Undeployed Layers, Research Gaps & Next Engineering Program

تقرير فجوات الطبقات غير المنشورة وبرنامج العمل البحثي والهندسي القادم

An exact 50-page bilingual Arabic-English audit of the gap between the MOUSSAIF AGI research corpus and a reproducible deployed cognitive runtime. It identifies which layers are specified, packaged, partially executable, unvalidated, or missing - then converts those findings into a prioritized E0-to-E5 research program.

تدقيق ثنائي اللغة من خمسين صفحة يفصل بين النظرية والمواصفة والتنفيذ والتحقق والتجربة والتعميم، ويحدد ما يجب بناؤه واختباره في المرحلة القادمة.

Main thesis

The next bottleneck is not more theory names. It is executable reference implementations, frozen baselines, telemetry, ablation, world-effect verification and cross-model replication.

العائق القادم هو الإثبات والتنفيذ والقياس، لا تكثير المصطلحات.

Scientific gate

Move core components from E0-E1 to E3, then require E4 controlled improvement and E5 replicated generalization.

نقل المكونات من E0-E1 إلى E3 ثم E4 وE5.

Evidence basis. Canonical MOUSSAIF1B 4-OS kernel and deployment prompt; MOUSSAIF Cognitive Methods Registry; HyperKnowledge OS V3; HyperLoop Runtime OS V3; Dual-Exponent HyperDeployment KB; Moussaif1B 20-Theories Research Paper; M-TAOS lineage. “Undeployed” means insufficient deployment evidence was located in the audited canonical corpus/ current context, not that no private implementation can exist elsewhere.

حالة بحثية موجهة نحو AGI وليست ادعاء بتحقق AGI.

Executive Gap Map & Decision Summary

SYSTEM AUDIT P0

الخريطة التنفيذية للفجوات وملخص القرار

E0 E1 E2 E3 E4 E5

ENGINEERING ASSESSMENT

The current MOUSSAIF corpus is unusually rich in architecture, terminology, schemas, skills, loops, research hypotheses and operational rules. The limiting factor is no longer idea generation. The limiting factor is deployment evidence: executable reference implementations, calibrated metrics, controlled baselines, reproducible experiments, real state stores, failure replay, and cross-environment verification. The next research cycle should therefore optimize for conversion from named constructs into measured mechanisms.

OBSERVED GAP / REQUIREMENTS

- High priority: router calibration, HyperLoop runtime, World-Effect verifier, evaluation harness, persistent world model, trace/metrics store.
- Research priority: MCD/QMCD operationalization, Cross-Model AGI Density, Context Compaction fidelity, Dual-Exponent HyperDeployment ablations.
- Infrastructure priority: MOUSSAIXT adapter contracts, environment discovery, transaction/rollback layer, reproducible benchmark runner.
- Governance priority: evidence ledger, falsification courts as executable tests, promotion gates and version lineage.
- Expected outcome: move core claims from E0-E1 toward E3-E4, then pursue E5 replication.

> **Decision rule:** Do not promote a construct because it has a large corpus, an elegant name or an “active” registry status. Promote it only when the mechanism is executable and its claimed benefit survives the appropriate evidence gate.

التقييم الهندسي

المنظومة غنية جدا بالنظريات والمواصفات وHyperSkills وHyperLoops والعقود المعرفية. الفجوة الأساسية لم تعد نقص الأفكار، بل الانتقال من الوصف إلى تنفيذ قابل للاختبار، ثم إلى مقارنة مضبوطة، ثم إلى تعميم قابل لإعادة الإنتاج. الأولوية القادمة هي إثبات الآليات لا زيادة الأسماء.

الفجوة والمتطلبات

- أولوية قصوى: معايرة Router وتشغيل HyperLoop والتحقق من World-Effect ومنصة تقييم دائمة.
- أولوية بحثية: تحويل MCD/QMCD إلى مقاييس قابلة للقياس واختبار Cross-Model AGI Density وضغط السياق.
- أولوية بنيوية: عقود MOUSSAIXT ومحولات البيئات وطبقة Transaction/Rollback.
- أولوية حوكمة: سجل الأدلة ومحاكم قابلة للتنفيذ وبوابات ترقية وإصدارات.
- الهدف: نقل النواة من E0-E1 إلى E3-E4 ثم تعميم E5.

> **قاعدة القرار:** لا تتم ترقية أي مفهوم بسبب حجم ملفه أو قوة اسمه؛ الترقية بالدليل والتنفيذ والاختبار فقط.

Maturity Model: E0-E5 and Deployment States

AUDIT METHOD P0

نموذج النضج: من E0 إلى E5 وحالات النشر

E0 E1 E2 E3 E4 E5

ENGINEERING ASSESSMENT

The audit separates the existence of a document from the existence of a working mechanism. A concept can be sophisticated yet remain undeployed. A registry can be internally consistent yet scientifically unvalidated. This report therefore uses two orthogonal axes: evidence maturity (E0-E5) and deployment state (Specified, Packaged, Executable, Verified, Benchmarked, Replicated).

OBSERVED GAP / REQUIREMENTS

- E0 - named construct / hypothesis.
- E1 - explicit mechanism, contract or operational principle.
- E2 - code/schema/artifact package exists and is internally coherent.
- E3 - execution or structural behavior is directly verified.
- E4 - controlled benchmark shows improvement versus frozen baseline.
- E5 - gains replicate across domains/models/ environments with bounded variance.

التقييم الهندسي

يفصل التدقيق بين وجود الوثيقة ووجود الآلية العاملة. قد يكون المفهوم متقدما نظريا لكنه غير منشور تنفيذيا. لذلك نستعمل محورين: رتبة الدليل E0-E5، وحالة النشر من Specified إلى Replicated.

الفجوة والمتطلبات

- E0: اسم أو فرضية.
- E1: آلية أو عقد تشغيلي واضح.
- E2: حزمة أو Schema أو تنفيذ أولي متماسك.
- E3: تشغيل متحقق مباشرة.
- E4: تفوق تجريبي مضبوط على Baseline ثابت.
- E5: تعميم متكرر عبر مجالات ونماذج وبيئات.

Specified Concept and contract exist.	**Executable** Reference mechanism runs.	**Verified** Observed execution and postconditions.
Benchmarked Controlled E4 comparison.	**Replicated** E5 across models/domains.	**Canonical** Only after lineage + regression gates.

Audit Scope, Evidence Rules & Non-Claims

AUDIT METHOD P0

نطاق التدقيق وقواعد الدليل وما لا يدعيه

E0 E1 E2 E3 E4 E5

ENGINEERING ASSESSMENT

The review is grounded in the canonical Drive corpus inspected for this report: MOUSSAIF1B four-OS kernel and deployment prompt, the cognitive methods registry, HyperKnowledge OS V3, HyperLoop Runtime OS V3, Dual-Exponent HyperDeployment KB, the 20-theory research paper, M-TAOS artifacts and related registries. “Undeployed” means that no sufficient execution/ verification evidence was located in the audited corpus or current artifact context; it does not prove that no private implementation exists elsewhere.

OBSERVED GAP / REQUIREMENTS

- Observed documents are evidence of specification, not automatically of runtime execution.
- A skill marked status: active is not considered E3 unless its dependencies run and its outputs are inspected.
- Large line counts and registries show corpus scale, not performance gains.
- Claims are downgraded when the required benchmark, trace, tool output or side effect cannot be observed.
- The report favors falsifiable closure tasks over speculative expansion.

التقييم الهندسي

يعتمد التقرير على الملفات المعيارية التي تم فحصها في Drive. كلمة "غير منشور" هنا تعني أن التدقيق لم يجد دليلا كافيا على تنفيذ وتشغيل وتحقق ضمن العينة المعيارية، وليس نفيا مطلقا لوجود كود خاص في بيئة أخرى.

الفجوة والمتطلبات

- الوثيقة دليل على المواصفة لا على التشغيل.
- status: active لا يكفي لإثبات التنفيذ الفعلي.
- حجم الملفات لا يساوي مكسب أداء.
- غياب Trace أو Benchmark أو Tool Evidence يخفض رتبة الادعاء.
- الأولوية لإغلاق فجوات قابلة للتكذيب بدل التوسع النظري.

Systemic Finding: Specification-Heavy, Evidence-Light Core

SYSTEMIC GAP | P0

النتيجة النظامية: مواصفات كثيفة ودليل تنفيذي أخف

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Across the architecture, the common pattern is strong naming, contracts, procedures, failure modes and expected outputs, but weaker evidence that the full chain runs end-to-end under reproducible conditions. This is not a weakness of the research direction; it is the normal transition point from architecture to engineering science. The most valuable next artifact is therefore not another megaprompt but a reference runtime plus benchmark corpus plus experiment ledger.

OBSERVED GAP / REQUIREMENTS

- Gap type A: declarative component without executable implementation.
- Gap type B: executable candidate without calibration or baseline.
- Gap type C: verified single environment without cross-model/domain replication.
- Gap type D: research equation without identifiable measurable variables.
- Gap type E: lifecycle stage exists in prose but not enforced by tooling.

التقييم الهندسي

النمط المتكرر هو قوة في التسمية والعقود والإجراءات وحالات الفشل، مع نقص نسبي في إثبات سلسلة تشغيل كاملة قابلة للتكرار. المرحلة التالية الطبيعية هي الانتقال من Architecture إلى Engineering Science عبر Runtime مرجعي وBenchmarks وسجل تجارب.

