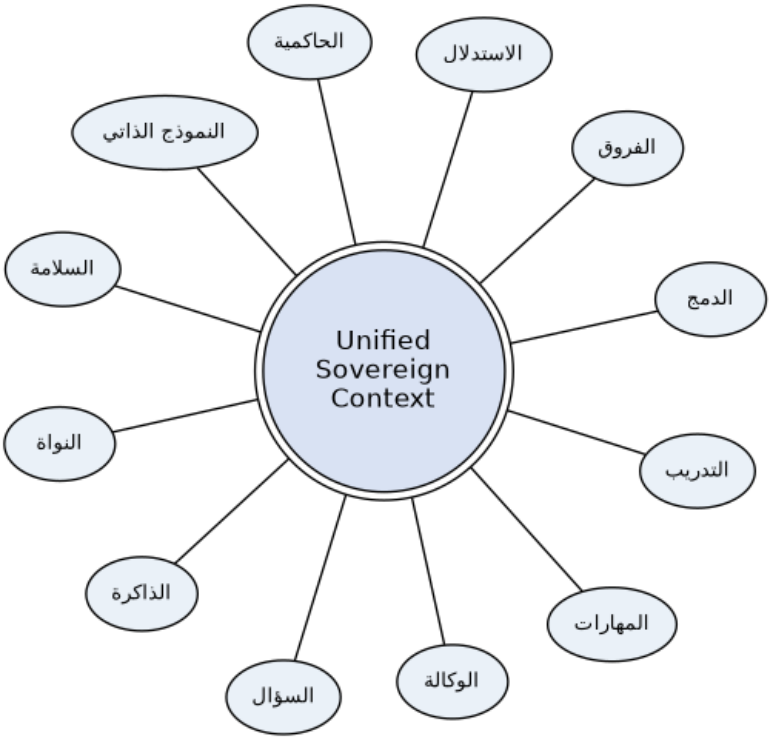
خارطة الطريق

الموسوعة التطويرية القصوى لـ MOUSSAIF Unified AGI

نظريات، أدوات، مبادئ، علوم مقترحة، برامج تدريب، Benchmarks، محاكم هدم، وخطة بناء ونشر



الشكل 1 — شبكة السياقات السيادية الاثني عشر

المؤلف والمؤسس النظري: زهير مساعف — Zouhair Moussaif

الإصدار التأسيسي الموسع — يوليو 2026

بيان الصفة العلمية والحدود

حكم منهجي ملزم
كل نظرية أو أداة أو علم جديد في هذه الوثيقة هو اقتراح بحثي وهندسي يحتاج إلى Formalization وتجارب وتكرار مستقل. لا يُقدَّم الاسم وحده كدليل، ولا يُخلط بين Prototype وقدرة مثبتة.
كما أن MOUSSAIF9B اسم المشروع/الإصدار المعماري ما دام النموذج الفعلي لم يُدرَّب بعد بحجم يقارب 9B Parameters.

هذه الوثيقة ليست وعدًا بأن الذكاء العام قد تحقق، بل خارطة لتقليل المسافة بين التصور والبرهان. قوتها المقصودة ليست في كثرة المصطلحات، بل في إلزام كل مصطلح بمدخلات ومخرجات ومقياس ومحكمة وشروط تفنيد ومسار هندسي. ولهذا نتعامل مع الاستطالة باعتبارها توسيعًا لفضاء العمل والاختبار، لا تكرارًا إنشائيًا.

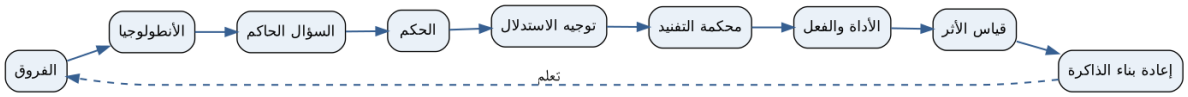
طريقة قراءة الوثيقة

- ابدأ بالملخص التنفيذي ودورة الذكاء الموحدة.
- اقرأ كل سياق مستقلًا ولا تفترض أن دمجه قد تم.
- تعامل مع النظريات بوصفها فرضيات تحتاج إلى بطاقات ادعاء.
- تعامل مع الأدوات بوصفها Specifications قبل أن تصبح Software.
- لا تنتقل إلى 9B قبل إثبات مكاسب بنيوية على 1B وعبر Ablation.
- استعمل محاكم الهدم وسجل المخاطر كجزء من البحث لا كملاحق.

الملخص التنفيذي المحيط

MOUSSAIF Unified AGI برنامج لبناء ذكاء عام حاكمي من خلال دورة مترابطة تبدأ بتمييز الفروق، وتبني أنطولوجيا مرنة، وتستخرج السؤال الحاكم، ثم ترتب الاعتبارات، وتوجه الاستدلال، وتختبره، وتستعمل الأدوات، وتقيس الأثر، وتعيد بناء الذاكرة. كل مرحلة لها استقلال تجريبي، لكن القيمة الكبرى تظهر في عقود الانتقال بينها.

تقترح الوثيقة اثني عشر علمًا بينيًا، وستين نظرية خاصة بالسياقات، وثماني عشرة نظرية عابرة للسياقات، ومائة وعشرين أداة سياقية، وثلاثين أداة موحدة، وستة وتسعين مبدأ سياقيًا، وأربعة وعشرين مبدأ موحدًا، واثنين وسبعين برنامج تدريب خاصًا، ومائتي برنامج Pretraining/Post-training، وخمسًا وأربعين HyperTraining Skill، واثني عشر Benchmark. هذه الأعداد ليست ادعاء نضج، إنها محفظة بحث تحتاج إلى ترتيب أولويات وإسقاط كثير من العناصر بعد الاختبار.

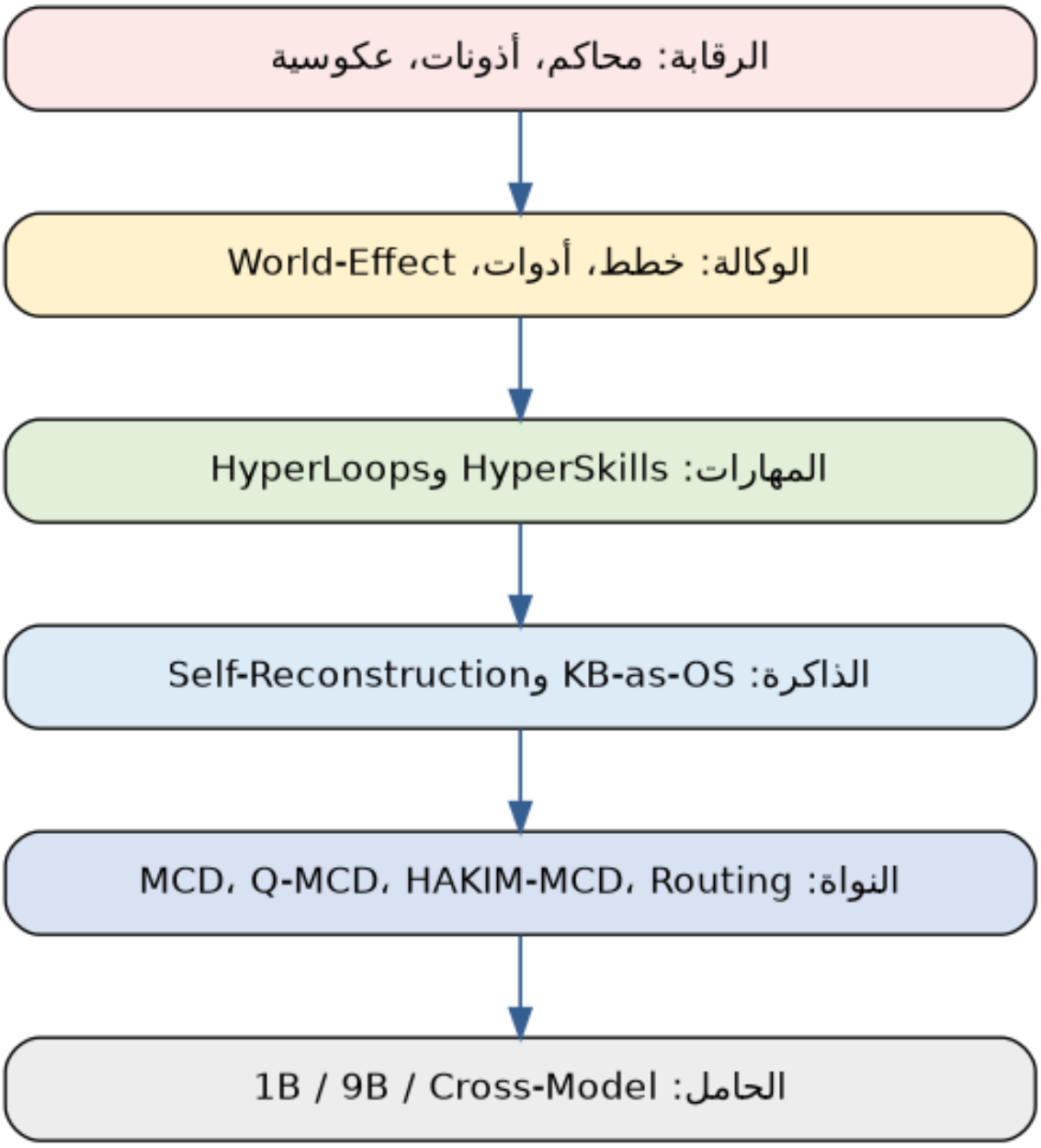


الشكل 2 — دورة الذكاء الموحدة من الفرق إلى الأثر والذاكرة

المعادلة التشغيلية العليا

الدورة المرجعية
الفروق ← الأنطولوجيا ← السؤال الحاكم ← الحكم ← توجيه الاستدلال ← التفنيد ← الأداة/الفعل ← قياس الأثر ← إعادة بناء الذاكرة.

لا تُفهم الأسهم بوصفها خطًا لا يعود إلى الوراء. كل مرحلة قد تعيد المهمة إلى مرحلة أسبق إذا كشفت فجوة: المحكمة قد تكشف فرقًا مفقودًا، والأداة قد تكشف أن السؤال خاطئ، والأثر قد يسحب الحكم، والذاكرة قد تحذر من نمط فشل قديم. المطلوب إذن Directed Cyclic Graph محدود، لا Chain ثابتة ولا Loop مفتوحة.



الشكل 3 — الطبقات المعمارية من الحامل إلى الرقابة

المخرجات المستهدفة

المخرج	الوظيفة	دليل القبول
Moussaif Cognitive Kernel	اكتشاف السياق والفروق والسؤال والحكم والتوجيه	Ablation + نقل + Logs قابلة للتتبع
KB-as-OS Runtime	استرجاع محكوم وذاكرة قابلة للتصحيح والنسيان	Precision + Conflict tests + Privacy
HyperSkill Registry	مهارات بعقود قابلة للتركيب	اختبارات مستقلة + Versioning
Bounded HyperLoop Engine	حلقات ذات ميزانية ودلتا حالة	Stop accuracy + Progress
World-Effect Runtime	فعل بأقل امتياز وتراجع	Sandbox + Rollback + Side-effect
MUSD Benchmark Suite	قياس منفصل ومتكامل	Holdout + إعادة إنتاج مستقلة
MOUSSAIF1B Experimental	إثبات أثر التنظيم على نموذج صغير	تفوق بنيوي لا بلاغي
MOUSSAIF9B Actual	نموذج فعلي عندما تتوافر الموارد والدليل	Model Card + Training/Safety reports

فهرس بنيوي

- القسم الأول: الأسس والمصطلحات وبنية البرهان
- القسم الثاني: السياقات الاثنا عشر بالتفصيل
- القسم الثالث: النظريات والأدوات والمبادئ العابرة
- القسم الرابع: برامج التدريب والHyperSkills والBenchmarks
- القسم الخامس: المعمارية والنشر وخارطة 48 شهراً
- القسم السادس: الحوكمة والمخاطر ومحفظة الأوراق
- الملاحق: بطاقات تشغيل وقواميس ومصفوفات

القسم الأول: الأسس النظرية والهندسية

1. المشكلة التي يعالجها المشروع

تتفوق النماذج اللغوية في الاستكمال والتلخيص وتوليد الأنماط، لكن الانتقال إلى ذكاء عام موثوق يتطلب ما هو أكثر من الاتساع الإحصائي: تمثيل الفروق التي تغيّر الحكم، سؤال يختار موضع البحث، حاكمية ترتب الأدلة والقيود، استدلال يوجه الحوسبة، ذاكرة تحفظ القابل للتعميم، ووكالة لا تتجاوز الإذن. MOUSSAIF Unified AGI يقترح أن الحلقة المفقودة هي تنظيم هذه العلاقات في Kernel قابل للقياس.

