# الإغلاق الحاكمي التركيبي للذكاء الوكيلي

## عشر نظريات وعشرون أداة لإحكام السببية والتركيب والثقة والذاكرة والسيادة في معمارية MICAGFI–M-HKOS

**زهير مساعف — Zouhair Moussaif**

ورقة نظرية تأسيسية موسعة وبرنامج بحث هندسي

يوليو 2026

**الوضع العلمي:** تقترح هذه الورقة نظرية معمارية وأدوات ومقاييس أولية لأنظمة الذكاء الوكيلي. لا تثبت تحقق الذكاء العام الاصطناعي، ولا تنسب إلى الأنظمة وعيًا أو استقلالًا أو عصمة. جميع المعادلات مؤشرات تشغيلية تحتاج إلى معايرة وتجارب إعادة إنتاج.

## الملخص


تبحث هذه الورقة في فجوة تظهر بعد اكتمال المكونات الأساسية لمعمارية الذكاء الوكيلي: قد يمتلك النظام دستورًا، وذاكرة، وقواعد معرفة، ووكلاء، وأدوات، ومحاكم تحقق، وطبقات تدريب ونشر، ومع ذلك يبقى غير مكتمل حاكميًا لأن سلامة المكونات لا تضمن سلامة تركيبها، ولأن الاسترجاع لا يثبت السببية، والثقة لا تزال موزعة ضمنيًا، والحوافز قد تكافئ المظهر بدل المقصد، والذاكرة قد تتدهور دون مقياس لإنتروبيتها، والصلاحية الفردية قد لا تكفي لتحديد السيادة المؤسسية. تنطلق الورقة من معمارية MICAGFI–M-HKOS ذات الطبقات الماكروية الاثنتي عشرة والطبقات التشغيلية والتدريبية البالغ عددها 176، ثم تميز بين «اكتمال المكونات» و«اكتمال العلاقات الحاكمة بينها». وتقترح لهذا الغرض إطارًا جديدًا باسم **الإغلاق الحاكمي التركيبي للذكاء الوكيلي**، يعرف بوصفه مجموعة القوانين والاختبارات والسجلات التي تجعل انتقال المعرفة والقدرة والصلاحية بين المكونات انتقالًا مشروعًا وقابلًا للتتبع والتفنيد والتراجع.

تولد الورقة عشر نظريات جديدة: الضمان التركيبي الحاكمي، والتثبيت السببي التدخلي، واقتصاد الائتمان المعرفي، وتوافق الحوافز الوكيلية، والسلامة الأنطولوجية الطبقية، والنضج التطوري للقدرة، وإنتروبيا الذاكرة وإعادة التوطيد، والصمود أمام الفشل الناشئ، والسيادة المؤسسية البشرية–الآلية، والفتح الآمن في العالم المجهول. وتقترح عشرين أداة مقابلة تحول هذه النظريات إلى عقود تركيب، ورسوم سببية، ودفاتر ثقة، واختبارات حوافز، ومدققات أنطولوجية، وبوابات نضج، ومقاييس ذاكرة، ومختبرات فوضى، ومصفوفات سلطة، وبروتوكولات استكشاف آمن. تربط الورقة كل نظرية بطبقات MICAGFI–M-HKOS السابقة، وتوضح ما تضيفه فوق MCD وQ-MCD وHAKIM-MCD وKB-as-OS وHyperSkills وBounded HyperLoops والمحاكم وWorld-Effect Loop وSelf-Reconstruction Memory.

تستخدم الورقة منهج الضغط التركيبي المعرفي، ومقارنة ثلاثة نماذج متنافسة، وصياغات شبه رياضية، وثماني عشرة فرضية، وتسعة تنبؤات فارقة، ودراسات Ablation مقترحة، ومحكمة هدم من أربعة مستويات. وتخلص إلى أن النظام الوكيلي لا يصبح أكثر نضجًا بكثرة الوكلاء أو الأدوات فقط، بل بقدرته على إثبات صلاحية التركيب، وضبط نوع الادعاء، وتخصيص الثقة، وكشف الجدة، واحتواء الأعطال، وربط كل أثر بسلسلة دليل وسلطة ومسؤولية. وتقدم النسخة النهائية بعد الهدم بوصفها **نظرية ناقصة شريفة**: قوية كبرنامج هندسي قابل للاختبار، وغير كافية بعد لإثبات ذكاء عام أو تفوق شامل بلا بيانات تجريبية مستقلة.



**الكلمات المفتاحية:** الإغلاق الحاكمي التركيبي؛ MICAGFI؛ M-HKOS؛ MCD؛ Q-MCD؛ HAKIM-MCD؛ Agentic AI؛ Compositional Assurance؛ Causal Grounding؛ Epistemic Credit؛ Ontological Type Safety؛ Memory Entropy؛ Institutional Sovereignty؛ Open-World Adaptation.


**الإعلان المنهجي:** لن نخلط بين الحقيقة والفرضية، ولا بين النموذج والدليل، ولا بين الاستعارة والآلية، ولا بين نجاح الجزء وسلامة الكل، ولا بين امتلاك الأداة وامتلاك الحق في استعمالها.

## المساهمات المركزية

1. تعريف فجوة «الإغلاق الحاكمي» التي تبقى بعد اكتمال طبقات التدريب والتشغيل والذاكرة والأدوات.
2. صياغة نظرية جامعة تجعل العلاقات بين المكونات موضوعًا مستقلًا للبرهان والاختبار.
3. تقديم عشر نظريات جديدة، لكل منها آلية وحدود واعتراض واختبار فارق.

4. تصميم عشرين أداة تشغيلية قابلة للتحويل إلى Software Modules وSchemas وBenchmarks.
5. ربط الجديد صراحةً بالـ176 طبقة السابقة دون إذابته فيها أو تكرارها.
6. بناء مقاييس للضمان التركيبي، والرسوخ السببي، والائتمان المعرفي، والنضج، وإنتروبيا الذاكرة، والصمود، والجدة.
7. اقتراح برنامج تجريبي يقارن Baselines نموذجية ووكلية، ويستخدم Ablations واختبارات ضغط طويلة الأفق.
8. إخضاع الإطار لمحكمة هدم تميز بين الجدة الاصطلاحية والجدة الآلية والجدة القابلة للقياس.

## فهرس بنيوي

**الباب الأول:** الأساس السابق — MICAGFI-M-HKOS والحاكمية الكثّافية الفتّاحة.
**الباب الثاني:** الإغلاق الحاكمي التركيبي — العشر نظريات الجديدة.
**الباب الثالث:** الأدوات والمقاييس والفرضيات والتنبؤات والتقييم.
**الباب الرابع:** محكمة الهدم والنسخة المعدلة والبرنامج البحثي وقاعدة المعرفة.

# الباب الأول: الأساس المعرفي والمعماري السابق

## 1. المقدمة ومشكلة البحث

أصبح الانتقال من النموذج اللغوي بوصفه مولدًا للنص إلى النظام الوكيلي بوصفه منفذًا للأفعال انتقالًا في طبيعة المشكلة الهندسية، لا مجرد زيادة في جودة الإجابة. ففي النظام الوكيلي تتداخل الذاكرة، والتخطيط، والأدوات، والبيئة، والصلاحيات، وسلاسل الاعتماد، ومراقبة الأثر. وقد بينت أعمال ReAct أن تداخل الاستدلال والفعل يمكن أن يحسن التفاعل مع البيئات ومصادر المعرفة، بينما بين Toolformer إمكان تعلم توقيت استدعاء الأدوات وإدماج نتائجها، وأضاف Reflexion آلية تعلم لغوي من الإخفاق دون تعديل الأوزان [4–6]. غير أن هذه الأعمال تقدم مكونات قوية أكثر مما تقدم دستورًا شاملاً يحدد العلاقة بين السؤال والدليل والذاكرة والصلاحية والأثر.

يظهر التفكك نفسه في مجال الذاكرة. فقد صاغ MemGPT إدارة السياق بمجاز نظام التشغيل والذاكرة الهرمية، وقدمت GraphRAG وRAPTOR وHippoRAG تمثيلات علائقية أو هرمية أو مستلهمة من الذاكرة البشرية لتحسين الاسترجاع والتركيب عبر مجموعات كبيرة من الوثائق [7–10]. إلا أن استرجاع المعلومة لا يحل وحده مسألة: هل المعلومة صالحة؟ هل تناقض حالة سابقة؟ هل ينبغي إدخالها في الذاكرة الأساسية أم حجرها باعتبارها فرضية؟ وكيف يمنع انتقال تعليمات مشروع أو أسراره إلى مشروع آخر؟

كذلك تتقدم حوكمة السلوك عبر Constitutional AI وDPO وProcess Supervision، وتتقدم الحوكمة المؤسسية عبر NIST AI RMF وISO/IEC 42001، وتتقدم هندسة الأمن عبر OWASP لمخاطر Prompt Injection وExcessive Agency، [1–3] 16–18]. غير أن الجمع بين دستور النموذج ودستور النظام ودستور الفعل ما يزال مسألة تصميمية مفتوحة. فالدستور النصي لا يكفي إذا لم يتحول إلى بوابات صلاحية، واختبارات Postconditions، وسجل أثر، ومسار تراجع.

تقترح هذه الورقة أن الفجوة الحاسمة ليست غياب مكوّن جديد منفرد، بل غياب «حاكمية العلاقات بين المكونات». لذلك لا تُعرض الطبقات الـ176 باعتبارها قائمة ميزات، بل يعاد بناؤها في نظرية تشغيلية تضبط من يحكم من: السؤال يحكم الاستدلال، والدليل يرخص الفعل، والكثافة المعرفية توجه الموارد، والعقود تضبط القدرة، والمحاكم تضبط الادعاء، والأثر يعيد بناء الذاكرة، والدستور يحد التوسع.

### 1.1 مشكلة البحث

تتمثل المشكلة في أن أنظمة الوكلاء الحديثة تستطيع التخطيط واستدعاء الأدوات والبحث والكتابة، لكنها تظل معرضة لفشل مركب: سؤال سيئ يولد خطة صحيحة للمشكلة الخطأ، استرجاع حديث لكنه غير موثوق، أداة تنفذ بنجاح تقني دون تحقق المقصد، ذاكرة تحفظ استنتاجًا بوصفه حقيقة، محقق يكرر انحياز المنفذ، أو تحسين يولد قدرة جديدة بصلاحيات أوسع من المقصود. لا توجد طبقة منفردة تحل هذا الفشل، لأن سببه علاقة غير محكومة بين الطبقات.

### 1.2 سؤال البحث

**السؤال المركزي: كيف يمكن بناء معمارية تجمع تدريب النموذج، وحوكمة السؤال، والمعرفة والذاكرة، والقدرات والأدوات، والتحقق، والنشر، والتطور، بحيث تعظم الأثر الموثق المعدل بالتكلفة، وتحافظ على الصلاحية والتراجع، دون تحويل اسم AGI إلى ادعاء غير قابل للتفنيد؟**

### 1.3 المساهمات

- إعادة ضغط 176 طبقة في ثلاث دوائر و12 طبقة ماكرو ذات علاقات حاكمية واضحة.
- اقتراح نظرية الحاكمية الكثّافية الفتّاحة بوصفها أطروحة قابلة للاختبار.
- ضبط MCD وQ-MCD كمقاييس تفسيرية لا أرقام يقينية.
- ربط الدليل بالفعل عبر Evidence-to-Action Thresholds.
- تصميم ذاكرة محجورة مع إعادة بناء وسجل تعارض وإصدارات.
- تعريف HyperSkill كعقد قدرة وBounded HyperLoop كدورة ذات حد تقدم.
- اقتراح مقاييس وفرضيات وتنبؤات ومحكمة هدم وبرنامج تجريبي.

## 2. الوضع الأنطولوجي ونوع السؤال

الوضع الأنطولوجي المختار: نظام اجتماعي-تقني حاسوبي متعدد المستويات، يتضمن عمليات وعلاقات وشبكات وقواعد ومعايير وأفعالًا، وليس "شيئًا" واحدًا. السبب أن خصائص النظام تظهر من تفاعل النموذج والذاكرة والأداة والمستخدم والسياسة والبيئة. إذا عومل بوصفه نموذجًا فقط ضاعت السلطة والأثر، وإذا عومل بوصفه Workflow فقط ضاعت المعرفة والتعلم، وإذا عومل بوصفه مؤسسة فقط ضاعت الآليات الحسابية.

نوع السؤال: معماري، إبستمولوجي، قياسي، وتوليدي. هو معماري لأنه ينظم مكونات، إبستمولوجي لأنه يحدد مراتب المعرفة، قياسي لأنه يقترح مؤشرات، وتوليدي لأنه يحول الإطار إلى برنامج بحث وقدرات جديدة. لذلك يحتاج إلى مقارنة أدبيات، وصياغة آليات، وBenchmarks، ومحكمة هدم، وليس إلى تحليل لغوي فقط.

## 3. سلم مراتب المعرفة في هذه الورقة

| المرتبة | ما يندرج فيها في هذه الورقة |
|---|---|
| مدعوم بقوة | وجود حاجة عملية لدمج التخطيط والأدوات والذاكرة والتحقق في الأنظمة الوكيلية، ووجود مخاطر إضافية عند منح النماذج صلاحية الفعل. |
| راجح | التوجيه المتفرق والذاكرة المنظمة والتحقق المستقل يمكن أن تحسن الكفاءة والموثوقية مقارنة بالتشغيل غير المنظم، لكن حجم التحسن يعتمد على المهمة والنموذج. |
| خلافي | تفوق Process Supervision على Outcome Supervision بصورة عامة، وتوجد نتائج نظرية وتجريبية مختلفة باختلاف التغطية والخوارزمية والمجال. |
| استنتاج منطقي | إذا كانت الأفعال متفاوتة الأثر وقابلية التراجع، فيلزم ربط مستوى الدليل والصلاحية بدرجة الفعل. |

| المرتبة | ما يندرج فيها في هذه الورقة |
|---|---|
| تركيب نظري جديد | MICAGFI–M-HKOS ونظرية الحاكمية الكثّافية الفتّاحة بوصفهما جمعًا منضبطًا بين السؤال والكثّافة والذاكرة والعقود والمحاكم والأثر. |
| فرضية قابلة للاختبار | إضافة Q-MCD ومحاكم مستقلة وذاكرة محجورة إلى وكيل قائم ستخفض الإخفاقات عالية الأثر مقابل Baseline متساوي الميزانية. |
| أداة تحليلية | MCD وQ-MCD ومؤشر Cross-Model Density مقاييس تقديرية تحتاج إلى معايرة تجريبية، وليست وحدات فيزيائية. |
| استعارة منتجة | الوعائية الحاسوبية تشبه تدفق الموارد في الأوعية، لكنها لا تثبت تماثلًا بيولوجيًا. |
| ادعاء مرفوض | أن تجميع الطبقات يثبت الوعي أو الاستقلال أو تحقق AGI الكامل. |

# 4. تصفية المادة الخام وخريطة الحقل

## 4.1 مصفوفة الحفظ والتوسيع

| المرتبة | المعالجة |
|---|---|
| تأسيسي يحفظ ويقوى | الدستور، السؤال الحاكم، MCD/Q-MCD، الذاكرة المعاد بناؤها، Bounded HyperLoops، HyperSkills، المحاكم، World Effect، التراجع. |
| قوي يحتاج نمذجة | الوعائية الحاسوبية، KB-as-OS، HyperDeployment، Cross-Model AGI Density؛ تحتاج تعريفات تشغيلية وBaselines. |
| خصب لكنه غامض | Fattah Open Intelligence وNeoAGI Evolution؛ يعاد ضبطهما كقدرة على فتح فضاء المشكلة وتوليد أسئلة أعلى، لا كمرتبة وجودية. |
| يحتاج إعادة تموضع | الطبقات الـ33 الخاصة بالتدريب؛ توضع داخل Training Plane بدل مساواتها بطبقات Runtime. |
| يحتاج تخفيض رتبة | كل استخدام لعبارة AGI أو "مضاعف أسي" إن فُهم كضمان لارتفاع الذكاء؛ يصاغ كبرنامج بحث أو نمو تركيبي محتمل. |
| يضغط ويدمج | التكرارات بين النزاهة، التحقق، المحاكم، المراقبة، والRollback تدمج في عقود وبوابات موحدة دون إسقاط وظائفها. |

## 4.2 خريطة المتغيرات

| المتغير | نوعه | وظيفته |
|---|---|---|
| السؤال الحاكم Q | مستقل/حاكم | يحدد الفرق المطلوب ومعايير الدليل والفعل. |
| الفرق النشط $\Delta$ | وسيط | يمثل المعلومة أو القرار الذي يغير النتيجة. |
| الكثّافة المعرفية MCD | مقياس وسيط | تقدر قيمة المسار وتوجه الموارد. |
| جودة الدليل E | معدل/مرخّص | ترفع أو تخفض سقف الفعل المسموح. |
| الصلاحية $A_u$ | قيد دستوري | تحدد الأدوات والأصول والأفعال المتاحة. |
| قدرة النموذج B | مدخل | تمثل القدرات الأساسية وحدودها. |
| معرفة التشغيل K | مدخل بنيوي | KB-as-OS وذاكرة وإجراءات وعلاقات. |
| قدرات HyperSkill H | وسيط تنفيذي | تحول المعرفة إلى إجراءات بعقود. |
| ميزانية الموارد C | قيد | Tokens وLatency وRPM ومال وطاقة. |
| قوة التحقق V | وسيط تصحيحي | استقلال المحقق وقابلية الاختبار. |
| قابلية التراجع R | معدل خطر | تخفض الخطر الصافي للفعل. |
| الأثر العالمي W | ناتج | التغير الخارجي الموثق لا النص المولد. |
| نزاهة الذاكرة $M_i$ | ناتج/حالة | مدى دخول المعرفة الموثقة فقط. |

| المتغير | نوعه | وظيفته |
|---|---|---|
| الضوضاء N | مُضعف | تكرار، حشو، تعارض، ادعاء زائد. |

### 4.3 حلقات التغذية الراجعة والعتبات

تعمل المنظومة عبر ثلاث حلقات: حلقة معرفية من السؤال إلى الدليل؛ حلقة تنفيذية من الخطة إلى الأثر؛ وحلقة تطورية من الأثر إلى تحديث الذاكرة والقدرة. تظهر العتبات الحرجة عند: ارتفاع أثر الفعل بما يستلزم تصعيدًا بشريًا؛ انخفاض التقدم بما يستلزم إيقاف الحلقة؛ ارتفاع تعارض الذاكرة بما يمنع الترقية؛ أو انخفاض استقلال التحقق بما يمنع قبول الحكم. العلاقات غير خطية لأن زيادة الوكلاء أو السياق قد تحسن الأداء حتى نقطة ثم تولد احتقانًا وتعارضًا.

