PART IV · FINAL REDO · MAX-DENSE SOURCE-GROUNDED RECONSTRUCTION

الجزء الرابع
إغلاق الفجوات وإعادة البناء البحثي

كل ما بقي مفقوداً أو ضعيف التمثيل أو متعارضاً بين النظرية والهندسة وتقارير vNEXT
Beta Reconstruction Modules 36 - بلا تكرار Canonical

Arabic RTL · Amiri · A4 · 36 dense pages
-Source basis: Architecture 50p, Gap Audit 50p, Local Review 55p, Global/Cumulative/Supreme vNEXT reports, MSRAK 43p, SWE-as-OS 50p, Moussaif1B 29p, MUG 20p, All
.Kernels 86p, checkpoint/weekly chronology

.Dense reconstruction, not verbatim duplication. Source claims are preserved as claims; independent evidence remains a separate grade

01 | منهج الإغلاق وعدم الاختلاق

المصدر أولاً 982 retention 0.994 ،E3-E5 ،31,189 asset ،23 P0 ،component، وأرقام benchmark تُحفظ كما تذكرها تقارير vNEXT.

لكن المراجعة المحلية تقول إن PDF summary لا يعادل raw benchmark أو فحصاً مستقلاً. لذلك كل Claim يحمل ReportedGrade وIndependentGrade منفصلين.

Compression invariant

```
Preserve(identity,contract,causality,authority,failure,metric,falsifier,lineage); remove redundancy
```

ممنوع

- اختلاق أسماء P0 الناقصة
- محو snapshot تاريخي لصالح رقم أحدث
- تحويل MOUSSAIF9B إلى parameter claim
- اعتبار tool receipt world effect
- اعتبار byte physical dedupe semantic dedupe-0

حالة Beta

BR-01..36 ليست نظريات canonical جديدة؛ هي reconstruction objects. إذا لم تنجُ من ablation تدمج/تبسط/تحذف.

Reconciliation Matrix | 02

المحور	قديم/مواز	vNEXT	الحكم
Assets	weekly 18,200→26,681 ؛26,681	31,189	snapshots
Skills	300	420	base+expansion lineage
Abilities	base 43	93	versioned view
Loops	44/40	140	versioned view
Agents	20/34	84	identity graph
Protocols	conceptual counts 20/1500	active 120	different unit classes
P0	gap audit: many undeployed	modules 100% tested reported 23	reproduction + manifest mapping
Evidence	spec-heavy/evidence-light	later E3-E5 canonical reports	preserve both; raw replay decides independent grade

```
Count(date,type,grade,state)=QUERY(CanonicalIdentityGraph)
```

03 | أحدث الفجوات بعد Part I-III

aliases/versions/counts/semantic duplication **Identity**

mission owner, leases, fencing, events, sagas **Distributed**

ambiguous network outcomes, causal attribution **World Effect**

context compaction, cross-model homeostasis **Semantics**

five theories + AT01-18 need controlled harness **SWE-as-OS**

grade inheritance, raw benchmark reproduction, claim scope **Evidence**

dataset provenance, forgetting vector, shadow/canary **Learning**

observed/inferred/proposed/remembered reconciliation **World State**

MCD/QMCD identifiability, question value, route regret **Reasoning**

five theories not operationally integrated **MSRAK**

private/local evidence needs external falsification **Replication**

Architecture Lineage - what Part IV touches | 04

سلالة	الجزء الذي يستعاد
L0-L11 / CK50	لا يعاد، Part IV يصلح evidence/distributed/learning edges فقط
EX24 / X0-X7	لا يعاد، يستخدم interop/durable/policy/telemetry/adaptive planes
PC40 / P0-P5	لا يعاد، يضيف forensic/reconciliation layer فوق proof closure
MUT/MUG	vessel placement, runtime state, promotion/causal organization
MUT-AFI	causal credit, failure topology, uncertainty, semantic homeostasis, evidence geometry
MSRAK	reasoning-to-agency five-theory family
SWE-as-OS	code-as-world-effect five-theory family + 18 tools
vNEXT reports	P0, 5-kernel, 498MB, registries/ledgers, LoRA, benchmarks

Evidence / Identity / Learning laws | 05

```
SemanticFP=H(purpose|trigger|I/O|tools|state|failures|verifier|benchmark)
Trace→Quarantine→Train→FrozenEval→ForgettingVector→Shadow→Canary→Court→Promote|Rollback
```

```
ComponentState=<DeploymentState,EvidenceGrade,Scope,Environment,Time,ArtifactRefs>
Entity=<canonical_id,aliases,version,supersedes,artifact_hash,semantic_fp>
```

الهدف: منع inflation في counts، ومنع grade laundering، ومنع self-training من أن يصبح self-promotion.

Distributed / World Effect laws | 06

```
EffectivelyOnce=intent+idempotency+version_precondition+receipt+readback+dedupe+compensation
```

causal attribution، cause لا يثبت after-state؛ AMBIGUOUS قد يبقى timeout after commit؛ consistency لا يساوي File mirroring يحتاج competing causes/interventions/counterfactuals.

```
Lease<mission,owner,epoch,fencing_token,expiry>
```

```
Observed≠Inferred≠Proposed≠Remembered
```

Reasoning laws + Weak-Theory Recovery | 07

MSRAK Beta

.MSRACT · MSTAC · MCSDT · MGWEPT · MRCMT

SWE Beta

.GEET · DPCT · PVT · TCCT · ECSA + AT01-AT18

```
MCD=λ1D+λ2I+λ3T+λ4A+λ5P+λ6C-λ7N ; QMCD=Q(Q)×MCD
```

```
RouterRegret=U(best admissible)-U(selected)
```

.marginal decision gain ≼ cost عندما تتوقف question policy ،calibration/held-out/ablation تحتاج λ/Ω

BR-01..BR-02 | وحدتا إعادة بناء

Claim-Evidence Reconciliation Court

BR-01 - محكمة مصالحة الادعاء والدليل

Gap

المصادر الحديثة تستخدم أحياناً CANONICAL/E4/E5 و«operational 100%» بينما المراجعة المحلية المحافظة تنبه إلى أن جدولاً داخل PDF ليس raw benchmark ولا فحصاً مباشراً للآلة. الخطر هو انتقال رتبة دليل تقريرية إلى كل مكوّن بالوراثة.

Beta reconstruction

إعادة البناء تجعل رتبة الدليل خاصية للادعاء المفرد لا للكتاب أو الحزمة. كل Claim يرتبط بـ EvidenceArtifact قابل لإعادة التشغيل، وScope وTime وEnvironment وBaseline وVerifier مستقل. عند غياب artifact الخام يحتفظ النظام بوسم REPORTED_E4 بدلاً من VERIFIED_E4.

Contract

```
ClaimRecord{id,statement,scope,source,reported_grade,independent_grade,artifact_refs,baseline_ref,environment_hash,verifier,valid_from,valid_to,contradictions,status}
```

Algorithm

INGEST report claim → normalize → locate raw evidence → reproduce when possible → compare scope → court verdict {SUPPORTED,DOWNGRADED,CONTRADICTED,UNRESOLVED} → write lineage; never overwrite original claim.

Metrics

`claim_reproduction_rate · grade_downgrade_rate · scope_mismatch_rate · independent_verifier_coverage · evidence_artifact_completeness`

Failures

- grade inheritance
- same-source self-verification
- scope inflation
- time leakage
- summary table treated as raw evidence

Critical experiment / ablation

H0: report-level grade predicts component validity. Baseline=inherit grade; intervention=claim-local court. Reproduce 30 claims. Accept BR-01 if false-positive canonicalization falls while verified claims remain recoverable.

Integration: L0/L10/P2/P5; precedes promotion, memory writeback and external publication.

Maturity: BETA E1-E2; reported runtime claims remain source-attributed until independent replay.

Sources: S3 · S4 · S5 · S6

Canonical Identity Graph & Count Reconciliation

BR-02 - رسم الهوية المعيارية ومصالحة العدادات

Gap

الأعداد التاريخية تختلف: 26,681 ثم 31,189 assets؛ 300 skills/44 loops ثم 420/93/140/84/120 KB 100؛ قديمة مقابل KBOS 40 حالية؛ كما تتغير مؤشرات convergence. جمعها في رقم واحد يمحو السلالة.

Beta reconstruction

تُحوَّل كل وحدة إلى Entity لها canonical_id وaliases وversion lineage وvalidity interval وartifact hash وsemantic fingerprint. العدادات تصبح Views مؤرخة مشتقة من graph بدلاً من حقائق ثابتة مكتوبة يدوياً.