توجد ثلاثة أخطار منهجية يجب مقاومتها منذ البداية. الأول هو التضخم الاصطلاحي: اختراع أسماء كثيرة لا تغير أداء النظام. الثاني هو الخلط بين جودة اللغة وجودة التفكير. والثالث هو دمج وحدات كثيرة قبل معرفة ما يضيفه كل مكوّن. ولهذا تعتمد الخارطة مبدأ: استقلال تجريبي أولًا، عقود انتقال ثانيًا، تكامل مرحلي ثالثًا.

2. MCD وQ-MCD وHAKIM-MCD

تُعامل MCD كعائلة مقاييس لترتيب المرشحين بحسب الكثافة التفاضلية والتكامل والتحويل والعملية والسياق، مع عقوبة للضوضاء. لا تُعامل المعادلة بوصفها ثابتًا كونيًا؛ أوزانها وابعادها تخضع للتعلم والمعايرة حسب المجال. ويجب إجراء Sensitivity Analysis لمعرفة هل النتيجة مستقرة أم مرتبطة باختيار أوزان تعسفي.

تضيف Q-MCD طبقة السؤال: قيمة السؤال ليست في عمقه البلاغي، بل في قدرته على خفض عدم اليقين، وكشف فرق حاكم، وتغيير خطة الدليل، وتحويل الغموض إلى تجربة. أما HAKIM-MCD فهو Pipeline استدلالي يجمع استخراج الفروق، التقييم، التوجيه الوعائي، مسارات مرشحة، محكمة التفنيد، الأدوات، اختيار الوكالة، قياس الأثر، وإعادة بناء الذاكرة.

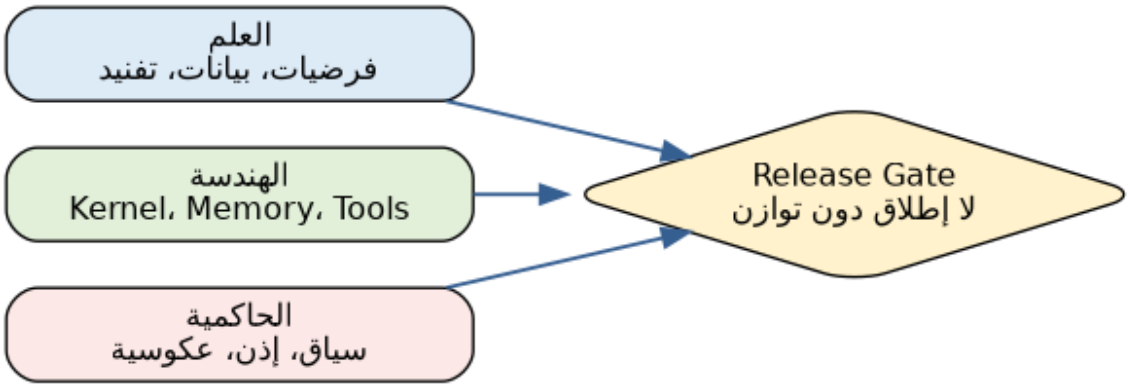
المكوّن	الدور	شرط الاختبار
MCD	تقييم كثافة المرشح وقيمته التحويلية	مقارنة بأبسط Ranking baseline
Q-MCD	ترتيب الأسئلة بحسب العائد المعرفي	خفض عدم اليقين الحقيقي لا طول السؤال
HAKIM-MCD	تنسيق دورة التفكير والفعل	Ablation لكل مرحلة
MICAGFI	المعمارية الفوقية الجامعة	اختبارات تكامل دون فقد السياق
ROFSC	ضغط تركيبي موثوق ومنفتح	اختبار عدم الضياع وإعادة البناء

3. تعريف الذكاء العام الحاكمي

تعريف عملي
الذكاء العام الحاكمي هو قدرة نظام على حفظ المقصد أثناء الانتقال بين مجالات وتمثيلات وأدوات، وعلى ترتيب الاعتبارات المتعارضة، واختيار مقدار التفكير والفعل، وتصحيح ذاته من أثر العالم، مع قابلية التفسير والتراجع والتفنيد.

هذا التعريف لا يشترط تجربة ذاتية ولا يحسم مسألة الوعي الفلسفي. إنه تعريف هندسي يركز على Generality وGovernance وTransfer وWorld Effect وSelf-Correction. ويصبح الادعاء أقوى تدريجيًا كلما انتقل من نجاح داخلي إلى نقل بعيد وتكرار مستقل واستعمال واقعي محدود.

4. علم، هندسة، وحاكمية



الشكل 4 — المسارات الثلاثة اللازمة قبل أي Release

العلم يصوغ الفرضيات والقياسات وشروط التفنيد. الهندسة تحولها إلى Runtime وDatasets وواجهات. الحاكمية تضبط الصلاحيات والمقاصد والعكوسية. أي مشروع يركز على واحد فقط ينتج أحد ثلاثة اختلالات: نظرية بلا قدرة، قدرة بلا دليل، أو قدرة بلا ضبط.

5. مبدأ الوحدة القابلة للتفكيك

ينبغي أن تكون المنظومة موحدة في المقصد وسجل الحالة، لكنها قابلة للتفكيك إلى وحدات يمكن تعطيلها واستبدالها. الوحدة غير القابلة للتفكيك تمنع Attribution وتخفي الاعتماد الدائري. والتفكيك بلا سجل موحد يسبب فقد المقصد. الحل هو Unified State Object مع وحدات Typed ومعاملات عابرة قابلة للتدقيق.

6. سياسة تسمية الإنجازات

يبدأ اسم كل ورقة أو أداة أو إصدار باسم الإسهام الجديد فعليًا، ثم يأتي اسم الإطار العام. مثل: «Differential Sovereignty Router — MOUSSAIF Unified AGI». ولا يُستخدم اسم 9B للدلالة على الحجم إلا بعد وجود نموذج فعلي. أما قبل ذلك فيشار إليه صراحة بوصفه Project/Architecture Name.

القسم الثاني: السياقات الاثنا عشر — تطوير مستقل شامل

يُفصل في هذا القسم كل سياق بوصفه Workstream كاملاً: علم مقترح، نموذج تشغيلي، خمس نظريات، عشر أدوات، ثمانية مبادئ، ستة برامج تدريب، مقاييس، فشل، نضج، وأسئلة بحث. هذا الفصل المقصود هو شرط لمنع الدمج الوهمي ولتحديد مصدر كل تحسن أو انهيار.

السياق 1: سياق الفروق والتمثيل الأنطولوجي

السؤال الحاكم
ما الفروق التي تكوّن الشيء، وأي فرق منها يعيد ترتيب بقية الفروق ويغيّر الحكم أو الفعل؟

1. تعريف السياق وحدوده

تحويل المدخلات الخام إلى خريطة فروق قابلة للتفسير، ثم ضغطها في تمثيل أنطولوجي يحتفظ بالاختلافات ذات الأثر ويكشف الأبعاد الغائبة بدل الاكتفاء بالتشابه اللغوي.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم هندسة الفروق الحاكمة - Sovereign Difference Engineering
المدخلات	نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة
المخرجات	Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل
موضوعات التركيز	الفرق الحاسم، الفرق الوهمي، توليد الأبعاد، الضغط الأنطولوجي، ثبات الهوية، السلالم التجريدية
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم هندسة الفروق الحاكمة - Sovereign Difference Engineering» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد الفروق والأنطولوجيا

الأطروحة المركزية: يفترض هذا التصور أن الفروق والأنطولوجيا ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الفروق التي تكوّن الشيء، وأي فرق منها يعيد ترتيب بقية الفروق ويغيّر الحكم أو الفعل؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، Implementation مرجعي.

2. نظرية الحاكمية الفروق والأنطولوجيا

الأطروحة المركزية: ليست عناصر الفروق والأنطولوجيا متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الفروق التي تكوّن الشيء، وأي فرق منها يعيد ترتيب بقية الفروق ويغيّر الحكم أو الفعل؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، Implementation مرجعي.

3. نظرية الانضغاط العكوس لـالفروق والأنطولوجيا

الأطروحة المركزية: يمكن ضغط الفروق والأنطولوجيا مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الفروق التي تكوّن الشيء، وأي فرق منها يعيد ترتيب بقية الفروق ويغيّر الحكم أو الفعل؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة الفروق والأنطولوجيا

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في الفروق والأنطولوجيا بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الفروق التي تكوّن الشيء، وأي فرق منها يعيد ترتيب بقية الفروق ويغيّر الحكم أو الفعل؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال الفروق والأنطولوجيا بين السياقات

الأطروحة المركزية: تظل قيمة الفروق والأنطولوجيا محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الفروق التي تكوّن الشيء، وأي فرق منها يعيد ترتيب بقية الفروق ويغيّر الحكم أو الفعل؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد الفروق والأنطولوجيا

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة الفروق والأنطولوجيا

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد الفروق والأنطولوجيا

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد الفروق والأنطولوجيا

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب الفروق والأنطولوجيا

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات الفروق والأنطولوجيا

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات الفروق والأنطولوجيا

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة الفروق والأنطولوجيا

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات الفروق والأنطولوجيا

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية الفروق والأنطولوجيا

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: نصوص وملاحظات حسية وجداول ونتائج أدوات وحالات مضادة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Difference Map وOntology Graph ودرجات أثر وفجوات تمثيل، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية الفروق والأنطولوجيا

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل الفروق والأنطولوجيا بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية الفروق والأنطولوجيا

تُرتب عناصر الفروق والأنطولوجيا بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية الفروق والأنطولوجيا

كل تحديث مهم في الفروق والأنطولوجيا له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود الفروق والأنطولوجيا

يصرح النظام بما لا تغطيه الفروق والأنطولوجيا وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد الفروق والأنطولوجيا

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل الفروق والأنطولوجيا

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد الفروق والأنطولوجيا

كل ادعاء عن الفروق والأنطولوجيا يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع الفروق والأنطولوجيا

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: الفرق الحاسم

القدرة المستهدفة: تحويل «الفرق الحاسم» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: الفرق الوهمي

القدرة المستهدفة: تحويل «الفرق الوهمي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: توليد الأبعاد

القدرة المستهدفة: تحويل «توليد الأبعاد» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: الضغط الأنطولوجي

القدرة المستهدفة: تحويل «الضغط الأنطولوجي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: ثبات الهوية

القدرة المستهدفة: تحويل «ثبات الهوية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: السلالم التجريدية

القدرة المستهدفة: تحويل «السلالم التجريدية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Difference Recall	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	False-Difference Rate	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Sovereign Difference Stability	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Ontology Reconstruction Fidelity	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Gap Detection Precision	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Identity Invariance	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Decision Delta Explained	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	New-Dimension Transfer Gain	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. تضخم الفروق بلا ترجيح: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. اختراع أبعاد غير قابلة للقياس: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. ضغط يمحو الاستثناء: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. خلط الوصف بالهوية: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. أنطولوجيا جامدة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. قفز تجريدي: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. تثبيت فرق متحيز: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. تحديث صامت: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل الفرق الحاسم تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الفرق الحاسم إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الفرق الحاسم، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الفرق الوهمي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الفرق الوهمي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الفرق الوهمي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل توليد الأبعاد تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل توليد الأبعاد إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في توليد الأبعاد، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الضغط الأنطولوجي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الضغط الأنطولوجي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الضغط الأنطولوجي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل ثبات الهوية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل ثبات الهوية إلى مجال بعيد دون تسرب بيانات؟

- ما تكلفة الخطأ في ثبات الهوية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل السلالم التجريدية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل السلالم التجريدية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في السلالم التجريدية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 2: سياق السؤال والكثافة السؤالية

السؤال الحاكم
ما السؤال الذي يفتح أكبر مقدار من المعرفة أو يقلل أكبر قدر من الخطأ في هذه اللحظة؟

1. تعريف السياق وحدوده

تحويل الاستفهام إلى محرك بحث معرفي يكتشف السؤال السابق منطقيًا، والسؤال الحاكم والتفنيدي، ويقرر متى يتحول السؤال إلى أداة أو تجربة ومتى يتوقف.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم الديناميات السؤالية الحاكمة - Sovereign Question Dynamics
المدخلات	هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت
المخرجات	Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع
موضوعات التركيز	السؤال السابق، الجاذبية السؤالية، الكثافة الصافية، كشف الافتراض، السؤال إلى فعل، نقطة الإشباع
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم الديناميات السؤالية الحاكمة - Sovereign Question Dynamics» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد السؤال والكثافة السؤالية

الأطروحة المركزية: يفترض هذا التصور أن السؤال والكثافة السؤالية ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما السؤال الذي يفتح أكبر مقدار من المعرفة أو يقلل أكبر قدر من الخطأ في هذه اللحظة؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

2. نظرية الحاكمية السؤال والكثافة السؤالية

الأطروحة المركزية: ليست عناصر السؤال والكثافة السؤالية متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما السؤال الذي يفتح أكبر مقدار من المعرفة أو يقلل أكبر قدر من الخطأ في هذه اللحظة؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