الفجوة والمتطلبات

- A: مكون وصفي بلا تنفيذ مرجعي.
- B: تنفيذ أولي بلا معايرة أو Baseline.
- C: نجاح محلي بلا تعميم.
- D: معادلة بلا متغيرات قابلة للقياس بوضوح.
- E: دورة حياة مكتوبة لكنها غير مفروضة برمجيا.

Gap 01 - HyperAbility as a Real Runtime Abstraction

SPECIFIED / NOT E3 | P0

الفجوة 01 - HyperAbility كطبقة قدرة تنفيذية حقيقية

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

HyperAbility is useful as the bridge between atomic HyperSkills and bounded HyperLoops, but in the audited corpus it is more formalized as an architecture-level abstraction than as a universally enforced runtime object. Without a typed ability layer, compositions risk becoming ad-hoc lists of skills rather than stable capabilities with policy, tooling, validation and fallback contracts.

OBSERVED GAP / REQUIREMENTS

- Define canonical HyperAbility schema and semantic versioning.
- Add composition constraints: compatible skills, shared state, authorization profile, latency/cost budgets.
- Create ability registry with dependency graph and collision detection.
- Separate capability identity from model/provider implementation.

التقييم الهندسي

HyperAbility ضرورية كطبقة تركيب بين HyperSkill وHyperLoop، لكنها تحتاج كيانا تنفيذيا موحدا لا مجرد مفهوم معماري. المطلوب تحويل القدرة إلى Object له Schema وسياسة وأدوات واختبارات وFallback.

الفجوة والمتطلبات

- تعريف Schema معياري للقدرة وإصداراتها.
- قيود تركيب للمهارات والحالة والتفويض والميزانية.
- سجل قدرات مع Graph للتبعيات والتعارض.
- فصل القدرة عن مزود النموذج.

BUILD NEXT / البناء التالي

- Implement ability compiler and registry validator.
- Build 10 reference abilities from existing HyperSkills.
- Expose invoke(), verify(), fallback(), rollback() contracts.

EXPERIMENTS / التجارب

- A/B compare skill-list execution vs typed HyperAbility execution on 50 tasks.
- Measure success, routing entropy, retries, context cost and recovery rate.

EXIT GATE / بوابة الخروج

- >=10 abilities execute reproducibly in two runtimes.
- Schema validation catches invalid compositions.
- At least one E4 benchmark shows measurable value.

Gap 02 - HyperSkills: Declarative Registry vs Executable Plugins

E1-E2 / PARTIAL | P0

الفجوة 02 - HyperSkills بين السجل الوصفي والإضافات التنفيذية

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

MOUSSAIF1B contains hundreds of richly specified skills with triggers, procedures, verification, risks and expected artifacts. The gap is that declarative skill objects are not automatically executable functions. A production skill fabric needs resolvable tool adapters, typed inputs/outputs, deterministic validation where possible, and runtime telemetry.

OBSERVED GAP / REQUIREMENTS

- Map each skill to implementation class: prompt-only, retrieval, deterministic code, tool orchestration, external side effect.
- Add executable entrypoint and dependency resolver.
- Distinguish “active in registry” from “runtime verified”.
- Create conformance tests for skill contracts.

التقييم الهندسي

سجل HyperSkills غني جدا، لكن وصف المهارة لا يجعلها Function قابلة للتنفيذ تلقائيا. يجب ربط كل Skill بأدوات فعلية ومدخلات ومخرجات Typed وValidator وTelemetry.

الفجوة والمتطلبات

- تصنيف المهارات حسب نوع التنفيذ.
- إضافة Entry Point وحل التبعيات.
- فصل Active في السجل عن Runtime Verified.
- اختبارات مطابقة لعقد كل Skill.

BUILD NEXT / البناء التالي

- Select 30 representative skills across four OS modules.
- Implement plugin wrappers and typed result envelopes.
- Record failure and unsupported-tool downgrade behavior.

EXPERIMENTS / التجارب

- Run contract tests and fault injection.
- Measure tool-resolution success, false trigger rate and verification coverage.

EXIT GATE / بوابة الخروج

- 30/30 reference skills have runnable entrypoints.
- >=95% contract-test pass rate in supported environments.

Gap 03 - Skill Compiler and Static Analysis Pipeline

SPECIFIED P0

الفجوة 03 - مترجم المهارات والتحليل الساكن

E0 E1 E2 E3 E4 E5

ENGINEERING ASSESSMENT

The research repeatedly proposes converting knowledge into HyperSkills, yet a true compiler stage is still needed. The compiler should transform a high-level skill specification into validated runtime metadata, adapter bindings, test stubs, risk policies, provenance links and machine-readable manifests.

OBSERVED GAP / REQUIREMENTS

- Formal grammar for skill definitions.
- Static checks for missing prerequisites, circular dependencies, unsafe side effects and unverifiable success criteria.
- Artifact hashing and provenance linkage.
- Generated test skeletons and fixture requirements.

التقييم الهندسي

تحويل المعرفة إلى HyperSkill يحتاج Compiler فعلي: يأخذ المواصفة ويولد Manifest وروابط الأدوات والاختبارات والسياسات والمصدر، ثم يرفض المهارات الناقصة قبل دخول Runtime.

الفجوة والمتطلبات

- قواعد نحوية رسمية للمواصفة.
- Static Checks للتبعيات والدورات والمخاطر ومعايير النجاح.
- Hash وProvenance.
- توليد قوالب اختبار تلقائية.

BUILD NEXT / البناء التالي

- Build CLI: moussaix skill compile/validate/test.
- Generate SBOM-like dependency manifest for each skill.

EXPERIMENTS / التجارب

- Feed malformed and adversarial skill specs.
- Measure rejection precision/recall and compiler determinism.

EXIT GATE / بوابة الخروج

- Compiler produces identical manifests from identical inputs.
- Unsafe or incomplete specs fail closed.

Gap 04 - Tool/Environment Adapter Layer

PARTIAL / FRAGMENTED | P0

الفجوة 04 - طبقة محولات الأدوات والبيئات

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The corpus correctly insists that the kernel must discover real tools rather than invent configuration. What is missing is a uniform adapter interface across ChatGPT connectors, Hermes/local tools, code agents, browsers, Drive, shell environments and model APIs. Without this, skills cannot be portable even when their logic is provider-neutral.

OBSERVED GAP / REQUIREMENTS

- Canonical capability descriptor: read/write/execute/ network/auth/streaming/transactionality.
- Discovery protocol and health checks.
- Secret-reference handling and scoped credentials.
- Adapter error taxonomy and retry semantics.

التقييم الهندسي

المعمارية تشترط اكتشاف الأدوات الحقيقية، لكن يلزم Adapter Contract موحد بين ChatGPT وHermes وCodex والبراوزر وDrive وShell وModel APIs حتى تصبح المهارات قابلة للنقل.

الفجوة والمتطلبات

- وصف موحد لقدرات الأداة.
- بروتوكول Discovery وHealth Check.
- إدارة مراجع الأسرار بصلاحيات محدودة.
- تصنيف أخطاء وإعادة المحاولة.

BUILD NEXT / البناء التالي

- Implement 5 adapters and a mock adapter.
- Build environment capability matrix at boot.

EXPERIMENTS / التجارب

- Port same 20-task suite across two environments.
- Measure substitution resilience and adapter-induced failures.

EXIT GATE / بوابة الخروج

- Same mission graph runs unchanged across at least two adapters.
- Unsupported capabilities downgrade cleanly.

Gap 05 - Cognitive Router: Calibration, Not Just Formula

E1 / UNCALIBRATED | P0

الفجوة 05 - معايرة Cognitive Router لا مجرد الصيغة

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The router scoring formula is explicit, but its weights are design priors, not yet demonstrated optimal coefficients. Relevance, expected utility, compatibility, evidence availability, historical success, cost, risk, context cost and redundancy require operational definitions and calibrated estimators.

OBSERVED GAP / REQUIREMENTS

- Define feature estimators and missing-value behavior.
- Log candidate set, selected route, outcomes and counterfactual alternatives.
- Learn or tune weights under safety constraints.
- Track route regret and calibration drift.

التقييم الهندسي

صيغة Router واضحة، لكن الأوزان الحالية Priors تصميمية. يجب تعريف كل متغير وقياسه ومعايرته تجريبيا بدل التعامل مع القيم كحقائق ثابتة.

الفجوة والمتطلبات

- تعريف Estimators للميزات.
- تسجيل المرشحين والمسار والنتائج والبدائل.
- معايرة الأوزان مع قيود السلامة.
- قياس Route Regret والانحراف.

BUILD NEXT / البناء التالي

- Implement router telemetry and offline replay.
- Freeze baseline heuristic and train/ tune candidate router.

EXPERIMENTS / التجارب

- Compare random, lexical, heuristic and calibrated routing.
- Metrics: task success, cost, latency, verification rate, regret.

EXIT GATE / بوابة الخروج

- Calibrated router beats frozen baseline without critical regressions.

Gap 06 - MCD: Measurable Variables and Identifiability

E0-E1 / RESEARCH | P0

الفجوة 06 - MCD: متغيرات قابلة للقياس وقابلية التعيين

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

MCD is central to the theory family, but D, I, T, A, P, C and N need precise observables. If multiple dimensions can explain the same routing decision, the score becomes non-identifiable and difficult to falsify. The next step is not a more complex formula; it is a measurement protocol.