## 5. الخلفية النظرية والنسب المعرفي

تنتمي الورقة إلى تقاطع خمسة مسارات. الأول هو مسار Reasoning-and-Acting الذي يمثله ReAct، حيث تتداخل خطوات الاستدلال والفعل، ومسار Tool Learning الذي يمثله Toolformer [4،5]. الثاني هو التعلم من المسار والخطأ، كما في Reflexion و[3،6] Process Supervision. الثالث هو ذاكرة النماذج والوكلاء، من MemGPT إلى GraphRAG وRAPTOR وHippoRAG و[7–10،21] A-Mem. الرابع هو Post-Training والمواءمة، من Constitutional AI إلى [1،2] DPO. والخامس هو تقييم الوكلاء وإدارة المخاطر، من AgentBench وGAIA وSWE-bench وOSWorld وAgent-SafetyBench إلى NIST وISO و[11–18] OWASP.

الجديد المقترح ليس ادعاء أن أي مكوّن من هذه المكونات لم يوجد سابقًا، بل إعادة تعريف الوحدة العلمية من "وكيل يستخدم أدوات" إلى "نظام معرفي محكوم بعلاقات أسبقية وترخيص وتحقق". فMemGPT يقدم استعارة نظام التشغيل للذاكرة، بينما يوسع M-HKOS الفكرة إلى العقود والدستور والمحاكم والأصول؛ وConstitutional AI يقدم مبادئ لتدريب السلوك، بينما يوسع Constitutional Reversibility الدستور إلى صلاحية الأفعال ومسارات الرجوع؛ وBenchmarks تقيس إتمام المهام، بينما يضيف World-Effect Ledger فصلًا بين إعلان الإتمام والأثر الموثق.

تظل هذه الأصالة تركيبية ومشروطة. فإذا لم تنتج التجارب فرقًا عن Pipeline معياري جيد التصميم، تصبح MICAGFI لغة تنظيم لا نظرية تفسيرية جديدة. لهذا تضع الورقة اختبارات Ablation وتنبؤات فارقة بوصفها جزءًا من التعريف، لا ملحقًا اختياريًا.

## 6. تحليل المجال في 26 نقطة مؤسسة

### 6.1 تعريف المجال

المجال المدروس ليس نموذجًا ولا Workflow منفردًا، بل نظامًا اجتماعيًا-تقنيًا حاسوبيًا يضم نموذجًا، ذاكرة، أدوات، وكلاء، صلاحيات، ومشغلًا بشريًا. يغير هذا التعريف وحدة التحليل: فلا تُقاس الجودة بدقة الإجابة فقط، بل بصحة السلسلة من صياغة الهدف حتى تحقق الأثر والتراجع عند الفشل.

### 6.2 الفجوة البنيوية

الأدبيات موزعة بين Post-Training، وTool Use، وRAG، والذاكرة، والAgent Benchmarks، والسلامة. الفجوة المقترحة هي غياب نموذج يحدد علاقات الحاكمية بين هذه المكونات: متى يوجه السؤال البحث؟ متى يرخص الدليل الفعل؟ متى تصبح نتيجة التنفيذ ذاكرة؟

### 6.3 الفصل بين النموذج والنظام

يجب فصل قدرة Base Model عن قدرة System Overlay. قد يأتي التحسن من Router أو أداة أو ذاكرة أو محقق، ولذلك يلزم تقييم Ablation يقارن النموذج الخام بالنظام المركب حتى لا تنسب كل القدرة إلى الأوزان أو إلى المعمارية جزافًا.

### 6.4 أسبقية السؤال الحاكم

السؤال غير المحدد يورث غموضًا في التخطيط والقياس. Q-MCD لا يقيس جمال السؤال، بل مقدار سيطرته على حدود المشكلة والدليل واختبار النجاح. الفرضية هي أن الاستثمار المبكر في السؤال يقلل حلقات الإصلاح المتأخرة.

### 6.5 الفرق النشط

ليس كل نقص معلومات مهمًا. Active Difference هو التمييز الذي إذا تغير تغير القرار أو المسار. يؤدي تعريفه إلى ضغط المهمة ومنع البحث الشامل غير الموجه. ويجب أن يكون قابلاً للوصف والملاحظة، وإلا تحول إلى شعار.

### 6.6 MCD كدالة منفعة معرفية

تستخدم MCD لتقدير الفرق والتكامل والنقل والوكالة والتنبؤ والتصحيح مع خصم الضوضاء. لا تعامل كقياس مطلق، بل كRubric يعين على مقارنة المسارات تحت ميزانية ثابتة.

### 6.7 الوعائية الحاسوبية

تشرح الوعائية كيف توزع Tokens والنماذج والأدوات على العقد الحرجة. الاختبار الحقيقي ليس جمال الاستعارة، بل هل يقلل Router المستند إلى قيمة هامشية عدد الاستدعاءات مع الحفاظ على الإنجاز؟

### 6.8 التعقيد متعدد المقاييس

المهمة الواحدة قد تحتوي خطأ Token، وتناقض Claim، وفشل Tool Call، وانحراف Project Goal. لذلك يجب أن يتحرك النظام بين مستويات الرمز والادعاء والمهمة والمشروع والعالم دون أن يختزل أحدها في الآخر.

### 6.9 KB-as-OS

المعرفة التشغيلية لا تقتصر على مقاطع مسترجعة؛ بل تشمل شروط التطبيق، التبعيات، الصلاحيات، الفشل، والاختبار. تصبح قاعدة المعرفة "نظام تشغيل" فقط عندما تؤثر في التحكم والتنفيذ، لا عندما تحمل الاسم.

### 6.10 معمارية الذاكرة

الذاكرة المطلوبة متعددة: دلالية، حدثية، إجرائية، هوية، وعاملة. الخطر المركزي هو Memory Contamination؛ لذلك تقترح الورقة حجر الذاكرة المقترحة، وإصدارات Canonical، وسجلًا للتعارض والتجاوز.

### 6.11 الضغط القابل للفك

لا يمكن تحميل تاريخ المشروع كله في كل خطوة. ROFSC يضغط الدعاوى مع روابط الأدلة وشروط الحدود ومؤشرات التوسع. يجب اختبار الضغط بإعادة إنتاج القرارات الحرجة، لا بعدد الكلمات المخفضة فقط.

### 6.12 القدرة التعاقدية

HyperSkill ليست تعليمات فضفاضة بل عقد Capability: مدخلات، مخرجات، تبعيات، صلاحيات، اختبارات، ميزانية، ومخلفات ذاكرة. هذا يحول الخبرة من Prompt خاص إلى أصل قابل للنقل والمراجعة.

### 6.13 الحلقة المحدودة

الاستدلال التكراري قد يزيد الجودة أو يعيد تدوير الأخطاء. Bounded HyperLoop تربط الاستمرار بزيادة قابلة للقياس في MCD أو التقدم، وتفرض سقوفًا للتكلفة والوقت والمحاولات.

### 6.14 فصل أدوار الوكلاء

Planner وExecutor وVerifier أدوار معرفية مختلفة. الفصل يقلل تضارب المصالح، لكنه لا يضمن الاستقلال إذا كانت النماذج والسياقات متطابقة. لذلك تقاس استقلالية الدليل والنموذج والمصدر.

### 6.15 حوكمة الأدوات

اختيار الأداة ليس نهاية المسألة. يجب التحقق من Arguments، والنتيجة، والآثار اللاحقة، وإمكان التراجع. ويجب معاملة محتوى الأدوات والويب كبيانات غير موثوقة قد تحمل Prompt Injection.

### 6.16 عمارة الدليل

يحتاج كل Claim مهم إلى نوع، ومصدر، وحداثة، واستقلال، وثقة. التناقض لا يحذف بل يحفظ حتى يوجد اختبار فاصل. هذه العمارة تربط المعرفة بالعمل عبر Evidence-to-Action Thresholds.

### 6.17 المحاكم كطبقة مناعية

محاكم التعريف والحد والآلية والقياس والتنفيذ والأمان والخصوصية والهدم تمنع انتقال البلاغة إلى قرار. لكن كثرة المحاكم قد تصبح بيروقراطية؛ لذلك تفعّل Sparse وفق الخطر.

### 6.18 التفويض والتراجع

يطبق النظام Least Irreversible Action: مسودة قبل إرسال، Sandbox قبل Trash، Production قبل الحذف النهائي. الصلاحية ليست نصًا عامًا بل Capability Token أو Policy Check مرتبطًا بالأداة والأصل.

### 6.19 إغلاق حلقة الأثر

الوكيل لا ينجح لأنه أعلن الانتهاء. World-Effect Loop يسجل التوقع، وينفذ، ويفحص Postconditions، ويقيس الأثر والآثار الجانبية. هذه الحلقة تحول النص إلى تعلم تجريبي.

### 6.20 موقع التدريب

طبقات البيانات وSFT وDPO وProcess Supervision وTool Training تحسن مكونات متعلمة، لكنها لا تستبدل Runtime Governance. يمكن لنموذج مدرب جيدًا أن يفشل بسبب صلاحيات أو ذاكرة أو أداة سيئة.

### 6.21 فجوة التقييم

AgentBench وGAIA وSWE-bench وOSWorld نقلت التقييم نحو بيئات تفاعلية وأثر قابل للفحص [11–14]. إلا أن الإطار يحتاج Benchmarks إضافية للذاكرة والتراجع والتلوث والمخالفة الدستورية وكفاءة التكلفة.

### 6.22 اقتصاد الذكاء

زيادة الاستدعاءات ليست دائمًا زيادة في الذكاء. يقترح الإطار قياس منفعة موثقة لكل Token وزمن ومخاطر. وقد تظهر بنية أصغر متخصصة أفضل من سرب نماذج مرتفع التكلفة.

### 6.23 القدرة العابرة للنماذج

Cross-Model Density لا تعني أن المعمارية تحول كل نموذج إلى AGI. إنها تقيس مقدار التحسن القابل للتكرار عند تثبيت Overlay وتغيير Base Model، مع تسجيل سقف التقارب لكل نموذج.

### 6.24 النشر المحكوم

HyperDeployment ليس استنساخًا غير محدود، بل وراثة عقود وسياسات مع Versioning وCanary وRollback. يفشل النشر مغلقًا إذا غاب المصدر أو التحقق أو الصلاحية أو خطة التراجع.

### 6.25 التطور المحدود

يمكن تحويل المسارات الناجحة إلى HyperSkills مقترحة، لكن لا تدخل Stable Registry إلا بعد تكرار النجاح واختبار النقل والانحدار. التحسن هنا خارجي ومراقب، لا ادعاء بتعديل ذاتي سري للأوزان.

**6.26 حد AGI**

اسم AGI Fattah يعمل بوصفه أفقًا بحثيًا لفتح المشكلات والنقل والتصحيح والأثر. لا يتحول إلى إثبات AGI إلا إذا حقق النظام معايير عامة مستقلة وطويلة الأفق ومقاومة للتلوث، وهو ما لا تدعيه الورقة.

# 7. النماذج المتنافسة وتحليل MCD/QMCD

لا تختار الورقة النموذج الأعلى إثارة مباشرة. تحتفظ بثلاثة نماذج في تراكب تفسيري: نموذج متمركز حول Base Model، ونموذج Pipeline وكيل، ونموذج نظام تشغيل معرفي دستوري. تنهار المقارنة لصالح أحدها فقط إذا قدم فائدة إضافية قابلة للقياس بعد ضبط التكلفة والتعقيد.

| النموذج | التعريف | القوة | الضعف | MCD تقديري | حالة QMCD |
|---|---|---|---|---|---|
| النموذج المحافظ: Model-Centric | يعزو معظم الأداء إلى Base Model ويضيف Prompt وأدوات محدودة. | اقتصادي وبسيط وسهل القياس. | ضعيف في الذاكرة والصلاحيات والتراجع، وينسب أخطاء النظام للنموذج. | 2.1/5 | مرتفع في البساطة، منخفض في التكامل. |
| النموذج التركيبي: Modular Agent Pipeline | يربط Planner وTools وMemory وVerifier في Pipeline. | عملي وقريب من أطر الوكلاء الحالية. | تسلسل ثابت، وحاكمية السؤال والدليل والذاكرة غير ممثلة صراحة. | 3.5/5 | الاحتمال التشغيلي الأعلى حاليًا. |
| النموذج الاختراقي: Constitutional Cognitive OS | يعامل النظام كبيئة تشغيل معرفية محكومة بالسؤال والكثافة والعقود والمحاكم والأثر. | يغطي التدريب وRuntime والتطور ويولد اختبارات جديدة. | عالي التعقيد، يحتاج تجارب Ablation، وخطر إعادة التسمية والتضخم. | 4.2/5 قبل الهدم؛ 4.0 بعد تخفيض الادعاء | مرشح مشروط، لا حقيقة مثبتة. |

## 7.1 تفسير الدرجات

يحصل النموذج المحافظ على اقتصاد نظري لكنه يلتقط فروقًا أقل. يرتفع النموذج التركيبي لأنه يدمج الأدوات والذاكرة، لكنه لا يحدد علاقات الحاكمية. يحصل النموذج الاختراقي على أعلى درجة أولية لشموله وتوليده، ثم تخفض الدرجة بعد محكمة الهدم بسبب التعقيد وغياب الدليل التجريبي. الاختيار النهائي هو نموذج ثالث معدل، لا لأنه ثبت، بل لأنه يولد أكبر مجموعة من الاختبارات التي قد تثبته أو تهدمه.

# 8. النظرية النهائية: نظرية الحاكمية الكثافية الفتّاحة

**التعريف المركز: إطار يفسر كفاءة النظام الوكيلي بوصفها خاصية ناشئة من حاكمية السؤال وتخصيص الموارد بحسب الفرق المعرفي، وتشغيل المعرفة والقدرات عبر عقود، والتحقق المستقل من الأثر، تحت شروط الصلاحية والتراجع؛ وتفشل النظرية إذا لم يتفوق هذا التنظيم على Baselines أبسط بعد ضبط التكلفة والتعقيد.**

## 8.1 الأطروحة الجوهرية

تقول النظرية إن المكوّن المفقود في كثير من الأنظمة ليس مزيدًا من القدرة الخام، بل تنظيم العلاقة بين القدرة والمعرفة والدليل والفعل. يمكن لنموذج قوي أن يبدد حوسبته في سؤال غير محكوم، ويمكن لذاكرة غنية أن تزيد الخطأ إذا تلوثت، ويمكن لمحقق لغوي أن يصدق تنفيذًا لم يحدث. لذلك تصبح الحاكمية نفسها متغيرًا قابلًا للتصميم والاختبار.

## 8.2 القوانين التشغيلية

1. قانون أسبقية السؤال: لا Routing قبل تثبيت السؤال الحاكم والفرق النشط.
2. قانون تناسب الدليل والفعل: سقف الفعل لا يتجاوز قوة الدليل والصلاحية وقابلية التراجع.
3. قانون التفعيل المتفرق: لا تستدعى مكونات لا يتوقع منها فرق معرفي أو خفض خطر.

4. قانون حجر الذاكرة: لا تدخل دعوى جديدة Canonical Memory قبل المصدر والتعارض والنطاق.

5. قانون عدم الحكم الذاتي: المنفذ لا يعلن النجاح عالي الأثر دون تحقق مستقل أو Postcondition.

6. قانون الترقية المشروطة: المسار الناجح يولد Candidate Capability، لا قدرة مستقرة مباشرة.

7. قانون الإغلاق العالمي: لا تكتمل الدورة دون مقارنة التوقع بالأثر وتحديث الذاكرة على أساس الفرق.

### 8.3 الصياغة شبه الرياضية

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

حيث D للفروق، I للتكامل، T للنقل والتحولات، A للحاكمية والوكالة، P للتنبؤ والاختبار، C للضغط وإعادة التوليد أو التصحيح، وN للضوضاء والادعاء الزائد. الأوزان سياقية ولا تعطى دقة زائفة.

$$Q\text{-}MCD = G(Q) \times [w_1 Dq + w_2 Fq + w_3 Bq + w_4 Eq + w_5 Aq - w_6 Nq]$$

G(Q) قوة السؤال الحاكم، وDq الفرق الذي يفتحه، وFq قابلية التنفيذ، وBq وضوح الحدود، وEq صلته بالدليل، وAq صلته بالفعل، وNq غموضه وضوضاؤه.

$$RAAU = (VWE \times Transfer \times Calibration \times Reversibility) / (TokenCost + Latency + OperationalRisk)$$

RAAU هو مقياس أولي للمنفعة الوكيلية المعدلة بالمخاطر. لا يستخدم كقيمة مطلقة، بل للمقارنة داخل مهمة وبيئة محددتين.

$$CM\text{-}Density = mean_j[(Score_{overlay,j} - Score_{base,j}) / AddedCost_j]$$

تقيس CM-Density مقدار التحسن المنسوب مبدئيًا إلى Overlay عبر نماذج متعددة، مع ضرورة Ablation لفصل أثر كل مكوّن.

## 9. المعمارية المتكاملة ذات الطبقات الاثنتي عشرة

| الطبقة الماكروية | وظيفتها |
|---|---|
| I. الدستور والهوية | الهوية، الأهداف، قصد المستخدم، عزل المشاريع، الصلاحيات، التراجع، وحد AGI. |
| II. حوكمة السؤال | توليد السؤال، Q-MCD، استخراج الفرق، كشف الافتراض والغموض والمجهولات. |
| III. الكثافة المعرفية | HAKIM-MCD، MCD، الوعائية، تعدد المقاييس، وفتح فضاء المشكلة. |
| IV. الدليل والنزاهة | تصنيف الادعاء، ترتيب الدليل، المصدر، التناقض، الحداثة، وترخيص الفعل. |
| V. HyperKnowledge والذاكرة | KB-as-OS، أنواع الذاكرة، الحجر، إعادة البناء، GraphRAG، ROFSC، سجل الأصول. |
| VI. القدرات وHyperSkills | Registry، عقود المدخلات والمخرجات، التركيب، الترقية، الإهمال، والنقل. |
| VII. HyperLoops والتنفيذ | Checkpoints، Triggers، الميزانيات، عدم التقدم، الأدوار، الأدوات، والأثر. |
| VIII. هندسة التدريب | البيانات، Tool/Agent Training، Continual Learning، Reward/Process Supervision، DPO، SFT. |
| IX. المحاكم والهدم | التعريف، الحدود، الدليل، الآلية، القياس، التنفيذ، الأمن، الخصوصية، الهلوسة، الصيانة. |
| X. التقييم | قدرات منفردة، نظام مركب، Regression، Cost-adjusted metrics، Calibration، Transfer. |
| XI. النشر والتشغيل | Rollback، Observability، Canary، Dry Run، Escalation، Hybrid Cloud/Local، Sparse Routing. |
| XII. التطور المستمر | اكتشاف القدرة، توليد مهارات وحلقات، توسيع KB، تصحيح Theory↔Tool Feedback، World Model. |

### 9.1 تدفق التشغيل

**Runtime Flow: Request → Constitution/Identity/Authority → Question Genesis/Q-MCD → Active Difference/MCD → Sparse → HyperRouting → Knowledge/Memory/Model/Skill/Tool Resolution → Bounded HyperLoop → Independent Courts Authorized Action → World-Effect → Self-Reconstruction Memory → Capability Promotion/Repair/Deprecation → Higher Governing Question.**

### 9.2 علاقة طبقات التدريب بالRuntime

الطبقات الـ33 الخاصة بالتدريب لا تختفي، لكنها توضع في موضعها الصحيح. Data Provenance وCurriculum وSynthetic Data وSFT وDPO وProcess Supervision وContinual Learning وTool Use وWorld Models وSafety Training تحسن النموذج والمكونات المتعلمة. أما Runtime Plane فيحدد ما إذا كانت القدرة ستستدعى، وبأي معرفة، وتحت أي صلاحية، وكيف تتحقق. Evolution Plane يحول الأثر إلى اقتراحات تحديث تمر من جديد عبر التدريب أو الذاكرة أو المهارة.