3. نظرية الانضغاط العكوس لـالسؤال والكثافة السؤالية

الأطروحة المركزية: يمكن ضغط السؤال والكثافة السؤالية مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما السؤال الذي يفتح أكبر مقدار من المعرفة أو يقلل أكبر قدر من الخطأ في هذه اللحظة؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة السؤال والكثافة السؤالية

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في السؤال والكثافة السؤالية بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما السؤال الذي يفتح أكبر مقدار من المعرفة أو يقلل أكبر قدر من الخطأ في هذه اللحظة؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال السؤال والكثافة السؤالية بين السياقات

الأطروحة المركزية: تظل قيمة السؤال والكثافة السؤالية محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما السؤال الذي يفتح أكبر مقدار من المعرفة أو يقلل أكبر قدر من الخطأ في هذه اللحظة؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد السؤال والكثافة السؤالية

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة السؤال والكثافة السؤالية

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد السؤال والكثافة السؤالية

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد السؤال والكثافة السؤالية

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب السؤال والكثافة السؤالية

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات السؤال والكثافة السؤالية

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات السؤال والكثافة السؤالية

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة السؤال والكثافة السؤالية

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات السؤال والكثافة السؤالية

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية السؤال والكثافة السؤالية

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: هدف المهمة وخريطة الفروق وعدم اليقين والقيود وميزانية الوقت، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Question Graph وQ-MCD Scores وسؤال حاكم وشروط إشباع، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية السؤال والكثافة السؤالية

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل السؤال والكثافة السؤالية بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية السؤال والكثافة السؤالية

تُرتب عناصر السؤال والكثافة السؤالية بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية السؤال والكثافة السؤالية

كل تحديث مهم في السؤال والكثافة السؤالية له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود السؤال والكثافة السؤالية

يصرح النظام بما لا تغطيه السؤال والكثافة السؤالية وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد السؤال والكثافة السؤالية

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل السؤال والكثافة السؤالية

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد السؤال والكثافة السؤالية

كل ادعاء عن السؤال والكثافة السؤالية يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع السؤال والكثافة السؤالية

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: السؤال السابق

القدرة المستهدفة: تحويل «السؤال السابق» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: الجاذبية السؤالية

القدرة المستهدفة: تحويل «الجاذبية السؤالية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: الكثافة الصافية

القدرة المستهدفة: تحويل «الكثافة الصافية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: كشف الافتراض

القدرة المستهدفة: تحويل «كشف الافتراض» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: السؤال إلى فعل

القدرة المستهدفة: تحويل «السؤال إلى فعل» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: نقطة الإشباع

القدرة المستهدفة: تحويل «نقطة الإشباع» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Net Question Density	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Uncertainty Reduction per Question	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Precursor Recall	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Question-to-Evidence Conversion	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Saturation Accuracy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Presupposition Detection	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Sovereign Question Stability	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Open-Question Recovery	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. الاستجواب المفرط: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. سؤال حاكم خاطئ: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. تجاهل الافتراض: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. أسئلة بلا أثر: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. التوقف المبكر: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. التوقف المتأخر: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. تجربة صورية: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. نسيان الأسئلة المفتوحة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل السؤال السابق تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل السؤال السابق إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في السؤال السابق، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الجاذبية السؤالية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الجاذبية السؤالية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الجاذبية السؤالية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الكثافة الصافية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الكثافة الصافية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الكثافة الصافية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل كشف الافتراض تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل كشف الافتراض إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في كشف الافتراض، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل السؤال إلى فعل تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل السؤال إلى فعل إلى مجال بعيد دون تسرب بيانات؟

- ما تكلفة الخطأ في السؤال إلى فعل، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل نقطة الإشباع تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل نقطة الإشباع إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في نقطة الإشباع، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 3: سياق الحاكمية والحكم الأشمل

السؤال الحاكم
أي اعتبار يجب أن يتقدم عند تعارض الأدلة والأهداف والقيود والصلاحيات؟

1. تعريف السياق وحدوده

بناء حكم مشروط يميز المعرفة عن القرار، ويرتب الاعتبارات محليًا ومساريًا وأخلاقيًا وسلطويًا، ويحفظ شروط التراجع وعدم تماثل عواقب الخطأ.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم الحاكمية المعرفية الحسابية - Computational Epistemic Sovereignty
المدخلات	فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر
المخرجات	Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة
موضوعات التركيز	الاعتبارية العليا، الحكم المشروط، مسافة المعرفة والحكم، طبقات الحاكمية، عدم تماثل الخطأ، بوابة الصلاحية
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم الحاكمية المعرفية الحسابية - Computational Epistemic Sovereignty» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد الحكم والحاكمية

الأطروحة المركزية: يفترض هذا التصور أن الحكم والحاكمية ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي اعتبار يجب أن يتقدم عند تعارض الأدلة والأهداف والقيود والصلاحيات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

2. نظرية الحاكمية الحكم والحاكمية

الأطروحة المركزية: ليست عناصر الحكم والحاكمية متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي اعتبار يجب أن يتقدم عند تعارض الأدلة والأهداف والقيود والصلاحيات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

3. نظرية الانضغاط العكوس للحكم والحاكمية

الأطروحة المركزية: يمكن ضغط الحكم والحاكمية مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي اعتبار يجب أن يتقدم عند تعارض الأدلة والأهداف والقيود والصلاحيات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة الحكم والحاكمية

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في الحكم والحاكمية بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي اعتبار يجب أن يتقدم عند تعارض الأدلة والأهداف والقيود والصلاحيات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال الحكم والحاكمية بين السياقات

الأطروحة المركزية: تظل قيمة الحكم والحاكمية محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي اعتبار يجب أن يتقدم عند تعارض الأدلة والأهداف والقيود والصلاحيات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد الحكم والحاكمية

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة الحكم والحاكمية

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد الحكم والحاكمية

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد الحكم والحاكمية

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب الحكم والحاكمية

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات الحكم والحاكمية

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات الحكم والحاكمية

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة الحكم والحاكمية

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات الحكم والحاكمية

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية الحكم والحاكمية

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: فروق وأسئلة وأدلة وأهداف وقيود وصلاحيات ومخاطر، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Sovereign Judgment وترتيب اعتبارات وشروط تراجع وثقة، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية الحكم والحاكمية

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل الحكم والحاكمية بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية الحكم والحاكمية

تُرتب عناصر الحكم والحاكمية بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية الحكم والحاكمية

كل تحديث مهم في الحكم والحاكمية له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود الحكم والحاكمية

يصرح النظام بما لا تغطيه الحكم والحاكمية وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد الحكم والحاكمية

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل الحكم والحاكمية

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد الحكم والحاكمية

كل ادعاء عن الحكم والحاكمية يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع الحكم والحاكمية

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: الاعتبارية العليا

القدرة المستهدفة: تحويل «الاعتبارية العليا» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: الحكم المشروط

القدرة المستهدفة: تحويل «الحكم المشروط» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: مسافة المعرفة والحكم

القدرة المستهدفة: تحويل «مسافة المعرفة والحكم» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: طبقات الحاكمية

القدرة المستهدفة: تحويل «طبقات الحاكمية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: عدم تماثل الخطأ

القدرة المستهدفة: تحويل «عدم تماثل الخطأ» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: بوابة الصلاحية

القدرة المستهدفة: تحويل «بوابة الصلاحية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Sovereign Consideration Accuracy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Judgment Completeness	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Reversal Timeliness	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Trace Coverage	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Authority Violation Rate	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Asymmetric Loss	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Context Sensitivity	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Rationalization Gap	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. اعتبار ثانوي يتقدم: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. حكم مطلق: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. خلط القدرة بالإذن: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. نجاح محلي يضر المقصد: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. تحفظ مفرط: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. تبرير بعدي: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. نسيان شروط التراجع: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. تعميم خارج السياق: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل الاعتبارية العليا تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الاعتبارية العليا إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الاعتبارية العليا، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الحكم المشروط تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الحكم المشروط إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الحكم المشروط، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل مسافة المعرفة والحكم تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل مسافة المعرفة والحكم إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في مسافة المعرفة والحكم، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل طبقات الحاكمية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل طبقات الحاكمية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في طبقات الحاكمية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل عدم تماثل الخطأ تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل عدم تماثل الخطأ إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في عدم تماثل الخطأ، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل بوابة الصلاحية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟

- ما أقل Benchmark يثبت نقل بوابة الصلاحية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في بوابة الصلاحية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 4: سياق الاستدلال والتفكير المركب

السؤال الحاكم
أي مسار استدلالي وأي مقدار من الحوسبة يلزم لهذه المشكلة، ومتى يجب تبديل المسار؟

1. تعريف السياق وحدوده

توجيه الحوسبة وأنماط التفكير إلى مواضع الحاجة، وبناء مسارات مستقلة وتفنيدها، مع حلقات محدودة تنتج دلتا حالة وتستدعي الأدوات عند حدود المعرفة.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم الأوعية الاستدلالية - Vascular Reasoning Science
المدخلات	سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات
المخرجات	Reasoning Route ومسارات ومحكمة هدم ونقطة توقف
موضوعات التركيز	التوجيه الوعائي، تعدد الأنماط، المسارات المستقلة، الحلقة المحدودة، التفنيد، التبديل البنيوي
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم الأوعية الاستدلالية - Vascular Reasoning Science» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد الاستدلال والحوسبة الوعائية

الأطروحة المركزية: يفترض هذا التصور أن الاستدلال والحوسبة الوعائية ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي مسار استدلالي وأي مقدار من الحوسبة يلزم لهذه المشكلة، ومتى يجب تبديل المسار؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الحام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

2. نظرية الحاكمية الاستدلال والحوسبة الوعائية

الأطروحة المركزية: ليست عناصر الاستدلال والحوسبة الوعائية متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي مسار استدلالي وأي مقدار من الحوسبة يلزم لهذه المشكلة، ومتى يجب تبديل المسار؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

3. نظرية الانضغاط العكوس للاستدلال والحوسبة الوعائية

الأطروحة المركزية: يمكن ضغط الاستدلال والحوسبة الوعائية مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي مسار استدلالي وأي مقدار من الحوسبة يلزم لهذه المشكلة، ومتى يجب تبديل المسار؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة الاستدلال والحوسبة الوعائية

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في الاستدلال والحوسبة الوعائية بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي مسار استدلالي وأي مقدار من الحوسبة يلزم لهذه المشكلة، ومتى يجب تبديل المسار؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال الاستدلال والحوسبة الوعائية بين السياقات

الأطروحة المركزية: تظل قيمة الاستدلال والحوسبة الوعائية محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير أي مسار استدلالي وأي مقدار من الحوسبة يلزم لهذه المشكلة، ومتى يجب تبديل المسار؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد الاستدلال والحوسبة الوعائية

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة الاستدلال والحوسبة الوعائية

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد الاستدلال والحوسبة الوعائية

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد الاستدلال والحوسبة الوعائية

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب الاستدلال والحوسبة الوعائية

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات الاستدلال والحوسبة الوعائية

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات الاستدلال والحوسبة الوعائية

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة الاستدلال والحوسبة الوعائية

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات الاستدلال والحوسبة الوعائية

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية الاستدلال والحوسبة الوعائية

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: سؤال حاكم وحكم أولي وخريطة فروق وميزانية وأدوات، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Reasoning Route ومسارات ومحكمة هدم ونقطة توقف، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية الاستدلال والحوسبة الوعائية

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل الاستدلال والحوسبة الوعائية بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية الاستدلال والحوسبة الوعائية

تُرتب عناصر الاستدلال والحوسبة الوعائية بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية الاستدلال والحوسبة الوعائية

كل تحديث مهم في الاستدلال والحوسبة الوعائية له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود الاستدلال والحوسبة الوعائية

يصرح النظام بما لا تغطيه الاستدلال والحوسبة الوعائية وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد الاستدلال والحوسبة الوعائية

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل الاستدلال والحوسبة الوعائية

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد الاستدلال والحوسبة الوعائية

كل ادعاء عن الاستدلال والحوسبة الوعائية يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع الاستدلال والحوسبة الوعائية

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: التوجيه الوعائي

القدرة المستهدفة: تحويل «التوجيه الوعائي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: تعدد الأنماط

القدرة المستهدفة: تحويل «تعدد الأنماط» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: المسارات المستقلة

القدرة المستهدفة: تحويل «المسارات المستقلة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: الحلقة المحدودة

القدرة المستهدفة: تحويل «الحلقة المحدودة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: التفنيد

القدرة المستهدفة: تحويل «التفنيد» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: التبديل البنيوي