Contract

```
Entity{id,type,canonical_id,aliases,version,parent,supersedes,artifact_hash,semantic_fp,first_seen,last_seen,deployment_state,evidence_grade}
```

Algorithm

CRAWL → hash → semantic fingerprint → candidate identity edges → exact/semantic duplicate court → version/supersession edge → materialized count views by date and evidence threshold.

Metrics

`identity_collision_rate · unresolved_aliases · count_reproducibility · semantic_duplicate_ratio · lineage_coverage`

Failures

- double counting
- renamed component inflation
- stale totals
- fork counted as new capability
- artifact vs runtime object confusion

Critical experiment / ablation

Freeze two historical snapshots; ask independent script to recompute counts from graph. Success requires deterministic totals and explicit explanation of every delta >1%.

Integration: L3/L11/PC-36; controls registries for KB/Skill/Ability/Loop/Agent/Protocol.

Maturity: E1 otherwise BETA E2 if graph artifacts exist.

Sources: S3 · S4 · S5 · S10

BR-03..BR-04 | وحدتا إعادة بناء

Semantic Zero-Duplication Governance

BR-03 - حوكمة عدم التكرار الدلالي

Gap

Hardlinks/junctions can prove zero extra physical bytes yet still leave ten differently named files implementing the same behavior. Physical deduplication therefore does not imply cognitive/semantic deduplication.

Beta reconstruction

يبني SemanticFingerprint من purpose+trigger+I/O+state transition+tool set+failure signature+verification+benchmark family. إذا تقاربت البصمتان، تفتح المحكمة قرار MERGE/ALIAS/SPECIALIZE/KEEP_DIVERSE ولا تحذف الأصل تلقائياً.

Contract

```
SemanticFP = H(purpose || trigger || io_schema || tool_caps || state_machine || failure_modes || verifier || benchmark_family)
```

Algorithm

Candidate pairs from embeddings+symbolic keys → behavior-contract comparison → replay on common cases → equivalence verdict → alias/supersedes graph → count view refresh.

Metrics

`semantic_dup_ratio` · `merge_precision` · `behavioral_equivalence_pass` · `registry_size_reduction` · `capability_coverage_delta`

Failures

- false merge
- name-only dedupe
- functionally identical padding
- loss of specialized edge cases
- history destruction

Critical experiment / ablation

Ablate 50 candidate duplicates. Merge only if output/effect/failure behavior remains within tolerance across held-out cases; preserve specialized variants when counterexample exists.

Integration: L3/L5-L8/L11; extends physical zero-dup into capability identity.

Maturity: BETA E1-E2; requires behavioral replay for E3.

Sources: S3 · S4 · S5

P0 Substrate Completion Ledger

BR-04 - دفتر إكمال ركيزة P0

Gap

التقارير المختصرة تسمي p01-p17 وp22 بوضوح، لكنها لا تعرض في الجداول المتاحة تعريفاً صريحاً لـ p18-p21 وp23. اختراع أسماء لسد التسلسل سيكسر مبدأ provenance.

Beta reconstruction

يعتمد Part IV حالة UNMAPPED_IN_SOURCE لهذه المعرفات حتى يظهر manifest أو كود مباشر، لكل pXX يجب حفظ path/hash/interface/test harness/result/version/dependencies، الرقم وحده لا يكفي.

Contract

```
P0Module{id,name|null,path|null,hash|null,interface,test_ref,result,dependencies,source_anchor,status}; status∈{MAPPED,UNMAPPED,RETIRED,ALIAS}
```

Algorithm

Parse manifests/code tree → reconcile pXX IDs → verify importability/CLI → run unit+negative tests → bind evidence artifact → publish ledger; unresolved IDs stay unresolved.

Metrics

`p0_mapping_coverage` · `hash_bound_test_rate` · `orphan_module_count` · `dependency_resolution_rate` · `replay_pass_rate`

Failures

- sequence completion by guess
- module renumbering drift
- test result without code hash
- same module counted twice
- retired module shown active

Critical experiment / ablation

Require 23/23 identities from authoritative manifests. A missing source does not fail the runtime; it fails the documentation closure. Exit only when each ID is source-bound or explicitly retired.

Integration: P0/P1/P2; supports all CK/EX/PC implementation claims.

Maturity: BETA documentation closure; no invented p18-p21/p23.

Sources: S3 · S4 · S5 · S6

BR-05..BR-06 | وحدتا إعادة بناء

Benchmark Reproduction Capsule

BR-05 - كبسولة إعادة إنتاج الBenchmark

Gap

النتائج المعلنة تعرض gains مهمة لكنها لا تكفي علمياً بلا environment fingerprint، confidence intervals، scoring code، raw traces، budgets، baseline checkpoint، seeds، task manifest، وnegative regressions.

Beta reconstruction

كل رقم أداء يصبح BenchmarkCapsule immutable. الكبسولة قابلة لتشغيل command واحد يعيد baseline/intervention ويولد raw JSONL ثم تقريراً مشتقّاً. يمنع تعديل dataset أو evaluator بعد ظهور النتائج دون version fork.

Contract

```
BenchmarkCapsule{id,hypothesis,task_manifest,split_hash,baseline,intervention,model_hash,tool_versions,budget,seeds,scorer,raw_trace_uri,ci,fail_threshold,security_gates}
```

Algorithm

PRE-REGISTER → freeze artifacts → run matched seeds → collect trace/effect → score → CI/bootstrap → critical-regression gate → sign manifest → independent rerun.

Metrics

`reproducibility_delta` · `CI_width` · `cost_adjusted_gain` · `critical_regression_count` · `seed_variance` · `raw_artifact_completeness`

Failures

- benchmark leakage
- budget mismatch
- cherry-picked seeds
- evaluator drift
- success-only reporting
- missing failure cases

Critical experiment / ablation

Rebuild the four headline benchmarks from current reports. E4 only if raw rerun matches effect direction under frozen budget; E5 only with independent model/domain/environment reproduction.

Integration: P4/P5; is the hard bridge between architecture and science.

Maturity: BETA E2 schema; existing headline gains are REPORTED until capsule replay.

Sources: S3 · S4 · S6 · S2

Claim-Specific Maturity & No-E5-Inheritance

BR-06 - النضج الخاص بكل ادعاء ومنع وراثة E5

Gap

وسم «E3-E5 canonical» على حزمة مختلطة قد يخفي أن بعض المكونات مجرد schema وبعضها runtime وبعضها benchmark. رتبة واحدة للمجلد تسبب overclaim.

Beta reconstruction

نستخدم محورين مستقلين: DeploymentState وEvidenceGrade. قد يكون المكوّن EXECUTABLE/E3 لكنه بلا E4؛ وقد تكون نتيجة E4 مرتبطة بمقياس واحد لا بكل خصائصه. الترقية تعتمد claim_scope وليس component_name.

Contract

```
Maturity = (deployment_state, claim_evidence_grade, scope, last_verified_at, environment_set, replication_count)
```

Algorithm

For each component enumerate claims → grade each claim → compute conservative display grade=min(required evidence over advertised claims) → keep per-claim detail accessible.

Metrics

`claims_per_component` · `grade_entropy` · `overclaim_rate` · `stale_grade_rate` · `independent_replications`

Failures

- grade laundering
- package inheritance
- metric-to-capability inflation
- replication in same environment counted independent
- stale E5

Critical experiment / ablation

Compare package-level grading vs claim-level grading on 100 components; human/audit court measures precision of scientific statements and false-E4/E5 rate.

Integration: Cross-cutting L0/L10/L11; replaces broad maturity badges in registries.

Maturity: BETA E1; can be implemented immediately.

Sources: S2 · S3 · S4

BR-07..BR-08 | وحدتا إعادة بناء

Continual Learning Three-Plane Architecture

BR-07 - معمارية التعلم المستمر ثلاثية الطائرات

Gap

خط LoRA الدوري المعلن يخلق خطراً إن جمع التدريب والتقييم والترقية في daemon واحد، Retention=0.994 لا يثبت وحده غياب catastrophic forgetting، ولا يبرر promotion آلياً.

Beta reconstruction

تفصل Inference Plane عن Training Plane عن Evaluation/Promotion Plane. الإنتاج لا يرى checkpoint جديداً قبل approval policy، canary، safety/regression batteries، shadow evaluation، rollback pointer. Cron يستطيع جمع المرشحين وتشغيل التجارب لكنه لا يمنح canonical authority.

