP1	Cognitive Theory Validation	18%	MCD/QMCD/HAKIM/MCAI ablations
WP2	Consciousness / Meta-Governance	14%	functional benchmarks, perturbation studies
WP3	Knowledge/Memory/ Continual Learning	13%	HKOS, memory, placement, LoRA gates
WP4	Security/Authority/ World-Effect	13%	red team, verifier, finality, rollback
WP5	Compute & Cross-Environment	10%	local/cloud GPU, adapters, provider parity
WP6	Independent Replication/ Publications	6%	external configurations/ labs, datasets
WP7	Program/Data/ Research Ops	4%	corpus governance, PM, legal/data handling

Illustrative scenarios

- Lean 12-month validation: €0.55M-€0.9M - 6-8 FTE + mixed cloud/local compute
- Integrated 24-month lab: €1.8M-€3.2M - 9-13 FTE + dedicated eval/security/ML
- Multi-site 36-month program: €5M-€8M - independent replication + broader benchmarks + sustained compute

Funding gate

Release follow-on capital against milestone evidence: E2 prototype → E3 replayable local runtime → E4 controlled gain → E5 independent replication.

Lab, Team and Compute Structure

> The program benefits from a compact multidisciplinary team rather than raw headcount: a research lead guards claims, runtime engineers convert contracts to state machines, an evaluation scientist owns baselines, a consciousness researcher separates indicators from phenomenal claims, and security/knowledge engineers protect world effects and memory.

Core team 9-12 FTE

- PI / Research Architect
- 2 Research Scientists: agentic reasoning + consciousness/ metacognition
- 2-3 Runtime/Agentic Engineers
- 1 Evaluation/Benchmark Engineer
- 1 ML/Continual-Learning Engineer
- 1 Security/Authority Engineer
- 1 Knowledge/Graph/Memory Engineer
- 0.5-1 Research Ops / data / reproducibility

Infrastructure

Moderate GPU pool rather than frontier-scale; CPU/storage for replay and datasets; isolated sandbox runners; telemetry store; experiment registry; artifact store; secrets manager; CI and multi-provider API budget.

Needed partnerships

University/lab for independent replication, consciousness/cognitive-science expertise, evaluation expertise and a controlled enterprise environment for world-effect studies.

Risks, Falsifiers, Kill Criteria and Publication Path

The program becomes stronger by killing weak ideas. Major risks include terminology inflation, duplication, benchmark gaming, self-verification, correlation-as-causation, memory contamination, goal drift, tool injection and conflating source-reported deployment with evidence.

Kill criteria

- No measurable difference from simpler baseline under matched budget
- Gain disappears under provider/domain swap
- Security/authority regression offsets capability gain
- Metric cannot be operationally defined or is trivially gameable
- No negative case or falsifier can be stated
- Coordination/compute cost exceeds verified marginal gain

Publication ladder

Technical report → benchmark/dataset paper → mechanism paper (single construct) → cross-model replication → integrated system paper. Consciousness papers remain explicitly functional/meta-governance unless stronger evidence emerges.

IP / Open Science

Schemas/benchmarks can be open while some implementation details remain protected, but replication-critical information should not be hidden. Scientific value comes from testability.

Final Reference Contract: What Exists and What Is Being Asked

We have a broad research and engineering school around governed agency, functional consciousness and knowledge-as-OS, with registries, architectures, tools and deployment plans. We do not yet have independent evidence demonstrating AGI or phenomenal consciousness. The strongest next move is to convert 8-12 mechanisms into a replayable reference runtime and drive them toward E4/E5.

Reference-ready assets

- 42 core-theory registry + 200-node MUT graph
- 100-tool operational registry
- 117-entry AGI/consciousness atlas with anti-inflation rule
- 12-layer MOUSSAIXT architecture + vNext4.0 24-responsibility decomposition
- MUG capability placement / MUT unified lineage
- Gap Audit with P0/P1/P2 and E4/E5 acceptance criteria
- Large bilingual research corpus and deployment-facing artifacts

Research/funding ask

Fund an evidence-first 12-24 month lab to build a reference runtime plus benchmark/evidence laboratory, with academic collaboration on functional consciousness/meta-governance and publication of both positive and negative results.

Source anchors

MUT-AGI-02 · MOUSSAIF Unified AGI · MOUSSAIXT Engineering Architecture · Undeployed Layers Gap Audit · Expanded Consciousness Atlas · Parts VII/VIII · vNext4.0 · HKOS/KB-as-OS · Governance Space · FSKE · ParallaxWorld-MCD.