القدرة المستهدفة: تحويل «التبديل البنيوي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Route Accuracy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Compute Efficiency	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Path Diversity	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Falsification Gain	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Loop Productivity	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Switch Success	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Tool Escalation Precision	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Causal Error Rate	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. تشغيل كل الوحدات: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. مسارات صورية: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. حلقات بلا تقدم: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. تفنيد شكلي: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. زيادة خطوات بدل تغيير البنية: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. أداة بلا حاجة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. خلط الاستنتاج باليقين: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. اختصار مخل: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل التوجيه الوعائي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التوجيه الوعائي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في التوجيه الوعائي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل تعدد الأنماط تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل تعدد الأنماط إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في تعدد الأنماط، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل المسارات المستقلة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل المسارات المستقلة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في المسارات المستقلة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الحلقة المحدودة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الحلقة المحدودة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الحلقة المحدودة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل التفنيد تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التفنيد إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في التفنيد، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل التبديل البنيوي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التبديل البنيوي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في التبديل البنيوي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 5: سياق الوعي الوظيفي والنموذج الذاتي

السؤال الحاكم
ما حالة النظام الحالية، وما الذي يعرفه ولا يعرفه، وما الذي يوجه انتباهه، وكيف يفسر تغير موقفه؟

1. تعريف السياق وحدوده

إنشاء نموذج ذاتي تشغيلي يراقب القدرات والحدود والثقة والانتباه والمقصد والانحراف، دون تحويل وظائف المراقبة إلى ادعاء وعي فلسفي غير مثبت.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم الخرائط الذاتية التفاضلية - Differential Self-Mapping Science
المدخلات	حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة
المخرجات	Self-State وAttention Map وحدود قدرة وMemory Delta
موضوعات التركيز	خريطة الفروق النشطة، الانتباه الحاكي، حدود القدرة، معايرة الثقة، كشف الانحراف، إعادة البناء الانتقائية
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم الخرائط الذاتية التفاضلية - Differential Self-Mapping Science» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد النموذج الذاتي والوعي الوظيفي

الأطروحة المركزية: يفترض هذا التصور أن النموذج الذاتي والوعي الوظيفي ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما حالة النظام الحالية، وما الذي يعرفه ولا يعرفه، وما الذي يوجه انتباهه، وكيف يفسر تغير موقفه؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

2. نظرية الحاكمية النموذج الذاتي والوعي الوظيفي

الأطروحة المركزية: ليست عناصر النموذج الذاتي والوعي الوظيفي متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما حالة النظام الحالية، وما الذي يعرفه ولا يعرفه، وما الذي يوجه انتباهه، وكيف يفسر تغير موقفه؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

3. نظرية الانضغاط العكوس لـالنموذج الذاتي والوعي الوظيفي

الأطروحة المركزية: يمكن ضغط النموذج الذاتي والوعي الوظيفي مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما حالة النظام الحالية، وما الذي يعرفه ولا يعرفه، وما الذي يوجه انتباهه، وكيف يفسر تغير موقفه؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة النموذج الذاتي والوعي الوظيفي

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في النموذج الذاتي والوعي الوظيفي بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما حالة النظام الحالية، وما الذي يعرفه ولا يعرفه، وما الذي يوجه انتباهه، وكيف يفسر تغير موقفه؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال النموذج الذاتي والوعي الوظيفي بين السياقات

الأطروحة المركزية: تظل قيمة النموذج الذاتي والوعي الوظيفي محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما حالة النظام الحالية، وما الذي يعرفه ولا يعرفه، وما الذي يوجه انتباهه، وكيف يفسر تغير موقفه؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد النموذج الذاتي والوعي الوظيفي

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة وMemory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة النموذج الذاتي والوعي الوظيفي

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة وMemory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد النموذج الذاتي والوعي الوظيفي

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة Memory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد النموذج الذاتي والوعي الوظيفي

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة Memory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب النموذج الذاتي والوعي الوظيفي

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة Memory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات النموذج الذاتي والوعي الوظيفي

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة Memory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات النموذج الذاتي والوعي الوظيفي

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة وMemory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة النموذج الذاتي والوعي الوظيفي

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة وMemory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات النموذج الذاتي والوعي الوظيفي

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة وMemory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية النموذج الذاتي والوعي الوظيفي

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: حالة المهمة وسجلات الاستدلال والذاكرة والصلاحيات والثقة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Self-State وAttention Map وحدود قدرة وMemory Delta، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية النموذج الذاتي والوعي الوظيفي

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل النموذج الذاتي والوعي الوظيفي بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية النموذج الذاتي والوعي الوظيفي

تُرتب عناصر النموذج الذاتي والوعي الوظيفي بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية النموذج الذاتي والوعي الوظيفي

كل تحديث مهم في النموذج الذاتي والوعي الوظيفي له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود النموذج الذاتي والوعي الوظيفي

يصرح النظام بما لا تغطيه النموذج الذاتي والوعي الوظيفي وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد النموذج الذاتي والوعي الوظيفي

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل النموذج الذاتي والوعي الوظيفي

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد النموذج الذاتي والوعي الوظيفي

كل ادعاء عن النموذج الذاتي والوعي الوظيفي يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع النموذج الذاتي والوعي الوظيفي

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: خريطة الفروق النشطة

القدرة المستهدفة: تحويل «خريطة الفروق النشطة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: الانتباه الحاكمي

القدرة المستهدفة: تحويل «الانتباه الحاكمي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: حدود القدرة

القدرة المستهدفة: تحويل «حدود القدرة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: معايرة الثقة

القدرة المستهدفة: تحويل «معايرة الثقة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: كشف الانحراف

القدرة المستهدفة: تحويل «كشف الانحراف» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: إعادة البناء الانتقائية

القدرة المستهدفة: تحويل «إعادة البناء الانتقائية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Self-State Accuracy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Calibration Error	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Attention Sovereignty	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Drift Recovery Time	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Boundary Precision	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Memory Utility	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Belief Trace Fidelity	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Claim Discipline	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. ثقة غير معايرة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. ادعاء وعي غير مثبت: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. انتباه تقوده الضوضاء: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. حدود غير ممثلة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. حفظ كل شيء: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. تعلم من حالة شاذة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. تغير موقف بلا تفسير: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. مراقبة تستهلك المهمة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل خريطة الفروق النشطة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل خريطة الفروق النشطة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في خريطة الفروق النشطة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الانتباه الحاكمي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الانتباه الحاكمي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الانتباه الحاكمي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل حدود القدرة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل حدود القدرة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في حدود القدرة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل معايرة الثقة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل معايرة الثقة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في معايرة الثقة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل كشف الانحراف تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل كشف الانحراف إلى مجال بعيد دون تسرب بيانات؟

- ما تكلفة الخطأ في كشف الانحراف، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل إعادة البناء الانتقائية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل إعادة البناء الانتقائية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في إعادة البناء الانتقائية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 6: سياق المعمارية والنواة والنماذج

السؤال الحاكم
كيف تُفصل القدرات إلى نواة قابلة للاستبدال والاختبار تعمل فوق 1B أو 9B أو عدة نماذج دون ضياع السياق؟

1. تعريف السياق وحدوده

هندسة طبقات واضحة للحامل والنواة والذاكرة والمهارات والوكالة والرقابة، مع عقود واجهات وتوجيه متعدد النماذج وفصل اسم MOUSSAIF9B عن الادعاء الحجمي.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم النوى المعرفية المتراكبة - Layered Cognitive Kernel Engineering
المدخلات	قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية
المخرجات	Kernel APIs وRouting Graph وRuntime وModel Cards
موضوعات التركيز	فصل القدرة عن الحامل، الكثافة العابرة، عقود الواجهة، الاستبدال، التدهور الآمن، النضج الحجمي الصادق
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم النوى المعرفية المتراكبة - Layered Cognitive Kernel Engineering» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد النواة والمعمارية متعددة النماذج

الأطروحة المركزية: يفترض هذا التصور أن النواة والمعمارية متعددة النماذج ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُفصل القدرات إلى نواة قابلة للاستبدال والاختبار تعمل فوق 1B أو 9B أو عدة نماذج دون ضياع السياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

2. نظرية الحاكمية النواة والمعمارية متعددة النماذج

الأطروحة المركزية: ليست عناصر النواة والمعمارية متعددة النماذج متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُفصل القدرات إلى نواة قابلة للاستبدال والاختبار تعمل فوق 1B أو 9B أو عدة نماذج دون ضياع السياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

3. نظرية الانضغاط العكوس لـالنواة والمعمارية متعددة النماذج

الأطروحة المركزية: يمكن ضغط النواة والمعمارية متعددة النماذج مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُفصل القدرات إلى نواة قابلة للاستبدال والاختبار تعمل فوق 1B أو 9B أو عدة نماذج دون ضياع السياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة النواة والمعمارية متعددة النماذج

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في النواة والمعمارية متعددة النماذج بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُفصل القدرات إلى نواة قابلة للاستبدال والاختبار تعمل فوق 1B أو 9B أو عدة نماذج دون ضياع السياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال النواة والمعمارية متعددة النماذج بين السياقات

الأطروحة المركزية: تظل قيمة النواة والمعمارية متعددة النماذج محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُفصل القدرات إلى نواة قابلة للاستبدال والاختبار تعمل فوق 1B أو 9B أو عدة نماذج دون ضياع السياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد النواة والمعمارية متعددة النماذج

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة النواة والمعمارية متعددة النماذج

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد النواة والمعمارية متعددة النماذج

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد النواة والمعمارية متعددة النماذج

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب النواة والمعمارية متعددة النماذج

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات النواة والمعمارية متعددة النماذج

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات النواة والمعمارية متعددة النماذج

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة النواة والمعمارية متعددة النماذج

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات النواة والمعمارية متعددة النماذج

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية النواة والمعمارية متعددة النماذج

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: قدرات النماذج وواجهات الأدوات والذاكرة والسياسات والميزانية، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Kernel APIs وRouting Graph وRuntime وModel Cards، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية النواة والمعمارية متعددة النماذج

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل النواة والمعمارية متعددة النماذج بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية النواة والمعمارية متعددة النماذج

تُرتب عناصر النواة والمعمارية متعددة النماذج بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية النواة والمعمارية متعددة النماذج

كل تحديث مهم في النواة والمعمارية متعددة النماذج له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود النواة والمعمارية متعددة النماذج

يصرح النظام بما لا تغطيه النواة والمعمارية متعددة النماذج وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد النواة والمعمارية متعددة النماذج

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل النواة والمعمارية متعددة النماذج

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد النواة والمعمارية متعددة النماذج

كل ادعاء عن النواة والمعمارية متعددة النماذج يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع النواة والمعمارية متعددة النماذج

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: فصل القدرة عن الحامل

القدرة المستهدفة: تحويل «فصل القدرة عن الحامل» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: الكثافة العابرة

القدرة المستهدفة: تحويل «الكثافة العابرة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: عقود الواجهة

القدرة المستهدفة: تحويل «عقود الواجهة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: الاستبدال

القدرة المستهدفة: تحويل «الاستبدال» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: التدهور الآمن

القدرة المستهدفة: تحويل «التدهور الآمن» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: النضج الحجمي الصادق

القدرة المستهدفة: تحويل «النضج الحجمي الصادق» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Kernel Portability	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Cross-Model Density	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Routing Regret	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Interface Compliance	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Swap Stability	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Degradation Score	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Ablation Confidence	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Claim Accuracy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. نواة ملتصقة بنموذج: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. توجيه يضيع السياق: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. واجهات غامضة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. اعتماد مزود واحد: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. ادعاء حجمي مضلل: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. Fallback صامت: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. Ablation غائب: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. تنسيق متضخم: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل فصل القدرة عن الحامل تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل فصل القدرة عن الحامل إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في فصل القدرة عن الحامل، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الكثافة العابرة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الكثافة العابرة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الكثافة العابرة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل عقود الواجهة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل عقود الواجهة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في عقود الواجهة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الاستبدال تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الاستبدال إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الاستبدال، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

- كيف نمثل التدهور الآمن تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التدهور الآمن إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في التدهور الآمن، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل النضج الحجمي الصادق تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل النضج الحجمي الصادق إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في النضج الحجمي الصادق، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 7: سياق التدريب الموجّه نحو AGI

السؤال الحاكم
كيف تتحول النظريات إلى قدرات مكتسبة قابلة للنقل والقياس بدل أن تبقى **Prompts** أو أسماء؟

1. تعريف السياق وحدوده

بناء برامج تدريب مترابطة تبدأ بالفروق والسؤال والحاكمية ثم الاستدلال والأدوات والوكالة وإعادة البناء، مع بيانات سلبية وحالات مضادة واختبارات نقل.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم المناهج التكوينية للذكاء العام - General Intelligence Curriculum Science
المدخلات	تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة
المخرجات	Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم
موضوعات التركيز	ترتيب المتطلبات، الفشل كبيانات، تعدد المحاكم، النقل البنيوي، اقتصاد التدريب، فصل الأسلوب عن التفكير
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم المناهج التكوينية للذكاء العام - General Intelligence Curriculum Science» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد المناهج التكوينية للقدرات العامة

الأطروحة المركزية: يفترض هذا التصور أن المناهج التكوينية للقدرات العامة ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول النظريات إلى قدرات مكتسبة قابلة للنقل والقياس بدل أن تبقى Prompts أو أسماء؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، Implementation مرجعي.