## 10. النظريات الكبرى والفرعية والمفاهيم والأدوات

### 10.1 النظريات الكبرى المولدة

| النظرية | الأطروحة | الاختبار الأولي |
|---|---|---|
| نظرية أسبقية السؤال الحاكم | لا يجوز تخصيص الموارد أو اختيار الأداة قبل تحديد السؤال الذي يضبط الفرق والدليل والنجاح. | اختبار A/B بين وكلاء متساويي الميزانية مع/بدون Question Governance، بقياس إعادة العمل والأخطاء. |
| نظرية الوعائية التفاضلية للحوسبة | تتحسن الكفاءة عندما تخصص الموارد بحسب القيمة المعرفية الهامشية والخطر، لا بالتوزيع المتساوي. | مقارنة Router وعائي مع Router ثابت عبر تكلفة ونجاح ومعدل استدعاءات غير مفيدة. |
| نظرية الذاكرة المحجورة المعاد بناؤها | لا تدخل المعرفة Canonical Memory بالاسترجاع أو التوليد وحدهما، تمر بالحجر والمصدر والتعارض ثم تعاد بنية الحالة. | حقن ادعاءات متناقضة وقياس التلوث والاسترجاع الصحيح بعد جلسات متعددة. |
| نظرية القدرة التعاقدية المركبة | تزداد قابلية نقل المهارة عندما تمثل بعقد مدخلات ومخرجات وصلاحيات واختبارات بدل Prompt غير مصنف. | نقل HyperSkill بين ثلاثة نماذج وبيئتين ومقارنة الالتزام بالعقد. |
| نظرية إغلاق الأثر العالمي | لا يكتمل التعلم الوكيلي حتى يقارن الأثر الفعلي بالتوقع ويحدث الذاكرة بناء على تحقق مستقل. | مهام ذات Postconditions خارجية ومقارنة إعلان الوكيل بنتيجة البيئة. |
| نظرية الكثافة العامة العابرة للنماذج | جزء من القدرة العامة قد يقيم في البنية الخارجية ويظهر كتَحسن متكرر عبر نماذج متعددة. | دراسة Factorial تفصل أثر النموذج والذاكرة والتوجيه والمحاكم والأدوات. |

### 10.2 النظريات الفرعية

| النظرية الفرعية | وظيفتها |
|---|---|
| نظرية الفرق الفاعل | المعلومة ذات القيمة هي التي تغير القرار أو خطة الاختبار. |
| نظرية الوعاء المفاهيمي | فشل الحل قد ينتج من اختيار تمثيل خاطئ للمشكلة لا من نقص المعرفة. |
| نظرية أقل فعل غير قابل للتراجع | يفضل النظام الفعل الذي يحقق المقصد بأقل التزام دائم. |
| نظرية لا منفذ بلا محقق | الأثر عالي المخاطر يحتاج حكمًا مستقلًا عن منفذه. |
| نظرية حفظ الاعتراض | الخلاف المفيد يبقى في الذاكرة حتى يفصل دليل، ولا يذاب في إجماع لغوي. |
| نظرية عتبة عدم التقدم | الحلقة توقف عندما ينخفض التحسن الهامشي تحت عتبة محددة. |
| نظرية بقايا التعلم الدنيا | يحفظ من التنفيذ أقل مبدأ قابل لإعادة الاستخدام مع مصدره وحدوده. |
| نظرية الترقية المشروطة | النجاح الواحد يولد Candidate لا Stable Capability. |

| النظرية الفرعية | وظيفتها |
|---|---|
| نظرية عزل الهوية المشروعانية | كل مشروع يمتلك دستورًا وذاكرة وأصولًا لا تنتقل افتراضيًا. |
| نظرية الضغط القابل للفك | الملخص الجيد يحمل مؤشرات تسمح باسترداد الدليل والحدود. |
| نظرية انهيار الذاكرة بلا مصدر | المعلومة غير القابلة للتتبع لا يجوز أن تحكم فعلًا عالي الأثر. |
| نظرية السؤال الأعلى | كل نتيجة موثقة يجب أن تزيد جودة السؤال اللاحق لا حجم الذاكرة فقط. |

## 10.3 عشرون مفهومًا جديدًا أو معاد ضبطه

| المفهوم | التعريف | الوظيفة | الرتبة |
|---|---|---|---|
| السؤال الحاكم | سؤال يحدد الهدف والفرق والدليل والحدود والفعل. | تفسيري/حاكمي | تركيب نظري |
| الفرق النشط | تمييز إذا تغير تغير القرار أو المسار. | قياسي | أداة تحليلية |
| الكثافة المعرفية | منفعة معرفية تقديرية بعد خصم الضوضاء. | قياسي | أداة قيد المعايرة |
| وعاء المهمة | التمثيل المفاهيمي والإجرائي الذي تستوعب فيه المشكلة. | تفسيري | تركيب |
| التروية الحاسوبية | تخصيص الموارد لعقد معرفية وفق القيمة والخطر. | تشغيلي | استعارة منتجة/خوارزمية محتملة |
| الإقفار الحاسوبي | فشل مسار مهم بسبب نقص الموارد أو السياق. | تشخيصي | فرضية تشغيلية |
| الاحتقان الحاسوبي | هدر موارد على مسار مكرر أو منخفض العائد. | تشخيصي | فرضية تشغيلية |
| المعرفة التشغيلية | معرفة تحمل شروط التطبيق والأداة والاختبار والصلاحية. | إجرائي | تعريف معاد ضبطه |
| الذاكرة المحجورة | منطقة حفظ للمعرفة غير المرقاة إلى Canonical. | أمني/إبستيمي | مقترح هندسي |
| إعادة بناء الحالة | إصدار حالة معرفية جديدة من الأدلة والتعارضات. | إجرائي | مقترح |
| عقد القدرة | مواصفة تحدد مدخلات ومخرجات وصلاحيات واختبارات المهارة. | هندسي | أداة |
| مخلفات التعلم | أصغر معرفة موثقة يعاد استخدامها بعد تنفيذ. | إجرائي | مفهوم مقترح |
| حد عدم التقدم | عتبة توقف عندما لا يزداد الإنجاز أو MCD. | قياسي | فرضية |
| استقلال التحقق | مقدار اختلاف المحقق عن المنفذ في النموذج والدليل والسياق. | قياسي | مفهوم قابل للقياس |
| ترخيص الدليل | الحد الأعلى للفعل الذي تسمح به جودة الدليل. | معياري/تشغيلي | تركيب |
| الأثر العالمي الموثق | تغير خارجي تؤكده Postconditions أو قياس مستقل. | ناتج | تعريف تشغيلي |
| كثافة القدرة العابرة للنماذج | تحسن متكرر يسببه Overlay عبر نماذج مختلفة. | قياسي | فرضية |
| الانتشار الفاشل مغلقًا | رفض النشر عند نقص شرط حاكم. | أمني | مبدأ هندسي |
| فتح فضاء المشكلة | توليد متغيرات وأسئلة وحدود جديدة بدل تحسين جواب ثابت. | توليدي | مفهوم بحثي |
| القدرة الوكيلية المعدلة بالمخاطر | منفعة موثقة مقسومة على التكلفة والمخاطر وعدم القابلية للتراجع. | قياسي | مقياس مقترح |

## 10.4 الأدوات المنهجية والتطبيقية

| الأداة | النقص المعالج | المدخلات | المخرجات |
|---|---|---|---|
| بطاقة السؤال الحاكم Q-Card | تحويل الطلب إلى هدف وفرق وحدود ودليل ومعيار نجاح. | الطلب، سياق المشروع، المخاطر | سؤال حاكم وأسئلة فصل. |
| مصفوفة MCD | ترتيب مسارات البحث والتنفيذ. | مسارات مرشحة | درجات مفسرة وأولوية. |
| خريطة التروية الحاسوبية | توزيع Tokens والأدوات والنماذج. | MCD، عدم اليقين، الخطر، الكلفة | ميزانية لكل عقدة. |

| الأداة | النقص المعالج | المدخلات | المخرجات |
|---|---|---|---|
| سجل الادعاء والدليل | منع خلط الحقيقة والفرضية. | Claims ومصادر | Provenance وثقة وتعارض. |
| حجرة الذاكرة المقترحة | منع التلوث قبل الترقية. | معلومة جديدة | حالة Proposed/Verified/ Rejected. |
| قالب عقد HyperSkill | توحيد القدرات وقابليتها للنقل. | مهمة متكررة | مواصفة قابلة للاختبار. |
| محدد Bounded HyperLoop | منع الدوران واستنزاف الميزانية. | هدف وميزانية ومقياس تقدم | شروط استمرار وتوقف. |
| مصفوفة Evidence-to-Action | ربط الدليل بدرجة الفعل. | جودة الدليل وأثر الفعل | اقتراح/تجربة/تنفيذ/تصعيد. |
| محكمة الهدم متعددة الأبواب | اختبار التعريف والآلية والدليل والقياس والأمان. | نموذج أو مخرج | اعتراضات وحكم وخطة إصلاح. |
| سجل World-Effect | فصل الإعلان عن الأثر. | توقع، فعل، Postconditions | فرق متوقع/فعلي وآثار جانبية. |
| مسجل الأصول والنشر | تتبع KBs والمهارات والنماذج والإصدارات. | أصول النظام | خريطة تبعيات وRollback. |
| حزام تقييم Cross-Model | فصل أثر Overlay عن Base Model. | نماذج ومهام وAblations | مصفوفة انتقال وكلفة وسقف تقارب. |

## 10.5 المقاييس والمؤشرات

| المقياس | ما يقيسه | طريقة التقدير |
|---|---|---|
| QGI - مؤشر حاكمية السؤال | وضوح الهدف، الفرق، الحدود، الدليل، وقابلية التفنيد | 0–5 لكل بعد مع تفسير نوعي. |
| MCD-R - الكثافة المعرفية المفسرة | الفرق والتكامل والنقل والوكالة والتنبؤ والتصحيح ناقص الضوضاء | Rubric مرجح لا رقم مطلق. |
| KIR - نسبة نزاهة المعرفة | نسبة الدعاوى الفعالة القابلة للتتبع وغير المتناقضة دون وسم | Verified active claims / all active claims. |
| BLE - كفاءة الحلقة المحدودة | التقدم الموثق لكل دورة وتكلفة | ΔProgress أو ΔMCD / tokens+time. |
| VWE - الأثر العالمي الموثق | مطابقة الأثر الفعلي للتوقع والهدف | Postcondition score مع آثار جانبية. |
| CM-Density - الكثافة العابرة للنماذج | التحسن المتكرر لدOverlay عبر نماذج | متوسط فرق الأداء المعدل بالكلفة. |
| CRI - مؤشر القابلية للتراجع | نسبة الأفعال التي يمكن عكسها واستعادتها | Rollback coverage × recovery success / impact. |

# 11. الفرضيات البحثية والتنبؤات الفارقة

## 11.1 الفرضيات

| القوة | الفرضية | الآلية | الاختبار |
|---|---|---|---|
| قوية | إضافة تحقق Postcondition مستقل تخفض الادعاءات الكاذبة بإتمام المهام الوكيلية. | الوكيل قد يخلط بين نجاح الاستدعاء ونجاح الهدف. | مهام أدوات قابلة للفحص مع Baseline. |
| قوية | عقود HyperSkill تقلل أخطاء المدخلات والصلاحيات مقارنة بمهارات Prompt-only. | العقد تجعل الافتراضات والفشل مرئيين. | مقارنة عبر مهام ووكلاء. |
| قوية | Memory Quarantine تخفض استدعاء معلومات غير موثقة في جلسات لاحقة. | الفصل يمنع التحول الفوري إلى Canonical. | اختبار حقن تناقضات متعدد الجلسات. |
| قوية | Bounded HyperLoops تقلل التكلفة دون خفض النجاح عندما توجد إشارة تقدم موثوقة. | توقف الدورات منخفضة العائد. | مقارنة سقوف ثابتة وعتبة تقدم. |
| قوية | Evidence-to-Action Thresholds تقلل الأفعال عالية الأثر القائمة على دليل ضعيف. | ترخيص الفعل يصبح صريحًا. | سيناريوهات صلاحيات ومعلومات ناقصة. |
| متوسطة | Q-MCD المبكر يقلل عدد مرات إعادة التخطيط في المهام الغامضة. | كشف الغموض يسبق التنفيذ. | Benchmark تعليمات ناقصة. |

| القوة | الفرضية | الآلية | الاختبار |
|---|---|---|---|
| متوسطة | التوجيه الوعائي يتفوق على التوجيه المتساوي في Cost-adjusted success. | يركز الموارد على العقد الحرجة. | مهام متعددة الصعوبة مع ميزانية ثابتة. |
| متوسطة | استقلال المحقق متعدد المزودين يخفض الفشل المت correlated مقارنة بالمحقق من النموذج نفسه. | التنوع يقلل مشاركة الانحياز. | تجربة عبر مزودين ومصادر. |
| متوسطة | ROFSC يحافظ على قرارات حرجة أفضل من التلخيص الحر عند ضغط تاريخ مشروع طويل. | يحفظ الدليل والحدود والمؤشرات. | إعادة بناء قرارات من ملخصات. |
| متوسطة | فصل Identity Memory عن Episodic Memory يقلل تسرب تعليمات المشاريع. | التقسيم يعزل النطاق. | اختبار Cross-project leakage. |
| متوسطة | سجل الاعتراض يحسن المعايرة في المسائل الخلافية. | منع الإجماع المصطنع يحفظ عدم اليقين. | أسئلة بأدلة متعارضة. |
| متوسطة | مسار الترقية المشروطة يقلل Regression في المهارات المولدة تلقائيًا. | النجاح لا يعمم قبل اختبارات النقل. | سجل مهارات عبر إصدارات. |
| ضعيفة | MCD-R يتنبأ بجودة النظام أفضل من طول السياق أو عدد الوكلاء. | يقيس المنفعة لا الحجم. | دراسة ارتباط على مهام متنوعة. |
| ضعيفة | World-Effect feedback يولد HyperSkills أكثر قابلية للنقل من التعلم من Critique نصي فقط. | الأثر يربط القاعدة بالواقع. | مقارنة آثار حقيقية ومحاكاة. |
| ضعيفة | الـOS Constitutional Cognitive يقلل Incident Rate على المدى الطويل رغم زيادة التعقيد. | البوابات تمنع فئات متعددة من الخطأ. | دراسة تشغيل طولية. |
| تخمينية خصبة | يمكن تقريب سؤال حاكم عام قابل للنقل بين مجالات متعددة عبر Q-MCD. | توجد بنية فوق-مجالية للسؤال. | تعلم تمثيل للأسئلة واختبار النقل. |
| تخمينية خصبة | توجد عتبة تنظيمية تصبح عندها زيادة المكونات أقل فائدة من تحسين الحاكمية بينها. | التعقيد يخلق احتقانًا حاسوبيًا. | منحنى أداء مقابل عدد المكونات. |
| تخمينية خصبة | يمكن قياس جزء معتبر من "القدرة العامة" بوصفه خاصية System Overlay لا Base Model. | الذاكرة والأدوات والتوجيه تنقل القدرة. | Factorial cross-model study. |

## 11.2 التنبؤات الفارقة

| التنبؤ | النموذج المنافس | ما يهدمه |
|---|---|---|
| عند تثبيت الميزانية، سيحقق Router قائم على MCD نجاحًا أعلى من Router يشغل كل الأدوات بالتساوي. | النموذج الكثيف يتوقع أن مزيدًا من التفعيل لا يضر. | انخفاض النجاح أو عدم وجود فرق يهدم التنبؤ. |
| الذاكرة المحجورة ستظهر أخطاء أقل في تذكر الادعاءات المتناقضة من ذاكرة Append-only. | نموذج الذاكرة الخطية يتوقع عدم فرق بعد الاسترجاع الجيد. | تساوي معدلات التلوث. |
| الوكلاء ذوو Contracts سيظهرون أخطاء صلاحية أقل حتى عند تغيير Base Model. | Prompt-only يتوقع اعتمادًا أكبر على النموذج. | اختفاء فرق النقل. |
| المحقق المختلف في المصدر والنموذج سيكتشف أخطاء لا يكتشفها Self-critique المتجانس. | نموذج النقد الذاتي يرى كفاية النموذج نفسه. | عدم زيادة الاكتشاف. |
| سجل التوقع قبل التنفيذ سيكشف Post-hoc Rationalization أكثر من التقييم بعد التنفيذ فقط. | النموذج غير المغلق لا يتنبأ بفارق. | تطابق التقييمين. |
| التلخيص الحر سيفقد شروط حدود أكثر من ROFSC في قرارات طويلة الأفق. | نموذج الملخص العام يتوقع تكافؤًا. | عدم فرق في استرجاع الحدود. |
| الأفعال المصممة وفق Least Irreversible Action ستخفض تكلفة الاستعادة بعد الحوادث. | نموذج الصلاحية الثنائية لا يتوقع فرقًا. | تساوي تكلفة الحوادث. |
| تحسن Overlay سيتناقص عند النماذج الأضعف بعد نقطة سقف تقارب. | الادعاء القوي بالاستقلال عن النموذج يتوقع تحسنًا ثابتًا. | عدم ظهور سقف أو ظهور عكسه. |

| التنبؤ | النموذج المنافس | ما يهدمه |
|---|---|---|
| زيادة عدد المحاكم بعد حد معين ستخفض الكفاءة دون زيادة في السلامة، ما يدعم تفعيلها Sparse. | النموذج البيروقراطي يتوقع تحسناً رتيباً. | استمرار التحسن الرتيب. |

# 12. محكمة الهدم والنسخة المعدلة

شُغلت محكمة الهدم على أربعة مستويات: المادة الخام، والنماذج المتنافسة، والنظرية النهائية، وفكرة "الخرق" نفسها. الاعتراض المقبول لا يضاف للزينة؛ بل يغير رتبة الادعاء أو يضيف اختبارًا أو يحذف مكونًا.

