MOUSSAIFXT · COGNITIVE KERNEL · PART II · MUT/MUG FRONTIER EXTENSION · 2026

النواة الإدراكية MOUSSAIFXT - الجزء الثاني

امتداد MUT/MUG والهندسة الحديثة الموثوقة

كتاب تكميلي يدمج آخر تحديثات MUT وMUG وMUT-AFI مع تقنيات Agentic موثوقة حتى أغسطس 2026: Structured Outputs، OpenTelemetry GenAI، Policy-as-Code، Durable Execution، A2A، MCP، Sandboxed Harnesses، الذاكرة المتعلمة، Test-Time Compute، أمن Prompt Injection، التقييم طويل الأفق، وvLLM الحديث - من دون كسر معمارية L0-L11 الأساسية.

زهير مساعف / Moussaif Zouhair · Arabic RTL · Amiri · A4

Research-grounded engineering extension. Internal constructs remain proposed until measured; external standards are cited separately.

العقد العلمي والتقني للجزء الثاني

هذا الجزء لا يعيد مكونات CK-01→CK-50 ولا النظريات الأساسية. وظيفته هي **تحديث النواة بعد آخر MUT/MUG/MUT-AFI وربطها بمعايير وتقنيات وبحوث حديثة موثوقة** من دون أن تصبح هذه التقنيات نفسها نظرية مساعفية أو مصدر سلطة داخلية.

Synthesis [S]

Reference implementation decisions proposed in this book.

External [X]

Standards, official docs, primary research and benchmarks.

Canonical [C]

MUT/MUG/MUT-AFI وMOUSSAIFXT Part I.

قاعدة الادعاء: لا يتحول MCP أو A2A أو Temporal أو vLLM إلى «دليل AGI». هي تقنيات بنية تحتية. ولا تتحول نظرية MUT-AFI إلى حقيقة لأنها دخلت Runtime schema؛ يلزم E4/E5.

```
Canonical Theory ≠ External Standard ≠ Reference Implementation ≠ Empirical Evidence
```

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

خريطة التحديث: ما الذي لم يتغير وما الذي أُضيف

L0 **Constitution**	L1 **Mission / Question**	L2 **Δ / MCD / QMCD**	L3-4 **HKOS / Evidence**
L5-7 **Skills / Abilities / Loops**	L8-9 **Agents / Tools**	L10 **Courts / Effect**	L11 **Memory / Evolution**

طبقات الامتداد المتعامدة

X1 Durable Execution Workflow persistence / recovery	**X0 Interoperability** MCP + A2A
X3 Telemetry & Eval OTel + long-horizon benchmarks	**X2 Policy & Trust** Policy-as-code / authorization
X5 Adaptive Inference Test-time compute + vLLM fabric	**X4 Adaptive Memory** Memory policy / graph evolution
X7 Capability Certification causal credit / E3→E5 promotion	**X6 Semantic Homeostasis** cross-model/provider continuity

القاعدة: X0-X7 لا تستبدل L0-L11؛ إنها planes تنفيذية تعبر الطبقات وتغلق فجوات production/evidence.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

آخر تحديثات MUT/MUG التي تغيّر تصميم XT

المصدر	الإضافة الحاكمة	أثرها في Part II
MUT-AGI-02	Ω_t state + pipeline + 100 tools + 19 HyperKB + sparse route	توحيد KernelState وruntime lifecycle
MUG×MOUSSAIF9B/1B	Kernel / Training Stack / Unified Theory كأجسام مختلفة	Vessel placement + selective internalization
MUT-AFI	20 نظريات تكميلية للفجوات السببية/التمثيل/الحاكمية/الفشل/النظم	ثماني طبقات هندسية جديدة
MOUSSAIFXT Part I	CK components + Local/SaaS/Hybrid + E0-E5 50	قاعدة لا يعاد بناؤها، هذا الكتاب patch/extension

```
Task Contract → Q-MCD → MCD → State → Sparse Route → Bounded Loop → Court → World Effect → Memory
```

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

دفتر المصادر الخارجية - أغسطس 2026

كل مصدر خارجي هنا إما معيار رسمي، توثيق منتج/مشروع أولي، أو ورقة بحث/benchmark أصلية. لا تُستخدم مقالات تلخيصية لبناء قرار معماري أساسي.

ID	المصدر ودوره
X01	[STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension.
X02	[STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration.
X03	[STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations.
X04	[DOC] OpenAI Agents SDK, Apr 2026 - sandbox-aware harness, memory, MCP, externalized state, snapshot/rehydration and compute separation.
X05	[DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.
X06	[STD/DOC] OPA and Cedar - policy-as-code engines separating policy decision from enforcement and context-aware authorization.
X07	[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving.
X08	[PPR] A-MEM (2025) - dynamically linked and evolving agentic memory.
X09	[PPR] Agentic Memory / AgeMem (2026) - learned unified LTM/STM memory operations as tool actions.
X10	[PPR] Scaling Test-time Compute for LLM Agents (2025) - diversified rollouts, reflection timing and verifier/merge choices matter.
X11	[PPR] Test-Time Scaling in Reasoning LLMs (2026) - protocol-matched compute accounting and reproducibility requirements.
X12	[BENCH] τ-bench - dynamic tool-agent-user interaction, end-state validation and pass^k reliability.
X13	[BENCH/SEC] AgentDojo - prompt-injection evaluation over tool-using agents and untrusted data.
X14	[BENCH] OSWorld 2.0 (2026) - long-horizon professional computer-use workflows with stateful artifacts and safety auditing.

مصفوفة التبني: Adopt / Pilot / Monitor

العنصر	القرار	السبب
MCP 2026-07-28	ADOPT	.Interop stateless + auth/routing + Tasks extension
A2A 1.0	ADOPT EDGE	.Remote agent federation; no direct canonical memory write
Temporal-style durable execution	ADOPT INTERFACE	.Long-running recovery and retries; keep engine replaceable
OPA/Cedar	ADOPT	.Policy decision/enforcement separation
OpenTelemetry GenAI	ADOPT CAREFULLY	.Common trace semantics; content recording policy-controlled
A-MEM/AgeMem	PILOT	.Promising learned memory; canonical writes remain governed
Adaptive test-time scaling	PILOT	.Measured gains but protocol/cost must be explicit
Speculative/disaggregated vLLM	PER-HARDWARE	.Serving optimization, not cognitive law
Blind agent swarm	REJECT DEFAULT	.Coordination tax and correlated failure

الفهرس التنفيذي - 24 محور امتداد

EX-01 تقارب حالة MUT مع Runtime النواة
MUT Runtime-State Convergence Layer

EX-02 طبقة MUG للحاكمية ثلاثية الأجسام
MUG Tri-Body Governance Integration

EX-03 طبقة الإسناد السببي للفروق
Causal Difference Credit Layer

EX-04 مخطط الأثر المضاد للواقع
Counterfactual World-Effect Planner

EX-05 الاستتباب الدلالي عبر النماذج
Cross-Model Semantic Homeostasis

EX-06 إغلاق القدرة وتعلم طوبولوجيا الفشل
Capability Closure & Failure Topology

EX-07 تحويل عدم اليقين إلى وكالة وديناميات الثقة
Uncertainty-to-Agency & Delegation Trust

EX-08 النمو القابل للعكس للاستقلالية ومناعة انحراف الهدف
Reversible Autonomy Growth & Goal-Drift Immunity

EX-09 هندسة الدليل وموازنة الحمل الإبستيمي
Evidence Geometry & Epistemic Load Balancing

EX-10 بيئة الأدوات والتعلم من أحكام المحاكم
Tool Ecology & Court-to-Learning

EX-11 ترقية الأبعاد ونقل المعرفة إلى مهارة
Dimension Promotion & Knowledge-to-Skill Transfer

EX-12 طبقة بروتوكول MCP الحديثة
MCP 2026 Tool/Data Interoperability Layer

EX-13 طبقة اتحاد الوكلاء A2A
A2A 1.0 Agent Federation Layer

EX-14 طبقة التنفيذ المتين طويل الأفق
Durable Execution Layer

EX-15 طبقة Policy-as-Code للتفويض
Policy-as-Code Authorization Layer

EX-16 طبقة الرصد الدلالي OpenTelemetry
OpenTelemetry GenAI Observability Layer

EX-17 طبقة المخرجات المقيدة بالعقد
Contract-Constrained Structured Generation

EX-18 طبقة الفصل بين Harness والحوسبة
Sandboxed Harness / Compute Separation

EX-19 طبقة إدارة الذاكرة المتعلمة
Adaptive Memory Governance Layer

EX-20 طبقة الحوسبة التكيفية وقت الاختبار
Adaptive Test-Time Compute Layer

EX-21 طبقة أمان الأدوات ومقاومة Prompt Injection
Tool Security & Prompt-Injection Resistance

EX-22 طبقة تقييم التفاعل طويل الأفق
Long-Horizon Agent Evaluation Layer

EX-23 نسيج الاستدلال المحلي عبر vLLM
vLLM Inference Fabric Layer

EX-24 معمارية MOUSSAIFXT v2 المرجعية
MOUSSAIFXT v2 Reference Integration Architecture

I

امتدادات MUT/MUG/MUT-AFI

إغلاق الفجوات الداخلية التي ظهرت بعد الكتاب الأول وتحويل آخر النظريات إلى طبقات هندسية منضبطة.

EX-01

تقارب حالة MUT مع Runtime النواة - الفرق البحثي

MUT Runtime-State Convergence Layer

الحالة	يمتد من
Canonical update + engineering synthesis	CK-04/05/06/12/28/35

الفجوة

Part I عرّف KernelState عملياً، بينما MUT-AGI-02 يحدد Ω_t وعملية /INTAKE→CONTRACT→QUESTION→STATE→ROUTE→PLAN→ACT→OBSERVE→VERIFY→COMMIT ROLLBACK→MEMORY. الفجوة هي توحيد الكائنين إلى state machine واحدة قابلة للهجرة والاختبار.