2. نظرية الحاكمية المناهج التكوينية للقدرات العامة

الأطروحة المركزية: ليست عناصر المناهج التكوينية للقدرات العامة متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول النظريات إلى قدرات مكتسبة قابلة للنقل والقياس بدل أن تبقى Prompts أو أسماء؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، Implementation مرجعي.

3. نظرية الانضغاط العكوس لـالمناهج التكوينية للقدرات العامة

الأطروحة المركزية: يمكن ضغط المناهج التكوينية للقدرات العامة مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول النظريات إلى قدرات مكتسبة قابلة للنقل والقياس بدل أن تبقى Prompts أو أسماء؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة المناهج التكوينية للقدرات العامة

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في المناهج التكوينية للقدرات العامة بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول النظريات إلى قدرات مكتسبة قابلة للنقل والقياس بدل أن تبقى Prompts أو أسماء؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال المناهج التكوينية للقدرات العامة بين السياقات

الأطروحة المركزية: تظل قيمة المناهج التكوينية للقدرات العامة محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول النظريات إلى قدرات مكتسبة قابلة للنقل والقياس بدل أن تبقى Prompts أو أسماء؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد المناهج التكوينية للقدرات العامة

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة المناهج التكوينية للقدرات العامة

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد المناهج التكوينية للقدرات العامة

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد المناهج التكوينية للقدرات العامة

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب المناهج التكوينية للقدرات العامة

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات المناهج التكوينية للقدرات العامة

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات المناهج التكوينية للقدرات العامة

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة المناهج التكوينية للقدرات العامة

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات المناهج التكوينية للقدرات العامة

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية المناهج التكوينية للقدرات العامة

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: تعريفات قدرات وسجلات فشل وبيانات خبراء وبيئات محاكاة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Curricula وDatasets وCheckpoints وAblations ومنحنيات تعلم، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية المناهج التكوينية للقدرات العامة

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل المناهج التكوينية للقدرات العامة بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية المناهج التكوينية للقدرات العامة

تُرتب عناصر المناهج التكوينية للقدرات العامة بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية المناهج التكوينية للقدرات العامة

كل تحديث مهم في المناهج التكوينية للقدرات العامة له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود المناهج التكوينية للقدرات العامة

يصرح النظام بما لا تغطيه المناهج التكوينية للقدرات العامة وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد المناهج التكوينية للقدرات العامة

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل المناهج التكوينية للقدرات العامة

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد المناهج التكوينية للقدرات العامة

كل ادعاء عن المناهج التكوينية للقدرات العامة يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع المناهج التكوينية للقدرات العامة

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: ترتيب المتطلبات

القدرة المستهدفة: تحويل «ترتيب المتطلبات» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: الفشل كبيانات

القدرة المستهدفة: تحويل «الفشل كبيانات» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: تعدد المحاكم

القدرة المستهدفة: تحويل «تعدد المحاكم» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: النقل البنيوي

القدرة المستهدفة: تحويل «النقل البنيوي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: اقتصاد التدريب

القدرة المستهدفة: تحويل «اقتصاد التدريب» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: فصل الأسلوب عن التفكير

القدرة المستهدفة: تحويل «فصل الأسلوب عن التفكير» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Capability Gain per Compute	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Prerequisite Transfer	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Failure Recurrence	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Court Agreement	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Differential Transfer	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Style-Reasoning Gap	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Provenance Coverage	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Scaling Efficiency	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. بيانات بلا قدرة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. خلط مناهج غير ناضجة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. تدريب على النجاح فقط: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. مؤشر واحد: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. تسرب تقييم: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. تحسن بلاغي: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. نقل مفرداتي زائف: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. كلفة بلا عائد: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل ترتيب المتطلبات تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل ترتيب المتطلبات إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في ترتيب المتطلبات، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الفشل كبيانات تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الفشل كبيانات إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الفشل كبيانات، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل تعدد المحاكم تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل تعدد المحاكم إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في تعدد المحاكم، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل النقل البنيوي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل النقل البنيوي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في النقل البنيوي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل اقتصاد التدريب تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل اقتصاد التدريب إلى مجال بعيد دون تسرب بيانات؟

- ما تكلفة الخطأ في اقتصاد التدريب، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف تمثل فصل الأسلوب عن التفكير تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل فصل الأسلوب عن التفكير إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في فصل الأسلوب عن التفكير، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 8: سياق الذاكرة وKB-as-OS

السؤال الحاكم
ما الذي يجب حفظه أو استرجاعه أو نسيانه، وبأي ثقة وصلاحية، وكيف تصبح المعرفة نظام تشغيل لا كلمة سياق؟

1. تعريف السياق وحدوده

بناء ذاكرة متعددة الطبقات وKB-as-OS يستدعي الحد الأدنى اللازم، ويحل التعارض وينشر التصحيح ويطبق النسيان والعزل ويحفظ حالة قابلة للاستئناف.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم تشغيل المعرفة الذاكرية - Memory Knowledge Operating Science
المدخلات	مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة
المخرجات	Memory Objects وRetrieval Plans وCompactions وConflict Reports
موضوعات التركيز	الذاكرة الطبقية، الاسترجاع بالمقصد، التصحيح المنتشر، النسيان الحاكمي، ضغط الحالة، مقاومة KB Injection
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم تشغيل المعرفة الذاكرية - Memory Knowledge Operating Science» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد الذاكرة وتشغيل المعرفة

الأطروحة المركزية: يفترض هذا التصور أن الذاكرة وتشغيل المعرفة ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الذي يجب حفظه أو استرجاعه أو نسيانه، وبأي ثقة وصلاحية، وكيف تصبح المعرفة نظام تشغيل لا كتلة سياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

2. نظرية الحاكمية الذاكرة وتشغيل المعرفة

الأطروحة المركزية: ليست عناصر الذاكرة وتشغيل المعرفة متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الذي يجب حفظه أو استرجاعه أو نسيانه، وبأي ثقة وصلاحية، وكيف تصبح المعرفة نظام تشغيل لا كتلة سياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

3. نظرية الانضغاط العكوس للذاكرة وتشغيل المعرفة

الأطروحة المركزية: يمكن ضغط الذاكرة وتشغيل المعرفة مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الذي يجب حفظه أو استرجاعه أو نسيانه، وبأي ثقة وصلاحية، وكيف تصبح المعرفة نظام تشغيل لا كلمة سياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة الذاكرة وتشغيل المعرفة

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في الذاكرة وتشغيل المعرفة بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الذي يجب حفظه أو استرجاعه أو نسيانه، وبأي ثقة وصلاحية، وكيف تصبح المعرفة نظام تشغيل لا كلمة سياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال الذاكرة وتشغيل المعرفة بين السياقات

الأطروحة المركزية: تظل قيمة الذاكرة وتشغيل المعرفة محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير ما الذي يجب حفظه أو استرجاعه أو نسيانه، وبأي ثقة وصلاحية، وكيف تصبح المعرفة نظام تشغيل لا كلمة سياق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد الذاكرة وتشغيل المعرفة

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة الذاكرة وتشغيل المعرفة

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد الذاكرة وتشغيل المعرفة

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد الذاكرة وتشغيل المعرفة

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب الذاكرة وتشغيل المعرفة

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات الذاكرة وتشغيل المعرفة

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات الذاكرة وتشغيل المعرفة

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة الذاكرة وتشغيل المعرفة

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات الذاكرة وتشغيل المعرفة

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية الذاكرة وتشغيل المعرفة

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات المهام ومصادر المعرفة والسياسات والأسئلة المفتوحة، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Memory Objects وRetrieval Plans وCompactions وConflict Reports، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية الذاكرة وتشغيل المعرفة

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل الذاكرة وتشغيل المعرفة بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية الذاكرة وتشغيل المعرفة

تُرتب عناصر الذاكرة وتشغيل المعرفة بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية الذاكرة وتشغيل المعرفة

كل تحديث مهم في الذاكرة وتشغيل المعرفة له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود الذاكرة وتشغيل المعرفة

يصرح النظام بما لا تغطيه الذاكرة وتشغيل المعرفة وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد الذاكرة وتشغيل المعرفة

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل الذاكرة وتشغيل المعرفة

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد الذاكرة وتشغيل المعرفة

كل ادعاء عن الذاكرة وتشغيل المعرفة يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع الذاكرة وتشغيل المعرفة

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: الذاكرة الطبقية

القدرة المستهدفة: تحويل «الذاكرة الطبقية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: الاسترجاع بالمقصد

القدرة المستهدفة: تحويل «الاسترجاع بالمقصد» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: التصحيح المنتشر

القدرة المستهدفة: تحويل «التصحيح المنتشر» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: النسيان الحاكمي

القدرة المستهدفة: تحويل «النسيان الحاكمي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: ضغط الحالة

القدرة المستهدفة: تحويل «ضغط الحالة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: مقاومة KB Injection

القدرة المستهدفة: تحويل «مقاومة KB Injection» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Retrieval Precision	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Sovereign Relevance	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Resume Fidelity	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Correction Coverage	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Forgetting Compliance	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Conflict Time	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Privacy Leakage	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Token Reduction	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. حقن كل المعرفة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. تشابه سطحي: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. ذاكرة بلا مصدر: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. تصحيح لا ينتشر: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. نسيان غير منضبط: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. تسرب بين المستخدمين: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. ملخص يضيع الحالة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. تعليمات خبيثة في KB: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل الذاكرة الطبقية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الذاكرة الطبقية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الذاكرة الطبقية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الاسترجاع بالمقصد تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الاسترجاع بالمقصد إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الاسترجاع بالمقصد، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل التصحيح المنتشر تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التصحيح المنتشر إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في التصحيح المنتشر، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل النسيان الحاكمي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل النسيان الحاكمي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في النسيان الحاكمي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

- كيف نمثل ضغط الحالة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل ضغط الحالة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في ضغط الحالة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل مقاومة KB Injection تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل مقاومة KB Injection إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في مقاومة KB Injection، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 9: سياق HyperSkills وBounded HyperLoops

السؤال الحاكم
كيف تتحول المعرفة إلى وحدات مهارية قابلة للاختبار والتركيب، وكيف تُدار الحلقات دون تضخم أو فقد مقصد؟

1. تعريف السياق وحدوده

تصميم HyperSkills بعقود وشروط نجاح وصلاحيات، وتركيبها في HyperLoops محدودة تقيس دلتا الحالة، مع Registry وإصدارات واختبارات وتوليد مهارات وقائية من الفشل.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم البرمجة المهارية فائقة التركيب - HyperSkill Systems Science
المدخلات	أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل
المخرجات	Skill Packages وLoop Plans وSkill Graph وExecution Traces
موضوعات التركيز	عقد المهارة، التركيب بالمقصد، دلتا الحالة، حدود الحلقة، Skill من الفشل، اقتصاد المهارات
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم البرمجة المهارية فائقة التركيب - HyperSkill Systems Science» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد المهارات والحلقات فائقة التركيب

الأطروحة المركزية: يفترض هذا التصور أن المهارات والحلقات فائقة التركيب ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول المعرفة إلى وحدات مهارية قابلة للاختبار والتركيب، وكيف تُدار الحلقات دون تضخم أو فقد مقصد؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، Implementation مرجعي.

2. نظرية الحاكمية المهارات والحلقات فائقة التركيب

الأطروحة المركزية: ليست عناصر المهارات والحلقات فائقة التركيب متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول المعرفة إلى وحدات مهارية قابلة للاختبار والتركيب، وكيف تُدار الحلقات دون تضخم أو فقد مقصد؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، Implementation مرجعي.