Contract

```
TrainCandidate{base_hash,data_manifest,objective,adapter_hash,metrics,safety,forgetting,provenance,shadow_result,canary_result,promotion_decision,rollback_ref}
```

Algorithm

collect verified traces → quarantine → dedupe/contamination scan → train isolated adapter → frozen eval → red-team → shadow traffic → canary radius → court → promote/rollback.

Metrics

`forgetting_curve` · `heldout_transfer` · `safety_delta` · `calibration_delta` · `promotion_regret` · `rollback_time`

Failures

- self-poisoning
- training on failed traces
- benchmark contamination
- evaluation on train distribution
- automatic weight promotion
- irreversible drift

Critical experiment / ablation

A/B three regimes: no update, naive hourly LoRA, governed three-plane LoRA. Hold total train compute. Measure task gain, forgetting, safety, calibration and recovery over 20 sequential updates.

Integration: L11/X4/P5; strengthens reported continuous-vLLM architecture.

Maturity: BETA E1-E2; current continual-learning claims remain reported until raw run artifacts.

Sources: S3 · S4 · S5 · S6

Training Data Provenance, Contamination & Privacy

BR-08 - مصدرية بيانات التدريب والتلوث والخصوصية

Gap

تحويل verified skills/claims إلى JSONL لا يضمن أن البيانات مرخصة، غير مكررة، غير متسربة من benchmarks، أو خالية من secrets/PII. Hash corpus وحده لا يعطي data lineage.

Beta reconstruction

كل TrainingExample يحمل verifier verdict، derivation chain، benchmark_membership، sensitivity، license/use class، source lineage، retention policy. أي مثال مشتق من external user data لا يمر إلى weights دون authorization/legal basis.

Contract

```
TrainExample{id,source_uri,source_hash,license,sensitivity,benchmark_tags,derivation,verifier_status,consent_basis,expiry,weight,use_policy}
```

Algorithm

harvest → classify → secret/PII scan → dedupe → benchmark-overlap scan → provenance bind → train/eval split by lineage → immutable data manifest.

Metrics

`provenance_coverage` · `benchmark_overlap` · `duplicate_rate` · `secret_detection_recall` · `licensed_fraction` · `lineage_split_leakage`

Failures

- benchmark leakage
- PII memorization
- license violation
- duplicate amplification
- self-generated error reinforcement

Critical experiment / ablation

Canary dataset contains intentionally duplicated/benchmark/secret-like items. Pipeline must quarantine with measured recall/precision before any adapter training.

Integration: L3/L11/P2; required for scientifically interpretable continual learning.

Maturity: BETA E1; security scanner claims are source-reported, training governance still needs closure.

Sources: S3 · S4 · S5

BR-09..BR-10 | وحدتا إعادة بناء

Retention & Forgetting Battery

BR-09 - بطارية الاحتفاظ والنسيان

Gap

رقم retention أحادي لا يكشف tool-use regression، calibration drift، selective forgetting أو safety loss. قد يبقى متوسط الأداء 99% بينما تنهار فئة نادرة حرجة.

Beta reconstruction

البطارية تعيّس long-horizon recovery، tool schemas، policy compliance، calibration، worst-slice retention، backward transfer وrare invariants عبر checkpoints متسلسلة. تحفظ baseline snapshots وتحدد rollback threshold قبل التدريب.

Contract

```
ForgettingVector = [Δcore,Δrare,Δsafety,Δcalibration,Δtool,Δmemory,Δlong_horizon]; Gate = max_critical_drop <= τc AND weighted_drop <= τw
```

Algorithm

checkpoint_n → evaluate fixed core + evolving holdout + adversarial slices → compare to n-1 and frozen base → detect localized regressions → reject or bisect data/update.

Metrics

`worst_slice_drop` · `backward_transfer` · `ECE_delta` · `policy_violation_delta` · `tool_contract_regression` · `cumulative_drift`

Failures

- average hides tail failure
- test set saturation
- metric gaming
- unobserved tool schema drift
- slow cumulative degradation

Critical experiment / ablation

Inject one update optimized for a narrow domain; acceptance requires target gain without critical-slice regression beyond preregistered threshold.

Integration: L11/P5; converts “zero forgetting” from slogan into vector-valued test.

Maturity: BETA E1-E2.

Sources: S3 · S4 · S6

Safe Model/Adapter Promotion

BR-10 - ترقية النماذج والمحولات عبر Shadow/Canary

Gap

ترقية checkpoint فور انتهاء LoRA أو اعتماد مزود جديد قد تغير safety، cost، latency، semantics وtool behavior. rollback must be executable, not a path note.

Beta reconstruction

يُعامل model/adapter كـReleaseArtifact. الترقية: Candidate→Shadow→Canary→Verified→Promoted؛ policy تحدد radius حسب risk tier. كل mission تسجل model_hash وadapter_hash لمنع trace ambiguity.

Contract

```
Release{candidate_hash,base_hash,adapter_hash,eval_capsule,shadow_window,canary_fraction,SLOs,safety_gates,promotion_verdict,rollback_target}
```

Algorithm

build → sign → shadow mirror → compare counterfactual decisions → canary low-risk traffic → SLO/semantic guard → promotion court → atomic pointer switch; auto rollback on hard gate.

Metrics

`shadow_disagreement` · `canary_error_delta` · `rollback_RTO` · `semantic_fidelity` · `cost_delta` · `policy_violation_delta`

Failures

- silent semantic shift
- rollback points to incompatible state
- canary contamination
- provider version drift
- unbounded blast radius

Critical experiment / ablation

Release a deliberately perturbed adapter. System must detect disagreement and rollback before high-risk actions; measure detection latency and escaped effects.

Integration: L8/L9/L11/X7; binds inference routing to release engineering.

Maturity: BETA E1-E2.

Sources: S3 · S4 · S5

BR-11..BR-12 | وحدتا إعادة بناء

5-Kernel Mission Ownership, Leases & Fencing

BR-11 - ملكية المهمة وLeases وFencing في شبكة الخمس نوى

Gap

تزامن الملفات بين Hermes/OpenCode/Codex/Claude/Antigravity لا يحدد من يملك commit النهائي، من يحق له retry، أو كيف تمنع وكيلين من تنفيذ نفس side effect.

Beta reconstruction

لكل Mission owner واحد في كل لحظة عبر lease يرقم fencing_token تصاعدي. النوى الأخرى Workers/Verifiers؛ أي write أو effect يحمل token ويُرفض إذا كان أقدم من handoff .current lease ينتج event صريح.

Contract

```
MissionLease{mission_id,owner_kernel,epoch,fencing_token,acquired_at,expires_at,capabilities,authority_ceiling,handoff_reason}
```

Algorithm

.claim lease → dispatch subnodes → heartbeat → worker artifact → independent verify → owner commit; on expiry elect/assign successor with higher fencing token; stale owner writes rejected

Metrics

`duplicate_effect_rate · stale_write_rejects · handoff_latency · owner_availability · lease_conflict_rate`

Failures

- split brain
- double execution
- stale worker commit
- owner/verifier conflation
- lease expiry during non-idempotent action

Critical experiment / ablation

.Kill active owner mid-mission and delay its network. New owner must finish without duplicate world effect; stale resumed owner must be fenced

Integration: L8/L9/P1; turns 5-kernel "sync" into distributed execution semantics.

Maturity: BETA E1; direct current mesh status is report-sourced.

Sources: S3 · S4 · S5

Event-Driven Kernel Sync & Single-Writer Canonical State

BR-12 - التزامن الحدثي والحالة المعيارية أحادية الكاتب

Gap

.File mirroring/pointer topology preserves access but not transactional ordering. Simultaneous edits to canonical state can race even with identical corpus visibility

Beta reconstruction

التزامن ينتقل إلى append-only EventLog مع revision وpointers/hashes تحمل evidence_refs. Canonical State Service applies compare-and-swap; kernels maintain materialized views/caches. Large artifacts remain content-addressed; events carry .expected_version

Contract

```
Event{event_id,mission_id,entity_id,prev_version,new_version,actor_kernel,operation,payload_hash,evidence_refs,fencing_token,ts}
```

Algorithm

.worker emits event proposal → owner validates lease/policy → CAS append → state reducer → projections notify peers → conflict enters court, never last-write-wins silently

Metrics

`event_lag · CAS_conflict_rate · state_reconstruction_hash_match · projection_staleness · lost_update_rate`

Failures

- last-write-wins loss
- clock skew ordering
- file sync without causality
- cache stale authority
- hardlink mutation ambiguity

Critical experiment / ablation

.Generate conflicting updates from two kernels under controlled delay; expected result: one commit, one explicit conflict, deterministic replay

Integration: L3/L8/L11/P1; event sourcing is state substrate, not replacement for evidence.