الأطروحة الهندسية

Thesis: اعتماد CanonicalRuntimeState موحد مع sub-objects versioned، وفصل mutable operational state عن append-only evidence/audit events.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

EX-01

تقارب حالة MUT مع Runtime النواة - المعمارية والعقد

MUT Runtime-State Convergence Layer

Inputs

MissionContract, QuestionObject, ActiveDifference, WorldState, PolicyDecision, RouteDecision, Budget, Evidence

Outputs

CanonicalRuntimeState v2 + StatePatch + ReplayCursor

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
INTAKE → CONTRACTED → QUESTIONED → ROUTED → PLANNED → EXECUTING → OBSERVED → VERIFIED → COMMITTED|ROLLED_BACK → MEMORIZED
```

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-01

تقارب حالة MUT مع Runtime النواة - البيانات والحالة

MUT Runtime-State Convergence Layer

Schema مرجعي

```
CanonicalRuntimeStateV2 { state_id, schema_version, mission, question, differences[], world_ref,
policy_ref, route_ref, dag_ref, claims[], evidence_refs[], tool_receipts[], court_verdicts[], effects[],
budgets, replay_cursor, residual_uncertainty }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-01

تقارب حالة MUT مع Runtime النواة - الخوارزمية والتكامل

MUT Runtime-State Convergence Layer

الخوارزمية

1. Freeze mission invariants before routing.
2. Every transition emits StatePatch, never opaque overwrite.
3. Observed state dominates remembered/generated state.
4. Negative results remain replayable.
5. Schema migration is explicit and versioned.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-01

تقارب حالة MUT مع Runtime النواة - الأمن والفشل

MUT Runtime-State Convergence Layer

أنماط الفشل

- state duplication
- schema drift بين MUT وXT
- commit without external readback
- silent loss during cross-model transfer
- replay executing side effects

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-01

تقارب حالة MUT مع Runtime النواة - القياس ومسار الترقية

MUT Runtime-State Convergence Layer

المقاييس

schema_migration_success	state_reconstruction_fidelity
lost_field_rate	replay_determinism
	commit_correctness

الاختبارات الحاسمة

- restart at every transition
- cross-model serialization
- rollback after partial tool effect
- old-schema migration
- negative-result replay

مسار التنفيذ

Implement as Postgres canonical rows + append-only event table + JSON Schema/Pydantic model; adapters project prompt/ model-specific state from this canonical object.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

EX-02

طبقة MUG للحاكمية ثلاثية الأجسام - الفرق البحثي

MUG Tri-Body Governance Integration

الحالة	يمتد من
Canonical MUG lineage	CK-01/12/34/36/41/50

الفجوة

MUG يقسم البرنامج إلى Moussaif1B Kernel وMOUSSAIF9B Training Stack وMUT بوصفها اللغة السببية. Part I عرضها كسلالة؛ المطلوب هنا عقد واضح لحركة الفرق بين runtime الخارجي والتدريب الوزني والنظرية.

الأطروحة الهندسية

Thesis: كل capability يجب أن يملك Vessel Decision يقرر بقاءه في Context/HyperKB/HyperSkill/HyperLoop/Adapter/Weight مع دليل ترقية وRollback.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-02

طبقة MUG للحاكمية ثلاثية الأجسام - المعمارية والعقد

MUG Tri-Body Governance Integration

Inputs

```
VerifiedTrace, CapabilityCandidate,
transfer_evidence, volatility, privacy,
compute_budget
```

Outputs

+ VesselPlacementDecision + PromotionPlan
RollbackTarget

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
OBSERVED_CAPABILITY → EXTERNALIZED → REPEATED → TRANSFER_TESTED → ADAPTER_CANDIDATE →
WEIGHT_CANDIDATE → PROMOTED|REVERTED
```

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-02

طبقة MUG للحاكمية ثلاثية الأجسام - البيانات والحالة

MUG Tri-Body Governance Integration

Schema مرجعي

```
VesselDecision { capability_id, evidence_rank, volatility, privacy_class, transfer_score, target_vessel,
promotion_gate, rollback_target, checkpoint_ref? }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-02

طبقة MUG للحاكمية ثلاثية الأجسام - الخوارزمية والتكامل

MUG Tri-Body Governance Integration

الخوارزمية

1. Start external by default.
2. Promote only patterns stable across tasks/domains.
3. Never hard-code volatile/private knowledge in weights.
4. Equal-budget ablation must isolate gain.
5. Checkpoint identity and parameter count must be explicit.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-02

طبقة MUG للحاكمية ثلاثية الأجسام - الأمن والفشل

MUG Tri-Body Governance Integration

أنماط الفشل

- project name mistaken for model size
- training eats mutable knowledge
- no forgetting audit
- capability promotion after one success
- unreversible weight update

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-02

طبقة MUG للحاكمية ثلاثية الأجسام - القياس ومسار الترقية

MUG Tri-Body Governance Integration

المقاييس

- `transfer_retention`
- `external_to_internal_gain`
- `rollback_readiness`
- `forgetting_regression`
- `verified_gain_per_compute`

الاختبارات الحاسمة

- context-vs-adapter-vs-weight ablation
- name-stripped transfer
- privacy leakage eval
- forgetting suite
- small-model proxy before scale-up

مسار التنفيذ

> Build a promotion registry connected to eval lab. Use lightweight adapters/LoRA only after repeatable external capability evidence.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

EX-03

طبقة الإسناد السببي للفروق - الفرق البحثي

Causal Difference Credit Layer

الحالة	يمتد من
MUT-AFI proposed layer	CK-06/33/34/36

الفجوة

MCD يحدد الفرق المهم وWorld-Effect يقرأ الأثر، لكن لا يكفي ذلك لإسناد النجاح إلى الفرق أو المكوّن الذي سببه.

الأطروحة الهندسية

> **Thesis:** أضف causal credit ledger يربط كل difference/route/tool intervention بالأثر النهائي عبر ablation أو counterfactual proxy.

> **عدم التكرار:** هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-03

طبقة الإسناد السببي للفروق - المعمارية والعقد

Causal Difference Credit Layer

Inputs

`DifferenceObject[], RouteTrace, interventions, WorldEffectRecord, confounders`

Outputs

CausalCreditRecord per component + uncertainty

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
OBSERVED → CANDIDATE_CAUSE → INTERVENTION/ABLATION → ATTRIBUTED → DISPUTED|ACCEPTED
```

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-03

طبقة الإسناد السببي للفروق - البيانات والحالة

Causal Difference Credit Layer

Schema مرجعي

```
CausalCredit { target_effect, candidate, intervention, counterfactual, delta_effect, confounders[],
uncertainty, verdict }
```

Persistence

state التشغيلي mutable عبر patches؛ append-evidence/audit only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-03

طبقة الإسناد السببي للفروق - الخوارزمية والتكامل

Causal Difference Credit Layer

الخوارزمية

1. Build causal candidate graph from trace.
2. Prefer real intervention when safe.
3. Otherwise matched counterfactual replay.
4. Subtract shared-root and co-activation effects.
5. Do not promote skill on correlational credit alone.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-03

طبقة الإسناد السببي للفروق - الأمن والفشل

Causal Difference Credit Layer

أنماط الفشل

- uncontrolled compute confound
- post-hoc storytelling
- selection bias
- mediator counted as root cause
- effect delayed beyond window

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-03

طبقة الإسناد السببي للفروق - القياس ومسار الترقية

Causal Difference Credit Layer

المقاييس

ablation_consistency	credit_calibration
false_causal_promotion	attribution_uncertainty
	delayed_effect_capture

الاختبارات الحاسمة

- single-component ablation
- matched-budget rerun
- mediator removal
- delayed readback
- cross-domain replication

مسار التنفيذ

> Store intervention metadata with every eval run; causal layer starts as offline analytics before influencing autonomous promotion.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

EX-04

مخطط الأثر المضاد للواقع - الفرق البحثي

Counterfactual World-Effect Planner

الحالة	يمتد من
MUT-AFI proposed layer	CK-27/28/33/34

الفجوة

.Part I verifies postconditions after action. New gap: estimating which reversible action is safest before mutation

الأطروحة الهندسية

Thesis: Plan candidate actions against explicit counterfactual world states and choose by expected verified utility, reversibility and causal identifiability.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-04

مخطط الأثر المضاد للواقع - المعمارية والعقد

Counterfactual World-Effect Planner

Inputs

```
WorldState, candidate_actions,
causal_model, constraints, risk_budget
```

Outputs

+ CounterfactualActionSet + selected_action
proof_obligation

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج postcondition readback + receipt.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
SIMULATE → COMPARE → RISK_GATE → AUTHORIZE → ACT → READBACK → UPDATE_MODEL
```

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-04

مخطط الأثر المضاد للواقع - البيانات والحالة

Counterfactual World-Effect Planner

Schema مرجعي

```
CFPlan { action, expected_state, expected_effect, reversibility, observation_plan, confounders,
utility_interval }
```

Persistence

state التشغيلي mutable عبر patches؛ append/audit-evidence؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-04

مخطط الأثر المضاد للواقع - الخوارزمية والتكامل

Counterfactual World-Effect Planner

الخوارزمية

1. Generate few structurally distinct actions.
2. Predict state delta, not prose benefit.
3. Attach observation/readback plan before authorization.
4. Prefer reversible, diagnostically informative action when utilities are close.
5. Update causal model from actual effect.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-04

مخطط الأثر المضاد للواقع - الأمن والفشل

Counterfactual World-Effect Planner

أنماط الفشل

overconfident simulation

world model hallucination

action irreversible

utility scalar hides tradeoff

observation cannot distinguish causes

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

قاعدة الهدم: إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-04

مخطط الأثر المضاد للواقع - القياس ومسار الترقية

Counterfactual World-Effect Planner

المقاييس

reversibility_weighted_utility	counterfactual_error
diagnostic_gain	prediction_interval_coverage
	rollback_success

الاختبارات الحاسمة

- synthetic causal environments
- known-simulator tasks
- paired actions
- high-uncertainty abstention
- post-hoc prediction scoring

مسار التنفيذ

> Start in sandbox/simulation and read-only planning; enable real effects only after prediction calibration.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-05

الاستتباب الدلالي عبر النماذج - الفرق البحثي

Cross-Model Semantic Homeostasis

الحالة	يمتد من
MUT-AFI proposed layer	CK-13/14/40/42

الفجوة

.Cross-model translation exists, but repeated handoffs can accumulate semantic drift even when schemas validate

الأطروحة الهندسية

:Thesis Maintain semantic invariants with round-trip translation tests, canonical hashes over structured state, and contradiction probes across model/provider changes.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-05

الاستتباب الدلالي عبر النماذج - المعمارية والعقد

Cross-Model Semantic Homeostasis

Inputs

CanonicalState, source/target capability profiles, invariant_set, translation_history

Outputs

TranslatedState + semantic_drift_score + repair_patch

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
SERIALIZE → TRANSLATE → ROUNDTRIP → PROBE → ACCEPT|REPAIR|REJECT
```

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-05

الاستتباب الدلالي عبر النماذج - البيانات والحالة

Cross-Model Semantic Homeostasis

Schema مرجعي