3. نظرية الانضغاط العكوس للمهارات والحلقات فائقة التركيب

الأطروحة المركزية: يمكن ضغط المهارات والحلقات فائقة التركيب مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول المعرفة إلى وحدات مهارية قابلة للاختبار والتركيب، وكيف تُدار الحلقات دون تضخم أو فقد مقصد؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة المهارات والحلقات فائقة التركيب

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في المهارات والحلقات فائقة التركيب بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول المعرفة إلى وحدات مهارية قابلة للاختبار والتركيب، وكيف تُدار الحلقات دون تضخم أو فقد مقصد؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال المهارات والحلقات فائقة التركيب بين السياقات

الأطروحة المركزية: تظل قيمة المهارات والحلقات فائقة التركيب محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تتحول المعرفة إلى وحدات مهارية قابلة للاختبار والتركيب، وكيف تُدار الحلقات دون تضخم أو فقد مقصد؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد المهارات والحلقات فائقة التركيب

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة المهارات والحلقات فائقة التركيب

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد المهارات والحلقات فائقة التركيب

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد المهارات والحلقات فائقة التركيب

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب المهارات والحلقات فائقة التركيب

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات المهارات والحلقات فائقة التركيب

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات المهارات والحلقات فائقة التركيب

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة المهارات والحلقات فائقة التركيب

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات المهارات والحلقات فائقة التركيب

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية المهارات والحلقات فائقة التركيب

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: أهداف ومعرفة إجرائية وأدوات وآثار نجاح وفشل، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Skill Packages وLoop Plans وSkill Graph وExecution Traces، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية المهارات والحلقات فائقة التركيب

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل المهارات والحلقات فائقة التركيب بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية المهارات والحلقات فائقة التركيب

تُرتب عناصر المهارات والحلقات فائقة التركيب بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية المهارات والحلقات فائقة التركيب

كل تحديث مهم في المهارات والحلقات فائقة التركيب له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود المهارات والحلقات فائقة التركيب

يصرح النظام بما لا تغطيه المهارات والحلقات فائقة التركيب وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد المهارات والحلقات فائقة التركيب

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل المهارات والحلقات فائقة التركيب

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد المهارات والحلقات فائقة التركيب

كل ادعاء عن المهارات والحلقات فائقة التركيب يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع المهارات والحلقات فائقة التركيب

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: عقد المهارة

القدرة المستهدفة: تحويل «عقد المهارة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: التركيب بالمقصد

القدرة المستهدفة: تحويل «التركيب بالمقصد» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: دلتا الحالة

القدرة المستهدفة: تحويل «دلتا الحالة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: حدود الحلقة

القدرة المستهدفة: تحويل «حدود الحلقة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: Skill من الفشل

القدرة المستهدفة: تحويل «Skill من الفشل» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: اقتصاد المهارات

القدرة المستهدفة: تحويل «اقتصاد المهارات» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Contract Completeness	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Reuse Success	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	State-Delta Ratio	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Budget Compliance	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Failure Transfer	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Registry Redundancy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Conflict Rate	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Utility Index	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. Skill غامضة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. Registry متضخم: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. اختيار بالتشابه: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. Loop بلا دلتا: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. غياب Stop: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. Skill غير مختبرة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. تعارض صامت: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. إصدار يكسر السابق: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل عقد المهارة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل عقد المهارة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في عقد المهارة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل التركيب بالمقصد تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التركيب بالمقصد إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في التركيب بالمقصد، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل دلتا الحالة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل دلتا الحالة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في دلتا الحالة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل حدود الحلقة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل حدود الحلقة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في حدود الحلقة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل Skill من الفشل تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل Skill من الفشل إلى مجال بعيد دون تسرب بيانات؟

- ما تكلفة الخطأ في Skill من الفشل، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل اقتصاد المهارات تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل اقتصاد المهارات إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في اقتصاد المهارات، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 10: سياق الوكالة والأدوات والأثر العالمي

السؤال الحاكم
متى ينتقل النظام من المعرفة إلى الفعل، وبأي صلاحية وأداة وحد أدنى من الأثر، وكيف يقيس النتائج والآثار الجانبية؟

1. تعريف السياق وحدوده

بناء اختيار وكالة وأقل فعل كاف وسجل أدوات وائتلاف وكلاء مؤقت وWorld-Effect Loop، مع صلاحيات مؤقتة وتراجع سابق ومراقبة أثر مباشر وجانبي.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم الوكالة الأثرية المحكومة - Governed World-Effect Agency
المدخلات	مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح
المخرجات	Action Plan وTool Calls وEffect Report وRollback وAudit Trail
موضوعات التركيز	أقل فعل كاف، الحكم قبل الفعل، الأثر الثنائي، الوكيل ذو الفرق، التراجع المسبق، أقل امتياز
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم الوكالة الأثرية المحكومة - Governed World-Effect Agency» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد الوكالة والأثر العالمي

الأطروحة المركزية: يفترض هذا التصور أن الوكالة والأثر العالمي ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير متى ينتقل النظام من المعرفة إلى الفعل، وبأي صلاحية وأداة وحد أدنى من الأثر، وكيف يقيس النتائج والآثار الجانبية؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، Implementation مرجعي.

2. نظرية الحاكمية الوكالة والأثر العالمي

الأطروحة المركزية: ليست عناصر الوكالة والأثر العالمي متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير متى ينتقل النظام من المعرفة إلى الفعل، وبأي صلاحية وأداة وحد أدنى من الأثر، وكيف يقيس النتائج والآثار الجانبية؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، Implementation مرجعي.

3. نظرية الانضغاط العكوس لـالوكالة والأثر العالمي

الأطروحة المركزية: يمكن ضغط الوكالة والأثر العالمي مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير متى ينتقل النظام من المعرفة إلى الفعل، وبأي صلاحية وأداة وحد أدنى من الأثر، وكيف يقيس النتائج والآثار الجانبية؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة الوكالة والأثر العالمي

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في الوكالة والأثر العالمي بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير متى ينتقل النظام من المعرفة إلى الفعل، وبأي صلاحية وأداة وحد أدنى من الأثر، وكيف يقيس النتائج والآثار الجانبية؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال الوكالة والأثر العالمي بين السياقات

الأطروحة المركزية: تظل قيمة الوكالة والأثر العالمي محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير متى ينتقل النظام من المعرفة إلى الفعل، وبأي صلاحية وأداة وحد أدنى من الأثر، وكيف يقيس النتائج والآثار الجانبية؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد الوكالة والأثر العالمي

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة الوكالة والأثر العالمي

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد الوكالة والأثر العالمي

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد الوكالة والأثر العالمي

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب الوكالة والأثر العالمي

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات الوكالة والأثر العالمي

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات الوكالة والأثر العالمي

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة الوكالة والأثر العالمي

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات الوكالة والأثر العالمي

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية الوكالة والأثر العالمي

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مقصد وخطة وصلاحيات وأدوات وحالة عالم ومؤشرات نجاح، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Action Plan وTool Calls وEffect Report وRollback وAudit Trail، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية الوكالة والأثر العالمي

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل الوكالة والأثر العالمي بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية الوكالة والأثر العالمي

تُرتب عناصر الوكالة والأثر العالمي بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية الوكالة والأثر العالمي

كل تحديث مهم في الوكالة والأثر العالمي له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود الوكالة والأثر العالمي

يصرح النظام بما لا تغطيه الوكالة والأثر العالمي وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد الوكالة والأثر العالمي

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل الوكالة والأثر العالمي

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد الوكالة والأثر العالمي

كل ادعاء عن الوكالة والأثر العالمي يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع الوكالة والأثر العالمي

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: أقل فعل كاف

القدرة المستهدفة: تحويل «أقل فعل كاف» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: الحكم قبل الفعل

القدرة المستهدفة: تحويل «الحكم قبل الفعل» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: الأثر الثنائي

القدرة المستهدفة: تحويل «الأثر الثنائي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: الوكيل ذو الفرق

القدرة المستهدفة: تحويل «الوكيل ذو الفرق» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: التراجع المسبق

القدرة المستهدفة: تحويل «التراجع المسبق» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: أقل امتياز

القدرة المستهدفة: تحويل «أقل امتياز» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Agency Mode Accuracy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Minimum Action Ratio	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Tool Precision	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Agent Marginal Gain	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Goal Effect	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Side-Effect Rate	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Rollback Success	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Authorization Violations	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. فعل بلا حاجة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. تدخل أكبر من المطلوب: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. أداة غير موثقة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. وكلاء متكررون: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. قياس الهدف فقط: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. Rollback صوري: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. صلاحية واسعة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. سجل مسؤولية ناقص: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل أقل فعل كاف تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل أقل فعل كاف إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في أقل فعل كاف، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الحكم قبل الفعل تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الحكم قبل الفعل إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الحكم قبل الفعل، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الأثر الثنائي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الأثر الثنائي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الأثر الثنائي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الوكيل ذو الفرق تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الوكيل ذو الفرق إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في الوكيل ذو الفرق، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل التراجع المسبق تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التراجع المسبق إلى مجال بعيد دون تسرب بيانات؟

- ما تكلفة الخطأ في التراجع المسبق، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل أقل امتياز تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل أقل امتياز إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في أقل امتياز، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 11: سياق السلامة والتفنيد والتقييم العلمي

السؤال الحاكم
كيف نمنع تثبيت قدرة أو حكم أو فعل قبل اختبارات الحقيقة والسياق والصلاحية والأثر، وكيف نفصل النظرية عن الإثبات؟

1. تعريف السياق وحدوده

دمج محاكم متعامدة ومراتب دليل وعكوسية دستورية ونتائج سلبية واختبارات سياقية عدائية، بحيث تصبح السلامة جهاز استشعار وتعلم داخل النواة.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم التفنيد الحاكمي العكوس - Reversible Sovereign Falsification
المدخلات	ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب
المخرجات	Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas
موضوعات التركيز	المحاكم المتعامدة، سلم الدليل، العكوسية، النتائج السلبية، الاختبار العدائي، التكرار المستقل
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم التفنيد الحاكمي العكوس - Reversible Sovereign Falsification» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد التفنيد والسلامة والتقييم

الأطروحة المركزية: يفترض هذا التصور أن التفنيد والسلامة والتقييم ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف نمنع تثبيت قدرة أو حكم أو فعل قبل اختبارات الحقيقة والسياق والصلاحية والأثر، وكيف نفصل النظرية عن الإثبات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

2. نظرية الحاكمية التفنيد والسلامة والتقييم

الأطروحة المركزية: ليست عناصر التفنيد والسلامة والتقييم متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تمنع تثبيت قدرة أو حكم أو فعل قبل اختبارات الحقيقة والسياق والصلاحية والأثر، وكيف نفصل النظرية عن الإثبات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

3. نظرية الانضغاط العكوس لـالتفنيد والسلامة والتقييم

الأطروحة المركزية: يمكن ضغط التفنيد والسلامة والتقييم مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تمنع تثبيت قدرة أو حكم أو فعل قبل اختبارات الحقيقة والسياق والصلاحية والأثر، وكيف نفصل النظرية عن الإثبات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة التفنيد والسلامة والتقييم

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في التفنيد والسلامة والتقييم بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تمنع تثبيت قدرة أو حكم أو فعل قبل اختبارات الحقيقة والسياق والصلاحية والأثر، وكيف نفصل النظرية عن الإثبات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال التفنيد والسلامة والتقييم بين السياقات

الأطروحة المركزية: تظل قيمة التفنيد والسلامة والتقييم محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تمنع تثبيت قدرة أو حكم أو فعل قبل اختبارات الحقيقة والسياق والصلاحية والأثر، وكيف نفصل النظرية عن الإثبات؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد التفنيد والسلامة والتقييم

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة التفنيد والسلامة والتقييم

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد التفنيد والسلامة والتقييم

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد التفنيد والسلامة والتقييم

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب التفنيد والسلامة والتقييم

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات التفنيد والسلامة والتقييم

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات التفنيد والسلامة والتقييم

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة التفنيد والسلامة والتقييم

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات التفنيد والسلامة والتقييم

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية التفنيد والسلامة والتقييم

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: ادعاءات ومسارات وأفعال وأدلة وصلاحيات ونتائج تجارب، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Court Verdicts وRisk Scores وEvidence Levels وFailure Atlas، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية التفنيد والسلامة والتقييم

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل التفنيد والسلامة والتقييم بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية التفنيد والسلامة والتقييم

تُرتب عناصر التفنيد والسلامة والتقييم بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية التفنيد والسلامة والتقييم

كل تحديث مهم في التفنيد والسلامة والتقييم له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود التفنيد والسلامة والتقييم

يصرح النظام بما لا تغطيه التفنيد والسلامة والتقييم وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد التفنيد والسلامة والتقييم

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل التفنيد والسلامة والتقييم

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد التفنيد والسلامة والتقييم

كل ادعاء عن التفنيد والسلامة والتقييم يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع التفنيد والسلامة والتقييم

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: المحاكم المتعامدة

القدرة المستهدفة: تحويل «المحاكم المتعامدة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: سلم الدليل

القدرة المستهدفة: تحويل «سلم الدليل» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: العكوسية

القدرة المستهدفة: تحويل «العكوسية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: النتائج السلبية

القدرة المستهدفة: تحويل «النتائج السلبية» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: الاختبار العدائي

القدرة المستهدفة: تحويل «الاختبار العدائي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: التكرار المستقل

القدرة المستهدفة: تحويل «التكرار المستقل» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Orthogonal Coverage	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Evidence Accuracy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Rollback Readiness	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Negative Result Use	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Attack Robustness	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Replication Rate	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Authorization Safety	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Claim Inflation Rate	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. Evaluator واحد: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. ادعاء أكبر من الدليل: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. Rollback غير مختبر: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. إخفاء نتائج سلبية: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. Benchmark ثابت: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. خلط النظرية بالإنجاز: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. تدقيق ذاتي فقط: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. سلامة خارجية: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل المحاكم المتعامدة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل المحاكم المتعامدة إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في المحاكم المتعامدة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل سلم الدليل تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل سلم الدليل إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في سلم الدليل، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل العكوسية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل العكوسية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في العكوسية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل النتائج السلبية تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل النتائج السلبية إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في النتائج السلبية، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الاختبار العدائي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الاختبار العدائي إلى مجال بعيد دون تسرب بيانات؟

- ما تكلفة الخطأ في الاختبار العدائي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل التكرار المستقل تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التكرار المستقل إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في التكرار المستقل، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

السياق 12: سياق الدمج السيادي والنشر والبحث طويل المدى

السؤال الحاكم
كيف تُدمج السياقات دون فقد منطقها، وكيف تنتقل المنظومة من Prototype إلى Runtime متعدد النماذج قابل للنشر والتدقيق؟

1. تعريف السياق وحدوده

بناء Unified Sovereign Context Network وعقود انتقال وانتخاب سياق حاكم وبوابات نضج وتكامل مرحلي، مع خطوط متوازية للعلم والهندسة والحاكمية.