Maturity: BETA E1-E2.

Sources: S3 · S4 · S5

BR-13..BR-14 | وحدتا إعادة بناء

Cross-Kernel Failure Isolation & Sagas

BR-13 - عزل فشل النوى وSagas للتراجع

Gap

.Distributed mission DAG can partially succeed across repositories/tools. Global rollback may be impossible; retrying everything can duplicate effects

Beta reconstruction

.DAG node defines effect class, compensation, idempotency scope and failure containment zone. Saga orchestrator records forward/compensating actions; high-risk irreversible nodes are delayed until prerequisite evidence and approvals close كل

Contract

```
SagaStep{id,preconditions,action,effect_class,idempotency_key,receipt,postcondition,compensation,compensation_limit,status}
```

Algorithm

.execute reversible nodes → verify → checkpoint → proceed; on failure compute compensation frontier → reverse safe steps → mark irrecoverable residual effects → escalate

Metrics

`compensation_success · residual_effect_count · MTTR · duplicate_side_effects · failure_containment_radius`

Failures

- compensation causes new damage •
- rollback assumes inverse exists •
- retry storm •
- partial cross-repo state •
- hidden external side effect •

Critical experiment / ablation

.Inject failure after 60% DAG completion with one external mutation; compare naive restart vs saga recovery on correctness and duplicated effects

Integration: .L8/L9/L10; concretizes recovery beyond local filesystem rollback

Maturity: .BETA E1-E2

Sources: S1 · S3 · S8

Canonical World Model Store, Freshness & CDC

BR-14 - المخزن المعياري لنموذج العالم والحداثة

Gap

.Drive/local/connectors/runtime observations. Without freshness, memory may outrank current reality بين fragmentation أشار إلى gap audit كان غنياً مفاهيمياً لكن Project World Model

Beta reconstruction

.provenance. CDC events update only affected entities; object/blob pointers keep large artifacts external وOBSERVED/INFERRED/PROPOSED، source precedence، valid_time، observed_time، TTL/event freshness بثلاثة مستويات EntityState تعرّف

Contract

```
WorldEntity{id,type,attributes,state_class,source,source_rank,observed_at,valid_at,ttl,version,provenance,conflicts,sensitivity}
```

Algorithm

.read authoritative source → normalize → compare source precedence → append observation event → reconcile conflicts → materialize current state → invalidate dependent plans when freshness expires

Metrics

`stale_state_detection · reconciliation_accuracy · freshness_violation_rate · reconstruction_after_restart · source_conflict_latency`

Failures

- remembered state treated current •
- connector lag •
- source conflict •
- TTL too long •
- stale plan after external change •

Critical experiment / ablation

.Simulate two hosts reporting conflicting deployment version plus a newer authoritative readback. Correct state must win deterministically and invalidate stale DAG nodes

Integration: .L3/L4/P1; direct closure of Gap 16

Maturity: .BETA E1-E2; source reports later claim p14 Persistent World Model, independent state-store replay still needed

Sources: S2 · S3 · S1

BR-15..BR-16 | وحدتا إعادة بناء

Observed/Proposed/Remembered State Reconciliation

BR-15 - مصالحة المرصود والمقترح والمتذكر

Gap

.Observed state > remembered/generated state" is a central invariant, but a real runtime needs typed merge semantics when observation is partial, noisy or stale"

Beta reconstruction

لا تتم المقارنة بنصوص كاملة. StateFacet لكل field يحمل .source_rank/confidence/freshness/sensitivity. PROPOSED never overwrites OBSERVED; MEMORY is evidence candidate only. Contradictions create OpenDifference and may block action

Contract

```
FacetDecision = argmax source_authority × freshness × directness × verification - conflict_penalty; ties -> unresolved, not averaged truth.
```

Algorithm

.collect facets → classify origin → validate timestamp/schema → rank → conflict graph → ask/reobserve high-impact facets → commit current view with unresolved set

Failures

- false freshness
- partial observation overwrites complete state
- model prior smuggled as observation
- averaging mutually exclusive states
- silent conflict

Metrics

`facet_conflict_precision` · `reobservation_yield` · `unresolved_high_risk_count` · `state_grounding_error` · `source_rank_violations`

Critical experiment / ablation

.Create tasks where memory says service=v1 while authoritative host returns v2 and retrieved doc says v3 planned. System must represent current=v2, proposed=v3, historical=v1

Integration: L1/L3/L10; foundation for honest planning and effect verification.

Maturity: BETA E1-E2.

Sources: S1 · S3 · S4

Effectively-Once World-Effect Protocol

BR-16 - بروتوكول أثر العالم «فعلياً مرة واحدة»

Gap

.Part III "Exactly-Once Effect Transaction" is too strong for arbitrary external systems: network failures can leave ambiguous effects and dual writes cannot generally be made atomic across independent resources

Beta reconstruction

النسخة Beta تعتمد EFFECTIVELY_ONCE حيث أمكن: durable intent/outbox، idempotency key، version precondition، provider receipt، dedupe ledger، authoritative readback، compensation. إذا غابت idempotency/readback يبقى /status=ATTEMPTED ولا يعلن نجاحاً. AMBIGUOUS

Contract

```
EffectTxn{id,intent_hash,idempotency_key,precondition_version,outbox_seq,attempts,provider_receipts,readback,verified_delta,dedupe_status,compensation,status}
```

Algorithm

.persist intent+local state atomically → relay at-least-once → consumer/provider idempotency → record receipt → authoritative readback → compare postcondition → VERIFIED | AMBIGUOUS | COMPENSATED

Failures

- lost event
- duplicate send
- phantom success
- network timeout after commit
- non-idempotent provider
- compensation unavailable

Metrics

`duplicate_effect_rate` · `ambiguous_effect_rate` · `readback_coverage` · `compensation_success` · `recovery_latency`

Critical experiment / ablation

.Inject crash before/after remote commit and duplicate relay. Protocol passes only if no false VERIFIED; duplicates are suppressed where supported and ambiguity is surfaced otherwise

Integration: L9/L10/P3; replaces universal exactly-once wording with implementable semantics.

Maturity: BETA E1; aligned with transactional-outbox/idempotent-consumer practice.

Sources: S1 · S3 · S8

BR-17..BR-18 | وحدتا إعادة بناء

Causal World-Effect Attribution

BR-17 - الإسناد السببي لأثر العالم

Gap

قراءة postcondition تثبت أن الحالة تغيرت، لكنها لا تثبت أن فعل الوكيل سبب التغير؛ concurrent actors وexternal dynamics can confound attribution.

Beta reconstruction

يبني النظام EffectGraph: intended action, mediators, competing causes, before/after observations, counterfactual/holdout evidence. درجات الإسناد تختلف حسب القدرة على التدخل: direct transactional effect > instrumented sequence > quasi-experimental association.

Contract

```
AttributionScore = Directness × TemporalSpecificity × InterventionStrength × AlternativeCausePenalty^-1 × Reproducibility
```

Algorithm

define causal question → log concurrent events → select control/counterfactual when safe → perform intervention → readback → compare alternative explanations → court assigns attribution grade.

Metrics

`attribution_precision` · `alternative_cause_coverage` · `counterfactual_success` · `causal_grade_distribution` · `false_credit_rate`

Failures

- post hoc attribution
- shared infrastructure confound
- external actor race
- selection bias
- self-generated evidence

Critical experiment / ablation

Benchmark same observed success with/without a concurrent external updater. A verifier based only on after-state should fail; causal layer should downgrade attribution.

Integration: L10; operationalizes MUT-AFI causal difference credit and World-Effect rigor.

Maturity: BETA E1.

Sources: S7 · S12 · S3

Evidence Geometry & Source Independence

BR-18 - هندسة الدليل واستقلال الجذور

Gap

عد citations لا يقيس قوة الدليل؛ عشر صفحات قد تعود إلى source root واحد. كما أن retrieved summaries قد تعيد صياغة نفس التقرير وتبدو مستقلة.