```
SemanticHomeostasis { invariants[], roundtrip_loss, contradiction_set, provider_delta, accepted_fields, quarantined_fields }
```

Persistence

state التشغيلي mutable عبر patches؛ append-only evidence/audit؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-05

الاستتباب الدلالي عبر النماذج - الخوارزمية والتكامل

Cross-Model Semantic Homeostasis

الخوارزمية

1. Translate typed objects not transcript.
2. Round-trip critical fields.
3. Ask target model invariant probes.
4. Compare action decisions under identical constraints.
5. Quarantine fields with high drift.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-05

الاستتباب الدلالي عبر النماذج - الأمن والفشل

Cross-Model Semantic Homeostasis

أنماط الفشل

- provider-specific hidden assumptions
- schema-valid semantic corruption
- tool affordance mismatch
- context truncation
- self-confirming evaluator

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-05

الاستتباب الدلالي عبر النماذج - القياس ومسار الترقية

Cross-Model Semantic Homeostasis

المقاييس

decision_equivalence	roundtrip_semantic_fidelity
provider_drift_rate	constraint_retention
	repair_success

الاختبارات الحاسمة

- provider swap mid-task
- small→large→small cycle
- context pressure
- tool set mismatch
- multilingual state transfer

مسار التنفيذ

> Make this mandatory for manual/provider switching and hybrid cloud/local failover.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-06

إغلاق القدرة وتعلم طوبولوجيا الفشل - الفرق البحثي

Capability Closure & Failure Topology

الحالة	يمتد من
MUT-AFI proposed layer	CK-22/23/24/36/47

الفجوة

.Skill/Ability/Loop tests can pass locally while the composed capability fails along hidden edges

الأطروحة الهندسية

Thesis: .Model capability as a dependency graph whose weakest critical edge and recurrent failure motifs govern promotion

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-06

إغلاق القدرة وتعلم طوبولوجيا الفشل - المعمارية والعقد

Capability Closure & Failure Topology

Inputs

`CapabilityGraph, component_evals, failure_events, dependency_edges`

Outputs

ClosureScore + FailureTopology + repair_priority

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
COMPOSE → STRESS → MAP_FAILURES → REPAIR → REPEAT → PROMOTE|QUARANTINE
```

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-06

إغلاق القدرة وتعلم طوبولوجيا الفشل - البيانات والحالة

Capability Closure & Failure Topology

Schema مرجعي

```
CapabilityClosure { nodes[], edges[], critical_edges[], closure_score, failure_clusters[],
repair_queue[] }
```

Persistence

state التشغيلي mutable عبر patches؛ append-evidence/audit only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-06

إغلاق القدرة وتعلم طوبولوجيا الفشل - الخوارزمية والتكامل

Capability Closure & Failure Topology

الخوارزمية

1. Stress composition, not only nodes.
2. Cluster failures by edge/condition.
3. Promote only if critical paths close.
4. Replay repaired motifs across tasks.
5. Retire redundant components.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-06

إغلاق القدرة وتعلم طوبولوجيا الفشل - الأمن والفشل

Capability Closure & Failure Topology

أنماط الفشل

too many micro-skills	averaging hides critical edge
failure taxonomy explosion	repair overfits one benchmark
	coordination overhead

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-06

إغلاق القدرة وتعلم طوبولوجيا الفشل - القياس ومسار الترقية

Capability Closure & Failure Topology

المقاييس

- `critical_edge_failure`
- `closure_rate`
- `composition_overhead`
- `repair_transfer`
- `failure_recurrence`

الاختبارات الحاسمة

- edge ablation
- dependency permutation
- cross-domain replay
- tool outage
- concurrency stress

مسار التنفيذ

> Integrate with HyperAbility compiler and eval lab; topology is updated only from verified failures.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-07

تحويل عدم اليقين إلى وكالة وديناميات الثقة - الفرق البحثي

Uncertainty-to-Agency & Delegation Trust

الحالة	يمتد من
MUT-AFI proposed layer	CK-03/12/25/26

الفجوة

Current router uses risk and uncertainty, but does not formalize how uncertainty changes autonomy, verification depth or delegation.

الأطروحة الهندسية

Thesis: Convert epistemic uncertainty into an authority ceiling and verification obligation; delegation trust is domain-local and updated from verified outcomes.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-07

تحويل عدم اليقين إلى وكالة وديناميات الثقة - المعمارية والعقد

Uncertainty-to-Agency & Delegation Trust

Inputs

`uncertainty_vector, authority_context, agent/tool history, action reversibility`

Outputs

+ AutonomyEnvelope + DelegationDecision VerificationDepth

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
ESTIMATE → BOUND → DELEGATE|ESCALATE → VERIFY → UPDATE_TRUST
```

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-07

تحويل عدم اليقين إلى وكالة وديناميات الثقة - البيانات والحالة

Uncertainty-to-Agency & Delegation Trust

Schema مرجعي

```
AutonomyEnvelope { domain, uncertainty, max_tier, required_verifiers, max_effect, expiry, trust_basis }
```

Persistence

state التشغيلي mutable عبر patches؛ append-only evidence/audit؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-07

تحويل عدم اليقين إلى وكالة وديناميات الثقة - الخوارزمية والتكامل

Uncertainty-to-Agency & Delegation Trust

الخوارزمية

1. Separate uncertainty types.
2. Lower authority as unknown unknowns/reversibility worsen.
3. Trust score is local to domain and tool.
4. Independent verifier for high effect.
5. Successful outcome updates trust only after readback.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-07

تحويل عدم اليقين إلى وكالة وديناميات الثقة - الأمن والفشل

Uncertainty-to-Agency & Delegation Trust

أنماط الفشل

- trust transfer across domains
- confidence-as-trust
- verifier correlation
- authority creep
- uncertainty ignored under deadline

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-07

تحويل عدم اليقين إلى وكالة وديناميات الثقة - القياس ومسار الترقية

Uncertainty-to-Agency & Delegation Trust

المقاييس

- false_escalation
- unsafe_autonomy_rate
- delegation_gain
- trust_calibration
- authority_creep

الاختبارات الحاسمة

- domain shift
- adversarial confidence
- new tool
- partial observability
- time-pressure tasks

مسار التنفيذ

> Bind AutonomyEnvelope into PolicyDecision and MissionDAG scheduling.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-08

النمو القابل للعكس للاستقلالية ومناعة انحراف الهدف - الفرق البحثي

Reversible Autonomy Growth & Goal-Drift Immunity

الحالة	يمتد من
MUT-AFI proposed layer	CK-02/03/04/36/49

الفجوة

A system may become more capable while silently broadening goals or authority.

الأطروحة الهندسية

Thesis: Autonomy increases only through reversible capability gates; invariant goal contract is tested continuously against proxy drift and hidden objective substitution.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

EX-08

النمو القابل للعكس للاستقلالية ومناعة انحراف الهدف - المعمارية والعقد

Reversible Autonomy Growth & Goal-Drift Immunity

Inputs

`MissionContract, capability_maturity, policy_invariants, action_trace, outcome_metrics`

Outputs

AutonomyPromotionDecision + DriftAlerts + rollback policy

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
BOUNDED → CANARY → REVIEWED → PROMOTED → MONITORED → REVOKED
```

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-08

النمو القابل للعكس للاستقلالية ومناعة انحراف الهدف - البيانات والحالة

Reversible Autonomy Growth & Goal-Drift Immunity

Schema مرجعي

```
AutonomyPromotion { capability, scope, current_tier, proposed_tier, evidence, drift_tests[], rollback, expiry }
```

Persistence

state التشغيلي mutable عبر patches؛ append-only evidence/audit؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-08

النمو القابل للعكس للاستقلالية ومناعة انحراف الهدف - الخوارزمية والتكامل

Reversible Autonomy Growth & Goal-Drift Immunity

الخوارزمية

1. Define immutable success conditions.
2. Canary narrow scope.
3. Run adversarial proxy tests.
4. Promote one authority dimension at a time.
5. Automatic revoke on drift signature.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-08

النمو القابل للعكس للاستقلالية ومناعة انحراف الهدف - الأمن والفشل

Reversible Autonomy Growth & Goal-Drift Immunity

أنماط الفشل

- scope creep
- goal proxy gaming
- evaluation contamination
- irreversible promotion
- policy migration drift

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-08

النمو القابل للعكس للاستقلالية ومناعة انحراف الهدف - القياس ومسار الترقية

Reversible Autonomy Growth & Goal-Drift Immunity

المقاييس

- promotion_precision
- goal_drift_rate
- canary_incident_rate
- revocation_latency
- scope_fidelity

الاختبارات الحاسمة

- reward hacking scenarios
- instruction conflict
- long-horizon goal retention
- policy update regression
- rollback drill

مسار التنفيذ

> Treat autonomy as versioned deployment, not a model personality setting.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

EX-09

هندسة الدليل وموازنة الحمل الإبستيمي - الفرق البحثي

Evidence Geometry & Epistemic Load Balancing

الحالة	يمتد من
MUT-AFI proposed layer	CK-19/21/30/32

الفجوة

Evidence count and retrieval recall do not capture independence, directness, contradiction mass or the cost of verifying each claim.

الأطروحة الهندسية

Thesis: Represent evidence as a graph with independence/shared roots/directness and allocate verification compute according to decision impact and epistemic load.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

EX-09

هندسة الدليل وموازنة الحمل الإبستيمي - المعمارية والعقد

Evidence Geometry & Epistemic Load Balancing

Inputs

```
ClaimGraph, sources, provenance_roots,
contradiction_edges,
verification_budget
```

Outputs

EvidenceGeometry + VerificationAllocation + residual_risk

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
COLLECT → ROOT → GEOMETRY → ALLOCATE → VERIFY → REWEIGH
```

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-09

هندسة الدليل وموازنة الحمل الإبستيمي - البيانات والحالة

Evidence Geometry & Epistemic Load Balancing

Schema مرجعي

```
EvidenceGeometry { claim, source_nodes[], shared_roots[], support_edges[], refute_edges[], directness, freshness, load, residual_risk }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-09

هندسة الدليل وموازنة الحمل الإبستيمي - الخوارزمية والتكامل

Evidence Geometry & Epistemic Load Balancing

الخوارزمية

1. Deduplicate common upstream roots.
2. Weight direct evidence over repetitions.
3. Preserve contradictory branches.
4. Allocate court time by impact × uncertainty.
5. Stop when marginal evidence value falls.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-09

هندسة الدليل وموازنة الحمل الإبستيمي - الأمن والفشل

Evidence Geometry & Epistemic Load Balancing

أنماط الفشل

independence illusion

citation stacking

verification budget starvation

source laundering

over-retrieval

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-09

هندسة الدليل وموازنة الحمل الإبستيمي - القياس ومسار الترقية

Evidence Geometry & Epistemic Load Balancing

المقاييس

`contradiction_coverage`	`independence_adjusted_support`
`residual_risk`	`verification_value_per_cost`
	`shared_root_penalty`

الاختبارات الحاسمة

- duplicate-source attack
- conflicting evidence set
- freshness shift
- budget reduction
- blind source labels

مسار التنفيذ

> Embed in EvidencePacket Assembler and Court Bus; use as ranking features, not truth oracle.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-10

بيئة الأدوات والتعلم من أحكام المحاكم - الفرق البحثي

Tool Ecology & Court-to-Learning

الحالة	يمتد من
MUT-AFI proposed layer	CK-22/29/32/36/39

الفجوة

Tools and courts are currently registries and gates; they do not yet co-adapt systematically from failures and verdicts.