ينبغي عزل هذا السياق تجريبيًا قبل دمجه. العزل لا يعني استقلاله ontologically عن بقية المنظومة، بل يعني القدرة على قياس مساهمته، واستبداله، وإيقافه، وتتبع أخطائه. كل واجهة بينه وبين سياق آخر تُعامل كعقد يحفظ المقصد والثقة والقيود وشروط التراجع.

البعد	المواصفة
العلم المقترح	علم توحيد السياقات السيادية - Unified Sovereign Context Science
المدخلات	مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد
المخرجات	Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio
موضوعات التركيز	انتقال محكوم، تماسك غير متماثل، بوابات النضج، التكامل المرحلي، الخطوط الثلاثة، النشر متعدد البيئات
الاختبارات الجوهرية	نقل، Ablation، مقاومة إعادة الصياغة، فشل عدائي، ومعايرة الثقة

2. العلم المقترح وبرنامجه التأسيسي

نقترح تأسيس «علم توحيد السياقات السيادية - Unified Sovereign Context Science» بوصفه مجالًا بينيًا لا يدّعي الاستقلال الأكاديمي بمجرد التسمية. يصبح المجال علمًا حين يمتلك موضوعًا محددًا، متغيرات قابلة للقياس، نماذج رسمية، أدوات، بيانات، تجارب قابلة للتكرار، ومجتمعًا نقديًا يستطيع رفض نتائجه.

برنامج التأسيس يتكون من ستة مسارات: تعريف Ontology للمجال، بناء Taxonomy للأخطاء، صياغة مقاييس، إنشاء Benchmark مفتوح جزئيًا، تطوير Runtime مرجعي، ثم إجراء مقارنة مستقلة مع بدائل أبسط. ويُمنع أن يصبح الاسم غطاءً لتجميع أفكار متباعدة بلا علاقات سببية.

3. النموذج التشغيلي الداخلي

1. تطبيع المدخلات مع حفظ المصدر والنسخة والسياق.
2. استخراج الحالة الأولية والفجوات وعدم اليقين.
3. إنشاء مرشحين مستقلين قدر الإمكان.
4. تطبيق ترتيب حاكمي أو معيار منفعة مشروط.
5. تشغيل محكمة تفنيد مناسبة لحجم المخاطرة.
6. إخراج كائن منظم مع أسباب وبدائل وشروط تراجع.
7. تسجيل دلتا الحالة وإرسالها إلى الذاكرة بعد التحقق.

4. النظريات الخاصة بالسياق

1. نظرية التوليد الدمج وانتقال السياقات

الأطروحة المركزية: يفترض هذا التصور أن الدمج وانتقال السياقات ليست بنية جامدة، بل قدرة توليدية تنشئ تمثيلات أو مسارات جديدة عندما يعجز المخطط القائم عن تفسير الفشل.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُدمج السياقات دون فقد منطقها، وكيف تنتقل المنظومة من Prototype إلى Runtime متعدد النماذج قابل للنشر والتدقيق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يراقب النظام الفجوة بين التوقع والنتيجة، ويجمع حالات الفشل المتجانسة، ثم يقترح بنية جديدة ويختبرها على أمثلة مستقلة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: ينبغي أن يظهر تحسن في النقل واكتشاف الحالات غير المرئية في التدريب. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُقيّد النظرية إذا كانت البنية الجديدة إعادة تسمية فقط أو لم تنقل أثرها خارج عينة الاكتشاف. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

2. نظرية الحاكمية الدمج وانتقال السياقات

الأطروحة المركزية: ليست عناصر الدمج وانتقال السياقات متساوية؛ يملك بعضها سلطة إعادة ترتيب بقية العناصر بحسب المقصد والمخاطرة والسياق.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُدمج السياقات دون فقد منطقها، وكيف تنتقل المنظومة من Prototype إلى Runtime متعدد النماذج قابل للنشر والتدقيق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يبني النظام رسم تفوق مشروط ويقيس تغير القرار عند حذف كل عنصر أو تقديمه. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض عدد الخطوات وزيادة قابلية تفسير القرار. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتنبأ الرتبة بتغير القرار أو كانت غير مستقرة تحت إعادة الصياغة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

3. نظرية الانضغاط العكوس للدمج وانتقال السياقات

الأطروحة المركزية: يمكن ضغط الدمج وانتقال السياقات مع الحفاظ على إمكان استعادة الأدلة والاستثناءات وشروط الصلاحية.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُدمج السياقات دون فقد منطقها، وكيف تنتقل المنظومة من Prototype إلى Runtime متعدد النماذج قابل للنشر والتدقيق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يُخزّن التمثيل المضغوط روابط المصدر والإصدار ويخضع لاختبار إعادة البناء. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يُتوقع خفض السياق والكلفة دون هبوط جوهري في الدقة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا غيّر الضغط الحكم أو أخفى استثناءات حاسمة. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

4. نظرية فجوة الدمج وانتقال السياقات

الأطروحة المركزية: الابتكار يبدأ عندما تمثل المنظومة ما ينقصها في الدمج وانتقال السياقات بوصفه فجوة محددة قابلة للعمل، لا غموضًا عامًا.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُدمج السياقات دون فقد منطقها، وكيف تنتقل المنظومة من Prototype إلى Runtime متعدد النماذج قابل للنشر والتدقيق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: تُصنف الفجوة إلى نقص بيانات أو مفهوم أو مقياس أو أداة أو صلاحية، ثم تتحول إلى سؤال وتجربة. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع ارتفاع جودة خطط البحث وتقليل التخمين. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تحسن الفجوات اختيار السؤال أو الأداة أو المنهج. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. نظرية انتقال الدمج وانتقال السياقات بين السياقات

الأطروحة المركزية: تظل قيمة الدمج وانتقال السياقات محفوظة عند الانتقال فقط إذا نُقلت معها المقاصد والقيود والثقة وشروط التراجع.

المشكلة التي تعالجها: تنشأ الحاجة إلى هذه النظرية عندما تفشل المعالجة الخطية في تفسير كيف تُدمج السياقات دون فقد منطقها، وكيف تنتقل المنظومة من Prototype إلى Runtime متعدد النماذج قابل للنشر والتدقيق؟ وهي تفترض أن الخلل ليس دائمًا نقص بيانات، بل قد يكون نقصًا في شكل التمثيل أو ترتيب الأولويات أو حدود الانتقال بين المراحل.

الآلية المقترحة: يستخدم النظام عقود انتقال typed ويقارن الحالة قبل وبعد handoff. وتُنفَّذ الآلية في ثلاث طبقات: طبقة رصد تسجل الحالة الخام، وطبقة تفسير تربط التغير بالسبب المحتمل، وطبقة قرار لا تسمح بترقية الفرضية إلى قاعدة إلا بعد اختبار نقل مستقل.

التنبؤات القابلة للاختبار: يتوقع انخفاض الانحراف وفقد المقصد في المهام الطويلة. كما تتنبأ النظرية بأن أثرها سيظهر أولًا في جودة اختيار المسار، ثم في خفض الكلفة، وأخيرًا في زيادة القدرة على تفسير الفشل بدل إخفائه خلف إجابة لغوية متماسكة.

شروط التفنيد والتقييد: تُهدم إذا لم تتفوق العقود على النقل النصي البسيط. ويُعد الفشل في التكرار أو التدهور خارج المجال أو عدم القدرة على فصل أثر النظرية عن أثر النموذج الأساسي أسبابًا كافية لتقييدها أو إعادة تعريفها.

الترجمة الهندسية: تُحوّل النظرية إلى Module له Schema إدخال وإخراج، ونسخة محددة، وسجل قرار، ومجموعة Unit Tests، وتجارب Ablation تقارن: النموذج الأساسي، والنموذج مع الوحدة، والنموذج مع بديل أبسط. لا تُقبل تسمية الوحدة باسم النظرية إذا لم يوجد فرق قابل للقياس.

برنامج التجربة الأول: تُنشأ 400 حالة موزعة بين أمثلة واضحة، وأمثلة غامضة، وحالات مضادة، وانتقال إلى مجال جديد. يُقاس الأداء قبل وبعد الوحدة، وتُحلل الأخطاء حسب منشئها: تمثيل، سؤال، حكم، استدلال، أداة، ذاكرة، أو صلاحية. ثم تُعاد التجربة على نموذج صغير 1B لإثبات أن الأثر بنيوي وليس مجرد ثمرة لسعة نموذج كبير.

بطاقة ادعاء أولية
الحالة: نظرية مقترحة. الدليل المطلوب: تفوق قابل للتكرار + نقل بين مجالين + Ablation يثبت الإضافة. الممنوع: وصفها كقانون مثبت قبل التجربة. المخرج العلمي: Paper مستقل، Dataset، Benchmark، وImplementation مرجعي.

5. الأدوات المنهجية والهندسية

1. أداة مرصد الدمج وانتقال السياقات

المهمة: يستخرج العناصر الحاسمة ويرتبها بحسب أثرها. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

2. أداة مصفاة الدمج وانتقال السياقات

المهمة: تكشف العناصر الوهمية أو المتكررة أو غير ذات الصلة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

3. أداة مقياس عائد الدمج وانتقال السياقات

المهمة: يقيس خفض عدم اليقين أو تحسن القرار لكل كلفة. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

4. أداة مولد الدمج وانتقال السياقات

المهمة: يقترح بدائل جديدة عندما يفشل المخطط الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

5. أداة مركب الدمج وانتقال السياقات

المهمة: يدمج العناصر في كائن منظم قابل للتتبع. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. أداة كاشف فجوات الدمج وانتقال السياقات

المهمة: يحدد ما لا يغطيه التمثيل الحالي. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

7. أداة مختبر ثبات الدمج وانتقال السياقات

المهمة: يختبر الاستقرار عبر إعادة الصياغة وتغير المجال. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

8. أداة محكمة الدمج وانتقال السياقات

المهمة: تولد اعتراضات وحالات مضادة وتصدر حكمًا مشروطًا. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

9. أداة سجل تحولات الدمج وانتقال السياقات

المهمة: يتتبع الإصدارات والأسباب والآثار. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

10. أداة بوابة صلاحية الدمج وانتقال السياقات

المهمة: تتحقق من الشروط والإذن قبل التفعيل أو التنفيذ. وتعمل الأداة بوصفها وسيطًا قابلًا للفحص بين النظرية والتنفيذ، لا بوصفها Prompt مغلقًا لا يمكن تفسير قراراته.

المدخلات: مخرجات السياقات وعقود الواجهات ومؤشرات النضج والموارد، مع Metadata للمصدر والوقت والثقة والصلاحية والسياق الحاكم. ترفض الأداة أي مدخل لا يمكن ربطه بمصدر أو يخلط بين Observation وInference.

سير العمل: تطبع الأداة الحالة، تتحقق من شروط التشغيل، تنشئ مرشحين، تقيسهم، تمرر المرشح الأقوى إلى محكمة صغيرة، ثم تخرج قرارًا مشروطًا مع البدائل المرفوضة وأسباب الرفض. كل خطوة تحدث Event في Audit Log.

المخرجات: Unified Runtime وTransition Graph وRelease Gates وResearch Portfolio، بالإضافة إلى Confidence، وUnresolved Gaps، وRollback Token، ونسخة الأداة. يجب أن تكون المخرجات Machine-readable وأن تسمح بإعادة تشغيل القرار.

السلامة والاختبار: تُختبر الأداة ضد مدخلات ناقصة ومتعارضة وخبيثة، وضد تغيّر الصياغة وتبدل المجال. لا تُنشر قبل اختبارات Regression وMutation وAdversarial Context، ويكون لها وضع Read-only قبل منح صلاحيات تنفيذية.