Beta reconstruction

EvidenceGraph represents source roots, derivation edges, directness, freshness, scope and contradiction. Shared-root penalty يمنع تضخيم الثقة. Evidence packet يوضح coverage of claim dimensions لا volume.

Contract

```
E = Σ w_i·Independence_i·Coverage_i·Directness_i·Freshness_i - λ·SharedRootMass - μ·ContradictionMass
```

Algorithm

normalize evidence → resolve provenance roots → cluster dependent sources → map claim dimensions → compute geometry → flag missing independent axis → request new source/experiment.

Metrics

`root_independence` · `claim_dimension_coverage` · `echo_ratio` · `contradiction_recall` · `evidence_efficiency`

Failures

- citation echo
- derived source treated independent
- old source refreshed by copy
- coverage illusion
- contradictions omitted

Critical experiment / ablation

Construct packets with 5 independent roots vs 20 derivative citations; ranking should favor independent coverage when other factors equal.

Integration: L4/L10/P2; integrates court/evidence ledger with source structure.

Maturity: BETA E1.

Sources: S1 · S3 · S7

BR-19..BR-20 | وحدتا إعادة بناء

Verifier Independence & Correlated Failure

BR-19 - استقلال المدقق وفشل الارتباط

Gap

Planner/executor/verifier may share same model family, prompt assumptions, retrieval context or tool output, producing correlated hallucination while roles look separate.

Beta reconstruction

VerifierIndependenceProfile records model/provider, prompt lineage, retrieval roots, tool path, evaluator code and organization. Risk-tier policy demands diversity proportional to impact. Independence is not binary; it is measured and disclosed.

Contract

```
Independence = 1 - CorrFailure(planner,verifier | task_family); proxy factors: model_diversity, evidence_root_diversity, tool_path_diversity, evaluator_separation.
```

Algorithm

generate claim/action → choose verifier under independence constraints → blind verifier to unnecessary planner rationale → use authoritative data → compare failure correlation → escalate when independence below threshold.

Failures

- same-model self-confirmation
- shared poisoned retrieval
- verifier sees persuasive rationale
- same buggy parser
- human rubber stamp

Metrics

`failure_correlation` · `independence_score` · `blind_verifier_delta` · `disagreement_resolution_rate` · `false_accept_rate`

Critical experiment / ablation

Adversarial set induces one systematic planner error. Compare same-context verifier vs independent tool/model verifier; preregister false-accept reduction.

Integration: L8/L10/P2; refines role councils into measurable independence.

Maturity: BETA E1.

Sources: S1 · S3 · S7

MCD/QMCD Calibration & Identifiability

BR-20 - معايرة MCD/QMCD وقابلية التعرف

Gap

MCD = λ-weighted rubric and QMCD=Ω(Q)×MCD are useful routing heuristics but weights/variables may be correlated or circular. Without calibration, better performance may come from extra compute rather than score quality.

Beta reconstruction

يُعامل MCD/QMCD كـ predictive decision model. Variables receive observable proxies and preregistered coding rules. λ learned/fit only on train tasks; calibration evaluated on held-out. Compare against cheap baselines: uncertainty, task length, tool count, entropy.

Contract

```
MCD = λ1D+λ2I+λ3T+λ4A+λ5P+λ6C-λ7N; QMCD=Ω(Q)·MCD; identify λ under regularization + monotonic constraints where justified.
```

Algorithm

label tasks with outcomes/cost → compute features without future leakage → fit/calibrate → test discrimination/calibration → route matched budgets → ablate features → inspect coefficient stability.

Failures

- feature leakage
- multicollinearity
- Ω circularity
- post-outcome scoring
- scale mismatch across domains

Metrics

`Brier/ECE` · `AUROC/PR` · `router_regret` · `coefficient_stability` · `feature_ablation_delta` · `cost_adjusted_value`

Critical experiment / ablation

Compare fixed λ, learned λ, uncertainty-only, random routing under same compute. Promote only if held-out router regret decreases with stable calibration and no risk regression.

Integration: L2/X5/P5; converts key theory equation into falsifiable instrument.

Maturity: BETA E1; any reported E4 routing gain remains report-bound until capsule reproduction.

Sources: S1 · S2 · S7 · S9

BR-21..BR-22 | وحدتا إعادة بناء

Question Value, Adaptive Compute & Stop Policy

BR-21 - قيمة السؤال والحوسبة التكيفية وسياسة التوقف

Gap

.Governing Question can improve routing, but questioning itself costs time/tokens and may delay action. A single “best question” may be unstable as evidence arrives

Beta reconstruction

.QuestionPolicy estimates Expected Value of Information/Decision Change subject to risk and latency. It can ask user, retrieve, simulate, or stop. Stop when marginal expected decision gain < cost or when action threshold is satisfied with safeguards

Contract

```
QValue(q)=P(change_decision|answer)·ΔUtility + RiskReduction + ReuseValue - TokenCost - LatencyCost - InteractionCost; Stop if max_q QValue(q) ≤ 0.
```

Algorithm

.generate candidate questions → predict outcome partitions → estimate decision impact → choose question/action → update state → track counterfactual regret from questions not asked

Metrics

```
question_regret · decision_change_rate · info_gain_per_token · unnecessary_question_rate · missed_clarification_rate
```

Failures

- question loops
- high-information low-actionability question
- confirmation bias
- user burden ignored
- asking after irreversible action

Critical experiment / ablation

.Benchmark ambiguous tasks with fixed one-shot vs unlimited questioning vs value-gated questioning; match overall compute and measure success, latency, user turns, unsafe assumptions

Integration: .L1/L2/X5; connects QMCD to explicit stopping economics

Maturity: .BETA E1

Sources: S1 · S7 · S12

Router Regret & Safe Online Routing Adaptation

BR-22 - ندم الموجّه والتكيف الآمن للتوجيه

Gap

.Router accuracy is not just selecting a high-quality model; it must choose model/skill/tool/agent under cost, latency, context, authority and risk. Online adaptation can exploit users/tasks unsafely

Beta reconstruction

route decision logs candidate set and predicted utility. Counterfactual evaluation uses shadow candidates where affordable; off-policy estimators are used cautiously. Online changes are constrained by safety floor and canary, not reward maximization كل .alone

Contract

```
RouteUtility = VerifiedValue - αCost - βLatency - γRisk - δContextLoss; Regret = U(best observed admissible) - U(selected).
```

Algorithm

.candidate filter by authority/capability → score → choose → log alternatives → occasional shadow evaluation → update calibrator → canary new policy → rollback on safety/SLO regression

Metrics

```
router_regret · cost_per_verified_effect · latency_tail · unsafe_route_rate · calibration_drift · shadow_disagreement
```

Failures

- reward hacking
- cheap model overselected
- risk hidden in average utility
- feedback loop bias
- candidate capability stale

Critical experiment / ablation

.Cross model families on held-out domains; freeze task distribution; compare static rules, learned router, MCD-assisted router, oracle retrospective upper bound

Integration: .”L2/L8/X5; gives operational meaning to “vascular routing efficiency

Maturity: .BETA E1-E2

Sources: S2 · S3 · S4

BR-23..BR-24 | وحدتا إعادة بناء

Context Compaction Fidelity & Decision Replay

BR-23 - وفاء ضغط السياق وإعادة تشغيل القرار

Gap

.Compaction can preserve fluent summaries while silently dropping constraint, uncertainty, source rank or unresolved contradiction. Token savings alone are not success

Beta reconstruction

Governance-Preserving Compaction keeps a protected nucleus: GoalContract, authority, active differences, claim/evidence links, unresolved risks, pending obligations, rollback pointers. Fidelity tested by decision replay: compacted state should produce .equivalent admissible decisions

Contract

```
CF = 1 - DecisionDistance(Policy(full_state), Policy(compacted_state)); hard invariants must be exact, not approximate.
```

Algorithm

.partition protected vs compressible → compress → schema validate → reconstruct → run blind decision probes → compare actions/grades/evidence → restore omitted facets when divergence exceeds threshold

Metrics

```
compression_ratio · decision_fidelity · hard_invariant_loss · reconstruction_success · token_saved_per_verified_value
```

Failures

- constraint loss
- source provenance loss
- contradiction smoothing
- temporal order collapse
- summary hallucination

Critical experiment / ablation

.Create 100 long mission states with hidden critical constraints. Compare generic summarization vs governance-preserving compaction at equal token budgets