الأطروحة الهندسية

Thesis: Treat tool adapters as an ecology with capability overlap, risk, latency and failure motifs; court verdicts produce structured training candidates, not free-form reflections.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

EX-10

بيئة الأدوات والتعلم من أحكام المحاكم - المعمارية والعقد

Tool Ecology & Court-to-Learning

Inputs

`ToolRegistry, ToolReceipts, CourtVerdicts, failures, costs`

Outputs

+ ToolEcologyGraph + SkillPatchCandidates training_examples

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
OBSERVE → CLUSTER → COMPARE_TOOLS → PATCH → CANARY → PROMOTE
```

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-10

بيئة الأدوات والتعلم من أحكام المحاكم - البيانات والحالة

Tool Ecology & Court-to-Learning

Schema مرجعي

```
ToolEcology { tools[], overlap_edges[], risk_profiles[], failure_signatures[], preferred_contexts[],
retirement_candidates[] }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-10

بيئة الأدوات والتعلم من أحكام المحاكم - الخوارزمية والتكامل

Tool Ecology & Court-to-Learning

الخوارزمية

1. Cluster receipts by task/failure.
2. Compare substitute tools under equal conditions.
3. Court verdict becomes typed patch.
4. Canary adapter/skill fix.
5. Retire dominated tools.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

EX-10

بيئة الأدوات والتعلم من أحكام المحاكم - الأمن والفشل

Tool Ecology & Court-to-Learning

أنماط الفشل

- self-training on wrong verdict
- tool proliferation
- hidden rate limits
- provider lock-in
- benchmark overfitting

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-10

بيئة الأدوات والتعلم من أحكام المحاكم - القياس ومسار الترقية

Tool Ecology & Court-to-Learning

المقاييس

- `failure_recurrence`
- `tool_substitution_gain`
- `registry_entropy`
- `court_patch_precision`
- `cost_per_verified_effect`

الاختبارات الحاسمة

- tool outage
- API version drift
- provider swap
- bad verifier injection
- rate-limit stress

مسار التنفيذ

> Use offline learning pipeline first; promotion to autonomous routing only after canary evidence.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-11

ترقية الأبعاد ونقل المعرفة إلى مهارة - الفرق البحثي

Dimension Promotion & Knowledge-to-Skill Transfer

الحالة	يمتد من
MUT-AFI proposed layer	CK-17/18/22/34/36

الفجوة

Open-dimensional reasoning can generate new dimensions endlessly; the kernel needs a promotion law that requires transfer and actionable gain.

الأطروحة الهندسية

Thesis: A candidate dimension survives only if it improves prediction/decision/transfer after ablation; stable knowledge pattern becomes skill only when procedure and validation generalize.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-11

ترقية الأبعاد ونقل المعرفة إلى مهارة - المعمارية والعقد

Dimension Promotion & Knowledge-to-Skill Transfer

Inputs

`residual_patterns, candidate_dimension, task_families, skill_candidates`

Outputs

DimensionVerdict + SkillPromotionRecord

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
RESIDUAL → CANDIDATE_DIM → ABLATE → TRANSFER → SKILLIZE → PROMOTE|MERGE|REJECT
```

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-11

ترقية الأبعاد ونقل المعرفة إلى مهارة - البيانات والحالة

Dimension Promotion & Knowledge-to-Skill Transfer

Schema مرجعي

```
DimensionPromotion { dimension, operational_definition, metric, independence_test, transfer_tasks[],
decision_gain, skill_candidate? }
```

Persistence

state التشغيلي mutable عبر patches؛ append/audit-evidence؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-11

ترقية الأبعاد ونقل المعرفة إلى مهارة - الخوارزمية والتكامل

Dimension Promotion & Knowledge-to-Skill Transfer

الخوارزمية

1. Define measurable dimension.
2. Ablate against existing ontology.
3. Test second domain with renamed surface.
4. Require decision/tool effect.
5. Compile to skill only after repeated procedure.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.

EX-11

ترقية الأبعاد ونقل المعرفة إلى مهارة - الأمن والفشل

Dimension Promotion & Knowledge-to-Skill Transfer

أنماط الفشل

- renaming old feature
- dimension inflation
- skillization too early
- transfer leakage
- ontology fragmentation

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIXT Engineering Architecture 50p; MUT-AFI Part V; MUG×MOUSSAIF9B/1B Architecture; MUT-AGI-02 Part II; MOUSSAIFXT Cognitive Kernel Part I.

EX-11

ترقية الأبعاد ونقل المعرفة إلى مهارة - القياس ومسار الترقية

Dimension Promotion & Knowledge-to-Skill Transfer

المقاييس

- `cross_domain_transfer`
- `dimension_independent_gain`
- `skill_transfer_retention`
- `decision_gain`
- `ontology_growth_rate`

الاختبارات الحاسمة

- name-stripped transfer
- second-domain holdout
- ontology ablation
- skill-vs-prompt baseline
- failure-mode portability

مسار التنفيذ

> Integrate with Concept Compiler and HyperSkill Factory; new ontology node stays candidate until transfer.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C].

II

المعايير والبنية التحتية الحديثة

ربط MOUSSAIFXT بمعايير البروتوكولات والتنفيذ المتين والسياسة والرصد والمخرجات المقيدة والعزل.

EX-12

طبقة بروتوكول MCP الحديثة - الفرق البحثي

MCP 2026 Tool/Data Interoperability Layer

الحالة	يمتد من
External standard adoption	CK-29/37/38/40

الفجوة

Part I tool adapters are custom. MCP 2026-07-28 introduces a stateless, routable, cacheable protocol core better suited to horizontally scaled tool/data services.

الأطروحة الهندسية

Thesis: Adopt MCP as an interoperability edge, not as the cognitive kernel itself: MOUSSAIFXT keeps canonical policy/state while MCP standardizes external tool/resource transport.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension.

EX-12

طبقة بروتوكول MCP الحديثة - المعمارية والعقد

MCP 2026 Tool/Data Interoperability Layer

Inputs

Kernel ToolRequest, MCP server metadata, policy context, trace context

Outputs

Normalized ToolReceipt + MCP transport metadata

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
DISCOVER(optional) → REQUEST → TOOL/TASK → RESULT → NORMALIZE
```

[STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension.

EX-12

طبقة بروتوكول MCP الحديثة - البيانات والحالة

MCP 2026 Tool/Data Interoperability Layer

Schema مرجعي

```
MCPAdapter { protocol_version, server_id, method, name, auth_scope, extensions[], trace_context,
timeout, cache_ttl, task_handle? }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension.

EX-12

طبقة بروتوكول MCP الحديثة - الخوارزمية والتكامل

MCP 2026 Tool/Data Interoperability Layer

الخوارزمية

1. Use stateless requests and ordinary load balancing.
2. Route/authorize using method/name metadata.
3. Cache deterministic list responses.
4. Treat long work via Tasks extension.
5. Do not adopt deprecated Roots/Sampling/Logging for new designs.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension.

EX-12

طبقة بروتوكول MCP الحديثة - الأمن والفشل

MCP 2026 Tool/Data Interoperability Layer

أنماط الفشل

- extension sprawl
- tool server treated as trusted authority
- MCP result mistaken for world effect
- OAuth issuer mistakes
- session assumptions retained

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension.

EX-12

طبقة بروتوكول MCP الحديثة - القياس ومسار الترقية

MCP 2026 Tool/Data Interoperability Layer

المقاييس

tool_catalog_cache_hit	interop_success
adapter_latency	auth_failure_precision
	normalized_receipt_coverage

الاختبارات الحاسمة

- multi-instance load balancing
- server restart
- OAuth issuer mismatch
- task cancellation
- tool schema change

مسار التنفيذ

> Build a server-side MCP gateway behind CK-29; allow native adapters and MCP side-by-side with identical ToolContract semantics.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension.

EX-13

طبقة اتحاد الوكلاء A2A - الفرق البحثي

A2A 1.0 Agent Federation Layer

الحالة	يمتد من
External standard adoption	CK-25/26/40/44

الفجوة

Agent Mesh assumes local orchestration. A2A 1.0 offers vendor/framework-neutral discovery and task exchange without exposing internal memory/tools.

الأطروحة الهندسية

Thesis: Use A2A for remote-agent federation while keeping authority, mission state and evidence canonical in MOUSSAIFXT.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration.

EX-13

طبقة اتحاد الوكلاء A2A - المعمارية والعقد

A2A 1.0 Agent Federation Layer

Inputs

`SubtaskContract, remote AgentCard, modalities, security context`

Outputs

RemoteTaskRecord + Artifacts + normalized evidence

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
DISCOVER → NEGOTIATE → SEND → STREAM/POLL → ARTIFACT → VERIFY → CLOSE
```

[STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration.

EX-13

طبقة اتحاد الوكلاء A2A - البيانات والحالة

A2A 1.0 Agent Federation Layer

Schema مرجعي

```
A2AFederation { agent_card, protocol_version, capabilities, task_id, context_id?, artifact_refs[], push/stream config, auth, trust_tier }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration.

EX-13

طبقة اتحاد الوكلاء A2A - الخوارزمية والتكامل

A2A 1.0 Agent Federation Layer

الخوارزمية

1. Discover capability from AgentCard.
2. Send minimal subtask contract.
3. Never share canonical memory wholesale.
4. Normalize messages/artifacts into evidence objects.
5. Verify before commit.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration.

EX-13

طبقة اتحاد الوكلاء A2A - الأمن والفشل

A2A 1.0 Agent Federation Layer

أنماط الفشل

- cross-agent prompt injection
- remote capability overclaim
- task ownership ambiguity
- resource leakage
- event ordering/cancellation issues

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration.

EX-13

طبقة اتحاد الوكلاء A2A - القياس ومسار الترقية

A2A 1.0 Agent Federation Layer

المقاييس

artifact_verifiability	remote_task_success
federation_latency	handoff_loss
	remote_trust_calibration

الاختبارات الحاسمة

- mixed frameworks
- protocol version negotiation
- agent card signature/identity
- cancel/resubscribe
- malicious artifact content

مسار التنفيذ

> Expose selected MOUSSAIFXT roles as A2A agents; remote agents receive scoped views and zero direct canonical-memory write access.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration.

EX-14

طبقة التنفيذ المتين طويل الأفق - الفرق البحثي

Durable Execution Layer

الحالة	يمتد من
External production technology adoption	CK-24/27/28/30/47

الفجوة

HyperLoops and Mission DAG define recovery semantics, but production long-running execution needs deterministic persistence, retries, timers and crash recovery.

الأطروحة الهندسية

Thesis: Map long-horizon workflows to durable orchestration while preserving cognitive state outside worker processes.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-14

طبقة التنفيذ المتين طويل الأفق - المعمارية والعقد

Durable Execution Layer

Inputs

MissionDAG, activities, timers, approvals, retry policies

Outputs

DurableWorkflowRun + replayable history + compensation state

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
SCHEDULED → RUNNING → WAITING → RETRYING → COMPENSATING → COMPLETED|FAILED
```

[DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-14

طبقة التنفيذ المتين طويل الأفق - البيانات والحالة

Durable Execution Layer

Schema مرجعي

```
DurableRun { workflow_id, node/activity, durable_state_ref, retry_policy, heartbeat, signal_waits,
compensation_stack, kernel_trace_id }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-14

طبقة التنفيذ المتين طويل الأفق - الخوارزمية والتكامل

Durable Execution Layer

الخوارزمية

1. Kernel decides; workflow engine persists.
2. LLM/tool calls become failure-prone activities.
3. Use deterministic workflow code for orchestration.
4. Externalize state and resume after crash.
5. Use compensation for irreversible APIs.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-14

طبقة التنفيذ المتين طويل الأفق - الأمن والفشل

Durable Execution Layer

أنماط الفشل

- retry duplicates side effects
- non-deterministic workflow code
- version migration
- history growth
- workflow engine becomes cognitive authority

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-14

طبقة التنفيذ المتين طويل الأفق - القياس ومسار الترقية

Durable Execution Layer

المقاييس

- `duplicate_effect_rate`
- `resume_success`
- `long_run_completion`
- `mean_recovery_time`
- `compensation_success`

الاختبارات الحاسمة

- kill worker mid-action
- network flakiness
- rate-limit storm
- human approval wait
- workflow version upgrade

مسار التنفيذ

> Temporal is a strong reference implementation; keep a portability interface so other durable engines can implement the same contract.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-15

طبقة Policy-as-Code للتفويض - الفرق البحثي

Policy-as-Code Authorization Layer

الحالة	يمتد من
External standard technology adoption	CK-02/03/29/31/46

الفجوة

Constitutional/authorization logic exists conceptually. Production systems need independently testable policy engines and consistent enforcement.

الأطروحة الهندسية

Thesis: Separate Policy Decision Point from enforcement; encode action permissions using OPA/Rego or Cedar-style principal-action-resource-context decisions.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[STD/DOC] OPA and Cedar - policy-as-code engines separating policy decision from enforcement and context-aware authorization.

EX-15

طبقة Policy-as-Code للتفويض - المعمارية والعقد

Policy-as-Code Authorization Layer

Inputs

principal/agent, action, resource/tool, context, policy bundle, external attributes

Outputs

allow|deny|obligations + rationale + policy_version

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
REQUEST → EVALUATE → ALLOW|DENY|OBLIGATION → ENFORCE → AUDIT
```

[STD/DOC] OPA and Cedar - policy-as-code engines separating policy decision from enforcement and context-aware authorization.

EX-15

طبقة Policy-as-Code للتفويض - البيانات والحالة

Policy-as-Code Authorization Layer

Schema مرجعي

```
PolicyQuery { principal, action, resource, context, mission_id, tenant, risk, tool_effect_class,
policy_version }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[STD/DOC] OPA and Cedar - policy-as-code engines separating policy decision from enforcement and context-aware authorization.

EX-15

طبقة Policy-as-Code للتفويض - الخوارزمية والتكامل

Policy-as-Code Authorization Layer

الخوارزمية

1. Authenticate identity first.
2. Ask policy engine before side effect.
3. Enforcer cannot override deny.
4. Policy version is traced with action.
5. Dry-run new policies before rollout.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[STD/DOC] OPA and Cedar - policy-as-code engines separating policy decision from enforcement and context-aware authorization.

EX-15

طبقة Policy-as-Code للتفويض - الأمن والفشل

Policy-as-Code Authorization Layer

أنماط الفشل

stale external data	policy bypass
inconsistent enforcement points	broad default allow
policy complexity hides intent	

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[STD/DOC] OPA and Cedar - policy-as-code engines separating policy decision from enforcement and context-aware authorization.

EX-15

طبقة Policy-as-Code للتفويض - القياس ومسار الترقية

Policy-as-Code Authorization Layer

المقاييس

false_deny	unauthorized_allow
coverage_of_side_effects	policy_latency
policy_regression_rate	

الاختبارات الحاسمة

- privilege escalation
- tenant boundary
- policy hot reload
- stale attribute
- deny-by-default outage

مسار التنفيذ

> Choose OPA for broad policy decisions or Cedar for authorization-centric PARC schemas; CK-02 remains semantic constitution above either engine.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[STD/DOC] OPA and Cedar - policy-as-code engines separating policy decision from enforcement and context-aware authorization.

EX-16

طبقة الرصد الدلالي OpenTelemetry - الفرق البحثي

OpenTelemetry GenAI Observability Layer

الحالة	يمتد من
External standard adoption	CK-44/45/48

الفجوة

.Part I has traces but custom names make cross-provider and cross-service comparison harder

الأطروحة الهندسية

Thesis: Adopt OpenTelemetry trace context and GenAI semantic conventions where stable enough, while maintaining MOUSSAIF-specific attributes as namespaced extensions.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations.

EX-16

طبقة الرصد الدلالي OpenTelemetry - المعمارية والعقد

OpenTelemetry GenAI Observability Layer

Inputs

model/tool/agent/workflow events, trace context, tenant policy

Outputs

OTel spans/events/metrics + redacted MOUSSAIF attributes

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
START_SPAN → EVENT/CHILD_SPAN → ERROR? → END → EXPORT
```

[STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations.

EX-16

طبقة الرصد الدلالي OpenTelemetry - البيانات والحالة

OpenTelemetry GenAI Observability Layer

Schema مرجعي

```
TraceEnvelope { trace_id, span_id, gen_ai.operation.name, provider, model, tool_name, workflow,
token_usage, cache_usage, moussaif.* }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations.

EX-16

طبقة الرصد الدلالي OpenTelemetry - الخوارزمية والتكامل

OpenTelemetry GenAI Observability Layer

الخوارزمية

1. Propagate traceparent across MCP/A2A/tool calls.
2. Emit low-cardinality span names.
3. Use events for checkpoints/state changes.
4. Redact prompt/tool payloads by default.
5. Version semantic convention choice.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations.

EX-16

طبقة الرصد الدلالي OpenTelemetry - الأمن والفشل

OpenTelemetry GenAI Observability Layer

أنماط الفشل

- high-cardinality explosion
- PII leakage in telemetry
- double instrumentation
- semantic-convention churn
- trace gaps across async tasks

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations.

EX-16

طبقة الرصد الدلالي OpenTelemetry - القياس ومسار الترقية

OpenTelemetry GenAI Observability Layer

المقاييس

- `trace_join_success`
- `critical_path_trace_coverage`
- `redaction_violations`
- `telemetry_cost`
- `cross_provider_comparability`

الاختبارات الحاسمة

- cross-service trace propagation
- async task resume
- tool failure
- PII redaction
- semconv version migration

مسار التنفيذ

> Instrument kernel, gateway and workers with OTel SDKs; keep GenAI content recording opt-in and policy-controlled.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations.

EX-17

طبقة المخرجات المقيدة بالعقد - الفرق البحثي

Contract-Constrained Structured Generation

الحالة	يمتد من
External reliable technology adoption	CK-17/18/21/29

الفجوة

.Schemas exist, but free-form model output can still produce invalid contracts and fragile parsing

الأطروحة الهندسية

Thesis: Use schema-constrained generation for runtime objects, while keeping semantic validation and court checks after syntax validity.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving. Also official structured tool-schema [DOC] .APIs

EX-17

طبقة المخرجات المقيدة بالعقد - المعمارية والعقد

Contract-Constrained Structured Generation

Inputs

JSON Schema/grammar, model profile, object type, constraints

Outputs

Schema-valid typed object + semantic validation result

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
DEFINE_SCHEMA → GENERATE_CONSTRAINED → PARSE → SEMANTIC_VALIDATE → ACCEPT|RETRY|REJECT
```

vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving. Also official structured tool-schema [DOC] APIs.

EX-17

طبقة المخرجات المقيدة بالعقد - البيانات والحالة

Contract-Constrained Structured Generation

Schema مرجعي

```
StructuredGeneration { schema_id, schema_version, model, backend, strictness, generated_object,
semantic_errors[] }
```

Persistence

state التشغيلي mutable عبر patches؛ append evidence/audit-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving. Also official structured tool-schema [DOC] APIs.

EX-17

طبقة المخرجات المقيدة بالعقد - الخوارزمية والتكامل

Contract-Constrained Structured Generation

الخوارزمية

1. Generate only critical runtime objects with strict schema.
2. Reject unsupported schema features at compile time.
3. Separate syntactic validity from truth.
4. Fallback to repair once, not infinite retries.
5. Cache schema artifacts.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving. Also official structured tool-schema [DOC] APIs.

EX-17

طبقة المخرجات المقيدة بالعقد - الأمن والفشل

Contract-Constrained Structured Generation

أنماط الفشل

- overly rigid schema
- schema-valid hallucination
- reasoning/structured output interaction
- backend incompatibility
- repair loop cost

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving. Also official structured tool-schema APIs.

EX-17

طبقة المخرجات المقيدة بالعقد - القياس ومسار الترقية

Contract-Constrained Structured Generation

المقاييس

- `semantic_valid_rate`
- `schema_valid_rate`
- `parse_latency`
- `repair_rate`
- `cross_model_schema_portability`

الاختبارات الحاسمة

- malformed tool arguments
- enum/range constraints
- nested objects
- provider swap
- adversarial text injection into JSON fields

مسار التنفيذ

> For local vLLM use structured_outputs JSON/grammar; for compatible APIs use strict structured function/tool schemas.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving. Also official structured tool-schema APIs.

EX-18

طبقة الفصل بين Harness والحوسبة - الفرق البحثي

Sandboxed Harness / Compute Separation

الحالة	يمتد من
External production pattern adoption	CK-28/31/38/40

الفجوة

Part I defines sandboxing, but long-horizon agents benefit when orchestration state is external to ephemeral compute.

الأطروحة الهندسية

Thesis: Keep cognitive/control state in trusted control plane and run model-generated code/actions inside disposable sandboxes with scoped credentials and snapshots.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[DOC] OpenAI Agents SDK, Apr 2026 - sandbox-aware harness, memory, MCP, externalized state, snapshot/rehydration and compute separation.

EX-18

طبقة الفصل بين Harness والحوسبة - المعمارية والعقد

Sandboxed Harness / Compute Separation

Inputs

`ExecutionPlan, files, dependencies, credential scopes, sandbox policy`

Outputs

SandboxRunReceipt + file diffs + snapshot + evidence

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
PROVISION → HYDRATE → EXECUTE → INSPECT → SNAPSHOT → DESTROY|REHYDRATE
```

[DOC] OpenAI Agents SDK, Apr 2026 - sandbox-aware harness, memory, MCP, externalized state, snapshot/rehydration and compute separation.

EX-18

طبقة الفصل بين Harness والحوسبة - البيانات والحالة

Sandboxed Harness / Compute Separation

Schema مرجعي

```
SandboxRun { image, filesystem_mounts, network_policy, credential_handles, command_log, artifact_refs, snapshot_ref, trace_id }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[DOC] OpenAI Agents SDK, Apr 2026 - sandbox-aware harness, memory, MCP, externalized state, snapshot/rehydration and compute separation.