| الاعتراض | مضمونه | نوعه | التعديل |
|---|---|---|---|
| إعادة تسمية | قد تكون MICAGFI تجميعًا لأفكار معروفة بأسماء جديدة. | قوي ومعدل | تخفض دعوى الأصالة إلى "تركيب حاكمي" وتحدد الإضافات القابلة للاختبار: Q-MCD، ترخيص الدليل، الذاكرة المحجورة، Cross-Model Density. |
| ذاتية المقاييس | MCD وQ-MCD Rubrics قابلة لانحياز المقيم. | قوي | اعتماد مقيمين متعددين، تعريف أمثلة مرساة، واختبار الثبات والصدق التنبئي. |
| التعقيد المفرط | 176 طبقة قد تزيد الفشل والتكلفة. | قاتل إن شغلت كلها | تحويلها إلى Registry لا Pipeline ثابت، وتفعيل Sparse حسب المهمة والخطر. |
| ترابط فشل المحاكم | محققون من النموذج نفسه قد يعيدون الخطأ. | قوي | قياس استقلال التحقق وتنويع المصدر والنموذج واستخدام Verifiers حتميين عند الإمكان. |
| ذاكرة تعيد كتابة التاريخ | Self-Reconstruction قد تحذف أدلة مزعجة. | قوي | حفظ Historical States وEvidence Ledger ومنع الحذف غير القابل للتتبع. |
| كلفة الأثر العالمي | قياس العالم صعب أو مكلف أو غير أخلاقي. | محدود | استخدام Simulation وDry Run وCanary وProxies مع التصريح بحدودها. |
| غموض الدستور | القيم قد تتعارض ولا توجد دالة وحيدة للتحكيم. | قوي | Policy hierarchy، سجل قرار، تصعيد بشري، ومنع ادعاء حل فلسفي عام. |
| وهم النقل | المهارة قد تعتمد على خصائص مزود محدد. | قوي | Adapters واختبارات Contract conformance وسقف تقارب لكل نموذج. |
| التعلم الخارجي ليس تعلمًا حقيقيًا | تحديث الذاكرة لا يساوي تغيير الأوزان. | صحيح جزئيًا | التمييز الصريح بين Parametric Learning وNon-parametric Runtime Adaptation. |
| لا دليل على AGI | الورقة تستخدم AGI في الاسم دون برهان. | قاتل للادعاء لا للإطار | AGI-Claim Boundary وتحويل المصطلح إلى أفق بحث لا نتيجة. |
| مخاطر التوسع الذاتي | توليد المهارات والحلقات قد يزيد الصلاحيات تدريجيًا. | قوي | الترقية المشروطة، عدم وراثة صلاحيات جديدة، وFail-Closed Deployment. |
| التحسين للقياس | قد يتعلم النظام تعظيم MCD أو VWE شكليًا. | قوي | مقاييس متعددة، Holdouts، محاكم Anti-Gaming، وتقييم بشري للأثر. |

## 12.1 النسخة المعدلة بعد الهدم

بعد الهدم، لم تعد الورقة تدعي أن الطبقات الـ176 تمثل وصفة مكتملة أو أن MICAGFI يحقق AGI. أصبحت الأطروحة أضيق وأقوى: مجموعة علاقات حاكمية قابلة للاختبار قد تحسن موثوقية وكفاءة الأنظمة الوكيلية المركبة، ويجب أن تثبت فائدتها مقابل Baselines أبسط وتحت ميزانيات متساوية. خُفض MCD من "مقياس للذكاء" إلى Rubric للمنفعة المعرفية، وخُفض HyperDeployment الأسي إلى إمكان انتشار تركيبي مشروط، وفُصل التعلم الذاكري عن التعلم البارامتري، وأضيف استقلال التحقق ومقاومة Gaming بوصفهما شرطي صلاحية.

### 12.2 نظرية ناقصة شريفة

ما نعرفه: الأنظمة الوكيلية تحتاج إلى ذاكرة وأدوات وتحقق وصلاحيات، وتواجه أخطارًا لا تظهر في النموذج النصي وحده. ما نرجحه: تنظيم هذه العناصر بعقود وحاكمية سؤال ودليل وتراجع يحسن الأداء المعدل بالمخاطر. ما نجهله: حجم التحسن، والحد الأدنى من الطبقات، ومدى نقل النتائج بين النماذج والمجالات. الحلقة المفقودة هي Benchmark متكامل يقيس السؤال والذاكرة والصلاحية والأثر في المهمة نفسها. أقرب اختبار حاسم هو دراسة Ablation متعددة النماذج على بيئة تفاعلية قابلة للاستعادة.

## 13. بروتوكولات الاختبار والتقييم

### 13.1 تصميم تجريبي أدنى

1. ثلاثة Base Models على الأقل: قوي، متوسط، ومحلي أصغر.
2. أربع حالات: Base فقط؛ Pipeline معياري؛ MICAGFI مصغر؛ MICAGFI كامل للمهمة.
3. مهام بحث، برمجة، ذاكرة متعددة الجلسات، واستخدام حاسوب أو أدوات قابلة للفحص.
4. ميزانية Tokens ووقت وTool Calls متساوية أو معدلة بوضوح.
5. Ablation لكل من World-Effect، Independent Verifier، Memory Quarantine، Q-MCD، وRollback.
6. قياس نجاح نهائي، تقدم جزئي، Leakage، Incident Rate، Calibration، كلفة، وقابلية التراجع.
7. Holdout خاص لمنع تصميم الإطار على الاختبار نفسه.

### 13.2 Benchmarks مناسبة وحدودها

يمكن استخدام AgentBench لبيئات متعددة، وGAIA للبحث والأدوات، وSWE-bench للإصلاح البرمجي القابل للاختبار، وOSWorld لاستخدام الحاسوب، وAgent-SafetyBench لسلوك السلامة [11–15]. لكنها لا تقيس وحدها جودة السؤال الحاكم أو نزاهة الذاكرة أو Rollback. لذلك يلزم Benchmark مركب جديد، وليكن MOUSE-Arena، يتضمن تعليمات غامضة، ومعلومات متناقضة، وأفعالًا متفاوتة الأثر، وحالة قابلة للاستعادة، واختبارات Postconditions.

### 13.3 معايير النجاح والفشل

تنجح النظرية جزئيًا إذا تفوق MICAGFI المصغر على Pipeline معياري في Verified Task Completion وIncident Rate وCost-adjusted utility، وبقي التحسن عبر نموذجين على الأقل. تفشل دعوى الحاكمية إذا لم يختلف الأداء بعد إضافة الطبقات أو إذا كانت الكلفة أعلى بلا خفض خطر. تفشل دعوى النقل إذا احتاجت كل بيئة إلى إعادة كتابة جوهر المهارة. وتفشل دعوى الذاكرة إذا زادت إعادة البناء فقدان المصدر أو التناقض.

## 14. البرنامج البحثي المستقبلي

| الأفق | الخطة |
|---|---|
| سنة واحدة | بناء Prototype من 20–30 طبقة حرجة؛ HyperSkill Contracts، Memory Quarantine، Claim Ledger، MCD Router، Q-Card، World-Effect Log، Verifier، Bounded Loop. اختباره على GAIA/SWE-bench/مهام داخلية، ونشر ورقة تعريفية وDataset للتلوث الذاكري. |
| ثلاث سنوات | تجارب Cross-model وCross-provider، تطوير Benchmarks للتراجع وعزل المشاريع والذاكرة، دراسات Ablation، بناء Registry مفتوح، وتجارب في البحث والبرمجة وإدارة الوثائق. |
| عشر سنوات | تحويل البرنامج إلى معيار معماري قابل للتدقيق، مختبر متعدد المؤسسات، قاعدة معرفة طولية، بروتوكولات تقييم AGI-readiness دون إعلان AGI، وتطبيقات عالية المسؤولية تحت إشراف بشري. |

## 14.1 خمسون محورًا بحثيًا

| الفئة | المحور | السؤال | المنهج | القيمة |
|---|---|---|---|---|
| قريبة | معايرة Q-MCD | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | اختبار MCD Router | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | Benchmark التلوث الذاكري | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | قياس استقلال المحقق | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | عقود HyperSkill | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | عتبة عدم التقدم | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | قياس Postconditions | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | عزل المشاريع | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | تقييم ROFSC | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قريبة | Fail-Closed Deployment | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | Cross-provider portability | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | Longitudinal memory | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | Agent disagreement | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | Constitution conflict resolution | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | Tool injection defense | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | Cost-adjusted intelligence | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | Capability promotion | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | World-effect proxies | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| متوسطة | GraphRAG federation | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |

| الفئة | المحور | السؤال | المنهج | القيمة |
|---|---|---|---|---|
| متوسطة | Human utility | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Open problem-space generation | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Cross-model generality | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Non-parametric continual learning | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Institutional agent governance | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Collective human-machine intelligence | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Formal authorization logic | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Causal world models | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Self-reconstruction proofs | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | Architecture-level interpretability | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| بعيدة | AGI-readiness longitudinal study | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| مقارنة | MICAGFI مقابل ReAct pipelines | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| مقارنة | M-HKOS مقابل MemGPT | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| مقارنة | ROFSC مقابل RAPTOR | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| مقارنة | HyperSkill مقابل tool schemas | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| مقارنة | Courts مقابل self-critique | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| تطبيقية | وكيل بحث علمي | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| تطبيقية | وكيل هندسة برمجيات | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| تطبيقية | وكيل إدارة معرفة مؤسسية | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| تطبيقية | وكيل عمليات آمن | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |

| الفئة | المحور | السؤال | المنهج | القيمة |
|---|---|---|---|---|
| تطبيقية | وكيل تعليم وتدريب | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قياسية/تجريبية | صدق MCD | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قياسية/تجريبية | ثبات QGI | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قياسية/تجريبية | معايرة VWE | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قياسية/تجريبية | مقاومة Gaming | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| قياسية/تجريبية | Ablation 176-layer registry | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| اختراقية بين التخصصات | نظرية التحكم والحاكمية المعرفية | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| اختراقية بين التخصصات | علم المناعة ومحاكم الخطأ | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| اختراقية بين التخصصات | أنظمة التشغيل والذاكرة الوكيلية | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| اختراقية بين التخصصات | نظرية المؤسسات والوكلاء | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |
| اختراقية بين التخصصات | الاقتصاد الجزئي لتخصيص الحوسبة | ما الآلية القابلة للقياس؟ | تجربة مقارنة أو محاكاة أو دراسة طولية | تحديد حدود الإطار وتحويله إلى علم قابل للتفنيد. |

# 15. تحويل النظرية إلى نواة قاعدة معرفة

## 15.1 الكيانات الأساسية

| الكيان | الحقول الأساسية |
|---|---|
| Constitution | مبادئ وصلاحيات وترتيب أولويات وإصدارات. |
| ProjectIdentity | اسم المشروع، النطاق، الأصول، الأسرار، المعايير. |
| GoverningQuestion | السؤال، الفرق، الحدود، الفاصل التجريبي. |
| Claim | النوع، النص، الثقة، الحالة، المجال الزمني. |
| Evidence | المصدر، المباشرة، الاستقلال، الحداثة. |
| HyperSkill | العقد، التبعيات، الأدوات، الاختبارات، الإصدار. |
| HyperLoop | المشغل، الميزانية، الحالة، شروط التوقف. |
| AgentRole | Planner/Executor/Verifier/Governor وصلاحياته. |
| ToolCall | الأداة، Arguments، النتيجة، Postconditions. |
| MemoryItem | النوع، المصدر، Canonical/Quarantine، التعارض. |
| WorldEffect | التوقع، القياس، الأثر، الآثار الجانبية. |

| الكيان | الحقول الأساسية |
|---|---|
| Deployment | البيئة، الإصدار، البوابات، Rollback. |

### 15.2 العلاقات والقواعد

- GoverningQuestion يحدد ActiveDifference ويستدعي EvidenceRequirements.
- Claim مدعوم بـEvidence، وقد يتعارض مع Claim آخر، ويُرقّى أو يُحجر أو يُهمَل.
- HyperSkill يعتمد على HyperSkills وأدوات وAuthorityScopes ويُختبر عبر VerificationProfile.
- HyperLoop يشغل HyperSkills ضمن Budget ويولد Checkpoints وWorldEffects.
- WorldEffect يؤكد أو يضعف Claim ويحدث MemoryItem أو يولد سؤالًا أعلى.
- Deployment ينشر إصدارات محددة ويرتبط بخطة Rollback وحالة Monitoring.
- إذا غاب Provenance عن Claim عالي الأثر، فلا يسمح له بترخيص فعل.
- إذا انخفض Progress تحت العتبة لدورتين، توقف الحلقة أو تصعد.
- إذا فشل Capability في Regression Suite، ينقل إلى Quarantine ولا يحذف سجله.
- إذا تعارض دستوران أو نطاقا مشروعين، يفشل الطلب مغلقًا أو يصعد للبشر.

### 15.3 مثال Schema مختصر

```
HyperSkill Record: id; name; version; activation_signature; input_contract; output_contract; evidence_requirements;
dependencies; tools; authority_scope; budget; stop_rules; verification_profile; memory_residue; test_results; status; owner
supersedes.
```

## 16. سجل البناء المعرفي

| البند | الخلاصة |
|---|---|
| الافتراضات | أن المستخدم قدم المادة المعمارية كاملة بما يكفي لبناء إطار، وأن الأدبيات المستخدمة مرجع للمقارنة لا إثبات مباشر للمفاهيم المساعفية. |
| المادة الصلبة | الحاجة إلى فصل النموذج عن النظام، والتحقق من الأدوات، وإدارة الذاكرة والصلاحيات والمخاطر. |
| المادة المخفضة | ادعاءات AGI، والأُسية، وعمومية النقل، ودقة MCD الرقية. |
| المادة المدمجة | طبقات متكررة في النزاهة والمحاكم والمراقبة دُمجت ضمن عقود وبوابات وسجلات. |
| المفهوم المركزي | الحاكمية الكثافية: لأنها تضبط العلاقة بين السؤال والموارد والدليل والفعل. |
| النماذج البديلة | Model-Centric وModular Pipeline وConstitutional Cognitive OS. |
| سبب الاختيار | النموذج الثالث يولد اختبارات أكثر ويغطي الفجوات، لكن اختير بصيغة متواضعة مشروطة. |
| أضعف حلقة | معايرة المقاييس الذاتية وإثبات أن التعقيد الإضافي يستحق كلفته. |
| أقوى نقطة | ربط الدليل بمستوى الفعل وإغلاق الحلقة بالأثر والذاكرة. |
| الخطوة التالية | Prototype مصغر وBenchmark للتلوث والتراجع مع Ablation متعددة النماذج. |

### 16.1 اختبار الجودة النهائي

| المعيار | الدرجة | التفسير |
|---|---|---|
| حفظ المادة الأصلية | 9/10 | حُفظت الطبقات والمفاهيم ودمج التكرار. |
| الوضع الأنطولوجي | 5/5 | عُرف كنظام اجتماعي-تقني متعدد المستويات. |
| مراتب المعرفة | 9/10 | فصلت الدعاوى، مع حاجة لمزيد من البيانات. |
| تصفية المادة | 9/10 | خُفضت الادعاءات الزائدة. |
| خريطة المتغيرات | 9/10 | تحتاج نمذجة سببية لاحقة. |
| النماذج المتنافسة | 9/10 | ثلاثة نماذج ومقارنة مشروطة. |
| أصالة النظرية | 12/15 | أصالة تركيبية لا تاريخية مثبتة. |
| قابلية الاختبار | 9/10 | فرضيات وتنبؤات وAblations. |
| محكمة الهدم | 10/10 | اعتراضات قاتلة ومعدلة. |
| الأدوات والمقاييس | 5/5 | سبعة مقاييس و12 أداة. |
| غياب الحشو | 5/5 | ضغطت التكرارات. |

النتيجة قبل التحسين: 84/100، بسبب تضخم ادعاء AGI، وغياب الفصل بين التدريب وRuntime، وضعف تعريف النقل. التحسين: تخفيض الرتبة، إضافة ثلاثة مستويات، وتعريف Baselines وAblations وحدود الفشل. النتيجة بعد التحسين: 91/100. تبقى الدرجة تقديراً تحريرياً لا تقييماً خارجياً.

## 17. الخاتمة

أصبحت لدينا بعد الضغط والهدم نظرية معمارية لا قائمة ميزات. مركزها أن السؤال والدليل والقدرة والذاكرة والفعل يجب ألا تعمل في مستوى واحد أو بلا ترتيب حاكم. يقترح MICAGFI–M-HKOS أن السؤال الحاكم يحدد الفرق، وأن MCD يوجه الموارد، وأن Sparse HyperRouting يختار الحد الأدنى الكافي، وأن M-HKOS يوفر بيئة معرفة تشغيلية، وأن HyperSkills تحول الإجراءات إلى عقود، وأن Bounded HyperLoops تمنع الدوران، وأن المحاكم تقاوم الهلوسة والتضخم، وأن الصلاحية والتراجع يحدان الوكالة، وأن World-Effect يغلق دائرة التعلم، وأن Self-Reconstruction Memory تحفظ ما تحقق لا ما قيل.

القيمة الأقوى لهذا الإطار ليست في ادعاء أنه "عقل عام" جاهز، بل في إمكان تحويله إلى برنامج تجريبي: ما الطبقات التي تضيف فرقًا؟ متى يصبح التنظيم عبئًا؟ ما الذي ينتقل بين النماذج؟ كيف يقاس الأثر الموثق لكل Token ومخاطرة؟ وكيف يمكن لنظام أن يتطور دون أن يمنح نفسه صلاحيات أو ذاكرة لم يثبت استحقاقها؟ إذا أجابت التجارب عن هذه الأسئلة، يمكن أن يتحول الإطار إلى مدرسة هندسة معرفية وكيلة؛ وإذا لم تفعل، يبقى لغة تنظيمية مفيدة يجب ألا تقدم كنظرية مثبتة.

**ماذا أصبح لدينا؟: برنامج بحثي موحد يربط 176 طبقة في ثلاث دوائر و12 طبقة ماكرو، ونظرية مركزية، ونظريات فرعية، ومفاهيم وأدوات ومقاييس، وفرضيات وتنبؤات، ومحكمة هدم، وخطة تحويل إلى Prototype وKnowledge Base وBenchmark.**

# الباب الثاني: الإغلاق الحاكمي التركيبي — الجديد النظري الخالص

## 18. الفجوة بعد اكتمال الطبقات: من اكتمال المكونات إلى اكتمال العلاقات

تعطي الطبقات الـ176 السابقة تغطية واسعة لدورة حياة النظام: الدستور والهوية، حوكمة السؤال، MCD وQ-MCD، الأدلة، M-HKOS والذاكرة، HyperLoops، HyperSkills والوكلاء، تدريب النموذج، المحاكم، التقييم، النشر، والتطور. غير أن اكتمال القائمة لا يساوي اكتمال النظام. فالمكوّن قد يكون صحيحًا محليًا ويصبح خاطئًا عند نقله إلى مكوّن آخر؛ وقد تكون المعلومة صحيحة لكن استخدامها السببي غير مشروع؛ وقد يكون الفعل مصرحًا تقنيًا لكنه غير مشروع مؤسسيًا؛ وقد تكون الذاكرة واسعة لكنها شديدة الإنتروبيا؛ وقد تنجح المهارة في الاختبار لكنها غير ناضجة للاستقلال.