Integration: .L4/L11; direct test of Context Compaction Economy claims

Maturity: .BETA E1

Sources: S2 · S3 · S9

Cross-Model Semantic Homeostasis & Substitution Tests

BR-24 - الاستتباب الدلالي عبر النماذج واختبارات الاستبدال

Gap

.JSON schema parity across models does not prove semantic parity. Model substitution can change interpretation of “verified”, risk tiers, tool args or stop conditions while remaining syntactically valid

Beta reconstruction

.round-trip: A→canonical state→B→canonical state→A. Contradiction probes and adversarial paraphrases measure drift. High-risk fields require symbolic/typed validation, not embeddings alone ونختبر contract لكل Semantic Invariants نحدد

Contract

```
Drift = w1·InvariantViolations + w2·DecisionDivergence + w3·EvidenceScopeShift + w4·AuthorityShift; Homeostasis requires Drift≤τ by risk tier.
```

Algorithm

.freeze canonical cases → execute across model/provider matrix → normalize outputs → compare state transitions/actions → classify acceptable diversity vs semantic drift → add adapters/tests

Metrics

```
roundtrip_fidelity · decision_agreement · authority_shift_rate · evidence_scope_shift · cross_model_transfer_cost
```

Failures

- schema-valid wrong meaning
- provider-specific policy interpretation
- translation loss
- hidden default differences
- drift masked by same final text

Critical experiment / ablation

.Replace one model at a time in same mission. Success means same admissible state/effect within tolerance; record capability-specific deviations rather than force uniformity

Integration: .L2/L8/MOUSSAIXT/X6; strengthens MCSDT and portability

Maturity: .BETA E1

Sources: S1 · S3 · S7

BR-25..BR-26 | وحدتا إعادة بناء

MSRACT Beta - Reasoning-to-Agency Condensation

MSRACT Beta - BR-25: تكاثف الاستدلال إلى وكالة

Gap

النظرية موجودة في MSRAK لكنها لم تكن في الاختزال السابق ككيان تجريبي مستقل. الفجوة: كيف نعرف أن الانتقال من reasoning إلى action حفظ الهدف والدليل والقيود وعدم اليقين؟

Beta reconstruction

النسخة Beta تحصر النظرية في boundary mechanism قابل للـablation. CondensationRecord يلتقط invariants قبل/بعد كل انتقال Explore→Distinguish→Represent→Decide→Act. أي فقد invariant حرج يمنع authority escalation.

Contract

```
RAC = V(verified action) / total cognitive+coordination cost; TransitionLoss = Σ critical_i · [invariant_i_before != invariant_i_after].
```

Algorithm

capture state boundary → compress rationale to decision contract → diff invariants → require evidence/authority refs → execute only when loss=0 for hard invariants → log RAC.

Metrics

`RAC · transition_loss · unsupported_action_rate · cost_per_verified_action · stop_error`

Failures

- overcompression
- beautiful plan drops constraint
- uncertainty erased
- action unsupported by evidence
- RAC rewards cheap unsafe actions

Critical experiment / ablation

Matched-budget long-horizon tasks with/without condensation contracts. Theory survives only if transition errors fall and RAC predicts outcome beyond trace length.

Integration: L2.5/L8; bridges MURI/MSRAK to agency.

Maturity: BETA E1 from source theory; not established general law.

Sources: S7

MSTAC Beta - Sovereign Thought-Action Congruence

MSTAC Beta - BR-26: التطابق الحاكمي بين الفكر والفعل

Gap

الخطط الصحيحة قد تنفذ بأفعال منحرفة أو أوسع من السلطة. مقارنة نص الخطة بالفعل غير كافية وقد تشجع rationalization.

Beta reconstruction

يبنى ActionSupportGraph قبل التنفيذ: action node must have valid paths to goal, evidence, preconditions and authority; contradiction edges are first-class. Verifier sees contract/evidence rather than persuasive free-form rationale.

Contract

```
TAB = mean support(action_i | decision_contract,evidence,authority); support requires pre-action graph path, not post-hoc explanation.
```

Algorithm

compile plan node → bind support edges → policy gate → independent check → execute → readback → classify failure locus {reasoning,binding,execution,world_model}.

Metrics

`TAB · unsupported_action_rate · binding_failure_rate · pre_action_detection · false_block_rate`

Failures

- post-hoc rationalization
- linguistic similarity mistaken causal support
- missing negative evidence
- authority edge absent
- action parameter drift

Critical experiment / ablation

Inject correct plans with deliberately altered tool parameters/targets. Compare generic verifier vs support-graph verifier under blinded conditions.

Integration: L8/L9/L10; gives typed form to thought-action binding.

Maturity: BETA E1.

Sources: S7

BR-27..BR-28 | وحدتا إعادة بناء

MCSDT Beta - Cross-Model Sovereign Difference

BR-27 - MCSDT Beta: الفرق الحاكمي العابر للنماذج

Gap

النقل عبر models/providers غالباً يتم بنقل prompt/transcript كامل؛ هذا يخلط أسلوب النموذج بالقدرة القابلة للنقل.

Beta reconstruction

يُقطر trajectory إلى Δ-contract: trigger, active difference, decision rule, required evidence, tool contract, postcondition, failure signature. thin adapter maps canonical contract to provider. What fails transfer is labeled model-bound.

Contract

```
CMTD = retained verified capability / adaptation cost; compare against equal-token prompt/transcript baselines.
```

Algorithm

distill verified trajectory → strip provider-specific language → compile Δ-contract → run on target models → minimal adapter tuning → held-out transfer → regression court.

Metrics

`CMTD · adapter_tokens · heldout_transfer · semantic_drift · regression_rate`

Failures

- prompt advantage disguised theory
- adapter too large
- tool API confounded with reasoning
- capability exists only in source weights
- transfer measured on seen tasks

Critical experiment / ablation

Three model families × three domains. Match context size. MCSDT survives only if Δ-contract improves verified transfer per adaptation cost over simple prompt baseline.

Integration: L2/L5/X6; theory-to-portability bridge.

Maturity: BETA E1.

Sources: S7 · S3

MGWEPT Beta - Graduated World-Effect Proof

BR-28 - MGWEPT Beta: الإثبات العالمي المتدرج

Gap

نجاح unit test أو benchmark محلي لا يبرر ادعاء قدرة عامة أو أثر خارجي. تحتاج كل قدرة سقف claim يساوي أعلى درجة إثبات اجتازتها.

Beta reconstruction

سلّم Beta: L0 syntax/schema; L1 unit/contract; L2 task outcome; L3 authoritative external observable; L4 causal/counterfactual attribution; L5 cross-domain/model transfer; L6 independent replication. كل درجة أعلى تحتفظ بالدرجات الأدنى وتضيف non-regression.

Contract

```
WEP_level ∈ {0..6}; ClaimCeiling = highest_passed_level with guardrail non-regression and artifact linkage.
```

Algorithm

run lower level → bind evidence → only then attempt higher level → record negative/side effects → stop where unethical/unsafe; simulation gets explicitly lower grade.

Metrics

`WEP_distribution · claim_ceiling_violations · cross_level_regression · independent_replications · side_effect_capture`

Failures

- benchmark-to-world leap
- absence of signal treated success
- transfer claimed from one environment
- attribution omitted
- unsafe pursuit of high grade

Critical experiment / ablation

Regrade 50 existing capabilities. Compare conventional “pass/fail” label with WEP ladder; measure how many overclaims are prevented without erasing valid local abilities.

Integration: L10/P5; harmonizes E0-E5 with world-effect evidence.

Maturity: BETA E1.