OBSERVED GAP / REQUIREMENTS

- Create variable dictionary with ranges, units and estimators.
- Define inter-rater annotation protocol for human labels.
- Test correlation/redundancy among dimensions.
- Specify monotonicity assumptions and edge cases.

التقييم الهندسي

MCD مركزية، لكن D وI وT وA وP وC وN يجب أن ترتبط بObservables. إن كان يمكن تبديل الأوزان والحصول على نفس القرار، تصبح الصيغة غير قابلة للتعيين والتكذيب.

الفجوة والمتطلبات

- قاموس متغيرات ووحدات وتقديرات.
- بروتوكول Annotation بشري.
- اختبار ترابط وتكرار الأبعاد.
- فرضيات Monotonicity وحالات حدية.

BUILD NEXT / البناء التالي

- Build MCD-labeled benchmark of 1,000 task fragments.
- Fit simple interpretable models before nonlinear variants.

EXPERIMENTS / التجارب

- Inter-rater reliability, predictive validity, ablation of each term.

EXIT GATE / بوابة الخروج

- Each dimension has independent measurable contribution or is merged/removed.

Gap 07 - QMCD: Governing Question Multiplier Validation

E0-E1 / RESEARCH | P1

الفجوة 07 - التحقق التجريبي من مضاعف QMCD للسؤال الحاكم

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

QMCD proposes that question quality multiplies useful cognition. This is a strong and testable hypothesis, but the multiplier Ω(Q) needs an empirical scale and intervention study. The core experiment should manipulate question quality while holding model, tools, retrieval budget and task constant.

OBSERVED GAP / REQUIREMENTS

- Operationalize diagnostic power, information gain, actionability and ambiguity penalty.
- Generate matched weak/medium/strong governing questions.
- Prevent leakage by blind question scoring.
- Estimate effect size on downstream success and cost.

التقييم الهندسي

QMCD تفترض أن جودة السؤال تضاعف الإدراك المفيد. المطلوب تجربة Intervention تغيّر جودة السؤال فقط مع تثبيت النموذج والأدوات والميزانية والمهمة.

الفجوة والمتطلبات

- قياس القوة التشخيصية ومكسب المعلومات وقابلية الفعل والغموض.
- إنشاء أسئلة ضعيفة/متوسطة/قوية متطابقة في المهمة.
- تقييم أعمى لمنع التسرب.
- قياس Effect Size على النجاح والكلفة.

BUILD NEXT / البناء التالي

- Build question-intervention dataset across 10 domains.

EXPERIMENTS / التجارب

- Randomized controlled prompts with fixed budgets.
- Test whether Ω(Q) predicts improvement beyond simple prompt length.

EXIT GATE / بوابة الخروج

- Statistically meaningful gain that replicates across models/domains.

Gap 08 - MCAI: Candidate Diversity and Selection Mechanism

E1 / PARTIAL | P1

الفجوة 08 - MCAI: تنوع المرشحين وآلية الاختيار

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

MCAI requires multiple candidate plans, perspectives or models and adversarial comparison. The undeployed layer is the candidate protocol: how candidates are made meaningfully diverse, how evidence is shared without inducing convergence, and how the selector avoids verbosity or majority bias.

OBSERVED GAP / REQUIREMENTS

- Define diversity operators: model, prompt, tool path, retrieval source, planning strategy.
- Blind selector with evidence-weighted rubric.
- Preserve unresolved disagreement rather than forcing synthesis.
- Measure marginal value of each additional candidate.

التقييم الهندسي

MCAI تحتاج بروتوكولا يضمن اختلافا حقيقيا بين المرشحين، ثم مقارنة لا تنحاز للأكثر طولاً أو للأغلبية. يجب فصل مرحلة التوليد عن التحكيم وقياس التنوع.

الفجوة والمتطلبات

- تعريف عوامل التنوع.
- Selector أعمى مبني على الدليل.
- حفظ الخلاف غير المحسوم.
- قياس القيمة الهامشية لكل Candidate إضافي.

BUILD NEXT / البناء التالي

- Implement candidate bus and selector API.

EXPERIMENTS / التجارب

- Compare 1, 2, 4 and 8 candidates under equal total compute.

EXIT GATE / بوابة الخروج

- MCAI configuration improves correctness/robustness per unit cost.

Gap 09 - HAKIM-MCD as an Integrated Executable Pipeline

E1 / PIPELINE SPEC | P0

الفجوة 09 - HAKIM-MCD كسلسلة تنفيذية متكاملة

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The HAKIM-MCD sequence is one of the clearest integrative contributions: Difference -> MCD -> Vascular Route -> Candidates -> Court -> Retrieval/Tools -> Agency -> World Effect -> Memory. The missing proof is an end-to-end orchestrator that exposes each stage boundary and allows stage ablation.

OBSERVED GAP / REQUIREMENTS

- Typed stage contracts and state transitions.
- Event log with timestamps, inputs, outputs and confidence.
- Stage bypass/ablation switches.
- Recovery rules when one stage is unavailable.

التقييم الهندسي

HAKIM-MCD توفر خطا تكامليا واضحا، لكن يجب تحويله إلى Orchestrator حقيقي مع Interfaces لكل مرحلة وإمكانية Ablation لمعرفة أي مرحلة تضيف قيمة.

الفجوة والمتطلبات

- عقود Typed وانتقالات حالة.
- Event Log لكل مرحلة.
- مفاتيح تعطيل/Ablation.
- قواعد Recovery عند غياب مرحلة.

BUILD NEXT / البناء التالي

- Build reference pipeline runner.

EXPERIMENTS / التجارب

- Full pipeline vs ablations on long-horizon tasks.

EXIT GATE / بوابة الخروج

- At least three stages show independent measurable contribution.

Gap 10 - Governing Question Quality Engine

E1 / NO CALIBRATED ENGINE | P1

الفجوة 10 - محرك جودة السؤال الحاكم

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The governing-question method is conceptually strong but currently needs a production scoring engine. Question generation, ranking and recursive refinement should be separated, instrumented and benchmarked so that the system can tell when reframing actually improves task value.

OBSERVED GAP / REQUIREMENTS

- Question candidate generator.
- Scoring model with interpretable features.
- Stop rule for recursive questioning.
- Link question score to downstream decision delta.

التقييم الهندسي

منهج السؤال الحاكم يحتاج Engine يولد ويرتب ويعيد صياغة الأسئلة ويقيس إن كان Reframing حسّن المهمة فعلا.

الفجوة والمتطلبات

- مولد أسئلة مرشحة.
- نموذج ترتيب بميزات مفهومة.
- Stop Rule للتساؤل المتكرر.
- ربط الدرجة بتغير القرار اللاحق.

BUILD NEXT / البناء التالي

- Create benchmark with known high-information questions.

EXPERIMENTS / التجارب

- Compare direct execution vs governing-question pass.

EXIT GATE / بوابة الخروج

- Improvement remains after controlling for extra tokens/latency.

Gap 11 - Vascular Compute Scheduler

E1 / NOT IMPLEMENTED AS SCHEDULER | P1

الفجوة 11 - مجدول الحساب الوعائي

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

“Vascular” routing proposes allocating deep compute to high-centrality, uncertain, novel or risky nodes instead of flooding all branches. To become engineering rather than metaphor, it needs a scheduler with explicit budgets, node priorities, queue policies and starvation protection.

OBSERVED GAP / REQUIREMENTS

- Node priority = expected information gain, dependency centrality, risk, novelty and reversibility cost.
- Budget allocator across reasoning, retrieval, tool calls and verification.
- Dynamic reprioritization after new evidence.
- Starvation and runaway-branch safeguards.

التقييم الهندسي

التوجيه الوعائي يجب أن يصبح Scheduler بميزانيات وأولويات وصفوف انتظار وحماية من Starvation، لا مجرد تشبيه لتوزيع الانتباه.

الفجوة والمتطلبات

- أولوية العقدة حسب مكسب المعلومات والمركزية والمخاطر والجدة.
- موزع ميزانية بين Reasoning وRetrieval وTools وVerification.
- إعادة ترتيب بعد دليل جديد.
- حماية من الجوع والانفجار.

BUILD NEXT / البناء التالي

- Implement budget scheduler over task DAG.

EXPERIMENTS / التجارب

- Compare uniform, greedy and vascular allocation at equal compute.

EXIT GATE / بوابة الخروج

- Better success/verification under fixed cost on sparse-difficulty tasks.

Gap 12 - Sparse HyperRouting: Search and Regret Measurement

E1 / ALGORITHM NEEDED | P0

الفجوة 12 - Sparse HyperRouting: البحث وقياس الندم

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Sparse HyperRouting should choose the minimal sufficient stack. The undeployed research problem is how to search a large skill/ability graph without missing rare but critical capabilities. Retrieval quality, route diversity and regret must be measurable.

OBSERVED GAP / REQUIREMENTS

- Hybrid semantic + symbolic candidate retrieval.
- Compatibility and authorization pruning.
- Beam width adaptation by uncertainty.
- Counterfactual replay to estimate route regret.

التقييم الهندسي

Sparse HyperRouting يجب أن يختار أقل Stack كافٍ دون فقد مهارة نادرة حاسمة. نحتاج خوارزمية Search وقياس Recall وRoute Regret.

الفجوة والمتطلبات

- استرجاع Semantic + Symbolic.
- Pruning حسب التوافق والتفويض.
- Beam متكيف مع عدم اليقين.
- Replay مضاد لتقدير Regret.

BUILD NEXT / البناء التالي

- Build skill graph index and route replay store.