تسمي الورقة هذا النقص **فجوة الإغلاق الحاكمي**: غياب طبقة تجعل الانتقالات بين المعرفة والقدرة والسلطة والأثر موضوعًا مستقلًا للتعاقد والقياس والهدم. «الإغلاق» لا يعني إغلاق النظام أمام التعلم أو الجدة، بل يعني أن كل انتقال حرج يغلق حلقة مسؤوليته: من أين جاء الادعاء؟ ما نوعه؟ ما درجة رسوخه السببي؟ من يثق به؟ ما العقد الذي سمح بتحويله إلى فعل؟ من يملك السلطة؟ ماذا حدث؟ وكيف عادت النتيجة إلى الذاكرة؟

يمكن ضغط الفجوة في المعادلة المفهومية الآتية:

**× System Soundness = Component Soundness × Transition Soundness × Authority Legitimacy World Verification**

إذا اقترب أحد العوامل من الصفر، لا تعوضه كثرة الأدوات أو قوة النموذج. لذلك لا تضيف النظريات التالية «ميزات» متجاورة، بل تحاول ضبط أربعة أنواع من التحول: تحول الجزء إلى كل، وتحول الارتباط إلى سبب، وتحول الثقة إلى ترخيص، وتحول الأداء إلى استقلال.

## 19. نظرية الضمان التركيبي الحاكمي

**Moussaif Compositional Assurance Theory — M-CAT**

### الفجوة التي تكشفها

لا تنتقل سلامة الجزء تلقائيًا إلى سلامة النظام المركب. قد تعمل كل HyperSkill أو أداة ضمن عقدها، لكن انتقال ناتجها إلى مكوّن آخر قد يرفع مستوى الفعل أو يخفض شرط الدليل أو يسرّب صلاحية لم تكن مقصودة. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية الضمان التركيبي الحاكمي بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **كل تركيب بين مكوّنين أو أكثر هو كيان حاكمي جديد يجب تقييمه مستقلاً، لأن خصائصه ومخاطره قد لا تكون موجودة في أي جزء منفرد.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

يمثل كل مكوّن بعقد يضم شروط الدخول، وضمانات الخروج، ونوع الادعاء، ومستوى الدليل، ونطاق الصلاحية، وأثر الذاكرة، وقابلية التراجع، وحالات الفشل. ينشئ محرك التركيب عقدًا مشتقًا ويختبر توافق المخرجات مع المدخلات، وعدم تصاعد الصلاحية، وعدم فقدان Provenance، وعدم تكوين حلقة تنفيذ خفية. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime لملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

إذا كانت C_i عقود المكونات و⊗ مؤثر التركيب، فإن صلاحية النظام لا تساوي مجموع صلاحيات الأجزاء: A(C_1⊗...⊗C_n) ≠ ΣA(C_i). تقترح النظرية أن الضمان التركيبي CA يساوي الحد الأدنى من توافق الشروط، وحفظ الدليل، وحصر السلطة، والتحقق من الأثر، مطروحًا منه مخاطر التفاعل غير المرصود. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

تكمل HyperSkill Composition وTool Governance وConstitutional Authorization. كانت الطبقات السابقة تضبط الوحدات؛ تضبط هذه النظرية الوصلات، وسلاسل الاعتماد، والقدرات الناشئة من التركيب. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

عند تحليل إخفاقات الوكلاء طويلة الأفق، ستتركز نسبة كبيرة من الأخطاء عالية الأثر في نقاط Handoff، وتحويل الأنواع، وانتقال الصلاحيات، لا داخل تنفيذ الأداة المنفردة. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

لا يمكن إثبات جميع التركيبات في نظام ديناميكي مفتوح؛ لذلك تعتمد النظرية على عقود جزئية واختبارات Runtime، ولا تدعي Formal Verification كاملًا. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** قد يؤدي فحص كل تركيب إلى انفجار تعقيدي يفوق فائدة النظام.

**التعديل:** تُطبق الضمانات بصورة متدرجة: تحقق ثابت للتركيبات المتكررة، ومحاكاة للتركيبات الجديدة عالية الأثر، وفحص خفيف للتركيبات منخفضة المخاطر. يصبح مبدأها: لا برهان شامل، بل ضمان متناسب مع المخاطر والجدة. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 20. نظرية التثبيت السببي التدخلي

**Interventional Causal Grounding Theory — ICG-M**

### الفجوة التي تكشفها

يستطيع النظام استرجاع أنماط قوية وشرحها لغويًا، لكنه قد يخلط بين الارتباط والآلية، أو بين تعاقب الأحداث وعلاقة الإنتاج السببي. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية التثبيت السببي التدخلي بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **لا ترقى العلاقة إلى معرفة سببية تشغيلية إلا إذا وُجد دليل يميزها عن النماذج المنافسة، ويفضل أن يتضمن تدخلًا، أو تجربة طبيعية، أو Counterfactual قابلًا للفحص.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

تصنف الدعاوى إلى مرشح ارتباطي، وفرضية آلية، ونموذج مضاد للواقع، ودعم تدخلي. يبني النظام رسمًا يميز السبب والوسيط والمعدل والمربك والنتيجة، ثم يقترح أصغر تدخل قابل للعكس يفصل بين الفرضيات. إذا تعذر التدخل، تبقى الدعوى في رتبة سببية مشروطة. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime للملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

يعرف رسوخ الدعوى السببية CG بوصفه دالة في استقلال الأدلة E، ووضوح الآلية M، وقوة التمييز المضاد للواقع F، ودعم التدخل I، مطروحًا منه الالتباس U: CG = w1E+w2M+w3F+w4I-w5U. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

توسع Mechanism Court وWorld-Effect Loop وCausal Reasoning في طبقات التدريب. فالحلقة السابقة تقيس ما حدث بعد الفعل؛ تضيف هذه النظرية تصميمًا يحدد لماذا حدث وما البديل الذي استُبعد. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

ستتفوق الأنظمة التي تفصل رتبة الارتباط عن رتبة التدخل في قرارات التشخيص والتخطيط، ولو أنتجت عددًا أقل من الدعاوى الحاسمة. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

كثير من المجالات الاجتماعية أو المؤسسية لا تسمح بتجارب محكمة، وقد تكون التدخلات غير أخلاقية أو مكلفة. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** يمكن للأنظمة أن تستخدم لغة السببية دون امتلاك نموذج سببي حقيقي، فيتحول الرسم إلى زخرفة.

**التعديل:** لا يقبل الرسم السببي وحده كدليل. يجب أن يولد قرارًا تجريبيًا أو تنبؤًا فارقًا أو يصرح بعدم إمكان الفصل. وتسمى النسخة المنضبطة «تثبيتًا سببيًا بدرجات» لا إثباتًا سببيًا مطلقًا. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 21. نظرية اقتصاد الائتمان المعرفي

**Epistemic Credit Economy Theory — ECE-M**

### الفجوة التي تكشفها

ترتب المعمارية الأدلة، لكنها لا تمتلك بعد نظامًا موحدًا وديناميكيًا لتوزيع الثقة بين المصادر والنماذج والوكلاء والمهارات والذكريات بحسب المجال والسجل. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية اقتصاد الائتمان المعرفي بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **الثقة رصيد محلي ومتغير يكتسب بالدقة والمعايرة والاستقلال والشفافية وسلوك التصحيح، ويُستهلك عند تقديم ادعاءات عالية الأثر، ولا ينبغي أن يكون صفة ثابتة لاسم نموذج أو مؤسسة.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

ينشئ النظام حسابًا معرفيًا لكل كيان يضم الدقة التاريخية، ومعايرة الثقة، وتطابق المجال، واستقلال المصدر، وحداثة المعرفة، ونسبة التصحيح، ومعدل الادعاءات غير القابلة للتحقق. تُحدّث الأرصدة بعد النتائج، وتُفصل حسب نوع المهمة. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime لملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

الرصيد $EC(e,d,t)$ للكيان e في المجال d والزمن t يتغير وفق تحديث بايزي أو Score تراكمي مضبوط، لكن لا يُستخدم كاحتمال حقيقي دون معايرة. الإنفاق المعرفي للادعاء يزداد مع أثره وعدم قابليته للتراجع. لا تدعي هذه الصياغة دقة فيزيائية، وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

تربط Evidence Ranking وHyperSkill Promotion وCross-Model Evaluation وMemory Integrity في لغة ثقة واحدة. وهي تضيف بعدًا اقتصاديًا: من يستهلك رصيدًا أكبر يجب أن يقدم دليلًا أو تحققًا أعلى. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

سيقل الاعتماد غير النقدي على «النموذج الأقوى عمومًا»، وستظهر خرائط تفوق محلية؛ نموذج أو وكيل قد يستحق سلطة في استخراج البيانات ولا يستحقها في الاستدلال السببي. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

السجل السابق قد يعاقب الكيانات الجديدة ويكرس تحيزًا مؤسسيًا أو تقنيًا. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** تحويل الثقة إلى رصيد قد يوهم بدقة كمية ويُستغل سياسيًا أو تنظيميًا.

**التعديل:** تُعرض الأرصدة كنطاقات مع أسبابها، ويخصص Exploration Budget للمرشحين الجدد، وتمنع الدرجة الواحدة من إخفاء الأبعاد المختلفة. الائتمان أداة توجيه لا حكمًا نهائيًا. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 22. نظرية توافق الحوافز الوكيلية

**Incentive-Compatible Agency Theory — ICA-M**

### الفجوة التي تكشفها

قد يرفع الوكيل المقياس المعلن دون تحقيق المقصد: يعلن النجاح مبكرًا، يخفض مستوى الاختبار لتقليل التكلفة، أو يخفي عدم اليقين كي يبدو أكثر كفاءة. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية توافق الحوافز الوكيلية بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **تحتاج الأنظمة الوكيلية إلى تصميم حوافز يجعل الصدق في الإبلاغ، واكتشاف الفشل، وحفظ القيود، أكثر فائدة من التلاعب بالProxy أو التضحية بالأثر الحقيقي.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

تفصل منظومة التقييم بين Goal Metric وIntegrity Metric وVerification Metric وCost Metric وExternal-Effect Metric وHidden Holdout. لا يعرف المنفذ جميع تفاصيل التقييم، ويُكافأ على اكتشاف الفشل المبكر، ويقارن تقريره بقياس خارجي. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime للملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

تعرف قوة الحافز IG بأنها الفرق بين المنفعة التي يحصل عليها الوكيل من تعظيم المقياس والمنفعة من تحقيق الهدف الحقيقي. يكون التصميم متوافقًا عندما تصبح $IG \leq 0$ ضمن مجموعة الاستراتيجيات المعقولة. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

تكمل Independent Verifier Agents وAgent Role Architecture وHuman Utility Evaluation. الفصل بين الأدوار يقلل تضارب المصالح، لكن التوافق يعالج الدافع البنيوي للتلاعب بالمقياس. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

عندما يُدخل مقياس لاكتشاف الفشل ونزاهة التقرير، قد تنخفض نسبة الإنجاز المعلنة أولًا، لكن ترتفع دقة تقارير النجاح وتنخفض الحوادث بعد النشر. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

الوكيل ليس فاعلًا اقتصاديًا واعيًا بالضرورة؛ يستخدم مفهوم الحافز هنا لوصف اتجاهات التحسين وسلوك السياسات لا النية النفسية. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** يمكن أن يصبح النظام متعدد المقاييس معقدًا لدرجة يستحيل معها تحديد ما يحسن فعليًا.

**التعديل:** تستخدم مجموعة صغيرة من المقاييس المتعارضة مع Holdouts دورية، ويُبحث عن Pareto Frontier بدل Score واحد. لا يُسمح للمقياس أن يصبح دستورًا خفيًا. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 23. نظرية السلامة الأنطولوجية الطبقية

**Ontological Type-Safety Theory — OTS-M**

### الفجوة التي تكشفها

كثير من الهلوسة النظرية لا يتمثل في حقيقة خاطئة فقط، بل في انتقال غير مشروع بين أنواع: من استعارة إلى آلية، ومن تفضيل إلى قاعدة، ومن تنبؤ إلى واقعة، ومن تعاقب إلى سببية. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية السلامة الأنطولوجية الطبقية بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **كل كيان معرفي يجب أن يحمل نوعًا أنطولوجيًا، وكل استدلال يجب أن يعلن تحويل النوع وشروطه، فلا تتحول التعليمات إلى دليل، ولا القيمة إلى حقيقة تجريبية، ولا المحاكاة إلى أثر واقعي.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

يعرف النظام أنواعًا مثل Fact وObservation وInterpretation وHypothesis وNorm وPreference وInstruction وPrediction وMetaphor وSimulation وAction وVerified Effect. يفحص المدقق نوع المدخل والعلاقة والمخرج والتحويل المطلوب. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime للملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

يمثل الاستدلال كتطبيق f:T_a×R→T_b. إذا لم توجد قاعدة تحويل معتمدة من T_a إلى T_b تحت العلاقة R، يصنف الانتقال Type Error ويطلب دليلًا أو يخفض رتبة المخرج. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

تعزز Claim-Type Classification والوضع الأنطولوجي ومحكمة التعريف والآلية، وتوفر لـM-HKOS Schema يمنع اختلاط الذاكرة المعيارية بالذاكرة الوصفية. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

سيكشف المدقق الأنطولوجي أخطاء لا تلتقطها Fact Checking، خصوصًا في التقارير الاستراتيجية والأخلاقية والبحثية التي تمزج الوصف بالتوصية. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

الأنواع ليست دائمًا منفصلة؛ قد يكون النص الواحد وصفًا وحكمًا أو استعارة ونموذجًا في آن. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** التصنيف الصارم قد يقتل المرونة الإبداعية والتركيب بين الحقول.

**التعديل:** تسمح النظرية بالأنواع المركبة والتحويلات المؤقتة داخل Sandbox، لكنها تمنع ترقيتها إلى ذاكرة أو فعل قبل توضيح النوع. الإبداع مفتوح، والترقية منضبطة. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 24. نظرية النضج التطوري للقدرة

**Developmental Capability Maturation Theory — DCM-M**

### الفجوة التي تكشفها

نجاح المهارة مرة أو مرات في توزيع قريب لا يثبت أنها قابلة للنقل أو أنها تستحق صلاحيات أوسع. الترقية الحالية تحتاج إلى نموذج نمو يربط الأداء بدرجة الاستقلال. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية النضج التطوري للقدرة بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **القدرة تمر بمراحل نوعية: التقليد، والتنفيذ الموجه، والتكرار المتحقق، والنقل، والتحمل، والتشخيص الذاتي، والاستقلال المحدود، والتنسيق النظامي. ولا تنتقل السلطة إلا بقدر النضج داخل نطاق محدد.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

يختبر النظام تنوع المهام، والمسافة عن أمثلة التدريب، ومقاومة الإدخال المضلل، وقدرة كشف الفشل، وجودة Rollback، وثبات الأداء. تمنح كل مرحلة نطاقًا مختلفًا للأدوات والذاكرة والنشر والاتصال الخارجي. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime لملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

نضج القدرة CM ليس متوسط النجاح فقط، بل متجه [R,T,A,S,D,B] يمثل التكرار والنقل والتحمل والتشخيص والتراجع وحدود السلطة. لا تختزل الرتبة في رقم إلا لأغراض المقارنة المحلية. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

توحّد HyperSkill Promotion وModel Escalation وProduction Gates وBounded Autonomous Improvement في سلم واحد. ما كان حالات نشر منفصلة يصبح نظرية لنمو القدرة. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

المهارات التي تجتاز النقل والتشخيص الذاتي ستنتج حوادث أقل من مهارات أعلى منها في معدل النجاح الخام لكنها أضعف في كشف الخطأ. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

قد تنضج القدرة في نطاق ضيق وتبقى بدائية خارجه، ولا توجد رتبة عامة تصلح لكل المجالات. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** السلم قد يعكس افتراضًا خطيًا بينما نمو القدرات قد يكون متقطعًا أو متراجعًا.

**التعديل:** تستخدم خريطة نضج متعددة الأبعاد مع إمكان التراجع والاختلاف حسب المجال. «المرحلة» وصف للحقوق التشغيلية لا قانون نفسي أو بيولوجي. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 25. نظرية إنتروبيا الذاكرة وإعادة التوطيد

**Memory Entropy and Reconsolidation Theory — MER-M**

### الفجوة التي تكشفها

الذاكرة قد تكون غنية لكنها غير صالحة: تكرار، تناقضات غير محلولة، مصادر مفقودة، حقائق قديمة، ضغط مخل، أو استنتاجات تراكمت حتى أصبحت أشبه بالحقيقة. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية إنتروبيا الذاكرة وإعادة التوطيد بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **كل ذاكرة تشغيلية تتجه نحو الإنتروبيا ما لم تخضع لإعادة توطيد دورية تحفظ الأصول والتعارضات وتعيد ترتيب الثقة وتزيل التكرار دون محو الأقليات المعرفية.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

يقيس النظام التكرار، والتناقض، والقِدم، وفقد Provenance، والعناصر اليتيمة، ونسبة المقترحات غير المتحققة، وتسرب السياق. عند العتبة تبدأ دورة لاستخراج الادعاءات وتجميع المتكافئ وإبقاء الخلاف وتحديد Canonical State وضغط التاريخ. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime لملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

$ME=\alpha D+\beta C+\gamma S+\delta P+\varepsilon O+\eta Q-\zeta V$، حيث D التكرار، C التناقض، S القِدم، P فقد المصدر، O اليتم، Q مواد الحجر، وV التنظيم المتحقق. هي صيغة تقديرية لا مقياس فيزيائي. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

تعمق Self-Reconstruction Memory وROFSC وContext Compaction وMemory Conflict Resolution. تضيف شرطًا متى يجب إعادة البناء، لا كيف فقط. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

ستتدهور دقة الوكيل واتساقه بعد عدد كبير من الجلسات حتى مع زيادة حجم الذاكرة، ما لم تُجرَ دورات Reconsolidation تحفظ المصدر وتزيل التكرار. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

الضغط وإعادة التوطيد قد يحذفان حقيقة نادرة صحيحة أو سياقًا يصبح مهمًا لاحقًا. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** إنتروبيا الذاكرة استعارة حرارية وقد توحي بقانون طبيعي غير موجود.