EX-18

طبقة الفصل بين Harness والحوسبة - الخوارزمية والتكامل

Sandboxed Harness / Compute Separation

الخوارزمية

1. Never place long-lived secrets in sandbox.
2. Mount only task-scoped files.
3. Use egress allow-lists where possible.
4. Persist state outside container.
5. Destroy compute after evidence capture.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[DOC] OpenAI Agents SDK, Apr 2026 - sandbox-aware harness, memory, MCP, externalized state, snapshot/rehydration and compute separation.

EX-18

طبقة الفصل بين Harness والحوسبة - الأمن والفشل

Sandboxed Harness / Compute Separation

أنماط الفشل

- sandbox escape
- secret exfiltration
- persistent hidden state
- dependency supply-chain
- artifact tampering

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

قاعدة الهدم: إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[DOC] OpenAI Agents SDK, Apr 2026 - sandbox-aware harness, memory, MCP, externalized state, snapshot/rehydration and compute separation.

EX-18

طبقة الفصل بين Harness والحوسبة - القياس ومسار الترقية

Sandboxed Harness / Compute Separation

المقاييس

- `sandbox_escape_tests`
- `credential_exposure`
- `artifact_integrity`
- `rehydration_success`
- `compute_isolation_cost`

الاختبارات الحاسمة

- prompt injection requests secret
- malicious package
- container crash
- network denial
- snapshot restore

مسار التنفيذ

> Reference current agent harnesses with native sandbox support, but preserve provider-neutral SandboxAdapter in MOUSSAIFXT.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[DOC] OpenAI Agents SDK, Apr 2026 - sandbox-aware harness, memory, MCP, externalized state, snapshot/rehydration and compute separation.

III

الذاكرة والاستدلال والأمن والتقييم

دمج نتائج البحث الحديثة في الذاكرة والحوسبة وقت الاختبار وأمن الوكلاء والتقييم طويل الأفق وServing.

EX-19

طبقة إدارة الذاكرة المتعلمة - الفرق البحثي

Adaptive Memory Governance Layer

الحالة	يمتد من
External research integration	CK-19/20/35/36

الفجوة

Part I memory lifecycle is rule-based. Research increasingly treats memory operations as agent actions and dynamic graphs.

الأطروحة الهندسية

Thesis: Keep canonical governance deterministic, but allow a learned MemoryPolicy to propose store/retrieve/update/summarize/discard actions under court and rollback constraints.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

A-MEM (2025) - dynamically linked and evolving agentic memory. [PPR] Agentic Memory / AgeMem (2026) - learned unified LTM/STM memory operations as [PPR] tool actions.

EX-19

طبقة إدارة الذاكرة المتعلمة - المعمارية والعقد

Adaptive Memory Governance Layer

Inputs

`working_context, episodic_memory, semantic_graph, task_state, memory_budget`

Outputs

MemoryActionProposal + retrieved bundle + promotion candidates

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
OBSERVE → PROPOSE_MEMORY_ACTION → POLICY_GATE → APPLY → EVALUATE → RETAIN|REVERT
```

A-MEM (2025) - dynamically linked and evolving agentic memory. [PPR] Agentic Memory / AgeMem (2026) - learned unified LTM/STM memory operations as [PPR] tool actions.

EX-19

طبقة إدارة الذاكرة المتعلمة - البيانات والحالة

Adaptive Memory Governance Layer

Schema مرجعي

```
MemoryAction { op:store|retrieve|update|summarize|discard|link, targets[], rationale, budget,
expected_gain, reversible }
```

Persistence

state التشغيلي mutable عبر patches؛ append-only evidence/audit؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[PPR] A-MEM (2025) - dynamically linked and evolving agentic memory. [PPR] Agentic Memory / AgeMem (2026) - learned unified LTM/STM memory operations as tool actions.

EX-19

طبقة إدارة الذاكرة المتعلمة - الخوارزمية والتكامل

Adaptive Memory Governance Layer

الخوارزمية

1. Learn proposals, not unrestricted canonical writes.
2. Dynamic linking may evolve representations.
3. Separate short-term and long-term objectives.
4. Evaluate memory effect on held-out tasks.
5. Quarantine destructive operations.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

A-MEM (2025) - dynamically linked and evolving agentic memory. [PPR] Agentic Memory / AgeMem (2026) - learned unified LTM/STM memory operations as [PPR] tool actions.

EX-19

طبقة إدارة الذاكرة المتعلمة - الأمن والفشل

Adaptive Memory Governance Layer

أنماط الفشل

- memory bloat
- self-poisoning
- feedback loop bias
- catastrophic discard
- retrieval echo chamber

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

قاعدة الهدم: إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

A-MEM (2025) - dynamically linked and evolving agentic memory. [PPR] Agentic Memory / AgeMem (2026) - learned unified LTM/STM memory operations as [PPR] tool actions.

EX-19

طبقة إدارة الذاكرة المتعلمة - القياس ومسار الترقية

Adaptive Memory Governance Layer

المقاييس

context_efficiency	task_gain_per_memory_op
memory_entropy	retrieval_precision
	harmful_write_rate

الاختبارات الحاسمة

- long-horizon hidden facts
- contradictory updates
- memory pressure
- domain shift
- adversarial memory injection

مسار التنفيذ

> Pilot A-MEM/AgeMem-style policies in shadow mode; policy proposes MemoryPatch, deterministic governance decides commit.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

A-MEM (2025) - dynamically linked and evolving agentic memory. [PPR] Agentic Memory / AgeMem (2026) - learned unified LTM/STM memory operations as [PPR] tool actions.

EX-20

طبقة الحوسبة التكيفية وقت الاختبار - الفرق البحثي

Adaptive Test-Time Compute Layer

الحالة	يمتد من
External research integration	CK-05/07/08/12/15

الفجوة

Part I has sparse routing but not a reproducible inference-protocol taxonomy for sequential reflection, parallel rollouts and verifier merging.

الأطروحة الهندسية

Thesis: Make inference-time compute an explicit budgeted policy with protocol identity, stop rule and exact accounting.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

Scaling Test-time Compute for LLM Agents (2025) - diversified rollouts, reflection timing and verifier/merge choices matter. [PPR] Test-Time Scaling in Reasoning [PPR] LLMs (2026) - protocol-matched compute accounting and reproducibility requirements.

EX-20

طبقة الحوسبة التكيفية وقت الاختبار - المعمارية والعقد

Adaptive Test-Time Compute Layer

Inputs

`QuestionObject, uncertainty, candidate_protocols, token/time/tool budget, verifier availability`

Outputs

InferenceProtocol + rollout set + merge/verifier record

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
SELECT_PROTOCOL → DELIBERATE/SAMPLE/SEARCH → VERIFY → MERGE → STOP
```

Scaling Test-time Compute for LLM Agents (2025) - diversified rollouts, reflection timing and verifier/merge choices matter. [PPR] Test-Time Scaling in Reasoning [PPR] LLMs (2026) - protocol-matched compute accounting and reproducibility requirements.

EX-20

طبقة الحوسبة التكيفية وقت الاختبار - البيانات والحالة

Adaptive Test-Time Compute Layer

Schema مرجعي

```
InferenceBudget { regime, max_tokens, max_rollouts, reflection_trigger, verifier, merge_method,
diversity_target, stop_rule, seed_info }
```

Persistence

state التشغيلي mutable عبر patches؛ append -evidence/audit only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

Scaling Test-time Compute for LLM Agents (2025) - diversified rollouts, reflection timing and verifier/merge choices matter. [PPR] Test-Time Scaling in Reasoning [PPR] .LLMs (2026) - protocol-matched compute accounting and reproducibility requirements

EX-20

طبقة الحوسبة التكيفية وقت الاختبار - الخوارزمية والتكامل

Adaptive Test-Time Compute Layer

الخوارزمية

1. Choose sequential vs leaf-level vs prefix-search explicitly.
2. Reflect only on triggers, not always.
3. Diversify rollouts structurally.
4. Verifier and merge method are part of evaluated system.
5. Report actual compute and uncertainty.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

Scaling Test-time Compute for LLM Agents (2025) - diversified rollouts, reflection timing and verifier/merge choices matter. [PPR] Test-Time Scaling in Reasoning [PPR] LLMs (2026) - protocol-matched compute accounting and reproducibility requirements.

EX-20

طبقة الحوسبة التكيفية وقت الاختبار - الأمن والفشل

Adaptive Test-Time Compute Layer

أنماط الفشل

self-consistency without diversity

unbounded reasoning

compute confounding benchmark

verifier bias

non-reproducible protocol

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

قاعدة الهدم: إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

Scaling Test-time Compute for LLM Agents (2025) - diversified rollouts, reflection timing and verifier/merge choices matter. [PPR] Test-Time Scaling in Reasoning [PPR] LLMs (2026) - protocol-matched compute accounting and reproducibility requirements.

EX-20

طبقة الحوسبة التكيفية وقت الاختبار - القياس ومسار الترقية

Adaptive Test-Time Compute Layer

المقاييس

protocol_regret	verified_gain_per_compute
reflection_precision	diversity_gain
	reproducibility_gap

الاختبارات الحاسمة

- fixed-budget comparisons
- reflection ablation
- verifier swap
- rollout diversity control
- exact/distributional replay

مسار التنفيذ

> Extend Vascular Compute scheduler with protocol registry and frozen inference-budget manifests.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

Scaling Test-time Compute for LLM Agents (2025) - diversified rollouts, reflection timing and verifier/merge choices matter. [PPR] Test-Time Scaling in Reasoning [PPR] LLMs (2026) - protocol-matched compute accounting and reproducibility requirements.

EX-21

طبقة أمان الأدوات ومقاومة Prompt Injection - الفرق البحثي

Tool Security & Prompt-Injection Resistance

الحالة	يمتد من
External security research integration	CK-02/29/31/32/46

الفجوة

Retrieved/tool data can contain instructions that hijack agents; Part I states retrieved text is data, not authority, but needs an executable security test plane.

الأطروحة الهندسية

Thesis: Label trust zones, separate data from instructions, require policy mediation for tool effects, and continuously test against adaptive prompt-injection scenarios.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[BENCH/SEC] AgentDojo - prompt-injection evaluation over tool-using agents and untrusted data.