EXPERIMENTS / التجارب

- Recall@k of required skills; cost vs dense routing; recovery from missed capability.

EXIT GATE / بوابة الخروج

- Sparse route matches/exceeds dense success with lower cost/context.

Gap 13 - Bounded HyperLoop Runtime State Machine

E1-E2 / NEEDS REFERENCE RUNTIME | P0

الفجوة 13 - آلة حالات تشغيل Bounded HyperLoop

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Bounded HyperLoops are among the most engineering-ready ideas in the corpus: explicit goal, state, stop condition, evidence threshold, failure threshold, rollback and memory writeback. The key missing artifact is a deterministic loop controller that enforces those fields rather than relying on agent prose.

OBSERVED GAP / REQUIREMENTS

- Finite-state model: INIT, RUN, VERIFY, REPAIR, ESCALATE, ROLLBACK, DONE.
- Hard iteration/time/token/tool budgets.
- Plateau detector and evidence threshold.
- Idempotent resume after interruption.

التقييم الهندسي

Bounded HyperLoop جاهزة هندسيا نسبيا. الفجوة هي State Machine تفرض الهدف والميزانية والتوقف والRollback والذاكرة برمجيا بدل الاعتماد على النص.

الفجوة والمتطلبات

- State Machine واضحة.
- حدود صلبة للدورات والزمن والتوكن والأدوات.
- كشف Plateau وعتبة دليل.
- استئناف Idempotent بعد الانقطاع.

BUILD NEXT / البناء التالي

- Implement loop engine and persistent checkpoints.

EXPERIMENTS / التجارب

- Fault injection: tool failure, stale data, contradictory verification, budget exhaustion.

EXIT GATE / بوابة الخروج

- No unbounded loops; deterministic safe exit; resumable execution.

Gap 14 - Dual-Exponent HyperDeployment: From Capacity Theory to Compiler

E0-E1 / RESEARCH+ENGINEERING | P1

الفجوة 14 - Dual-Exponent HyperDeployment من نظرية السعة إلى Compiler

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Dual-Exponent HyperDeployment distinguishes nominal combinatorial capacity from effective realized capacity. The major gap is translating C_nom and C_eff into a deployment compiler that never enumerates the exponential space, but uses sparse election and measurable quality factors.

OBSERVED GAP / REQUIREMENTS

- Operational estimators for Qk, Qs, Ql, Pv, Wr, redundancy and conflict loss.
- Sparse plan compiler with branch budget.
- Capability graph pruning and duplicate collapse.
- Measure whether the formula predicts realized success.

التقييم الهندسي

Dual-Exponent تفرق بين السعة الاسمية والفعالة، لكن يجب تحويلها إلى Compiler يستخدم Sparse Election ولا يحاول استكشاف الفضاء الأسي، مع قياس عوامل Qk/Qs/Ql/Pv/Wr وغيرها.

الفجوة والمتطلبات

- تقدير عوامل السعة الفعالة.
- Compiler بخطة Sparse وBranch Budget.
- Pruning وDedup للقدرات.
- اختبار Predictive Validity للصيغة.

BUILD NEXT / البناء التالي

- Implement minimal HyperDeployment planner.

EXPERIMENTS / التجارب

- Ablate factors and compare predicted vs observed capacity.

EXIT GATE / بوابة الخروج

- Formula is revised or retained based on predictive evidence, not naming.

Gap 15 - Knowledge-Integrity AutoDeployment CI/CD

E1 / LIFECYCLE NOT ENFORCED | P0

الفجوة 15 - CI/CD للنشر التلقائي مع سلامة المعرفة

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The proposed lifecycle Compile -> Static Checks -> Sandbox -> Benchmark -> Regression -> Canary -> Promote/Quarantine/Rollback is exactly what the system needs. The gap is enforcement: a generated skill or KB must be unable to mark itself operational before passing machine gates.

OBSERVED GAP / REQUIREMENTS

- Immutable artifact states with signed transition events.
- CI checks for schema, provenance, tests, security and redundancy.
- Canary registry and rollback trigger.
- Promotion decision must reference benchmark evidence.

التقييم الهندسي

دورة Compile -> Sandbox -> Benchmark -> Canary ممتازة، لكن يجب فرضها Tooling بحيث لا يستطيع Artifact إعلان نفسه Operational قبل عبور البوابات.

الفجوة والمتطلبات

- حالات Artifact غير قابلة للتلاعب مع أحداث انتقال.
- CI لل Schema والمصدر والاختبار والأمان.
- Canary وRollback Trigger.
- الترقية مرتبطة بدليل Benchmark.

BUILD NEXT / البناء التالي

- Create Git/Drive-compatible pipeline and promotion ledger.

EXPERIMENTS / التجارب

- Attempt unauthorized state jumps and regression promotion.

EXIT GATE / بوابة الخروج

- Invalid promotion is technically blocked; rollback tested.

Gap 16 - Project World Model: Canonical Store and Freshness

E1-E2 / STORE FRAGMENTED | P0

الفجوة 16 - Project World Model: مخزن معياري وسياسة حداثة

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The Project World Model defines projects, repositories, deployments, assets, constraints, business objectives, provenance, freshness and sensitivity. The gap is a single transactional canonical store that reconciles Drive, local workspaces, connectors and runtime observations without conflating remembered, proposed and current state.

OBSERVED GAP / REQUIREMENTS

- Canonical entity IDs and schemas.
- Source precedence and conflict resolution.
- Freshness TTL/event policies.
- Change-data capture and provenance graph.

التقييم الهندسي

نموذج العالم محدد جيدا، لكن يحتاج Store معياري Transactional يجمع Drive والبيئات المحلية والموصلات ويفصل Current/Proposed/ Historical.

الفجوة والمتطلبات

- معرفات وSchemas معيارية.
- قواعد أسبقية المصدر وحل التعارض.
- Freshness Policies.
- Change Data Capture وProvenance Graph.

BUILD NEXT / البناء التالي

- Implement local reference store (SQLite/Postgres + object pointers).

EXPERIMENTS / التجارب

- Simulate conflicting state from two environments.
- Measure stale-state detection and reconciliation accuracy.

EXIT GATE / بوابة الخروج

- Runtime can reconstruct current state after restart without hallucinated fields.

Gap 17 - GraphRAG: Retrieval Quality and Graph Maintenance

E1-E2 / CORPUS EXISTS | P0

الفجوة 17 - GraphRAG: جودة الاسترجاع وصيانة الرسم

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The corpus is designed for GraphRAG, but a knowledge corpus is not itself a GraphRAG system. The next layer is an actual graph schema, ingestion pipeline, embeddings/ indexes, edge provenance, retrieval planner and evaluation set.

OBSERVED GAP / REQUIREMENTS

- Node types: project, concept, claim, source, skill, ability, loop, agent, tool, artifact, experiment.
- Versioned edges and temporal validity.
- Hybrid graph/vector retrieval.
- Retrieval court for missing/irrelevant evidence.

التقييم الهندسي

وجود HyperKB لا يساوي GraphRAG. المطلوب Graph Schema وIngestion وIndexes وEdge Provenance ومخطط Retrieval وتقييم.

الفجوة والمتطلبات

- أنواع عقد واضحة.
- Edges بإصدارات وصلاحية زمنية.
- Graph + Vector Retrieval.
- محكمة Retrieval.

BUILD NEXT / البناء التالي

- Build reference graph from canonical subset.

EXPERIMENTS / التجارب

- Evaluate Recall@k, citation/ provenance correctness, temporal accuracy and latency.

EXIT GATE / بوابة الخروج

- GraphRAG beats flat retrieval on multi-hop/temporal tasks.

Gap 18 - Knowledge Atom, Chunking and Provenance Contract

E1 / STANDARDIZATION NEEDED | P1

الفجوة 18 - عقد الذرة المعرفية والتقطيع والمصدر

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The research already emphasizes semantic chunks and provenance. What remains is a universal atom schema shared by all KBs so that agents can retrieve the smallest sufficient evidence without losing context or version lineage.

OBSERVED GAP / REQUIREMENTS

- Atom fields: identity, claim/purpose, scope, source, timestamp, confidence, sensitivity, dependencies, supersedes, retrieval cues.
- Chunk boundary tests for tables/code/Arabic-English bilingual material.
- Deduplication by semantic and provenance identity.

التقييم الهندسي

يلزم Atom Schema موحد لكل HyperKB حتى يمكن استرجاع أصغر دليل كاف مع السياق والإصدار والمصدر.

الفجوة والمتطلبات

- حقول موحدة للذرة المعرفية.
- اختبارات حدود Chunk للجداول والكود والثنائية اللغوية.
- Dedup بالمعنى والمصدر.

BUILD NEXT / البناء التالي

- Publish JSON Schema and ingestion validator.

EXPERIMENTS / التجارب

- Reconstruction fidelity from atoms vs full document.

EXIT GATE / بوابة الخروج

- >=95% continuation fidelity on held-out reconstruction tasks.

Gap 19 - Self-Reconstruction Memory as Tested Minimal State

E1 / NO FIDELITY BENCHMARK | P0

الفجوة 19 - ذاكرة إعادة البناء كحالة دنيا مختبرة

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Self-Reconstruction Memory proposes storing the minimal trajectory needed to resume correctly rather than full traces. The missing science is measuring “minimal sufficient.” Over-compression loses causal context; under-compression recreates context pollution.