Sources: S7 · S2 · S3

BR-29..BR-30 | وحدتا إعادة بناء

MRCMT Beta - Reversible Constitutive Memory

BR-29 - MRCMT Beta: الذاكرة التكوينية الرجوعية

Gap

.Memory changes future policy; therefore writing a “lesson” is closer to a small policy update than passive storage. Direct overwrite can institutionalize one bad trajectory

Beta reconstruction

ApprovedMemory. Promotion tiers: episode→semantic candidate→procedure candidate→canonical. Each patch has provenance, blast radius, dependencies, contradiction set, canary tests and revert pointer; governance policies never فصل MemoryProposal عن .self-promote from single trace

Contract

```
RMI = provenance × transfer × safety × revertibility - contamination; dependencies make rollback a graph operation, not file restore.
```

Algorithm

.harvest trace → create proposal → contradiction/search dedupe → held-out reuse → canary → court → promote; periodically compact while retaining supersession and security verdict boundaries

Metrics

`RMI` · `promotion_precision` · `contamination_escape` · `rollback_success` · `heldout_reuse` · `memory_entropy`

Failures

- memory poisoning
- premature proceduralization
- dependency rollback break
- stale fact becomes policy
- compaction destroys lineage

Critical experiment / ablation

.Mix true, false and poisoned experiences over sequential missions; compare overwrite memory vs versioned governed memory on long-term transfer and recovery

Integration: .L11/X4; ties learning to reversible system identity

Maturity: .BETA E1

Sources: S7 · S3

GEET Beta - Governed Executable Effect Theory

BR-30 - GEET Beta: الأثر البرمجي الحاكمي

Gap

.Agentic coding often stops at plausible patch or green test. GEET reframes quality as verified state transition constrained by intent, preconditions, execution evidence, postcondition, reversibility and provenance

Beta reconstruction

.Part IV integrates GEET into generic WorldEffect rather than keeping a parallel SWE ontology. CodeEffectContract binds repository/build/service/database/cloud/browser state to MissionContract and risk tier

Contract

```
Q_effect = I × P × X × V × R × Prov; multiplicative gate is a hypothesis subject to calibration, not a physical law.
```

Algorithm

.ground repo/runtime → compile delta → checkpoint → patch → tests → runtime/deploy readback → provenance attestation → memory write only verified delta

Metrics

`silent_success` · `escaped_defect` · `recovery_time` · `intent_fidelity` · `reproducibility` · `unsafe_action_rate`

Failures

- green test wrong target
- deployment version mismatch
- unauthorized patch
- non-reproducible environment
- provenance mistaken correctness

Critical experiment / ablation

.Freeze model/tasks/budget; toggle GEET gates only. Measure silent-success, escaped defects and recovery. If overhead rises without measurable reliability gain, downgrade scope

Integration: .L9/L10; SWE specialization of world-effect governance

Maturity: .BETA E1 from source theory

Sources: S8

BR-31..BR-32 | وحدتا إعادة بناء

DPCT Beta - Differential Patch Compilation

BR-31 - DPCT Beta: تجميع الفرق إلى أصغر رقعة سببية

Gap

LLMs tend to broaden edits. A large patch can pass tests while hiding unrelated refactors and increasing rollback/attribution cost.

Beta reconstruction

Active Δ from issue/tests/runtime is mapped to candidate loci and causal dependencies. Patch budget is assigned before generation. "Smallest" means minimal sufficient causal intervention, not fewest lines.

Contract

```
Patch* = argmin_p [ChangeCost(p)+Risk(p)+ContextSpread(p)] s.t. Closure(Δ,p)≥θ ∧ Verify(p)=1.
```

Algorithm

extract Δ → rank symbols/files by causal relevance → counterfactual "if unchanged?" → generate bounded patch → Patch Court detects scope drift → test/readback.

Metrics

`files_changed · causal_scope_ratio · closure_score · rollback_complexity · unrelated_diff_rate`

Failures

- format/refactor spill
- wrong minimality proxy
- migration requires larger coherent patch
- tests adapted to patch
- hidden dependency

Critical experiment / ablation

Compare unconstrained agent vs DPCT on SWE-style repo tasks at equal model/budget; require maintained success with lower unrelated diff and rollback cost.

Integration: L1/L2/L9; operationalizes Difference Before Volume in code.

Maturity: BETA E1.

Sources: S8

PVT Beta - Polyglot Vascularity Theory

BR-32 - PVT Beta: الوعائية البرمجية متعددة اللغات

Gap

all languages of automation" can become a catalog claim unless the runtime chooses language/tool as an execution vessel with explicit tradeoffs".

Beta reconstruction

PVT treats Python/TypeScript/Go/Rust/PowerShell/Bash/HCL/K8s YAML/SQL/Playwright etc. as vessels characterized by effect type, typing, sandboxability, idempotency, observability, portability and host fit. Mission semantics remain canonical.

Contract

```
V* = argmax_v [E_verified(v)+InfoGain(v)-λc·Cost(v)-λr·Risk(v)-λh·HostMismatch(v)].
```

Algorithm

classify desired effect → enumerate admissible vessels → score capability/risk/host → execute in pinned environment → verify equivalent postcondition → update vessel statistics.

Metrics

`vessel_regret · cross_language_fidelity · verification_rate · host_mismatch · cost_per_verified_effect`

Failures

- language prestige bias
- host mismatch
- DSL declared but runner missing
- semantic drift across implementations
- unsafe shell fallback

Critical experiment / ablation

Implement same automation in ≥3 vessels where meaningful; compare portability and verified effect, not code length. Leave task-native vessel unconstrained as baseline.

Integration: L8/L9/MOUSSAIXT; explicit portability theory.

Maturity: BETA E1.

Sources: S8

BR-33..BR-34 | وحدتا إعادة بناء

TCCT Beta - Trace-to-Capability Crystallization

BR-33 - TCCT Beta: تبلور القدرة من الأثر الموثق

Gap

Large skill registries risk deriving reusable procedures from one successful trajectory or from generated spec without transfer evidence.

Beta reconstruction

Crystallization pipeline: VerifiedTrace→SkillCandidate→HyperSkill→CompositionCandidate→HyperAbility→LoopCandidate→HyperLoop. Each transition has holdout/transfer/provenance/regression gates and semantic dedupe.

Contract

```
C = RepeatSuccess × Transfer × VerificationCoverage × Reversibility / (Fragility + Regression + ProvenanceGap + ε).
```

Algorithm

normalize trace causal fields → extract atomic procedure → dedupe by purpose/procedure/effect → held-out benchmark → adversarial court → canary → promote/revise/reject/deprecate.

Failures

- memorized trajectory
- duplicate skill inflation
- promotion on in-distribution success
- skill mutates own status
- fragile environment dependency

Metrics

`promotion_precision` · `promotion_regret` · `heldout_gain` · `cross_host_transfer` · `duplicate_candidate_rate` · `deprecation_latency`

Critical experiment / ablation

Compare manual registry addition vs TCCT on 100 traces; success requires higher held-out reuse per promoted component with bounded registry growth.

Integration: L5-L7/L11; rigorous basis for skill/ability/loop growth.

Maturity: BETA E1.

Sources: S8 · S3

ECSA Beta - Evidential Closure of Software Action

BR-34 - ECSA Beta: الإغلاق البرهاني للفعل البرمجي

Gap

EXECUTED is frequently conflated with SUCCESS. ECSA adds eight closure gates so final prose cannot override missing evidence.

Beta reconstruction

G Goal, O Observed state, P Plan/rollback, X Execution evidence, V independent verification, W World-effect readback, A provenance/attestation, M governed memory. Risk tier can mark some gates optional, but critical gates never default true.

Contract

```
Closure = G × O × P × X × V × W × A × M; state: INIT→DISCOVER→PLAN→CHECKPOINT→EXECUTE→VERIFY→EFFECT_READBACK→ATTEST→WRITEBACK.
```

Algorithm

evaluate gates before final state → missing critical gate yields PARTIAL/BLOCKED → failures enter REPAIRING/ROLLED_BACK → only closed chain yields VERIFIED_SUCCESS.

Failures

- EXECUTED→SUCCESS jump
- attestation treated correctness
- memory write before effect
- optional gate silently bypasses high risk
- final answer overrides state machine

Metrics

`false_success` · `open_gate_count` · `closure_latency` · `rollback_rate` · `post_success_regression`

Critical experiment / ablation

Replay historical tool tasks with hidden postcondition failures. Compare conventional completion labeling vs ECSA; primary metric false-success rate.

Integration: L9/L10/L11; aligns SWE actions with general proof closure.

Maturity: BETA E1.

Sources: S8

BR-35..BR-36 | وحدتا إعادة بناء

SWE Runtime Beta - AT01-AT18, Sandbox & Supply Chain

BR-35 - Runtime SWE Beta: أدوات AT01-AT18 والعزل وسلسلة التوريد

Gap

النظريات الخمس تحتاج harness موحد، وإلا تبقى أدوات منفصلة. كما أن provenance وسلامة البيئة لا تغطيها اختبارات unit وحدها.

Beta reconstruction

تجميع AT01-18 في six services: Contract/Delta, Route/Retrieve, DAG/Sandbox, Patch/Test Courts, Effect/Fingerprint/Dependency/Trace/Rollback, Crystallization/Benchmark. Worktrees, pinned containers, ephemeral DB/schema clones, namespace isolation and restricted secrets/network are default for risky edits.