EX-21

طبقة أمان الأدوات ومقاومة Prompt Injection - المعمارية والعقد

Tool Security & Prompt-Injection Resistance

Inputs

`tool output, retrieved content, trust labels, proposed action, policy context`

Outputs

SanitizedEvidence + SecurityVerdict + blocked_instructions + action obligations

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
INGEST_UNTRUSTED → CLASSIFY → SANITIZE → REASON → POLICY_GATE → EXECUTE|BLOCK
```

[BENCH/SEC] AgentDojo - prompt-injection evaluation over tool-using agents and untrusted data.

EX-21

طبقة أمان الأدوات ومقاومة Prompt Injection - البيانات والحالة

Tool Security & Prompt-Injection Resistance

Schema مرجعي

```
UntrustedContent { source, trust_zone, content_ref, embedded_instruction_detected, requested_effects[], sanitization, security_case_id }
```

Persistence

state التشغيلي mutable عبر patches؛ append-only evidence/audit؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[BENCH/SEC] AgentDojo - prompt-injection evaluation over tool-using agents and untrusted data.

EX-21

طبقة أمان الأدوات ومقاومة Prompt Injection - الخوارزمية والتكامل

Tool Security & Prompt-Injection Resistance

الخوارزمية

1. Tool/retrieval content never grants authority.
2. Keep hidden/system policy outside untrusted text channel.
3. Use allow-listed tool capabilities.
4. Require explicit user/owner intent for sensitive exfiltration.
5. Run adversarial benchmark in CI.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[BENCH/SEC] AgentDojo - prompt-injection evaluation over tool-using agents and untrusted data.

EX-21

طبقة أمان الأدوات ومقاومة Prompt Injection - الأمن والفشل

Tool Security & Prompt-Injection Resistance

أنماط الفشل

- data exfiltration
- indirect prompt injection
- instruction smuggling in files
- trusted-source spoofing
- defense breaks task utility

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[BENCH/SEC] AgentDojo - prompt-injection evaluation over tool-using agents and untrusted data.

EX-21

طبقة أمان الأدوات ومقاومة Prompt Injection - القياس ومسار الترقية

Tool Security & Prompt-Injection Resistance

المقاييس

benign_task_success	attack_success_rate
false_block	exfiltration_rate
	security_regression

الاختبارات الحاسمة

- AgentDojo-style tasks
- malicious email/document
- tool description poisoning
- cross-tenant secret
- multi-step injection

مسار التنفيذ

> Build a security court + CI benchmark; do not rely on one prompt-based defense.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[BENCH/SEC] AgentDojo - prompt-injection evaluation over tool-using agents and untrusted data.

EX-22

طبقة تقييم التفاعل طويل الأفق - الفرق البحثي

Long-Horizon Agent Evaluation Layer

الحالة	يمتد من
External benchmark integration	CK-33/45/48/49

الفجوة

Short unit tests cannot expose hidden-state loss, policy inconsistency, user interaction failures or skipped verification over hundreds of actions.

الأطروحة الهندسية

Thesis: Adopt state-based and long-horizon benchmarks with repeated-trial reliability, end-state verification and safety audits.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[τ-bench - dynamic tool-agent-user interaction, end-state validation and pass^k reliability. [BENCH] OSWorld 2.0 (2026) - long-horizon professional computer use workflows with stateful artifacts and safety auditing.

EX-22

طبقة تقييم التفاعل طويل الأفق - المعمارية والعقد

Long-Horizon Agent Evaluation Layer

Inputs

benchmark task, initial state, policy rules, run budget, seeds, agent config

Outputs

EvalRecord + end-state diff + pass^k/partial score + failure taxonomy

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
SETUP → RUN → INTERACT → END_STATE → VERIFY → CLASSIFY_FAILURE → REPLAY
```

[BENCH] τ-bench - dynamic tool-agent-user interaction, end-state validation and pass^k reliability. [BENCH] OSWorld 2.0 (2026) - long-horizon professional computer use workflows with stateful artifacts and safety auditing.

EX-22

طبقة تقييم التفاعل طويل الأفق - البيانات والحالة

Long-Horizon Agent Evaluation Layer

Schema مرجعي

```
LongHorizonEval { benchmark, task_id, initial_state_hash, agent_manifest, budget, repetitions,
end_state_diff, safety_events, trace_ref }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

τ-bench - dynamic tool-agent-user interaction, end-state validation and pass^k reliability. [BENCH] OSWorld 2.0 (2026) - long-horizon professional computer [BENCH] use workflows with stateful artifacts and safety auditing.

EX-22

طبقة تقييم التفاعل طويل الأفق - الخوارزمية والتكامل

Long-Horizon Agent Evaluation Layer

الخوارزمية

1. Score state, not prose.
2. Repeat trials for reliability.
3. Preserve agent/config version.
4. Classify ask/verify/constraint-tracking failures.
5. Record safety separately from completion.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[BENCH] OSWorld 2.0 (2026) - long-horizon professional computer use workflows with stateful artifacts and safety auditing. [BENCH] τ-bench - dynamic tool-agent-user interaction, end-state validation and pass^k reliability.

EX-22

طبقة تقييم التفاعل طويل الأفق - الأمن والفشل

Long-Horizon Agent Evaluation Layer

أنماط الفشل

- dynamic environment drift
- benchmark leakage
- single-run cherry picking
- cost mismatch
- partial success hidden

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

-τ-bench - dynamic tool-agent-user interaction, end-state validation and pass^k reliability. [BENCH] OSWorld 2.0 (2026) - long-horizon professional computer [BENCH] use workflows with stateful artifacts and safety auditing.

EX-22

طبقة تقييم التفاعل طويل الأفق - القياس ومسار الترقية

Long-Horizon Agent Evaluation Layer

المقاييس

verified_completion	pass^k
constraint_retention	partial_progress
safety_incident_rate	

الاختبارات الحاسمة

- τ-bench-style policy interaction
- OSWorld2.0 long workflows
- internal mission benchmark
- replay after model update
- cross-provider matched budget

مسار التنفيذ

> Create an Eval Harness with adapters for τ-bench/OSWorld-like and internal world-effect tasks; promotion requires repeated trials.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

-τ-bench - dynamic tool-agent-user interaction, end-state validation and pass^k reliability. [BENCH] OSWorld 2.0 (2026) - long-horizon professional computer [BENCH] use workflows with stateful artifacts and safety auditing.

EX-23

نسيج الاستدلال المحلي عبر vLLM - الفرق البحثي

vLLM Inference Fabric Layer

الحالة	يمتد من
External production technology adoption	CK-13/15/41/44

الفجوة

Part I has a vLLM gateway but not a 2026 deployment profile spanning structured outputs, caching, speculative decoding, MoE parallelism and disaggregated serving.

الأطروحة الهندسية

Thesis: Treat inference serving as a resource fabric with capability telemetry; kernel routing optimizes verified-value/latency/cost, not GPU knobs directly.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving.

EX-23

نسيج الاستدلال المحلي عبر vLLM - المعمارية والعقد

vLLM Inference Fabric Layer

Inputs

model profile, request shape, prefix hash, latency/throughput target, hardware topology

Outputs

InferenceRoute + engine config + cache/speculation metadata

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
PROFILE → PLACE → SERVE → OBSERVE → REBALANCE → DEGRADED/FALLBACK
```

[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving.

EX-23

نسيج الاستدلال المحلي عبر vLLM - البيانات والحالة

vLLM Inference Fabric Layer

Schema مرجعي

```
InferenceProfile { model, tp, pp, dp, ep, prefix_cache, structured_backend, speculation, disaggregation, kv_state, queue_depth }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving.

EX-23

نسيج الاستدلال المحلي عبر vLLM - الخوارزمية والتكامل

vLLM Inference Fabric Layer

الخوارزمية

1. Enable prefix caching for repeated canonical prefixes.
2. Use cryptographic hash in multi-tenant risk-sensitive cache.
3. Use structured outputs for runtime objects.
4. Speculative decode for latency-bound low/medium QPS.
5. Use TP/PP/DP/EP/disaggregation based on model/hardware.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving.

EX-23

نسيج الاستدلال المحلي عبر vLLM - الأمن والفشل

vLLM Inference Fabric Layer

أنماط الفشل

- bad parallelism choice
- cache collision/leak
- KV transfer bottleneck
- speculator mismatch
- kernel conflates serving with capability

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

> **قاعدة الهدم:** إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving.

EX-23

نسيج الاستدلال المحلي عبر vLLM - القياس ومسار الترقية

vLLM Inference Fabric Layer

المقاييس

itl	ttft
cache_hit	tokens_per_second
verified_value_per_gpu_second	queue_delay

الاختبارات الحاسمة

- same task across serving configs
- cache isolation
- MoE load skew
- disaggregated prefill/decode
- failover to alternate model

مسار التنفيذ

> Expose vLLM metrics to CK-12 router; keep inference config in provider adapter, not in cognitive theory objects.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[DOC] vLLM 2026 stable docs - structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP and disaggregated serving.

IV

معمارية v2 المرجعية

تجميع كل الامتدادات في topology واحدة قابلة للتنفيذ من دون framework lock-in.

EX-24

معمارية MOUSSAIFXT v2 المرجعية - الفرق البحثي

MOUSSAIFXT v2 Reference Integration Architecture

الحالة	يمتد من
Reference synthesis	All CK + EX layers

الفجوة

The new research/standards are useful only if they integrate without replacing the canonical 12-layer stack or creating framework lock-in.

الأطروحة الهندسية

Thesis: Keep L0-L11 canonical and add eight orthogonal extension planes: Interop, Durable Execution, Policy, Telemetry, Adaptive Memory, Adaptive Inference, Semantic Homeostasis, Capability Certification.

عدم التكرار: هذا المحور لا يعيد المكوّن الأصلي؛ يضيف transition أو plane أو evidence gate لم يكن كاملاً في Part I.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p. [STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension. [STD/DOC] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration. [DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations. [DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-24

معمارية MOUSSAIFXT v2 المرجعية - المعمارية والعقد

MOUSSAIFXT v2 Reference Integration Architecture

Inputs

All canonical contracts + external adapters + eval evidence

Outputs

Deployable v2 topology + extension manifests + maturity ledger

حدود الطبقة

- لا تتجاوز Constitution/Authority.
- لا تملك canonical memory مباشرة إلا عبر MemoryPatch.
- كل side effect يحتاج receipt + postcondition readback.
- كل external technology adapter قابل للاستبدال.

آلة الحالة

```
BOOT → DISCOVER → CONTRACT → COGNITIVE_ROUTE → DURABLE_EXECUTE → FEDERATE → VERIFY → TRACE → MEMORY → CERTIFY
```

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p. [STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension. [STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration. [STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations. [DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-24

معمارية MOUSSAIFXT v2 المرجعية - البيانات والحالة

MOUSSAIFXT v2 Reference Integration Architecture

Schema مرجعي