OBSERVED GAP / REQUIREMENTS

- Canonical resume capsule: previous state, active difference, decision, evidence, effect, residual uncertainty, next question.
- Memory compactor plus provenance pointers.
- Reconstruction test after cold restart.

التقييم الهندسي

Self-Reconstruction Memory تحتاج معيارا لقياس الحد الأدنى الكافي. الضغط الزائد يفقد السببية، والضغط الضعيف يعيد تلوث السياق.

الفجوة والمتطلبات

- Resume Capsule معياري.
- Compactor مع Provenance Pointers.
- اختبار Cold Restart.

BUILD NEXT / البناء التالي

- Create memory benchmark with interrupted long tasks.

EXPERIMENTS / التجارب

- Compare full trace, summary, capsule and no-memory conditions.

EXIT GATE / بوابة الخروج

- Capsule retains near-full continuation success at materially lower context cost.

Gap 20 - Context Compaction Economy: Fidelity Metric

E1 / METRIC NEEDED | P1

الفجوة 20 - اقتصاد ضغط السياق ومقياس الأمانة

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Context Compaction is currently governed by a strong qualitative principle: preserve decision-changing information. The gap is a quantitative fidelity measure that penalizes omission of constraints, provenance, unresolved contradictions and active dependencies.

OBSERVED GAP / REQUIREMENTS

- Define continuation fidelity and decision-equivalence metrics.
- Compression ratio alone is insufficient.
- Adversarial tests for hidden constraint loss.
- Multi-resolution L0-L4 representations with recoverable links.

التقييم الهندسي

ضغط السياق يحتاج Metric يقيس حفظ المعلومات التي تغير القرار، خصوصا القيود والمصدر والتناقضات والتبعيات.

الفجوة والمتطلبات

- قياس Continuation Fidelity وDecision Equivalence.
- Compression Ratio وحده غير كاف.
- اختبارات فقد قيد خفي.
- تمثيل متعدد الدقة مع روابط استرجاع.

BUILD NEXT / البناء التالي

- Build compaction benchmark from 100 project states.

EXPERIMENTS / التجارب

- Blind agents resume from compact vs full state.

EXIT GATE / بوابة الخروج

- >=95% decision equivalence at target token reduction.

Gap 21 - Cross-Model AGI Density: Transfer Experiments

E0-E1 / UNVALIDATED | P1

الفجوة 21 - Cross-Model AGI Density وتجارب النقل

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Cross-Model AGI Density is one of the most interesting research hypotheses: a canonical kernel should preserve task state and competence when the underlying model/ provider changes. The idea needs rigorous transfer metrics such as TransferGain, TranslationLoss and SubstitutionResilience under matched tasks.

OBSERVED GAP / REQUIREMENTS

- Provider-neutral task representation.
- Canonical context/ability package injected into multiple models.
- Control for model capability and token budget.
- Measure state loss, action divergence and recovery.

التقييم الهندسي

فرضية Cross-Model AGI Density مهمة: هل تحافظ النواة على الحالة والكفاءة عند تغيير النموذج؟ المطلوب تجارب نقل مضبوطة ومقاييس TransferGain وTranslationLoss وSubstitutionResilience.

الفجوة والمتطلبات

- تمثيل مهمة محايد للمزود.
- حزمة Canonical عبر نماذج متعددة.
- ضبط قدرة النموذج والميزانية.
- قياس فقد الحالة واختلاف الأفعال والتعافي.

BUILD NEXT / البناء التالي

- Run same suite on at least three model families.

EXPERIMENTS / التجارب

- Kernel-on vs kernel-off transfer study.

EXIT GATE / بوابة الخروج

- Replicated improvement in transfer robustness = E5 candidate.

Gap 22 - Model Adapter Registry and Capability Negotiation

SPECIFIED / NO UNIVERSAL CONTRACT | P1

الفجوة 22 - سجل محولات النماذج والتفاوض على القدرات

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Model-neutral architecture requires more than API wrappers. The kernel must know context limits, tool-call semantics, structured-output reliability, latency, cost, reasoning modes, safety constraints and failure patterns. Capability negotiation should influence routing before execution.

OBSERVED GAP / REQUIREMENTS

- Adapter manifest with tested capabilities and version.
- Capability probes at startup.
- Fallback selection and format translation.
- No hidden provider-specific assumptions in HyperSkills.

التقييم الهندسي

حياد النموذج يحتاج Registry يصف Context وTool Calls وStructured Output والكلفة والLatency وأنماط الفشل، ويستعملها Router قبل التنفيذ.

الفجوة والمتطلبات

- Manifest قدرات وإصدار.
- Probes عند الإقلاع.
- Fallback وترجمة Format.
- منع افتراضات مزود داخل HyperSkill.

BUILD NEXT / البناء التالي

- Implement adapters for 3 model families plus mock.

EXPERIMENTS / التجارب

- Conformance tests and forced provider swap mid-task.

EXIT GATE / بوابة الخروج

- State survives provider swap with bounded translation loss.

Gap 23 - Multi-Agent Mesh: Real Orchestration Semantics

E1 / ROLES > RUNTIME | P0

الفجوة 23 - شبكة الوكلاء ودلالات التنسيق الفعلية

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The corpus defines agent councils and meshes, but named roles do not yet establish concurrency semantics, message protocols, authority boundaries or shared-state conflict rules. A true mesh needs deterministic orchestration primitives.

OBSERVED GAP / REQUIREMENTS

- Agent identity, lease, scope and authority token.
- Task queue and message envelope.
- Shared-state read/write isolation.
- Supervisor/escalation and dead-agent recovery.

التقييم الهندسي

أسماء الوكلاء والأدوار لا تكفي. يجب تحديد Message Protocol وAuthority وShared State وConcurrency حتى تصبح الشبكة Runtime حقيقي.

الفجوة والمتطلبات

- هوية وScope وAuthority لكل وكيل.
- Task Queue وMessage Envelope.
- عزل القراءة/الكتابة.
- Supervisor وRecovery.

BUILD NEXT / البناء التالي

- Implement 5-agent reference mesh.

EXPERIMENTS / التجارب

- Race-condition, duplicate-work and disagreement tests.

EXIT GATE / بوابة الخروج

- No silent state corruption; reproducible coordination traces.

Gap 24 - Concurrency, Conflict Arbitration and Duplicate Work

UNDEPLOYED CONTROL LAYER | P0

الفجوة 24 - التزامن والتحكيم ومنع العمل المكرر

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Large agent/task counts can create coordination loss, exactly the failure that HyperDeployment theory warns about. The undeployed control plane should prevent two agents from mutating the same object inconsistently, spending duplicated tool budgets, or promoting conflicting conclusions.

OBSERVED GAP / REQUIREMENTS

- Object locks or optimistic concurrency with version checks.
- Dedup keys for tasks and side effects.
- Conflict-resolution court and merge policy.
- Budget ownership and cancellation propagation.

التقييم الهندسي

زيادة الوكلاء قد تولد Coordination Loss. نحتاج Control Plane يمنع تضارب الكتابة وتكرار الكلفة وترقية نتائج متناقضة.

الفجوة والمتطلبات

- Locks أو Optimistic Concurrency.
- Dedup Keys للمهام والآثار.
- محكمة تعارض وMerge Policy.
- ملكية الميزانية وإلغاء متسلسل.

BUILD NEXT / البناء التالي

- Implement concurrency service in reference runtime.

EXPERIMENTS / التجارب

- Stress with duplicated events and simultaneous writes.

EXIT GATE / بوابة الخروج

- Exactly-once or explicitly idempotent semantics for critical mutations.

Gap 25 - Plan DAG Executor and Dependency Semantics

E1 / EXECUTOR SPEC | P0

الفجوة 25 - منفذ Plan DAG ودلالات التبعيات

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The executor loop is well specified, but a production Plan DAG needs typed dependencies, preconditions, postconditions, resource locks, retries, compensation actions and observability. Otherwise the plan remains a sequence of prose steps.

OBSERVED GAP / REQUIREMENTS

- DAG node schema and edge semantics.
- Precondition evaluator and postcondition verifier.
- Compensation/rollback per node.
- Partial-order parallelism with safety constraints.

التقييم الهندسي

Plan DAG يحتاج تبعيات Typed وشروطا قبل/بعد وLocks وRetries وCompensation وObservability حتى لا يبقى قائمة نصية.

الفجوة والمتطلبات

- Schema للعقد والحواف.
- فاحص Preconditions وPostconditions.
- Rollback لكل عقدة.
- Parallelism آمن.

BUILD NEXT / البناء التالي

- Implement DAG runner and durable execution log.

EXPERIMENTS / التجارب

- Failure at each node; resume after crash; reorder safe nodes.

EXIT GATE / بوابة الخروج

- Deterministic resume and verified final state.

Gap 26 - Authorization Policy Engine

E1 / NEEDS ENFORCEMENT | P0

الفجوة 26 - محرك سياسة التفويض

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

T0-T4 authorization tiers are clearly defined. The gap is converting them into a policy engine that can evaluate requested action, target, sensitivity, reversibility and side-effect class before tool invocation. Textual reminders are weaker than enforceable gates.

OBSERVED GAP / REQUIREMENTS

- Machine-readable policy rules.
- Action classification and scope minimization.
- Approval token binding to exact step/target.
- Audit trail and expiration.

التقييم الهندسي

مستويات T0-T4 واضحة، لكن يجب فرضها بمحرك Policy قبل Tool Call، لا عبر تذكير نصي فقط.