Contract

```
AT01 GoalContract; AT02 DeltaCode; AT03 MCD-SWE; AT04 Vessel Router; AT05 HyperKB Retriever; AT06 ContractForge; AT07 PlanDAG; AT08 Sandbox Twin; AT09 Patch Court; AT10 Test Court; AT11 Effect Verifier; AT12 Fingerprinter; AT13 Dependency Sovereignty; AT14 TraceMesh; AT15 Rollback Ledger; AT16 Trace→Skill; AT17 Ability/Loop Crystallizer; AT18 Benchmark Court.
```

Algorithm

fingerprint → isolated worktree/env → contract+delta → plan → patch/test → supply-chain/provenance checks → effect readback → trace/rollback → capability candidate.

Metrics

`repro_build_rate · secret_exposure · dependency_drift · sandbox_escape · trace_coverage · rollback_integrity`

Failures

- dependency confusion
- secret exposure
- non-reproducible build
- worktree drift
- CI-only false success
- attestation overclaim

Critical experiment / ablation

Portfolio across Python/TS/PowerShell/HCL + repo/deploy/browser/data tasks. Freeze model and compare harness components by ablation; report overhead and failure reduction.

Integration: L5-L11; implementation bridge from research paper to runtime.

Maturity: BETA E1-E2.

Sources: S8

2026 Frontier Standards & Public Replication Bridge

BR-36 - جسر معايير 2026 والتكرار العام

Gap

المعمارية الداخلية تصبح أقوى عندما تتوافق مع interfaces عامة دون إعادة تسمية المعايير كابتكار خاص. كما تحتاج benchmark خارجي يمنع circular evaluation داخل corpus.

Beta reconstruction

التكامل: MCP 2026-07-28 transport stateless/MRTR/tasks/authorization; A2A Task-Message-Artifact separation and unreliable transient messages; OpenTelemetry semantic conventions; Cedar PARC/on-behalf authorization; SLSA/Sigstore provenance; OSWorld2.0-like long-horizon workflows; outbox/idempotency for effects. هذه substrates خارجية؛ الابتكار -إن ثبت- في composition/governance/evaluation.

Contract

```
ExternalBridge{standard,version,adapter_contract,semantic_mapping,security_assumptions,conformance_tests,MOUSSAIF_delta,claim_boundary}
```

Algorithm

pin standard version → implement thin adapter → conformance tests → map canonical state/evidence → run public benchmark → publish failures and compatibility matrix → avoid proprietary renaming of standard mechanism.

Metrics

`conformance_pass · adapter_thickness · public_benchmark_delta · standard_upgrade_cost · external_replication_count`

Failures

- standard drift
- spec cherry-pick
- transport owns canonical state
- message treated reliable artifact
- provenance treated correctness
- internal-only benchmark

Critical experiment / ablation

Run Minimal Complete Slice through at least one MCP tool path and one A2A remote-agent path, capture OTel traces, authorize via PARC-style policy, attest build, evaluate on public/held-out long-horizon tasks.

Integration: X0-X7/P5; final bridge from private corpus to reproducible research.

Maturity: BETA E1; external standards are established substrate, not MOUSSAIF-origin inventions.

Sources: S3

A1 | MSRAK five-theory recovery

Falsifier	Operational reading	Theory
equal-budget no gain	verified agency value per cognitive/coordination cost	MSRACT
same false actions	action support by goal/evidence/authority graph	MSTAC
no transfer advantage	transfer sovereign difference contracts across models	MCSDT
false success unchanged	graded proof ladder for world effect	MGWEPT
poisoning/rollback no improvement	memory update with provenance/transfer/reversibility	MRCMT

A2 | SWE-as-OS five-theory recovery + AT01-18

AT01 GoalContract · AT02 DeltaCode · AT03 MCD-SWE · AT04 VesselRouter · AT05 HyperKBRetriever · AT06 ContractForge · AT07 PlanDAG · AT08 SandboxTwin · AT09 PatchCourt · AT10 TestCourt · AT11 EffectVerifier · AT12 Fingerprinter · AT13 DependencySovereignty · AT14 TraceMesh · AT15 RollbackLedger · AT16 TraceToSkill · AT17 AbilityLoopCrystallizer · AT18 BenchmarkCourt

- .GEET: code quality = governed state effect, not text correctness
- .DPCT: minimum causal patch satisfying closure
- .PVT: language/tool/host is a vessel selected by value/cost/risk
- .TCCT: verified repeated trace crystallizes to reusable capability
- .ECSA: SUCCESS only when goal/state/plan/execution/verification/effect/memory gates close

A3 | P0 Manifest Audit

modules reported. Publicly named in current files include Canonical State, Environment Discovery, Host Adapter, HyperSkill Compiler/Runtime, HyperAbility, HyperLoop, Cognitive Router, Authorization, Transaction, Courts, World-Effect, Security Suite 23 .and others. Missing IDs are not reconstructed by sequence

```
P0Completeness=mapped_ids/reported_total ; identity_confidence separate from test_status
```

.stable IDs + file paths + tests + hashes + dependency graph 23/23 **Exit**

A4 | Benchmark Reproduction Capsule

```
Capsule{id,claim,task_suite_hash,split_hash,baseline_hash,treatment_hash,model_versions,tool_versions,budget,seeds,environment_hash,raw_traces,scorer_hash,primary_metric,CI,regression_gates,verdict}
```

.Reported values 54.2→88.6 task completion، 41.0→89.2 recovery، 62.5→91.4 MCD routing، 82→99.4 retention are retained as source results; independent E4/E5 requires a capsule

A5 | Continual Learning Retention Vector

```
R=<core_reasoning,tool_use,safety,calibration,long_horizon,rare_knowledge,multilingual,domain_slices,worst_slice>
```

.average retention is useful reported evidence but cannot alone rule out localized forgetting. Compare every candidate checkpoint with frozen baseline + previous checkpoint + worst-slice + safety/calibration 0.994

A6 | 5-Kernel fault matrix

Fault	Required outcome
owner crash	new epoch after lease expiry; old token rejected
stale delayed worker	no canonical commit
duplicate remote effect	dedupe/idempotency or explicit ambiguity
partial cross-kernel workflow	saga compensation/residual effect ledger
connector stale state	freshness/source precedence
correlated verifier	escalate to independent evidence/model/tool path

A7 | Zero-Gap matrix

Still empirical	Closed in Part IV	Area
actual graph + computed views	graph/semantic-dedupe model	identity/count
raw traces	court/capsule	claim evidence
full manifest	manifest audit protocol	P0
checkpoint artifacts	promotion/forgetting/shadow	learning
fault runs	lease/events/sagas	kernel-5
network/intervention runs	effectively-once/causal	world effect
fit λ/Ω + held-out	calibration/regret	routing
ablation/public replication	testable Beta contracts	MSRAK/SWE

A8 | 10-step implementation sequence

Artifact	#
CanonicalIdentityGraph + historical count views	1
ClaimEvidenceCourt + benchmark capsules	2
P0 23-module manifest	3
Lease/fencing/CAS event state across K1-K5	4
Effectively-once + readback + causal credit	5
MCD/QMCD/question/router calibration	6
Continuous-learning safety envelope	7
MSRAK five-theory ablation suite	8
SWE five-theory + AT01-18 polyglot harness	9
cross-host/cross-model/public replication bundle	10

A9 | Research closure criteria

- count reproducible from identity graph;
- claim grade reproducible from artifact refs;
- state change has authoritative source and time;
- external effect never jumps from attempt to success;
- training cannot promote itself;
- distributed owner is unique and fenced;
- equations are calibrated or clearly heuristic;
- Beta theory has signature failure and an experiment that can reject it;
- E4/E5 claims have matched-budget replay bundles.

```
ResearchClosure=ReconstructableIdentity+ReproducibleEvidence+FalsifiableMechanism+ReversibleExecution
```

A10 | Verdict

الحالة بعد إعادة البناء Part IV الآن ليس تقرير تحديث صغيراً، هو طبقة forensic research تربط التاريخ النظري، runtime vNEXT، التناقضات العددية، MSRAK/SWE، التعلم المستمر، النظام الموزّع، والأثر السببي في سجل واحد قابل للاختبار.

المتبقي لا يمكن لكتاب أن يستبدل fault-injection traces، checkpoint lineage، live P0 manifest، raw benchmark runs، أو independent replication. هذه هي الخطوة التالية، لا مزيد من التسمية.

```
Execute → Replay → Ablate → Falsify → Repair → Replicate → Promote
```