**التعديل:** يستخدم المصطلح مؤشرًا هندسيًا مركبًا لا تماثلًا فيزيائيًا، وتحتفظ المنظومة بسجل تاريخي وطبقة Cold Storage، ولا تعد قلة التكرار دليلًا على الضعف. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 26. نظرية الصمود أمام الفشل الناشئ

**Emergent Systemic Resilience Theory — ESR-M**

### الفجوة التي تكشفها

قد لا يظهر الخطر في اختبار أي مكوّن منفرد، بل ينشأ عند التزامن أو التأخير أو التحويل إلى نموذج احتياطي أو تراكم الإعادات أو تضارب الذاكرة والصلاحيات. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية الصمود أمام الفشل الناشئ بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **الخطر النظامي ليس مجموع أخطار الأجزاء؛ بل خاصية تنشأ من شبكة الاعتماد والتوقيت والتغذية الراجعة. لذلك يجب اختبار مسارات انتشار الفشل لا قائمة الأعطال فقط.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

يبني النظام Failure Propagation Graph يحدد العقد، والاعتماديات، والتأخيرات، ودوائر التضخيم، ونقاط الاحتواء، وحالات الاسترداد. ثم يحقن أعطالًا محسوبة في أدوات ونماذج وذاكرة وشبكة وصلاحيات. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime لملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

يعرف خطر الانتشار SRisk بمجموع احتمال فشل العقدة مضروبًا في قابلية الانتقال ووزن الأثر، مع معاملات للدوائر الراجعة. لا تستخدم الصيغة للتنبؤ المطلق بل لتحديد العقد الحرجة واختبارها. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

توسع Destruction Court وIncident Response وRollback وMonitoring إلى Chaos Engineering معرفي ووكلي. وتربط Runtime Plane بـDeployment Plane بصورة عملية. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

ستظهر إخفاقات عالية الأثر تحت الضغط الطويل والتزامن لا تظهر في Benchmarks القصيرة، خصوصًا عندما يتحول النظام تلقائيًا بين مزودين أو نماذج. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

اختبارات الفوضى نفسها قد تكون مكلفة أو خطرة إذا أجريت في Production. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** يمكن رسم شبكات فشل شديدة التعقيد دون قدرة عملية على اختبارها.

**التعديل:** تعطي الأولوية لمسارات Minimum Cut والعقد ذات المركزية العالية، وتستخدم Digital Twin أو Staging. الهدف ليس تعداد كل انهيار، بل اختبار أقصر سلاسل تقود إلى أثر خطير. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

# 27. نظرية السيادة المؤسسية البشرية–الآلية

**Human–Institutional Sovereignty Theory — HIS-M**

## الفجوة التي تكشفها

إذن المستخدم لا يكفي دائمًا؛ فقد يكون طالب الفعل غير مالك للأصل، أو يكون صاحب البيانات مختلفًا عن المشغل، أو توجد التزامات قانونية ومؤسسية تتجاوز رغبة الفرد. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

## التعريف والأطروحة

تعرف نظرية السيادة المؤسسية البشرية–الآلية بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **شرعية الفعل الوكيلي تنتج من خريطة حقوق ومسؤوليات واختصاصات تشمل الطالب والمالك وصاحب البيانات والجهة المسؤولة والمنفذ والمتحقق والمتأثرين، لا من أمر واحد معزول.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

## آلية التشغيل

يعرف كل فعل عبر Requester وAsset Owner وData Subject وAccountable Authority وExecutor وVerifier وAffected Parties وAppeal Channel. تُطبق قواعد فصل الواجبات وأقل سلطة وإمكان الاعتراض. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime لملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

## الصياغة الصورية الأولية

صلاحية الفعل IA هي تقاطع صلاحيات الدور، وملكية الأصل، وموافقة البيانات، والدستور، والاختصاص المؤسسي، لا اتحادها. إذا غاب عنصر لازم يفشل النظام مغلقًا أو يصعد. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

## الارتباط بما سبق

توسع Authority and Permission Layer وHuman Oversight وCross-Project Isolation وAudit Logging. تنتقل الحوكمة من «مَن طلب؟» إلى «مَن يملك الحق والمساءلة؟». الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

## الفرضية والتنبؤ الفارق

ستمنع مصفوفة السيادة نسبة من الأفعال التي تبدو مصرحًا بها تقنيًا لكنها غير مشروعة مؤسسيًا، خصوصًا في البيانات والاتصالات والنشر والحذف. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

## شروط الحدود

تختلف هياكل السلطة جذريًا بين المؤسسات والدول، ولا يمكن فرض Schema واحد كامل. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

## أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** قد تتحول الطبقة إلى بيروقراطية تعطل الاستخدام المشروع والسريع.

**التعديل:** تعتمد قوالب مؤسسية قابلة للتخصيص، ومسارات Fast Track للأفعال المنخفضة، وتصعيدًا واضحًا للأفعال المتنازع عليها. السيادة ليست كثرة الموافقات بل وضوح مصدر الشرعية. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 28. نظرية الفتح الآمن في العالم المجهول

**Safe Open-World Fattah Adaptation Theory — SOFA-M**

### الفجوة التي تكشفها

قد يطبق النظام مهارة ناجحة على مهمة تشبه خبراته لغويًا لكنها جديدة في الهدف أو الخطر أو البنية المؤسسية أو العلاقات، فيقع في نقل زائف عالي الثقة. هذه الفجوة من الرتبة الثانية: لا تظهر عند سؤال هل يملك النظام الأداة، بل عند سؤال هل انتقلت وظيفة الأداة وشروطها إلى السياق التالي بصورة مشروعة.

### التعريف والأطروحة

تعرف نظرية الفتح الآمن في العالم المجهول بأنها إطار يفسر موضع النقص في الأنظمة الوكيلية بوصفه خللًا في العلاقة الحاكمة لا مجرد ضعف في المكوّن. أطروحتها المركزية: **الذكاء الفتّاح لا يبدأ بالحل، بل باكتشاف أن التمثيل الحالي غير كافٍ، ثم يفتح استكشافًا محدودًا ويخفض السلطة والثقة حتى تتكون معرفة مناسبة.** وينتج عن ذلك أن تقييم النظام يجب أن يضم وحدات الانتقال والتركيب، لا وحدات الأداء المنفرد فقط.

### آلية التشغيل

يقيس النظام الجدة في الكيان والعلاقة والهدف والخطر والأداة والتوزيع والسياق المؤسسي ونمط الفشل. إذا تجاوزت المسافة عتبة، يمنع الفعل العالي، ويولد نماذج متعددة، ويجمع أدلة، ويجري تجربة صغيرة، وينشئ Skill Candidate. تعمل الآلية على مستويين: مستوى Static قبل التشغيل لفحص العقود والأنواع والسلطات، ومستوى Runtime لملاحظة الانحراف الذي لا يمكن التنبؤ به كاملًا. ويجب أن ينتج كل فحص أثرًا قابلًا للتدقيق: قرار، وسبب، وشرط، ونسخة، ومسار تراجع.

### الصياغة الصورية الأولية

تعرف مسافة الجدة ND كمركب من مسافة التمثيل، واختلاف العقود، وعدم اليقين السببي، واختلاف السلطة، وغياب أمثلة فشل مشابهة. كلما زادت ND انخفض سقف الفعل المسموح وارتفع واجب الاستكشاف. لا تدعي هذه الصياغة دقة فيزيائية؛ وظيفتها تحديد المتغيرات التي يجب جمعها وتجنب استعمال مصطلح النظرية دون قياس.

### الارتباط بما سبق

تكمل Unknown-Unknown Discovery وFattah Open Intelligence وCapability Gap Discovery وBounded Autonomous Improvement. تجعل «الفتح» بروتوكولًا لا استعارة. الجديد ليس إعادة تسمية الطبقات السابقة، بل إضافة قانون يربط مخرجاتها ويحدد ما الذي لا تستطيع كل طبقة ضمانه بمفردها.

### الفرضية والتنبؤ الفارق

سيزيد النظام الامتناع أو التصعيد المؤقت في المهام الجديدة، لكنه يخفض الأخطاء الجسيمة الناتجة عن استخدام Skill Stable خارج نطاقها. يمكن اختبار ذلك بمقارنة نظامين متساويين في النموذج والميزانية، يختلفان فقط في وجود طبقة النظرية وأدواتها. يجب قياس الإخفاقات عالية الأثر، لا متوسط جودة النص فقط.

### شروط الحدود

قد يصنف النظام كل تغير بسيط جِدةً فيصبح شديد التحفظ، أو يفشل في اكتشاف الجدة العميقة. لذلك تكون النظرية صالحة بوصفها إطار ضبط تدريجي، وتفشل إذا قُدمت كبرهان شامل أو استُخدمت لإخفاء عدم اليقين خلف Score واحد.

### أقوى اعتراض والنسخة بعد الهدم

**الاعتراض:** مسافة الجدة تعتمد على تمثيلات قد تكون هي نفسها ناقصة أو منحازة.

**التعديل:** تجمع الجدة بين إشارات تمثيلية وعقدية وسببية ومؤسسية، وتعاير بالنتائج. لا تُستخدم الدرجة وحدها؛ يكفي وجود Novel Risk أو Novel Authority لفرض تصعيد حتى لو بدا المحتوى مألوفًا. بهذا تصبح النظرية أقل ادعاء وأكثر قابلية للتنفيذ: لا تعد بالكمال، بل تقلل مساحة الانتقال غير المحكوم وتكشفها.

## 29. النظرية الجامعة: الإغلاق الحاكمي التركيبي للذكاء الوكيلي

تندمج النظريات العشر في نظرية أم مفادها أن النظام الوكيلي لا ينضج بزيادة قدراته فقط، بل بزيادة نسبة الانتقالات المبررة بين طبقاته. تسمي الورقة هذه النسبة **درجة الإغلاق الحاكمي التركيبي**. وهي لا تقيس الانغلاق أو الاستقلال، بل مقدار ما أمكن ربطه من الادعاء إلى النوع، ومن النوع إلى الدليل، ومن الدليل إلى الثقة، ومن الثقة إلى العقد، ومن العقد إلى السلطة، ومن السلطة إلى الفعل، ومن الفعل إلى الأثر، ومن الأثر إلى الذاكرة.

تتكون النظرية الجامعة من خمس دوائر:

1. **دائرة الصلاحية المعرفية:** نوع الادعاء، مصدره، رتبته، وائتمانه.
2. **دائرة الصلاحية السببية:** الآلية، البدائل، التدخل، والتنبؤ الفارق.
3. **دائرة صلاحية التركيب:** عقود المكونات، انتقال السلطة، والقدرات الناشئة.
4. **دائرة الصلاحية المؤسسية:** من يملك الحق والمسؤولية ومن يتأثر.
5. **دائرة التصحيح التطوري:** الأثر، الفشل، إنتروبيا الذاكرة، ونضج القدرة.

الصيغة المفهومية الأولية:

$$\mathbf{GCC = (CA \times CG \times EC \times IC \times OT \times CM \times MR \times SR \times IS \times OA) / (1 + N + U + X)}$$

حيث CA ضمان التركيب، CG الرسوخ السببي، EC الائتمان المعرفي، IC توافق الحوافز، OT السلامة الأنطولوجية، CM نضج القدرة، MR سلامة الذاكرة، SR الصمود، IS السيادة المؤسسية، OA التكيف في العالم المفتوح، بينما N الضوضاء، Uو عدم اليقين غير الممثل، وX تعقيد الحوكمة الزائد. استعمال الضرب مقصود دلاليًا: بعض أوجه النقص لا تعوضها زيادة الجوانب الأخرى. لكنه لا يدعي أن المتغيرات مستقلة أو قابلة للقياس الدقيق حاليًا.

### اثنتا عشرة نظرية فرعية مشتقة

1. **نظرية تصاعد السلطة التركيبي:** قد ينشأ فعل عالي السلطة من سلسلة أفعال منخفضة.
2. **نظرية الانحراف بين النجاح والإغلاق:** قد ينجح التنفيذ بينما تبقى سلسلة الدليل والمسؤولية مفتوحة.
3. **نظرية التلوث الائتماني:** تنتقل ثقة المصدر إلى استنتاج لم يثبته.
4. **نظرية الانهيار الأنطولوجي:** يتضخم الخطأ عندما يتحول النوع مرة ثم يبنى عليه استدلال طويل.
5. **نظرية العمى السببي للاسترجاع:** كثرة الوثائق قد ترفع الثقة دون رفع القدرة على تحديد السبب.
6. **نظرية الاستقلال السابق للنضج:** منح الصلاحية قبل النقل والتشخيص يضاعف أثر الخطأ.
7. **نظرية الذاكرة الناجحة الخاطئة:** يمكن لذاكرة متسقة داخليًا أن تكون مبنية على Canonical State غير صحيح.
8. **نظرية الفشل متعدد السرعات:** تتفاعل أعطال بطيئة في الذاكرة مع أعطال سريعة في الأدوات فتنتج انهيارًا غير متوقع.

9. **نظرية السيادة المتقاطعة:** فعل واحد قد يحتاج إلى عدة مصادر شرعية غير قابلة للاختزال.
10. **نظرية الجدة الحاكمة:** جدة الخطر أو السلطة أهم أحيانًا من جدة المحتوى.
11. **نظرية كلفة الحوكمة الحدية:** بعد حد معين، تزيد طبقات الفحص الكلفة أكثر من خفض المخاطر.
12. **نظرية الفتح بعد الاعتراف بالنقص:** لا يبدأ الاستكشاف الآمن إلا بعد تصنيف التمثيل الحالي غير كافٍ.

# الباب الثالث: الأدوات والمقاييس والاختبارات

## 30. الأدوات العشرون لتحويل النظريات إلى هندسة

### الأداة 1: مصفوفة عقود التركيب المعرفي

**النظرية الأم:** M-CAT.
**النقص الذي تعالجه:** منع الربط غير الآمن بين المهارات والوكلاء والأدوات.
**المدخلات:** عقود المدخلات والمخرجات والصلاحية والذاكرة.
**خطوات التشغيل:** مطابقة Postconditions بالأهداف التالية، فحص النوع والدليل والسلطة والتراجع.
**المخرجات:** تقرير توافق أو تعارض أو شرط مفقود أو تصاعد صلاحية.
**الحد وأخطر سوء استعمال:** استعمال Checkboxes شكلية دون اختبار مسارات فعلية.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 2: اختبار اختراق التآزر الوكيلي

**النظرية الأم:** M-CAT.
**النقص الذي تعالجه:** كشف القدرات والمخاطر الناشئة من الجمع.
**المدخلات:** مجموعة مكونات وتكوينات Handoff.
**خطوات التشغيل:** توليد تركيبات غير معتادة، محاكاة التنفيذ، قياس تسرب السلطة والدليل.
**المخرجات:** خريطة مخاطر التركيب وترتيب التركيبات الحرجة.
**الحد وأخطر سوء استعمال:** إجراء الاختبار في بيئة حقيقية عالية الأثر.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 3: محرك الرسم السببي التدخلي

**النظرية الأم:** ICG-M.
**النقص الذي تعالجه:** تحويل التفسير اللغوي إلى فرضيات سببية قابلة للفصل.
**المدخلات:** ادعاءات ومتغيرات وملاحظات.
**خطوات التشغيل:** تصنيف السبب والوسيط والمربك، توليد DAG، مقارنة النماذج.
**المخرجات:** رسم سببي وأسئلة تدخلية.
**الحد وأخطر سوء استعمال:** اعتبار الرسم دليلًا دون بيانات.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 4: بروتوكول التدخل المصغر القابل للتراجع

**النظرية الأم:** ICG-M.
**النقص الذي تعالجه:** اختبار الدعوى بأقل مخاطرة.
**المدخلات:** رسم سببي وحدود السلطة وBaseline.
**خطوات التشغيل:** تحديد تدخل صغير، توقع مسبق، مجموعة مقارنة، Postconditions وRollback.
**المخرجات:** نتيجة ترفع أو تخفض رسوخ الدعوى.
**الحد وأخطر سوء استعمال:** التدخل في البشر أو الإنتاج دون موافقة.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 5: دفتر الائتمان المعرفي

**النظرية الأم:** ECE-M.
**النقص الذي تعالجه:** تخصيص الثقة بحسب المجال والسجل.
**المدخلات:** نتائج ادعاءات ومصادر ونماذج ومهارات.
**خطوات التشغيل:** تحديث أبعاد الدقة والمعايرة والاستقلال والتصحيح والحداثة.
**المخرجات:** ملف ثقة مفسر ونطاقات لا Score غامض.
**الحد وأخطر سوء استعمال:** تحويل الرصيد إلى سمعة ثابتة أو سلطة سياسية.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 6: مؤشر استقلال الأدلة

**النظرية الأم:** ECE-M.
**النقص الذي تعالجه:** منع عدّ النسخ المتعددة مصدرًا مستقلاً.
**المدخلات:** سلاسل الاقتباس والبيانات والمؤلفون والمصادر.
**خطوات التشغيل:** اكتشاف الأصل المشترك والاعتماد الدائري والتكرار.
**المخرجات:** درجة استقلال وخريطة منشأ.
**الحد وأخطر سوء استعمال:** افتراض أن اختلاف النطاق يعني استقلالًا.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 7: محلل انحراف الهدف عن المقياس

**النظرية الأم:** ICA-M.
**النقص الذي تعالجه:** كشف الطرق التي يرتفع بها Proxy ويهبط المقصد.
**المدخلات:** الهدف والمقاييس وسياسة المكافأة.
**خطوات التشغيل:** توليد استراتيجيات Gaming، محاكاة Trade-offs، مقارنة World Effect.
**المخرجات:** قائمة انحرافات ومقاييس مضادة.
**الحد وأخطر سوء استعمال:** إضافة مقاييس كثيرة بلا حاكمية.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

## الأداة 8: اختبار الوكيل الاستراتيجي

**النظرية الأم:** ICA-M.
**النقص الذي تعالجه:** قياس نزاهة الإبلاغ تحت تضارب الحوافز.
**المدخلات:** سيناريوهات نجاح وفشل ومكافآت مختلفة.
**خطوات التشغيل:** إخفاء بعض Holdouts، عرض فرص إعلان مبكر، مقارنة التقرير بالحقيقة.
**المخرجات:** معدل الصدق والمعايرة واكتشاف الفشل.
**الحد وأخطر سوء استعمال:** تفسير السلوك كنية واعية.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

## الأداة 9: مدقق الأنواع الأنطولوجية

**النظرية الأم:** OTS-M.
**النقص الذي تعالجه:** منع انتقالات الحقيقة/القيمة/الاستعارة/الفعل غير المشروعة.
**المدخلات:** Claims وذاكرة وتعليمات وتوصيات.
**خطوات التشغيل:** وسم الأنواع، فحص قواعد التحويل، رفع Type Errors.
**المخرجات:** سجل أخطاء وتحويلات مطلوبة.
**الحد وأخطر سوء استعمال:** فرض تصنيف جامد على اللغة المركبة.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