الفجوة والمتطلبات

- قواعد Policy قابلة للآلة.
- تصنيف الأفعال وتقليل Scope.
- ربط الموافقة بالخطوة والهدف.
- Audit وExpiration.

BUILD NEXT / البناء التالي

- Implement pre-tool policy hook.

EXPERIMENTS / التجارب

- Attempt confused-deputy, stale approval and broadened-scope attacks.

EXIT GATE / بوابة الخروج

- Unauthorized privileged actions cannot execute through compliant adapters.

Gap 27 - Reversible Side-Effect Transaction Layer

PARTIAL / TOOL-SPECIFIC | P0

الفجوة 27 - طبقة معاملات قابلة للتراجع للآثار الخارجية

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Constitutional reversibility is a core design law, but tools differ radically in rollback support. The system needs a transaction abstraction that classifies operations as reversible, compensatable, checkpointable or irreversible, and selects safer execution patterns accordingly.

OBSERVED GAP / REQUIREMENTS

- Transaction descriptor per tool action.
- Pre-image snapshot/checkpoint when possible.
- Compensation action registry.
- Irreversible-action escalation.

التقييم الهندسي

قابلية التراجع قانون أساسي، لكن الأدوات تختلف. نحتاج Transaction Abstraction تصنف العملية إلى Reversible/Compensatable/Checkpointable/Irreversible وتختار التنفيذ الآمن.

الفجوة والمتطلبات

- وصف معاملة لكل Tool Action.
- Snapshot قبل التعديل.
- سجل Compensation.
- تصعيد الأفعال غير القابلة للتراجع.

BUILD NEXT / البناء التالي

- Implement transactional wrappers for files, records and messages where supported.

EXPERIMENTS / التجارب

- Rollback drills and compensation failure tests.

EXIT GATE / بوابة الخروج

- Rollback success rate measured and reported by tool class.

Gap 28 - Claim/Evidence Ledger as a First-Class Runtime Object

E1 / FRAGMENTED | P0

الفجوة 28 - سجل الادعاء والدليل ككائن Runtime أساسي

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Evidence discipline is one of the strongest parts of the research, but the runtime needs a claim ledger that binds every material claim to source, date, scope, confidence, assumptions, counter-evidence and validation status. Without this, epistemic labels can be lost between agents and memory compaction.

OBSERVED GAP / REQUIREMENTS

- Claim IDs and provenance edges.
- Evidence type/grade and freshness.
- Contradiction links and supersession.
- Machine gate preventing “verified” without evidence pointer.

التقييم الهندسي

انضباط الدليل قوي نظريا، لكن يلزم Claim Ledger دائم يربط كل ادعاء بمصدر وتاريخ وScope وثقة وافتراضات ودليل مضاد وحالة تحقق.

الفجوة والمتطلبات

- معرفات Claim وروابط المصدر.
- درجة الدليل وحداثته.
- روابط التناقض والاستبدال.
- بوابة تمنع Verified بلا Evidence Pointer.

BUILD NEXT / البناء التالي

- Implement claim store and API.

EXPERIMENTS / التجارب

- Trace claims through multi-agent, compressed-memory and report-generation flows.

EXIT GATE / بوابة الخروج

- No material verified claim loses provenance in end-to-end benchmark.

Gap 29 - Falsification / Demolition Courts as Executable Tests

E1 / PROMPT-LIKE | P0

الفجوة 29 - محاكم التفنيد والهدم كاختبارات قابلة للتنفيذ

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The courts are currently powerful reasoning protocols. The next step is to compile them into adversarial test suites with generated counterexamples, invariants and failure verdict schemas. Courts should produce reproducible evidence, not just critical prose.

OBSERVED GAP / REQUIREMENTS

- Court type -> required tests, fixtures and evidence.
- Automated adversarial case generation.
- Verdict schema: PASS, REPAIR, QUARANTINE, BLOCK.
- Link court failures to repair actions.

التقييم الهندسي

محاكم الهدم قوية كمنهج، لكن يجب تحويلها إلى Test Suites تولد Counterexamples وتفحص Invariants وتصدر Verdict قابل لإعادة الإنتاج.

الفجوة والمتطلبات

- كل Court مرتبط باختبارات وFixtures.
- توليد حالات خصمية.
- Verdict معياري.
- ربط الفشل بإصلاح.

BUILD NEXT / البناء التالي

- Implement 5 executable courts first: provenance, temporal, integrity, security, world-effect.

EXPERIMENTS / التجارب

- Seed known defects and measure detection.

EXIT GATE / بوابة الخروج

- High detection rate with bounded false positives.

Gap 30 - World-Effect Verifier

E1 / TOOL EVIDENCE INCONSISTENT | P0

الفجوة 30 - متحقق World-Effect

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

World-Effect is the decisive boundary between convincing text and accomplished work. The gap is a generic verifier framework that knows how to confirm that a file exists and opens, a deployment serves correctly, a message was sent, a record changed, a benchmark ran, or an external metric moved.

OBSERVED GAP / REQUIREMENTS

- Effect schema: target state, observation method, confidence, timestamp, independent verifier.
- Tool-specific verifiers and postcondition checks.
- Observed vs inferred vs claimed status.
- Idempotency and stale-observation protection.

التقييم الهندسي

World-Effect هو الحد بين النص والعمل. المطلوب Framework يثبت الأثر حسب نوعه: ملف، نشر، رسالة، سجل، Metric، Benchmark خارجي.

الفجوة والمتطلبات

- Schema للأثر وطريقة الملاحظة.
- Verifiers حسب الأداة.
- Observed/Inferred/Claimed.
- حماية من Observation قديم.

BUILD NEXT / البناء التالي

- Implement verifier plugins for 5 effect classes.

EXPERIMENTS / التجارب

- Inject false-positive tool responses and stale state.

EXIT GATE / بوابة الخروج

- No VERIFIED_SUCCESS without an acceptable observation path.

Gap 31 - Observability, Trace Store and Metrics Semantics

E1 / NEEDS BACKEND | P0

الفجوة 31 - الرصد وسجل الآثار ودلالات المقاييس

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The architecture requires traces, but a scalable runtime needs a structured telemetry backend: events, spans, state transitions, tool calls, route decisions, evidence, costs, errors, repairs and world effects. Raw logs alone are insufficient for scientific comparison.

OBSERVED GAP / REQUIREMENTS

- Canonical event schema and correlation IDs.
- Metrics definitions with units and aggregation rules.
- Privacy/sensitivity redaction.
- Query layer for experiment analysis.

التقييم الهندسي

المعمارية تحتاج Backend للرصد يجمع Events وSpans وانتقالات الحالة وTool Calls والRouting والدليل والكلفة والإصلاح والأثر.

الفجوة والمتطلبات

- Event Schema وCorrelation IDs.
- تعريف Metrics ووحداتها.
- Redaction للحساسية.
- Query Layer للتحليل.

BUILD NEXT / البناء التالي

- Implement local telemetry stack and export format.

EXPERIMENTS / التجارب

- Replay identical run and compare trace determinism.

EXIT GATE / بوابة الخروج

- Every benchmark result is trace-linked and reproducible.

Gap 32 - Real Failure Replay Engine

SPECIFIED IN EVOLUTION LAB | P0

الفجوة 32 - محرك إعادة تشغيل الفشل الحقيقي

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The Evolution Lab explicitly depends on replaying real failures without removing causal conditions. The gap is a replay engine that snapshots relevant inputs, tool responses, state and timing while safely redacting secrets. Without replay, learning candidates cannot be reliably compared against historical failures.

OBSERVED GAP / REQUIREMENTS

- Failure bundle format and causal-condition manifest.
- Mock/stub external effects for safe replay.
- Deterministic seeds where possible.
- Replay divergence detector.

التقييم الهندسي

Evolution Lab تعتمد على Real Failure Replay. المطلوب Engine يحفظ المدخلات والحالة وTool Responses والعوامل السببية مع Redaction للأسرار.

الفجوة والمتطلبات

- صيغة Failure Bundle.
- Mocks للآثار الخارجية.
- Seeds حتمية.
- كشف اختلاف Replay.

BUILD NEXT / البناء التالي

- Build replay corpus from 50 real/ constructed failures.

EXPERIMENTS / التجارب

- Candidate vs baseline on frozen replay set.

EXIT GATE / بوابة الخروج

- Replays reproduce target failure or explain divergence.

Gap 33 - Frozen Baseline and Benchmark Suite

REQUIRED / NOT SUFFICIENTLY EVIDENCED | P0

الفجوة 33 - Baseline ثابت وحزمة Benchmarks

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The research paper itself says many performance claims remain below E4 until controlled experiments exist. A frozen baseline suite is therefore the single most important scientific infrastructure gap. It must cover correctness, tool use, long-horizon tasks, retrieval, recovery, authorization and cost.

OBSERVED GAP / REQUIREMENTS

- Versioned benchmark tasks with hidden tests.
- Baselines: direct model, RAG-only, ReAct-like, simple tool agent, MOUSSAIF ablations.
- Fixed budgets and environment snapshots.
- Statistical protocol and confidence intervals.

التقييم الهندسي

الورقة نفسها تربط E4 بالتجارب المضبوطة. لذا Benchmark Suite ثابتة هي أهم فجوة علمية: صحة، أدوات، مهام طويلة، Recovery، Retrieval، Cost، Authorization.

الفجوة والمتطلبات

- مهام بإصدارات واختبارات مخفية.
- Baselines متعددة.
- ميزانيات وبيئات ثابتة.
- بروتوكول إحصائي.