```
XTv2Manifest { kernel_version, extension_planes[], protocol_versions, policy_engine, trace_semconv,
durable_engine, inference_fabric, eval_profiles, promotion_policy }
```

Persistence

state التشغيلي mutable عبر patches؛ evidence/audit append-only؛ snapshots للتسريع لا لاستبدال الأصل.

Versioning

schema_version + component_version + source/protocol version. أي migration تنتج سجل فرق واختبار رجوع.

Invariant: syntax validity لا تساوي semantic validity، والـadapter لا يرفع رتبة الادعاء.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p. [STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension. [STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration. [STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations. [DOC] Temporal Durable Execution workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-24

معمارية MOUSSAIFXT v2 المرجعية - الخوارزمية والتكامل

MOUSSAIFXT v2 Reference Integration Architecture

الخوارزمية

1. Canonical state remains provider-neutral.
2. MCP/A2A are edges, not authority.
3. Durable engine persists workflow, not meaning.
4. Policy engine enforces decisions, constitution defines semantics.
5. Telemetry and eval produce evidence for promotion.

نقاط التكامل

Plane	العقد
KernelState	read canonical / return patch
Policy	authorize before effect
Telemetry	one trace across model/tool/agent
Memory	commit only verified delta
Eval	versioned config + frozen budget

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p. [STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension. [STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration. [STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations. [DOC] Temporal Durable Execution workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-24

معمارية MOUSSAIFXT v2 المرجعية - الأمن والفشل

MOUSSAIFXT v2 Reference Integration Architecture

أنماط الفشل

framework lock-in

extension explosion

protocol version drift

duplicated state across engines

false sense of maturity

الحراسة

- Fail closed عند غموض الصلاحية.
- لا learning من self-report بلا verification.
- لا retry تلقائي mutation إلا مع idempotency/compensation.
- لا telemetry حساس افتراضياً.

قاعدة الهدم: إذا أضاف المحور latency/complexity من دون مكسب متكرر في verified effect أو safety أو recovery أو portability، يجب تبسيطه أو عزله.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p. [STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension. [STD] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration. [STD/DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations. [DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

EX-24

معمارية MOUSSAIFXT v2 المرجعية - القياس ومسار الترقية

MOUSSAIFXT v2 Reference Integration Architecture

المقاييس

cross_host_fidelity	end_to_end_verified_completion
recovery_success	protocol_portability
E3_to_E4_promotion_precision	security_regression

الاختبارات الحاسمة

- local→cloud failover
- MCP tool + A2A remote agent in one mission
- policy change mid-run
- sandbox crash
- model swap
- long-horizon eval

مسار التنفيذ

> Build v2 as extension manifests around Part I kernel, not a rewrite. Start with one minimal complete slice: mission → MCP tool → durable workflow → world-effect readback → memory → trace → eval.

E2	schema/adapter موجود ومختبر.
E3	تشغيل replayable مع state/effect evidence.
E4	تحسن مضبوط على baseline وميزانية مساوية.
E5	تكرار عبر provider/domain/environment مستقل.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p. [STD] Model Context Protocol specification 2026-07-28 - stateless core, MRTR, header routing, cacheable lists, authorization hardening, Tasks extension. [STD/DOC] Agent2Agent Protocol v1.0.x - AgentCard discovery, task/message/artifact model, interoperable remote-agent collaboration. [DOC] OpenTelemetry Semantic Conventions 1.43.0 + GenAI conventions - common trace/event/metric vocabulary for model/tool/agent/workflow operations. [DOC] Temporal Durable Execution - workflows persist state and activities retry/recover; suited to long-running stateful agents.

A

الملاحق التنفيذية

عقود وخرائط عملية لتحويل Part II إلى backlog بناء واختبار، لا إلى توسع نظري فقط.

ملحق A - معمارية XT v2 المختصرة

```
User / System Goal
  ↓
MOUSSAIFXT Canonical Kernel (L0-L11)
  ├─ X0 Interop Gateway: MCP / A2A
  ├─ X1 Durable Execution Adapter
  ├─ X2 Policy Decision Point
  ├─ X3 OpenTelemetry / Eval Bus
  ├─ X4 Adaptive Memory Policy (governed)
  ├─ X5 Test-Time Compute + vLLM Fabric
  ├─ X6 Semantic Homeostasis
  └─ X7 Capability Certification
        ↓
Authorized Tool / Remote Agent / Sandbox
        ↓
Observed World Effect
        ↓
Court → MemoryPatch → Eval → Promotion
```

> كل extension plane يملك adapter interface؛ kernel لا يعتمد مباشرة على vendor SDK.

.MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p [C]

ملحق B - بروتوكولات ومنافذ الإصدار المقترحة

سطح	Protocol / Pattern	ملاحظة
Tool/data edge	MCP 2026-07-28	stateless; extension/version pinned
Remote agent edge	A2A 1.0.x	scoped subtask, normalized artifacts
Tracing	W3C Trace Context + OpenTelemetry	content redacted by default
Policy	OPA/Rego or Cedar	decision separate from enforcement
Durability	Temporal adapter interface	workflow code not cognitive authority
Local inference	OpenAI-compatible vLLM edge	structured outputs + cache-aware routing

ملحق C - Events الجديدة

- interop.mcp.requested
- interop.mcp.completed
- federation.a2a.task_created
- federation.a2a.artifact_received
- durable.workflow_checkpoint
- durable.workflow_resumed
- policy.decision
- policy.obligation
- semantic.roundtrip_checked
- causal.ablation_completed
- memory.action_proposed
- memory.action_committed
- inference.protocol_selected
- inference.rollout_completed
- security.injection_detected
- eval.long_horizon_completed
- capability.promotion_candidate
- capability.promotion_revoked

جميع الأحداث تحمل trace_id, mission_id, state_id, version, source/protocol version, actor/tenant and evidence references.

ملحق D - Threat Model 2026

التهديد	السطح	الضابط
Indirect prompt injection	MCP/A2A/retrieval	data≠authority + policy gate + security court
Tool replay side effects	durable retries	idempotency + compensation
Cross-tenant cache leak	prefix/KV cache	secure hashing + tenant isolation
Remote agent overclaim	A2A AgentCard	local trust tier + verify artifact
Telemetry PII leak	OTel	redaction/content-off by default
Memory self-poisoning	learned memory	proposal-only + quarantine + rollback
Goal drift	long-horizon autonomy	immutable mission invariants + revocation

E - Benchmark Stack ملحق

طبقة	Benchmark style	المقياس
Tool/user policy	τ-bench-like	pass^k + end-state correctness
Computer use	OSWorld2.0-like	binary completion + partial progress + safety
Prompt injection	AgentDojo-like	attack success + benign utility
Memory	long-horizon hidden fact / context pressure	task gain + memory efficiency
Routing	matched-budget provider/model set	router regret + verified gain
Durability	fault injection	resume / duplicate effect / recovery
World Effect	stateful external environment	false-done + causal attribution

ملحق F - خريطة الإصدار المرحلية

1. **R2.0:** CanonicalState v2 + protocol/version registry.
2. **R2.1:** MCP gateway + OTel traces + structured outputs.
3. **R2.2:** Policy engine + sandbox execution + durable workflow adapter.
4. **R2.3:** A2A federation + semantic homeostasis.
5. **R2.4:** adaptive test-time compute + vLLM telemetry routing.
6. **R2.5:** governed adaptive memory shadow mode.
7. **R2.6:** causal credit + capability closure + promotion ledger.
8. **R2.7:** long-horizon benchmark suite + E4 candidates.

> لا تنتقل إلى R2.5+ قبل أن يكون replayable، policy-enforced، observable، durable minimal slice، وله World-Effect readback.

ملحق G - Definition of Done للنواة v2

```
XTv2Ready = CanonicalStatePass ∧ PolicyPass ∧ DurablePass ∧ InteropPass ∧ TracePass ∧ EffectPass
∧ RecoveryPass ∧ EvalPass
```

- Kernel restart يحافظ على المهمة دون فقد الدليل أو الصلاحية.
- MCP tool وA2A agent يعملان عبر نفس Tool/Artifact normalization.
- كل mutation ذو idempotency أو compensation.
- trace end-to-end متصل عبر model→tool/agent→effect.
- policy version قابل لإعادة بناء كل قرار.
- model/provider swap لا يكسر mission invariants.
- security regression suite يمر.
- long-horizon eval repeated، لا single demo.

ملحق H - ما لم يُثبت بعد

غير معاير

الأوزان الدقيقة لـMCD/QMCD وthresholds autonomy/promotion.

غير مثبت

أن هذه الطبقات مجتمعة تحقق AGI.

يحتاج E4/E5

مكاسب مقارنة متكررة عبر models/domains/environments.

قيد البحث

learned memory, test-time compute policy, causal credit.

الهدف التالي: تحويل هذا Part II إلى implementation backlog وbenchmark ledger، لا إنتاج Part III نظري جديد قبل النتائج.

ملحق I - قائمة المصادر الخارجية المختارة

1. MCP Specification 2026-07-28.
2. Agent2Agent Protocol v1.0.x / official specification and SDK releases.
3. OpenTelemetry Semantic Conventions 1.43.0 and GenAI conventions.
4. OpenAI Agents SDK - sandbox-aware harness update, Apr 2026.
5. Temporal Durable Execution documentation.
6. Open Policy Agent and Cedar policy-language documentation.
7. vLLM stable docs: structured outputs, prefix caching, speculative decoding, DP/TP/PP/EP, disaggregated serving.
8. A-MEM: Agentic Memory for LLM Agents, 2025.
9. Agentic Memory / AgeMem, 2026.
10. Scaling Test-time Compute for LLM Agents, 2025.
11. Test-Time Scaling in Reasoning LLMs, 2026.
12. τ-bench, 2024.
13. AgentDojo, 2024.
14. OSWorld 2.0, 2026.

Status differs: standards/docs vs preprints/benchmarks. The book labels adoption accordingly.

الخاتمة - اتجاه التمديد بعد Part II

بعد هذا الامتداد، MOUSSAIFXT لا يحتاج طبقات نظرية أكثر بقدر ما يحتاج **تشغيلاً مقارناً**. البنية الجديدة تجعل آخر MUT/MUG/MUT-AFI قابلة للربط مع المعايير الحديثة من دون أن تسلم لها السلطة أو تربط النواة بمزود واحد.

```
MUT/MUG Theory → Canonical Contract → Extension Plane → Durable / Policy-Enforced Execution → Observable Effect → Benchmark Evidence → Promotion
```

المعيار العملي للمرحلة التالية: تشغيل نفس missions عبر Local/SaaS/Hybrid، مع OTel trace، durable execution، policy engine، MCP/A2A edges، vLLM/cloud provider swaps، ثم مقارنة reliability/cost/safety/recovery تحت budgets متساوية.

[C] MOUSSAIFXT Cognitive Kernel Part I; MUT-AGI-02 Part II; MUG×MOUSSAIF9B/1B Architecture; MUT-AFI Part V; MOUSSAIXT Engineering Architecture 50p.