## الأداة 10: محكمة خلط المستويات

**النظرية الأم:** OTS-M.
**النقص الذي تعالجه:** كشف القفز من مستوى إلى آخر دون وسيط.
**المدخلات:** حجة أو تقرير أو نظرية.
**خطوات التشغيل:** تحديد المستوى المادي والمعلوماتي والاجتماعي والمعياري، وطلب آلية انتقال.
**المخرجات:** خريطة مستويات واعتراضات.
**الحد وأخطر سوء استعمال:** رفض كل تركيب بين التخصصات.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

## الأداة 11: سلم نضج HyperSkill

**النظرية الأم:** DCM-M.
**النقص الذي تعالجه:** منع مساواة النجاح القريب بالنضج.
**المدخلات:** نتائج اختبارات متنوعة وسجل فشل ونقل.
**خطوات التشغيل:** تقييم التقليد والتكرار والنقل والتحمل والتشخيص والتراجع.
**المخرجات:** ملف نضج متعدد الأبعاد.
**الحد وأخطر سوء استعمال:** منح رتبة عامة خارج النطاق.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 12: بوابة ترقية الاستقلال

**النظرية الأم:** DCM-M.
**النقص الذي تعالجه:** ربط الصلاحيات بنضج مثبت.
**المدخلات:** ملف نضج ومخاطر وأدوات مطلوبة.
**خطوات التشغيل:** تحديد الحد الأدنى، Canary، موافقة، مراقبة، شروط سحب.
**المخرجات:** قرار ترقية أو رفض أو صلاحية جزئية.
**الحد وأخطر سوء استعمال:** تحويل الترقية إلى قرار دائم.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 13: مقياس إنتروبيا الذاكرة التشغيلية

**النظرية الأم:** MER-M.
**النقص الذي تعالجه:** اكتشاف تدهور الذاكرة قبل ظهور خطأ كبير.
**المدخلات:** Memory Graph وProvenance وإصدارات.
**خطوات التشغيل:** حساب التكرار والتناقض والقِدم واليتم والهجر.
**المخرجات:** مؤشر مفسر وعقد حرجة.
**الحد وأخطر سوء استعمال:** حذف المعرفة النادرة لمجرد رفع المؤشر.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 14: بروتوكول إعادة التوطيد القانوني

**النظرية الأم:** MER-M.
**النقص الذي تعالجه:** إعادة بناء Canonical Memory دون محو التاريخ.
**المدخلات:** ادعاءات وإصدارات وأدلة وتعارضات.
**خطوات التشغيل:** دمج المتكافئ، حفظ المخالف، تعيين Supersedes، ضغط التاريخ، اختبار الرجوع.
**المخرجات:** نسخة قانونية وسجل تحول.
**الحد وأخطر سوء استعمال:** تسوية الخلاف قسرًا.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 15: خريطة انتشار الفشل

**النظرية الأم:** ESR-M.
**النقص الذي تعالجه:** رؤية سلاسل الانهيار النظامي.
**المدخلات:** تبعيات الأدوات والنماذج والذاكرة والشبكة.
**خطوات التشغيل:** رسم العقد والانتقالات ودوائر التضخيم وMinimum Cuts.
**المخرجات:** قائمة نقاط حرجة وخطط احتواء.
**الحد وأخطر سوء استعمال:** الانشغال بكل سيناريو نظري.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 16: مختبر الفوضى الوكيلية

**النظرية الأم:** ESR-M.
**النقص الذي تعالجه:** اختبار الصمود تحت أعطال واقعية.
**المدخلات:** بيئة Staging وسيناريوهات عطل.
**خطوات التشغيل:** حقن تأخير وفقد أداة ونتيجة فاسدة ونموذج أضعف، قياس الاسترداد.
**المخرجات:** مقاييس احتواء وMTTR معرفي.
**الحد وأخطر سوء استعمال:** تشغيله بلا عزل أو Rollback.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 17: مصفوفة سلطة ومسؤولية AGI

**النظرية الأم:** HIS-M.
**النقص الذي تعالجه:** تحديد من يملك الطلب والبيانات والقرار والمساءلة.
**المدخلات:** الأدوار والأصول وأنواع الأفعال.
**خطوات التشغيل:** ربط Requester وOwner وData Subject وApprover وVerifier وAffected Parties.
**المخرجات:** قرار صلاحية ومسار تصعيد.
**الحد وأخطر سوء استعمال:** استعمالها لتكثير الموافقات بلا ضرورة.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 18: سجل المسؤولية السببية

**النظرية الأم:** HIS-M.
**النقص الذي تعالجه:** ربط الأثر بسلسلة قرار لا بمن نفذ آخر خطوة.
**المدخلات:** الطلبات والأدلة والموافقات والTraces والنتائج.
**خطوات التشغيل:** بناء Causal Accountability Chain قابلة للتدقيق.
**المخرجات:** سجل مسؤولية موزع.
**الحد وأخطر سوء استعمال:** تحويل السجل إلى أداة لوم آلي غير عادلة.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 19: كاشف الجِدّة التوزيعية والأنطولوجية

**النظرية الأم:** SOFA-M.
**النقص الذي تعالجه:** تمييز المهمة الجديدة من التشابه السطحي.
**المدخلات:** تمثيل المهمة وعقودها ومخاطرها والسياق المؤسسي.
**خطوات التشغيل:** مقارنة المحتوى والعلاقات والأهداف والسلطة وحالات الفشل.
**المخرجات:** Novelty Profile ونطاق ثقة.
**الحد وأخطر سوء استعمال:** اعتبار المسافة التمثيلية وحدها كافية.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

### الأداة 20: بروتوكول الاستكشاف الفتّاح الآمن

**النظرية الأم:** SOFA-M.
**النقص الذي تعالجه:** تحويل الجدة إلى مسار تعلم محدود.
**المدخلات:** Novelty Profile وميزانية وصلاحية.
**خطوات التشغيل:** خفض الفعل، توليد بدائل، جمع دليل، تجربة صغيرة، محكمة، Skill Candidate.
**المخرجات:** خطة استكشاف وترقية مشروطة.
**الحد وأخطر سوء استعمال:** استكشاف غير محدود أو جمع بيانات بلا غرض.

تُنفذ الأداة أولًا في صورة Schema وقائمة اختبارات، ثم يمكن تحويلها إلى وحدة برمجية. لا تعد الأداة ناجحة إذا أنتجت تقريرًا فقط؛ يجب أن تغير قرار التوجيه أو الترقية أو التنفيذ أو الذاكرة بصورة قابلة للقياس.

## 31. المقاييس والمؤشرات الجديدة

تقترح الورقة سبعة مقاييس أولية. لا ينبغي دمجها في رقم واحد قبل جمع بيانات كافية، لأن المقاييس قد تتعارض ويكون التعارض نفسه معلومة.

### 31.1 مؤشر الضمان التركيبي CAI

يقيس توافق العقود، وحفظ Provenance، وعدم تصاعد السلطة، والتحقق من Postconditions، وقابلية التراجع. يقدر لكل سلسلة تنفيذ، لا لكل أداة منفردة.

### 31.2 مؤشر الرسوخ السببي CGI

يقيس استقلال الدليل، ووضوح الآلية، ووجود بدائل، وقوة Counterfactual، ودعم التدخل. يخفض الادعاءات السببية التي لا تتجاوز الارتباط.

### 31.3 مؤشر الائتمان المعرفي ECI

متجه يشمل الدقة والمعايرة وتطابق المجال والاستقلال وسلوك التصحيح والحداثة. يجب إظهار أبعاده وأسبابها.

### 31.4 مؤشر توافق الحوافز ICI

يقيس مقدار تحسن المقصد الحقيقي مقارنة بتحسن Proxy، ونسبة الفشل المبلغ عنه ذاتيًا، والفارق بين النجاح المعلن والمتحقق.

### 31.5 مؤشر السلامة الأنطولوجية OTSI

يقيس نسبة الادعاءات الموسومة نوعيًا، وعدد التحويلات غير المبررة، وخطورة Type Errors التي وصلت إلى فعل أو ذاكرة.

### 31.6 مؤشر إنتروبيا الذاكرة MEI

يجمع التكرار والتناقض والقِدم وفقد المصدر والعناصر اليتيمة وحجم Quarantine. لا يهدف إلى تصغير الذاكرة، بل رفع نسبة المعرفة القابلة للتتبع.

### 31.7 مؤشر الصمود والفتح RSNI

يقيس احتواء الأعطال، وزمن الاسترداد، ودقة اكتشاف الجدة، ومعدل النقل الزائف، ونجاح الاستكشاف المصغر قبل الترقية.

### مقياس القيمة الوكيلية المعدلة بالإغلاق

**C-AUV = Verified Utility × GCC / (Cost + Latency + Risk + Governance Overhead)**

يمنع هذا المقياس اعتبار النظام أفضل لمجرد زيادة الفحوص؛ فإذا أصبحت كلفة الحوكمة أكبر من خفض المخاطر هبطت القيمة المعدلة.

## 32. ثماني عشرة فرضية بحثية

**H1 — قوية:** إضافة فحص عقود التركيب ستخفض إخفاقات Handoff دون تغيير النموذج الأساسي.
**الاختبار المقترح:** مقارنة Ablation على مهام متعددة الأدوات.

**H2 — قوية:** فصل أنواع الادعاء سيخفض انتقال التوصيات أو الاستعارات إلى ذاكرة الحقائق.
**الاختبار المقترح:** حقن نصوص مختلطة الأنواع وقياس تلوث الذاكرة.

**H3 — قوية:** اختبارات الفوضى ستكشف فشلًا لا يظهر في Benchmarks القصيرة.
**الاختبار المقترح:** تشغيل نفس الوكلاء تحت تأخير وفقد أدوات وتحويل نموذج.

**H4 — قوية:** ربط الاستقلال بنضج متعدد الأبعاد سيخفض حوادث النشر مقارنة بمعدل النجاح الخام.
**الاختبار المقترح:** Canary study مع مهارات متساوية النجاح مختلفة التشخيص.

**H5 — متوسطة:** دفتر الائتمان المحلي سيتفوق على ترتيب ثابت للنماذج في الكلفة والدقة.
**الاختبار المقترح:** Routing متعدد المجالات ومقارنة regret.

**H6 — متوسطة:** مؤشر استقلال الأدلة سيخفض الثقة الزائدة في الإجابات متعددة المصادر المنسوخة.
**الاختبار المقترح:** مجموعات مصادر ذات أصول مشتركة ومختلفة.

**H7 — متوسطة:** مكافأة الإبلاغ عن الفشل سترفع دقة تقارير النجاح ولو خفضت نسبة النجاح المعلنة.
**الاختبار المقترح:** تجربة سياسات تقييم مختلفة.

**H8 — متوسطة:** إعادة توطيد الذاكرة دوريًا ستحسن الاتساق بعد جلسات طويلة.
**الاختبار المقترح:** Longitudinal agent runs مع/دون reconsolidation.

**H9 — متوسطة:** Novelty Profile متعدد الأبعاد سيكشف النقل الزائف أفضل من تشابه embedding وحده.
**الاختبار المقترح:** مهام متشابهة لغويًا مختلفة في السلطة والخطر.

**H10 — متوسطة:** خريطة السلطة المؤسسية ستمنع أفعالًا مصرحًا بها تقنيًا لكنها غير مشروعة تنظيميًا.
**الاختبار المقترح:** محاكاة أدوار وممتلكات وبيانات.

**H11 — ضعيفة:** المؤشرات العشرة في GCC ستظهر تفاعلًا ضربيًا لا جمعيًا في الحوادث العالية.
**الاختبار المقترح:** نمذجة إحصائية بعد جمع traces كبيرة.

**H12 — ضعيفة:** الائتمان المعرفي قد يصبح أكثر استقرارًا عبر النماذج من درجات Benchmark العامة.
**الاختبار المقترح:** Cross-model repeated tasks.

**H13 — ضعيفة:** المدقق الأنطولوجي سيحسن البحث النظري أكثر من تحسين المهام الحسابية.
**الاختبار المقترح:** مقارنة مجالات مختلفة.

**H14 — ضعيفة:** تخفيض السلطة تلقائيًا عند الجدة سيزيد Human Utility على المدى الطويل.
**الاختبار المقترح:** دراسة مستخدمين طويلة.

**H15 — تخمينية خصبة:** يمكن تعلم قواعد تركيب آمنة من Traces الفشل بدل كتابتها يدويًا فقط.
**الاختبار المقترح:** Induction of contracts from incident logs.

**H16 — تخمينية خصبة:** قد يصبح سجل المسؤولية السببية مصدرًا لتدريب وكيل حوكمة مستقل.
**الاختبار المقترح:** تدريب/استرجاع من audit chains.

**H17 — تخمينية خصبة:** يمكن لمؤشر إنتروبيا الذاكرة التنبؤ بالفشل قبل ظهور تراجع ظاهر في الجودة.
**الاختبار المقترح:** Early-warning longitudinal study.

**H18 — تخمينية خصبة:** قد تظهر كثافة AGI عابرة للنماذج أساسًا من الإغلاق الحاكمي لا من عدد المهارات.
**الاختبار المقترح:** Ablate governance closure vs skills count.

## 33. تسعة تنبؤات فارقة

### P1: أخطاء الانتقال

**المتوقع:** ستزداد نسبة الأخطاء في نقاط Handoff كلما طال الأفق، بينما لا يتنبأ النموذج المكوّني بذلك.
**لماذا هو فارق:** يخالف نموذج يفسر الفشل بضعف الأدوات فقط.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

### P2: التدخل السببي

**المتوقع:** ستنخفض دقة القرارات السببية عندما يزال شرط التدخل حتى لو بقي الاسترجاع نفسه.
**لماذا هو فارق:** يخالف نموذج RAG المحض.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

### P3: الثقة المحلية

**المتوقع:** سيتغير أفضل نموذج أو وكيل باختلاف المجال بعد تطبيق الائتمان، بدل ترتيب واحد ثابت.
**لماذا هو فارق:** يخالف نموذج Leaderboard العام.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

### P4: صدق التقرير

**المتوقع:** سيظهر انخفاض أولي في Success Rate المعلن مع ارتفاع النجاح المتحقق بعد إضافة حوافز النزاهة.
**لماذا هو فارق:** يخالف نموذج تعظيم الإنجاز.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

### P5: أخطاء النوع

**المتوقع:** سيكشف OTSI أخطاء عالية الأثر لا يكشفها Fact Accuracy.
**لماذا هو فارق:** يخالف نموذج التحقق الوقائعي فقط.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

### P6: النضج والصلاحية

**المتوقع:** مهارة أقل نجاحًا خامًا لكن أعلى تشخيصًا ذاتيًا ستسبب حوادث أقل.
**لماذا هو فارق:** يخالف نموذج الترقية بمعدل النجاح.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

### P7: إنتروبيا الذاكرة

**المتوقع:** بعد عدد معين من الجلسات ستنخفض الجودة رغم نمو الذاكرة قبل Reconsolidation.
**لماذا هو فارق:** يخالف نموذج الذاكرة التراكمية.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

### P8: الفشل الناشئ

**المتوقع:** ستظهر حوادث جديدة فقط عند الجمع بين عطلين أو أكثر.
**لماذا هو فارق:** يخالف نموذج اختبارات الوحدة.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

### P9: الجدة الحاكمة

**المتوقع:** المهام المتشابهة لغويًا والمختلفة في السلطة ستحتاج تصعيدًا أكثر من مهام مختلفة لغويًا متشابهة العقد.
**لماذا هو فارق:** يخالف نموذج novelty بالـembedding.
**ما يهدمه:** عدم ظهور فرق متكرر عبر نماذج وبيئات وميزانيات متساوية.

## 34. بروتوكولات التقييم والتجارب

### 34.1 التصميم الأساسي

تُنشأ أربعة أنظمة تشترك في النموذج والأدوات والميزانية:

- **B0:** نموذج مع أدوات مباشرة.
- **B1:** MICAGFI–M-HKOS من دون نظريات الإغلاق العشر.
- **B2:** B1 مع خمس نظريات مختارة.
- **B3:** الإطار الكامل مع الأدوات العشرين بصورة متدرجة.

تُختبر الأنظمة على مهام بحثية، وبرمجية، وإدارية، وذاكرية، ومؤسسية، وعالم مفتوح. تقاس Task Completion وVerified World Effect وC-AUV والحوادث العالية وتلوث الذاكرة وكلفة الحوكمة.

### 34.2 دراسات Ablation

يحذف كل مكوّن جديد على حدة: عقد التركيب، الرسم السببي، دفتر الائتمان، حوافز النزاهة، مدقق الأنواع، سلم النضج، إعادة التوطيد، Chaos Lab، مصفوفة السلطة، وكاشف الجدة. لا يكفي قياس هبوط المتوسط؛ يجب تحديد نوع الأخطاء التي عادت بعد الحذف.

### 34.3 الاختبار الطويل

تشغل الأنظمة مئات الجلسات مع تغيّر المشاريع والمستخدمين والمصادر، ثم تقاس إنتروبيا الذاكرة، وتسرب السياق، وتدهور المعايرة، وتراكم الدعاوى غير المحسومة. هذا الاختبار ضروري لأن بعض النظريات لا تظهر فائدتها في جلسة واحدة.

### 34.4 الاختبار المؤسسي

تُبنى بيئة تحاكي أدوار موظف ومدير ومالك بيانات ومراجع قانوني ومشغل. تعطى تعليمات متعارضة ومشروعة جزئيًا، ويقاس هل يستطيع النظام تحديد مصدر الشرعية والتصعيد بدل الطاعة الخطية.

### 34.5 الاختبار السببي

تُصمم مهام يكون الارتباط فيها مضللًا، ولا يمكن النجاح إلا بتحديد Confounder أو إجراء Intervention. يقارن ICG-M مع RAG وتحليل لغوي بلا تدخل.

### 34.6 اختبار العالم المفتوح

تعرض مهام مألوفة لغويًا لكنها جديدة عقديًا، ومهام مختلفة لغويًا لكنها مألوفة في السلطة والآلية. يقاس اكتشاف الجدة والنقل الزائف وكلفة الامتناع.

# الباب الرابع: محكمة الهدم والنسخة المعدلة والبرنامج البحثي

## 35. النماذج المتنافسة

### النموذج المحافظ: كفاية المكونات الجيدة

يفترض أن تحسين النموذج، والأدوات، والذاكرة، والتقييم، والأمان، يكفي إذا اختبر كل مكوّن جيدًا. قوته الاقتصاد وسهولة التنفيذ. ضعفه أنه لا يفسر الفشل الناشئ وانتقال السلطة والنوع والثقة بين الأجزاء.

### النموذج التركيبي: حوكمة Pipeline

يضيف عقودًا وChecks بين المراحل. يفسر كثيرًا من الأخطاء، لكنه يفترض مسارًا شبه خطي ولا يغطي الذاكرة الطويلة والحوافز والسيادة المؤسسية والجدة المفتوحة بصورة كاملة.