BUILD NEXT / البناء التالي

- Build v0.1 benchmark of 200 tasks.

EXPERIMENTS / التجارب

- Run repeated trials across at least two models.

EXIT GATE / بوابة الخروج

- Public/internal benchmark report with reproducible configs.

Gap 34 - Candidate Promotion, Canary and Automatic Rollback

E1 / POLICY NOT ENFORCED | P0

الفجوة 34 - ترقية المرشحين وCanary وRollback تلقائي

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The Evolution Lab forbids self-promotion and specifies regression gates. The gap is an actual promotion controller that prevents candidate skills, prompts, router weights or adapters from becoming canonical without signed benchmark results and canary evidence.

OBSERVED GAP / REQUIREMENTS

- Candidate registry and immutable baseline pointer.
- Critical regression dimensions with zero tolerance.
- Canary exposure percentage and rollback thresholds.
- Promotion decision record with approver/evidence.

التقييم الهندسي

Evolution Lab تمنع Self-Promotion، لكن يلزم Promotion Controller يفرض Benchmarks وCanary قبل اعتماد أي Skill أو Weight أو Adapter.

الفجوة والمتطلبات

- سجل Candidate وBaseline ثابت.
- صفر Regression في الأبعاد الحرجة.
- Canary وThresholds للتراجع.
- سجل قرار الترقية ودليله.

BUILD NEXT / البناء التالي

- Implement promotion service around artifact registry.

EXPERIMENTS / التجارب

- Inject candidate with hidden safety regression.

EXIT GATE / بوابة الخروج

- Unsafe candidate cannot promote; rollback works automatically.

Gap 35 - Training/Adaptation Layer: From Trace Corpus to Proven Improvement

E0-E1 / CAUTIOUS | P2

الفجوة 35 - طبقة التدريب والتكييف من Trace إلى تحسن مثبت

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The corpus discusses trace datasets and bounded adapters, but the system correctly avoids claiming that loading KB files changes model weights. The research gap is to test whether distilled traces, retrieval policies or lightweight adapters measurably improve task performance without damaging generality or safety.

OBSERVED GAP / REQUIREMENTS

- Trace cleaning and leakage control.
- Train/evaluate split by task family and time.
- Adapter vs retrieval-only vs prompt-only comparisons.
- Regression gates and forgetting tests.

التقييم الهندسي

المنظومة لا تدعي أن تحميل KB يغير الأوزان. الفجوة البحثية هي اختبار Distillation أو Retrieval Policies أو Adapters خفيفة وقياس أثرها مع حماية التعميم والسلامة.

الفجوة والمتطلبات

- تنظيف Trace ومنع Leakage.
- تقسيم زمني ومجالي.
- مقارنة Adapter/Retrieval/Prompt.
- اختبارات Regression ونسيان.

BUILD NEXT / البناء التالي

- Create small reproducible adaptation experiment first.

EXPERIMENTS / التجارب

- Measure gain, cost, calibration, robustness and safety.

EXIT GATE / بوابة الخروج

- No promotion unless gain survives held-out and cross-domain tests.

Gap 36 - Hypothesis-to-Experiment Research Compiler

E1 / MANUAL | P1

الفجوة 36 - مترجم الفرضية إلى تجربة

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The theory construction pipeline already calls for definitions, mechanisms, metrics, hypotheses, counter-models and empirical programs. The missing layer is a research compiler that takes a theory object and generates falsifiers, experiments, required datasets, baselines, measurements and preregistered acceptance criteria.

OBSERVED GAP / REQUIREMENTS

- Theory schema with construct definitions and causal assumptions.
- Automatic experiment-plan template.
- Alternative hypothesis and confound register.
- Result ingestion that updates theory status.

التقييم الهندسي

Pipeline النظريات موجودة، لكن نحتاج Research Compiler يحول Theory Object إلى Falsifiers وتجارب وDatasets وBaselines وAcceptance Criteria مسبقة.

الفجوة والمتطلبات

- Theory Schema وتعريف Construct.
- قالب تجربة تلقائي.
- سجل فرضيات بديلة وConfounds.
- إدخال النتائج وتحديث رتبة النظرية.

BUILD NEXT / البناء التالي

- Implement for MCD, QMCD and Context Compaction first.

EXPERIMENTS / التجارب

- Independent reviewer checks whether generated tests could falsify the theory.

EXIT GATE / بوابة الخروج

- Every promoted theory has at least one executable falsification plan.

Gap 37 - Theory Redundancy, Equivalence and Lineage Audit

PARTIAL / NEEDED | P1

الفجوة 37 - تدقيق تكرار النظريات والتكافؤ والسلالة

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

The corpus already warns against duplicate naming. As the theory family grows, semantic equivalence becomes a major risk: a new name may restate sparse routing, active inference, scheduling, memory compression or verification under different vocabulary. A formal lineage court is needed before adding theory names.

OBSERVED GAP / REQUIREMENTS

- Concept graph of definitions, mechanisms, predictions and unique observables.
- Equivalence classes and “delta required” rule.
- External prior-art comparison when publishing scientifically.
- Deprecation/merge protocol.

التقييم الهندسي

مع نمو العائلة، خطر إعادة تسمية فكرة قديمة كبير. يجب محكمة Lineage/Equivalence قبل قبول اسم نظرية جديد.

الفجوة والمتطلبات

- Graph للتعريفات والآليات والتنبؤات.
- Equivalence Classes وDelta Required.
- مقارنة Prior Art عند النشر.
- Merge/Deprecation Protocol.

BUILD NEXT / البناء التالي

- Build theory registry with 20 current core theories.

EXPERIMENTS / التجارب

- Pairwise semantic/mechanism comparison and reviewer adjudication.

EXIT GATE / بوابة الخروج

- Every surviving theory states a non-redundant mechanism/prediction.

Gap 38 - Mathematical Identifiability and Dimensional Analysis

E0-E1 / FORMALIZATION GAP | P1

الفجوة 38 - قابلية التعيين الرياضي والتحليل البعدي

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Several formulas are useful as design heuristics, but scientific maturation requires dimensional consistency, identifiable parameters, domain constraints and explicit null models. This applies especially to MCD, QMCD and effective HyperDeployment capacity.

OBSERVED GAP / REQUIREMENTS

- State whether each quantity is ordinal, interval, probability, count or normalized score.
- Prove or test invariances under rescaling.
- Sensitivity analysis for lambda weights and multiplicative factors.
- Specify null/degenerate cases.

التقييم الهندسي

بعض الصيغ مفيدة كـHeuristics، لكن النضج العلمي يحتاج Dimensional Consistency وIdentifiable Parameters وNull Models وحدود مجال واضحة.

الفجوة والمتطلبات

- تحديد نوع كل كمية.
- اختبار Invariance عند تغيير المقياس.
- Sensitivity للأوزان والعوامل.
- حالات Null وDegenerate.

BUILD NEXT / البناء التالي

- Create formal appendix and executable notebooks.

EXPERIMENTS / التجارب

- Parameter recovery from synthetic data; perturbation tests.

EXIT GATE / بوابة الخروج

- Equations make predictions that differ from simpler null models.

Gap 39 - Causal Attribution and Component Ablation Program

NOT YET CENTRALIZED | P0

الفجوة 39 - الإسناد السببي وبرنامج Ablation

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

A large integrated system can improve while the wrong component receives credit. Causal attribution requires ablations: remove the governing-question stage, MCD, GraphRAG, a court, memory, verification or multi-agent composition while holding other factors constant.

OBSERVED GAP / REQUIREMENTS

- Factorial or fractional-factorial designs for interacting components.
- Matched compute budgets.
- Report negative and null results.
- Interaction effects between retrieval, routing, loops and verification.

التقييم الهندسي

النظام الكبير قد يتحسن دون معرفة المكون المسؤول. يلزم Ablation Program يعطل كل طبقة على حدة مع تثبيت الباقي.

الفجوة والمتطلبات

- تصاميم Factorial للمكونات المتفاعلة.
- ميزانيات متطابقة.
- نشر النتائج السلبية والمحايدة.
- قياس Interaction Effects.

BUILD NEXT / البناء التالي

- Integrate ablation switches into HAKIM pipeline runner.

EXPERIMENTS / التجارب

- Run on frozen benchmark with multiple seeds/models.

EXIT GATE / بوابة الخروج

- Core components retain only if marginal value is demonstrated.

Gap 40 - Security Threat Model for Agentic OS

E1 / NEEDS RED-TEAM | P0

الفجوة 40 - نموذج تهديد أمني لنظام وكيل

E0 | **E1** | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

Authorization rules exist, but an agentic OS also needs a threat model for prompt injection, poisoned KB content, malicious tools, credential misuse, cross-project data leakage, unsafe automation, compromised adapters and rollback tampering.

OBSERVED GAP / REQUIREMENTS

- Trust boundaries and data-flow diagrams.
- Content provenance and quarantine.
- Tool allowlists/scoped credentials.
- Prompt-injection detection and instruction-source precedence.
- Tamper-evident audit logs.

التقييم الهندسي

قواعد التفويض موجودة، لكن يلزم Threat Model يغطي Prompt Injection وPoisoned KB وMalicious Tools وتسرب البيانات وسوء استخدام Credentials والتلاعب بالRollback.

الفجوة والمتطلبات

- Trust Boundaries وData Flows.
- Quarantine للمحتوى.
- Allowlists وScoped Credentials.
- حماية من Prompt Injection.
- Audit غير قابل للتلاعب.