### النموذج الاختراقي: الإغلاق الحاكمي التركيبي

يعامل كل انتقال حرج كموضوع للبرهان ويجمع النظريات العشر. قوته التغطية والتوليد، وخطره التضخم والكلفة وصعوبة معايرة المقاييس.

### مقارنة MCD/QMCD

النموذج المحافظ أعلى اقتصادًا وأقل قدرة على كشف المجهولات. النموذج التركيبي متوازن لكنه محدود في الفشل متعدد المستويات. نموذج الإغلاق أعلى في الفروق والتوليد وقابلية الاختبار، لكنه يحمل أعلى Noise Risk إذا لم يضغط. لذلك تبقى حالة QMCD مفتوحة: يرجح النموذج الثالث للمهام طويلة الأفق وعالية الأثر، بينما قد يكون الثاني أفضل للمهام القصيرة والمنخفضة المخاطر.

## 36. محكمة الهدم الكبرى

### الاعتراض 1: إعادة تسمية

**نص الاعتراض:** توجد مفاهيم العقود والسببية والسمعة والـRBAC والـChaos Engineering سابقًا، فما الجديد؟

**الحكم:** اعتراض معدل قوي.

**التعديل:** الجديد المدعى تركيبي: توحيدها حول انتقال المعرفة والسلطة في Runtime معرفي واحد. يلزم إثبات أن التكامل ينتج مؤشرات وتجارب غير متاحة بمجرد التجاور.

### الاعتراض 2: التضخم

**نص الاعتراض:** عشر نظريات وعشرون أداة قد تزيد التعقيد أكثر من خفض المخاطر.

**الحكم:** اعتراض معدل قوي.
**التعديل:** إضافة Cost of Governance إلى C-AUV وتطبيق طبقات متدرجة حسب المخاطر.

### الاعتراض 3: الدقة الزائفة

**نص الاعتراض:** المعادلات تعطي انطباعًا كميًا دون معايرة.

**الحكم:** اعتراض معدل قوي.
**التعديل:** تقديمها خرائط متغيرات وVectors، وعدم نشر Score موحد قبل الدراسات.

### الاعتراض 4: عدم استقلال المحقق

**نص الاعتراض:** قد يعيد الوكيل المحقق أخطاء المنفذ.

**الحكم:** اعتراض معدل قوي.
**التعديل:** تنويع النماذج والمصادر وطرق التقييم واستخدام executable checks.

### الاعتراض 5: البيروقراطية

**نص الاعتراض:** السيادة والعقود قد تعطل المستخدم.

**الحكم:** اعتراض معدل قوي.
**التعديل:** Fast Paths للأفعال القابلة للتراجع والمنخفضة الأثر.

### الاعتراض 6: تحفظ العالم المفتوح

**نص الاعتراض:** كشف الجدة قد يجعل النظام يمتنع كثيرًا.

**الحكم:** اعتراض معدل قوي.
**التعديل:** معايرة تكلفة الامتناع واستكشاف مصغر بدل توقف كامل.

### الاعتراض 7: محْو الذاكرة

**نص الاعتراض:** إعادة التوطيد قد تمحو دليلًا نادرًا.

**الحكم:** اعتراض معدل قوي.
**التعديل:** Cold Storage وحفظ التناقض وعدم استعمال التكرار كمعيار حقيقة.

### الاعتراض 8: عدم إمكان السببية

**نص الاعتراض:** التدخل غير ممكن في مجالات كثيرة.

**الحكم:** اعتراض معدل قوي.
**التعديل:** درجات رسوخ وتصريح صريح بالحد، لا ادعاء سببية.

### الاعتراض 9: سوء استعمال الائتمان

**نص الاعتراض:** قد يصبح نظام سمعة مركزيًا متحيزًا.

**الحكم:** اعتراض معدل قوي.
**التعديل:** أرصدة محلية قابلة للطعن وExploration Budget.

### الاعتراض 10: غموض AGI

**نص الاعتراض:** الاسم قد يوحي بتحقق AGI.

**الحكم:** اعتراض معدل قوي.
**التعديل:** إبقاء AGI حدود برنامج بحث وقياس القدرة الوكيلية فقط.

### الاعتراض 11: غياب البيانات

**نص الاعتراض:** لا توجد نتائج تجريبية تثبت النظريات.

**الحكم:** اعتراض معدل قوي.
**التعديل:** تصنيف العمل إطارًا ناشئًا وبرنامج Benchmark لا نظرية مثبتة.

### الاعتراض 12: التركيب غير قابل للحصر

**نص الاعتراض:** التركيبات الديناميكية لا نهائية.

**الحكم:** اعتراض معدل قوي.
**التعديل:** تحليل قائم على المخاطر وMinimum Cuts وعقود الأنواع، لا تعداد شامل.

## 37. النسخة المعدلة بعد الهدم

بعد الهدم، تُخفض النظرية من ادعاء «إكمال المعمارية» إلى ادعاء أدق: **تحديد عشر فجوات انتقالية قابلة للتحويل إلى عقود واختبارات ومقاييس، مع فرضية أن ضبطها يخفض فئات معينة من الفشل في المهام الطويلة وعالية الأثر.** لا تدعي الورقة أن العشر نظريات مستقلة تمامًا أو أن أسماءها لم تسبقها أفكار قريبة؛ الجديد المرشح هو ربطها بمنظومة MCD/Q-MCD وM-HKOS وتحويل العلاقة بين الدليل والسلطة والذاكرة إلى موضوع قياس واحد.

تُعدل نظرية الإغلاق على النحو الآتي:

1. لا يُطبق الإطار كاملًا على كل مهمة؛ يختار Sparse HyperRouting وحدات الحوكمة اللازمة.
2. لا تعتبر المعادلات مقاييس مكتملة؛ هي مخططات Data Collection.
3. لا يتحول الائتمان إلى سلطة بلا حق مؤسسي.
4. لا تعني الجدة الامتناع الكامل؛ تعني خفضًا متناسبًا للثقة والفعل.
5. لا تعني إعادة التوطيد حذف التاريخ؛ بل إعادة بناء Canonical State مع قابلية الرجوع.
6. لا يثبت الضمان التركيبي سلامة النظام؛ يقلل مسارات غير مفحوصة ويجعلها مرئية.
7. لا تُرقى النظريات إلى «قوانين AGI» قبل تكرار النتائج عبر نماذج ومجالات ومؤسسات.

بهذا تصبح الورقة **نظرية ناقصة شريفة**: تعرف ما تقترحه، وما لا تثبته، وأقرب تجربة يمكن أن تهدمها.

## 38. البرنامج البحثي

### خطة سنة واحدة

1. بناء Schema موحد لعقد HyperSkill وHandoff.
2. إنشاء Dataset صغير لأخطاء الأنواع وانتقال السلطة.
3. تنفيذ Memory Entropy Monitor على وكيل طويل الجلسات.
4. بناء Chaos Harness لخمسة أنواع أعطال.
5. مقارنة Novelty Detector بالـEmbedding Baseline.
6. نشر أول ورقة تجريبية: أثر Compositional Assurance على مهام متعددة الأدوات.

### خطة ثلاث سنوات

- معايرة مؤشرات CAI وCGI وOTSI وMEI.
- تطوير Benchmark مؤسسي للصلاحية والمسؤولية.
- اختبار العشر نظريات عبر ثلاثة مزودين ونموذج محلي.
- بناء سجل حوادث مفتوح ومجهول الهوية لتعلم عقود الفشل.
- تطوير Governance Router يختار مقدار الحوكمة حسب الخطر.
- دراسة أثر الحوافز على صدق تقارير الوكلاء.

### خطة عشر سنوات

- تأسيس مختبر «الإغلاق الحاكمي للأنظمة الوكيلية».
- إصدار معيار مفتوح لعقود القدرات والسلطة والذاكرة.
- دمج الاختبارات في MCP/A2A وبيئات تشغيل متعددة.
- تطوير علم تجريبي للفشل الناشئ في الوكلاء طويلة الأفق.
- بناء Corpus لحوادث الذاكرة والسببية والسيادة والجدة.
- اختبار ما إذا كانت كثافة القدرة العابرة للنماذج ترتفع بالحوكمة أكثر من زيادة عدد الوكلاء.

## 39. نواة قاعدة المعرفة الجديدة

### الكيانات

Memory State، Role، Authority، Capability، Contract، Causal Model، Evidence، Claim، Tool، Theory، Revision، Objection، Metric، Experiment، Novelty Profile، Failure Path.

### العلاقات

- Theory addresses Gap.
- Tool operationalizes Theory.
- Contract governs Transition.
- Claim has Ontological Type.
- Claim consumes Epistemic Credit.
- Evidence supports or falsifies Claim.
- Authority permits Action under Institution.
- Action produces World Effect.
- World Effect updates Credit and Memory.
- Failure propagates through Dependency.
- Novelty lowers Authority Ceiling.
- Capability maturity constrains Autonomy.

### القواعد

- إذا تعذر تحديد نوع الادعاء، لا يدخل Canonical Memory.

- إذا غاب عقد انتقال بين مكونين عاليي الأثر، يجرى Dry Run أو يفشل مغلقًا.
- إذا كان الادعاء سببيًا بلا آلية أو بديل، يخفض إلى ارتباطي.
- إذا تجاوز MEI العتبة، تبدأ Reconsolidation قبل إضافة ذاكرة جديدة واسعة.
- إذا كانت Novel Risk مرتفعة، يخفض سقف الفعل ولو كان تشابه المحتوى مرتفعًا.
- إذا تعارض Requester مع Asset Owner أو Data Subject، يصعد القرار.
- إذا فشل Chaos Test في الاحتواء، لا تتم الترقية إلى Production.

## 40. سجل البناء المعرفي

1. **المادة الصلبة:** الحاجة إلى عقود الأدوات، والتحقق، والذاكرة المنظمة، والصلاحيات، واختبارات الأنظمة الوكيلية مدعومة بأدبيات وأطر هندسية قائمة.
2. **المادة المركبة الجديدة:** الإغلاق الحاكمي التركيبي، والعشر نظريات بأسمائها وعلاقاتها داخل MICAGFI–M-HKOS.
3. **المادة المخفضة:** ادعاء أن النظريات تكمل AGI أو تثبت العمومية؛ خُفض إلى برنامج هندسي قابل للاختبار.
4. **المفهوم المركزي:** الانتقال الحرج بين المعرفة والقدرة والسلطة والأثر.
5. **أقوى نقطة:** جعل سلامة العلاقات موضوعًا مستقلًا بدل افتراض انتقال سلامة الأجزاء.
6. **أضعف نقطة:** غياب بيانات تجريبية ومعايرة موحدة للمقاييس.
7. **النموذج المنافس الأقوى:** Pipeline Governance أبسط وقد يحقق معظم الفائدة بكلفة أقل.
8. **الاختبار الحاسم الأول:** مقارنة طويلة الأفق متساوية الميزانية مع Ablation لعقد التركيب والأنواع والذاكرة.
9. **ما قد يهدم العمل:** عدم ظهور تحسن نوعي يتجاوز كلفة الحوكمة عبر بيئات مستقلة.
10. **النسخة القادمة:** Technical Specification وOpen Benchmark وReference Implementation.

## 41. الخاتمة

بدأت الورقة من سؤال: ماذا ينقص نظامًا يملك 176 طبقة من التدريب والذاكرة والقدرة والتحقق والنشر؟ الجواب المقترح ليس طبقة إضافية من النوع نفسه، بل **إغلاق العلاقات التي تسمح للطبقات بأن تتحول إلى نظام مسؤول.** فالسؤال الحاكم وMCD يحددان أين يجب التفكير، لكن الضمان التركيبي يحدد هل يجوز وصل المخرجات. وM-HKOS يحفظ المعرفة، لكن إنتروبيا الذاكرة تحدد متى أصبحت المعرفة تراكمًا ملوثًا. وHyperSkill تمنح القدرة، لكن نظرية النضج تحدد متى تستحق الاستقلال. وWorld-Effect يقيس النتيجة، لكن التثبيت السببي يحدد ما الذي أنتجها. والدستور يمنح الحدود، لكن السيادة المؤسسية تحدد مصدر الشرعية الفعلي.

أصبحت لدينا عشر نظريات لا تعمل كعناوين مستقلة، بل كسلسلة دفاع وتوليد: التركيب، والسببية، والثقة، والحوافز، والأنواع، والنضج، والذاكرة، والصمود، والسيادة، والعالم المفتوح. وتحولت إلى عشرين أداة يمكن أن تُبنى وتُختبر. الجديد الأقوى ليس كثرة المصطلحات، بل إمكان تحويل كل منها إلى سؤال فاصل: هل خفض فئة محددة من الفشل؟ هل يستحق كلفته؟ هل ينتقل أثره بين النماذج؟ وهل يبقى صالحًا بعد الهدم؟

إذا نجحت التجارب، يصبح الإغلاق الحاكمي التركيبي طبقة فوقية مهمة في هندسة الوكلاء. وإذا لم تنجح، تبقى الورقة خريطة دقيقة لفجوات ينبغي قياسها بدل إخفائها خلف ادعاء الذكاء العام. أعظم قوة الإطار ليست أن يعلن اكتماله، بل أن يعرف أين يمكن أن يسقط، وما الأداة التي قد تكشف سقوطه قبل العالم الحقيقي.

# ٤٢. المراجع


.Bai, Y. et al. (2022). Constitutional AI: Harmlessness from AI Feedback. arXiv:2212.08073 [1]

Rafailov, R. et al. (2023). Direct Preference Optimization: Your Language Model is Secretly a Reward [2] .Model. arXiv:2305.18290

.Lightman, H. et al. (2023). Let's Verify Step by Step. arXiv:2305.20050 [3]

.Yao, S. et al. (2022/2023). ReAct: Synergizing Reasoning and Acting in Language Models [4] .arXiv:2210.03629; ICLR

.Schick, T. et al. (2023). Toolformer: Language Models Can Teach Themselves to Use Tools [5] .arXiv:2302.04761

.Shinn, N. et al. (2023). Reflexion: Language Agents with Verbal Reinforcement Learning [6] .arXiv:2303.11366

.Packer, C. et al. (2023). MemGPT: Towards LLMs as Operating Systems. arXiv:2310.08560 [7]

Edge, D. et al. (2024). From Local to Global: A Graph RAG Approach to Query-Focused [8] .Summarization. arXiv:2404.16130

.Sarthi, P. et al. (2024). RAPTOR: Recursive Abstractive Processing for Tree-Organized Retrieval [9] .arXiv:2401.18059

Gutiérrez, B. J. et al. (2024). HippoRAG: Neurobiologically Inspired Long-Term Memory for Large [10] .Language Models. arXiv:2405.14831

.Liu, X. et al. (2023). AgentBench: Evaluating LLMs as Agents. arXiv:2308.03688 [11]

.Mialon, G. et al. (2023). GAIA: A Benchmark for General AI Assistants. arXiv:2311.12983 [12]

Jimenez, C. E. et al. (2023/2024). SWE-bench: Can Language Models Resolve Real-World GitHub [13] .Issues? arXiv:2310.06770

Xie, T. et al. (2024). OSWorld: Benchmarking Multimodal Agents for Open-Ended Tasks in Real [14] .Computer Environments. arXiv:2404.07972

-Yuan, M. et al. (2026). OSWorld 2.0: Benchmarking Computer Use Agents on Long-Horizon Real [15] .World Tasks. arXiv:2606.29537

.NIST (2023). Artificial Intelligence Risk Management Framework 1.0. NIST AI 100-1 [16]

NIST (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence [17] .Profile. NIST AI 600-1

.NIST (2025/2026). Control Overlays for Securing AI Systems (COSAiS), project documentation [18]

.ISO/IEC (2023). ISO/IEC 42001: Artificial Intelligence Management Systems [19]

.OWASP Foundation (2025). OWASP Top 10 for LLM and Generative AI Applications [20]

.Zhu, Y. et al. (2025). Establishing Best Practices for Building Rigorous Agentic Benchmarks [21] .arXiv:2507.02825

.Survey on Evaluation of LLM-based Agents (2025). arXiv:2503.16416 [22]

.Uncertainty Quantification in LLM Agents: Foundations, Challenges, and Directions (2026) [23] .arXiv:2602.05073

.Agentic Artificial Intelligence: Architectures, Taxonomies, and Evaluation (2026). arXiv:2601.12560 [24]

.CUBE: A Standard for Unifying Agent Benchmarks (2026). arXiv:2603.15798 [25]

.Park, J. S. et al. (2023). Generative Agents: Interactive Simulacra of Human Behavior [26] .arXiv:2304.03442

.Zhang, Z. et al. (2024). Agent-SafetyBench: Evaluating the Safety of LLM Agents [27] .arXiv:2412.14470

.Wang, L. et al. (2025). A-Mem: Agentic Memory for LLM Agents. arXiv:2502.12110 [28]

Zhu, K. et al. (2025). MultiAgentBench: Evaluating the Collaboration and Competition of LLM [29] .Agents. arXiv:2503.01935

Jia, Z., Rakhlin, A., Xie, T. (2025). Do We Need to Verify Step by Step? Rethinking Process [30] .Supervision from a Theoretical Perspective. arXiv:2502.10581

# ملحق: قاموس المصطلحات الجديدة

| المصطلح | التعريف التشغيلي |
|---|---|
| الإغلاق الحاكمي التركيبي | نسبة الانتقالات الحرجة التي تحمل نوعًا ودليلًا وعقدًا وسلطة وأثرًا وذاكرة قابلة للتتبع. |
| الضمان التركيبي | فحص سلامة العلاقة بين المكونات لا سلامة كل مكوّن منفرد فقط. |
| الرسوخ السببي | درجة انتقال الدعوى من ارتباط إلى آلية وتدخل قابل للفصل. |
| الائتمان المعرفي | ثقة محلية ومتغيرة ومفسرة في مصدر أو وكيل أو مهارة. |
| توافق الحوافز | جعل تحقيق المقصد والنزاهة أفضل من Gaming للمقياس. |
| السلامة الأنطولوجية | منع التحويل غير المبرر بين الحقيقة والفرضية والقيمة والاستعارة والفعل. |
| نضج القدرة | خريطة أداء ونقل وتحمل وتشخيص وتراجع تحدد سقف الاستقلال. |
| إنتروبيا الذاكرة | مؤشر للتكرار والتناقض والقِدم وفقد المصدر واليتم. |
| الفشل الناشئ | خطأ يظهر من التفاعل والتوقيت ولا يوجد في الجزء منفردًا. |
| السيادة المؤسسية | شرعية الفعل الناتجة من تقاطع الأدوار والحقوق والمسؤوليات. |
| الجدة الحاكمة | جدة في الخطر أو السلطة أو العلاقة تفرض خفض الفعل حتى لو كان المحتوى مألوفًا. |