BUILD NEXT / البناء التالي

- Run structured red-team suite.

EXPERIMENTS / التجارب

- Attack retrieval, tool adapters, memory and approval tokens.

EXIT GATE / بوابة الخروج

- Critical security regressions = zero before E4 promotion.

Gap 41 - MOUSSAIXT Cross-Environment Runtime and Sync

FORMALIZED / NOT ONE VERIFIED RUNTIME | P0

الفجوة 41 - تشغيل MOUSSAIXT العابر للبيئات والمزامنة

E0 | E1 | E2 | E3 | E4 | E5

ENGINEERING ASSESSMENT

MOUSSAIXT is best treated as the cross-environment shell binding cloud ChatGPT tooling, Drive memory, Hermes/local execution, coding agents and model providers. The undeployed layer is a verified runtime protocol for boot, environment discovery, state sync, capability negotiation, conflict resolution and provenance-preserving writeback.

OBSERVED GAP / REQUIREMENTS

- Boot manifest and environment identity.
- Incremental sync with vector clocks/version IDs.
- Canonical state ownership rules.
- Offline/online reconciliation.
- Cross-environment trace correlation.

التقييم الهندسي

MOUSSAIXT كغلاف عابر للبيئات يحتاج Runtime Protocol موحد للإقلاع وDiscovery والمزامنة والتفاوض على القدرات وحل التعارض والWriteback مع Provenance.

الفجوة والمتطلبات

- Boot Manifest وهوية البيئة.
- Incremental Sync بإصدارات.
- قواعد ملكية الحالة.
- Reconciliation عند العودة Online.
- ربط Traces عبر البيئات.

BUILD NEXT / البناء التالي

- Implement two-node reference deployment: cloud + local.

EXPERIMENTS / التجارب

- Concurrent edits, disconnect/ reconnect, provider swap, partial tool availability.

EXIT GATE / بوابة الخروج

- No silent state loss; conflicts are surfaced and resolvable.

Priority Matrix: What to Build First

ROADMAP P0

مصفوفة الأولويات: ماذا نبني أولا

E0 E1 E2 E3 E4 E5

ENGINEERING ASSESSMENT

The correct dependency order is infrastructure before ambitious theory validation. E4 experiments are impossible without a reliable executor, trace store, frozen benchmarks and effect verification. The work is therefore grouped into P0 foundation, P1 research validation and P2 advanced adaptation.

OBSERVED GAP / REQUIREMENTS

- P0: reference runtime, adapter layer, skill compiler, router telemetry, HyperLoop state machine, world model, evidence ledger, executable courts, world-effect verifier, trace store, replay, benchmark suite, security.
- P1: MCD/QMCD measurement, MCAI, vascular scheduler, HyperDeployment predictors, context compaction, cross-model transfer, theory compiler, formalization.
- P2: learned router, lightweight adapters/training, large-scale distributed agent mesh, broader public benchmarks.

التقييم الهندسي

الترتيب الصحيح يبدأ بالبنية التي تجعل التجارب ممكنة. لا معنى لـ E4 بلا Executor وTrace Store وBenchmark وWorld-Effect Verifier. لذلك نقسم العمل إلى P0 ثم P1 ثم P2.

الفجوة والمتطلبات

- P0: Runtime وAdapters وCompiler وHyperLoop وWorld وModel وEvidence وCourts وVerification وTelemetry وBenchmarks وSecurity.
- P1: MCD/QMCD وMCAI وVascular وHyperDeployment وCompaction وCross-Model والتFormalization.
- P2: Learned Router وAdapters تدريبية وتوزيع واسع وBenchmarks عامة.

P0 - Foundation	P1 - Validation	P2 - Scale
Required to make evidence possible.	Turns theories into falsifiable experiments.	Only after P0/P1 are stable.

90-Day Build Program: Three Engineering Sprints

ROADMAP P0

برنامج بناء 90 يوما: ثلاث مراحل هندسية

E0 E1 E2 E3 E4 E5

ENGINEERING ASSESSMENT

A practical next phase can be organized into three 30-day sprints. Sprint 1 creates the measurement substrate. Sprint 2 creates the executable cognitive runtime. Sprint 3 performs the first controlled research campaign. The goal is not production AGI; it is a reproducible research platform that can accept or reject MOUSSAIF hypotheses with evidence.

OBSERVED GAP / REQUIREMENTS

- Days 1-30: schemas, canonical world store, telemetry, benchmark v0.1, replay bundles, adapters, security baseline.
- Days 31-60: skill compiler, HyperAbility runtime, HyperLoop engine, router, Plan DAG, courts, world-effect verifier.
- Days 61-90: HAKIM end-to-end runner, MCD/QMCD operationalization, ablations, cross-model trials, context-compaction study, E4 report.

التقييم الهندسي

يمكن تنظيم المرحلة القادمة في ثلاث دورات من 30 يوما: أولا منصة القياس، ثانيا Runtime معرفي قابل للتنفيذ، ثالثا حملة تجارب مضبوطة. الهدف منصة بحثية تستطيع قبول أو رفض فرضيات مساعف بالدليل.

الفجوة والمتطلبات

- 1-30: Schemas وWorld Store وTelemetry وBenchmark وReplay وAdapters وأمن.
- 31-60: Skill Compiler وHyperAbility وHyperLoop وRouter وDAG وCourts وWorld-Effect.
- 61-90: HAKIM Runner وMCD/QMCD وAblations وCross-Model وCompaction وتقرير E4.

BUILD NEXT / البناء التالي

- Weekly artifact freeze and benchmark regression run.
- No new core theory name unless it passes lineage/delta review.

EXPERIMENTS / التجارب

- Track E-level change per component, not page/file count.

EXIT GATE / بوابة الخروج

- At day 90, at least 10 core components reach E3 and 3-5 hypotheses have E4 evidence or falsification.

Acceptance Criteria for E4 and the Path to E5

RESEARCH GATES P0

معايير قبول E4 والطريق إلى E5

E0 E1 E2 E3 E4 E5

ENGINEERING ASSESSMENT

E4 should be a hard gate: the system must beat a frozen baseline on predefined metrics under matched budgets and repeated trials without critical regressions. E5 requires replication across models, domains and environments. A component that fails should be revised, merged or removed rather than protected by terminology.

OBSERVED GAP / REQUIREMENTS

- Pre-register hypothesis, primary metric, secondary metrics, failure threshold and stopping rule.
- Use hidden/held-out tasks and repeated seeds.
- Report confidence intervals and cost/latency, not success alone.
- Require security, authorization and truthfulness non-regression.
- Replicate by an independent configuration or model family for E5.

التقييم الهندسي

E4 يجب أن تكون بوابة صعبة: تحسن على Baseline ثابت بمقاييس محددة وميزانيات متطابقة وتكرار، بلا Regression حرج. E5 تتطلب تعميم النتيجة عبر نماذج ومجالات وبيئات.

الفجوة والمتطلبات

- تسجيل الفرضية والمقياس وFailure Threshold وStop Rule مسبقا.
- Held-out Tasks وSeeds متعددة.
- Confidence Intervals مع الكلفة والLatency.
- عدم تراجع الأمن والتفويض والصدق.
- تكرار مستقل أو عبر Model Family للوصول إلى E5.

Next Canonical Target: MOUSSAIF Research Gap Closure Program v1

CONCLUSION P0

الهدف المعياري التالي: برنامج إغلاق فجوات أبحاث MOUSSAIF v1

E0 E1 E2 E3 E4 E5

ENGINEERING ASSESSMENT

The next canonical system should be a research-grade reference implementation, not a larger corpus. Its identity can be: MOUSSAIF AGI OS Research Gap Closure Program v1 - a minimal end-to-end runtime plus benchmark/evidence laboratory. It should be deliberately smaller than the knowledge archive, because its purpose is to prove which parts of the archive deserve to become operational doctrine.

OBSERVED GAP / REQUIREMENTS

- Core deliverables: runtime repo, schemas, 30 executable skills, 10 HyperAbilities, bounded-loop engine, HAKIM runner, world model, evidence ledger, courts, verifier, telemetry, replay, benchmark v0.1.
- Research deliverables: MCD/QMCD protocol, ablation matrix, cross-model transfer study, compaction study, HyperDeployment predictor validation.
- Governance deliverables: authorization engine, security red-team suite, promotion/canary/rollback service.
- Success is measured in E-level advancement and reproducibility, not token count, file size or number of coined terms.

التقييم الهندسي

الهدف التالي ليس ملفا أضخم، بل Reference Implementation صغيرة نسبيا ومشددة: Runtime كامل من السؤال إلى الأثر، مع Benchmark/Evidence Lab. وظيفتها تحديد أي أجزاء من الأرشيف تستحق التحول إلى عقيدة تشغيلية وأيها يحتاج تعديل أو حذف.

الفجوة والمتطلبات

- المخرجات الأساسية: Runtime Repo وSchemas و30 Skill تنفيذية و10 HyperAbilities ومحرك HyperLoop وHAKIM وWorld Model وEvidence وCourts وTelemetry وBenchmark.
- المخرجات البحثية: بروتوكول MCD/QMCD وAblations وCross-Model وCompaction وHyperDeployment Validation.
- الحوكمة: Authorization Engine وRed-Team وPromotion/Canary/Rollback.
- النجاح يقاس بارتفاع E-Level وقابلية الإعادة، لا بحجم الملفات أو عدد المصطلحات.