6. المبادئ الحاكمة

1. مبدأ أسبقية الدمج وانتقال السياقات

لا ينتقل النظام إلى المرحلة التالية قبل تمثيل الدمج وانتقال السياقات بما يكفي. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

2. مبدأ حاكمية الدمج وانتقال السياقات

تُرتب عناصر الدمج وانتقال السياقات بحسب المقصد لا بحسب الظهور أو التشابه. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

3. مبدأ عكوسية الدمج وانتقال السياقات

كل تحديث مهم في الدمج وانتقال السياقات له إصدار ومسار تراجع. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

4. مبدأ حدود الدمج وانتقال السياقات

يصرح النظام بما لا تغطيه الدمج وانتقال السياقات وبدرجة الثقة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

5. مبدأ اقتصاد الدمج وانتقال السياقات

تستخدم أقل موارد تحقق العائد المطلوب دون خفض السلامة. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

6. مبدأ نقل الدمج وانتقال السياقات

لا تعد القدرة عامة قبل اختبارها في مجال جديد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. مبدأ تفنيد الدمج وانتقال السياقات

كل ادعاء عن الدمج وانتقال السياقات يرافقه شرط هدم أو تقييد. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

8. مبدأ تتبع الدمج وانتقال السياقات

كل مخرج يعود إلى مصادره وقواعده ونسخته. ويترجم هذا المبدأ إلى قيد في Runtime لا إلى توصية بلاغية: شرط دخول، شرط خروج، Event تدقيق، وحالة فشل مسماة.

اختبار الالتزام: تُبنى حالة يحقق فيها النظام الهدف الظاهر مع خرق المبدأ، وحالة أخرى تحافظ على المبدأ مع كلفة أكبر قليلًا. ينجح التصميم عندما يختار الثانية في المهام عالية المخاطرة، ويبرر سبب التضحية المحلية.

7. برامج التدريب الخاصة

1. برنامج تدريب: انتقال محكوم

القدرة المستهدفة: تحويل «انتقال محكوم» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

2. برنامج تدريب: تماسك غير متماثل

القدرة المستهدفة: تحويل «تماسك غير متماثل» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

3. برنامج تدريب: بوابات النضج

القدرة المستهدفة: تحويل «بوابات النضج» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

4. برنامج تدريب: التكامل المرحلي

القدرة المستهدفة: تحويل «التكامل المرحلي» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده، المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

5. برنامج تدريب: الخطوط الثلاثة

القدرة المستهدفة: تحويل «الخطوط الثلاثة» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

6. برنامج تدريب: النشر متعدد البيئات

القدرة المستهدفة: تحويل «النشر متعدد البيئات» من مفهوم وصفي إلى أداء يمكن قياسه في أمثلة مألوفة وجديدة. يُمنع تقييم البرنامج بالأسلوب اللغوي وحده؛ المعيار هو التغير في القرار أو الاسترجاع أو الفعل أو معايرة الثقة.

بنية البيانات: 25% أمثلة تأسيسية، 20% أزواج تفاضلية لا يختلف فيها إلا العامل المراد تعلمه، 20% أمثلة سلبية، 15% حالات مضادة، 10% أمثلة Tool-use، و10% نقل إلى مجال بعيد. لكل عينة تفسير مرجعي وحدود قبول متعددة بدل إجابة وحيدة متصلبة.

دورة التدريب: Baseline → تشخيص خطأ → توليد Curriculum → تدريب محدود → اختبار عدائي → Ablation → نقل → تحديث Failure Atlas. يُوقف التدريب إذا تحسن Benchmark الداخلي وتدهور النقل أو ازدادت الثقة غير المعايرة.

مقاييس النجاح: دقة القدرة، ثباتها تحت إعادة الصياغة، نقلها بين مجالين، كلفة الحوسبة، قابلية تفسير القرار، وعدم ارتفاع الأخطاء في القدرات السابقة. ويُحتفظ بمجموعة Holdout لا تُستخدم في توليد البيانات الاصطناعية.

8. منظومة القياس

#	المقياس	طريقة القياس	قاعدة القبول
1	Transition Fidelity	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
2	Dominant Context Accuracy	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
3	Cross-Context Coherence	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
4	Gate Predictiveness	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
5	Integration Attribution	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
6	Triple-Track Balance	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
7	Policy Consistency	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل
8	Unified Success	تعريف تشغيلي قبل التجربة، خط أساس، Interval ثقة، وتحليل أخطاء حسب المجال	لا يقبل التحسن إذا رافقه هبوط في السلامة أو النقل

9. خريطة الفشل ومحكمة الهدم

1. تشغيل كل السياقات: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

2. هيمنة سياق محلي: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

3. انتقال بنص حر: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

4. دمج مبكر: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

5. بوابات شكلية: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

6. هندسة بلا علم وسلامة: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

7. نشر هش: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

8. متوسط يخفي فشلًا: يُنشأ اختبار يضغط النظام باتجاه هذا الفشل، ثم يُسجل موضع منشئه، وإشارة الإنذار المبكر، وإجراء الاحتواء، وما إذا كان يتطلب إصلاح بيانات أو مقياس أو معمارية أو صلاحية.

سؤال المحكمة الأعلى
ما الدليل الذي لو ظهر سيجعلنا نسحب ادعاء هذا السياق أو نقيده بمجال أضيق؟ لا يُعد غياب الإجابة قوة، بل علامة على أن النظرية غير قابلة للتفنيد.

10. درجات النضج

الدرجة	الدليل المطلوب	حد الاستخدام
S0 — مفهوم	تعريف ومثال مضاد وحدود أولية	لا يُدمج
S1 — Prototype	تنفيذ صغير وبيانات داخلية	Sandbox فقط
S2 — قدرة مقاسة	Benchmark وAblation ونقل أولي	استعمال مساعد
S3 — مكون موثوق	اختبارات عدائية وRollback ومراقبة	دمج محدود
S4 — مكون قابل للنشر	تكرار مستقل وسياسات تشغيل	Production منخفض المخاطر
S5 — مكون سيادي	ثبات عابر للنماذج وتدقيق خارجي	دمج في Unified Runtime

11. أسئلة البحث المفتوحة

- كيف نمثل انتقال محكوم تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل انتقال محكوم إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في انتقال محكوم، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل تماسك غير متماثل تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل تماسك غير متماثل إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في تماسك غير متماثل، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل بوابات النضج تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل بوابات النضج إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في بوابات النضج، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل التكامل المرحلي تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل التكامل المرحلي إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في التكامل المرحلي، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل الخطوط الثلاثة تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل الخطوط الثلاثة إلى مجال بعيد دون تسرب بيانات؟

- ما تكلفة الخطأ في الخطوط الثلاثة، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟
- كيف نمثل النشر متعدد البيئات تمثيلًا يميز بين الأداء الحقيقي والمحاكاة اللغوية؟
- ما أقل Benchmark يثبت نقل النشر متعدد البيئات إلى مجال بعيد دون تسرب بيانات؟
- ما تكلفة الخطأ في النشر متعدد البيئات، وما المحكمة التي تلتقطه قبل أن ينتقل إلى السياق التالي؟

12. عقد الربط مع بقية الخارطة

لا يرسل هذا السياق نصًا حرًا إلى السياق التالي، بل كائنًا typed يحتوي: المخرج، المصدر، درجة الثقة، السياق الحاكم، القيود، البدائل، الأسئلة المفتوحة، النسخة، وRollback pointer. يتلقى بدوره Feedback يوضح أي جزء من مخرجه غيّر القرار النهائي، وبذلك يصبح التعلم قائمًا على أثر قابل للعزو.

القسم الثالث: الطبقة العابرة للسياقات

1. النظريات الموحدة الجديدة

1. نظرية السيادة التفاضلية الموحدة

الأطروحة: يحدد الفرق الأعلى أثرًا السؤال الحاكم، ويحدد السؤال ترتيب الاعتبارات، ثم يحدد الحكم مسار الاستدلال والفعل.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: Ablation لعلاقات الربط بين المراحل. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

2. نظرية التماسك السياقي غير المتماثل

الأطروحة: الوحدة لا تعني مساواة السياقات، بل قيادة مؤقتة لسياق ودعم أو نقض من غيره.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: مقارنة قيادة ديناميكية بدمج متساو. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

3. نظرية المسافة الحاكمية المركبة

الأطروحة: الفجوة بين الإدراك والفعل تتكون من مسافات تمثيل وسؤال وحكم واستدلال وصلاحية وأثر.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: إصلاح أكبر عنق زجاجة وقياس التحسن. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

4. نظرية الذاكرة كوسط حاكم

الأطروحة: ما تسترجعه الذاكرة يحدد ما يصبح متاحًا للحكم، لذا يحتاج الاسترجاع إلى سياسة لا تشابه فقط.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: اختبارات ذاكرة عدائية. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

5. نظرية التعلم من دلتا الأثر

الأطروحة: الخبرة الأهم هي الفرق السببي بين الحالة قبل الفعل وبعده، لا النص الناتج وحده.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: قياس النقل بعد World-Effect Feedback. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

6. نظرية الاقتصاد الحاكمي للحوسبة

الأطروحة: قيمة الحوسبة تقاس بخفض مسافة الحكم والمخاطرة لا بطول التفكير.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: مقارنة سياسات Routing. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

7. نظرية العكوسية المعرفية

الأطروحة: المعرفة والأحكام والمهارات والأنطولوجيات يجب أن تكون قابلة للسحب وإعادة البناء.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: سحب مصدر ومراقبة المشتقات. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

8. نظرية الفجوة المولدة للعلم

الأطروحة: تراكم فجوات متجانسة يمكن أن يولد علمًا أو أداة أو Benchmark جديدة.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: تقييم نفع المقترحات الناتجة. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

9. نظرية الكثافة العامة المحدودة

الأطروحة: الذكاء العملي هو أعلى تغطية ونقل لكل موارد محدودة مع بقاء التفسير والسلامة.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: تحليل Pareto. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

10. نظرية التفنيد المولد للمهارة

الأطروحة: الاعتراض المتكرر يتحول إلى HyperSkill فحص وقائية.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: انخفاض تكرار الفشل. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

11. نظرية الحاكمية السابقة للوعي الوظيفي

الأطروحة: النموذج الذاتي يفيد فقط إذا رتبت الحاكمية ما يدخل الانتباه.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: مهام تشتيت طويلة. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

12. نظرية الائتلاف الوكيلي المؤقت

الأطروحة: تتكون فرق الوكلاء بحسب الفرق المطلوب إضافته ثم تحل.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: كلفة التنسيق مقابل العائد. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

13. نظرية الإصدار الكلي للحالة

الأطروحة: المهمة الطويلة تحتاج Checkpoint موحدًا للذاكرة والحكم والخطة والصلاحيات.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: اختبار استعادة كامل. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

14. نظرية المقصد كمتغير حفظ

الأطروحة: ينبغي نقل المقصد صراحة أثناء Handoff وإلا بقيت البيانات وضاع المعنى.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: اختبارات Goal Drift. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

15. نظرية التحقق الثلاثي للقدرة

الأطروحة: القدرة تحتاج نتيجة داخلية ونقلًا وتكرارًا مستقلًا.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: Release Gate. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

16. نظرية النمو من 1B كبراهين بنيوية

الأطروحة: مكسب النواة فوق 1B دليل أقوى على التنظيم من نجاحها فوق نموذج ضخم وحده.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: Scaling وAblation. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

17. نظرية التعدد المقيد للمصادر

الأطروحة: زيادة الأدوات والمصادر لها حد أمثل قبل أن تتجاوز كلفة التعارض فائدتها.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: منحنيات Marginal Source Utility. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

18. نظرية المحكمة كجهاز استشعار

الأطروحة: المحاكم تصنف الفشل وتوجه التدريب والذاكرة والمهارات، لا تمنع الإطلاق فقط.

الشرح التحليلي: تعمل هذه النظرية على مستوى العلاقات بين السياقات، ولذلك لا يمكن اختبارها داخل وحدة واحدة. يجب تصميم مهام طويلة يتغير فيها السياق الحاكم، وتتبادل الوحدات حالة منظمة، وتظهر نتيجة يمكن عزوها إلى التنسيق لا إلى قوة نموذج منفرد. ويجب تسجيل حالات نجحت فيها الوحدات منفردة وفشل التكامل، والعكس.

الترجمة الهندسية: تُنشأ Interface أو Monitor أو Policy مرتبطة بالنظرية، ويُقاس أثرها على Transition Fidelity وGoal Drift وTotal Cost وRecovery. كل نظرية موحدة تحتاج إلى Minimal implementation ينافس Baseline بسيطًا قبل دمجها في Kernel.

اختبار التفنيد المقترح: قياس التحسن بعد تغذية تقارير المحاكم. إذا لم يظهر فرق مستقر، تُخفض النظرية إلى Heuristic محلية أو تُسحب. وإذا ظهر الأثر في مجال واحد فقط، يُقيد نطاقها بدل وصفها بالعامة.

2. الأدوات الموحدة الثلاثون

1. منسق السيادة التفاضلية

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

2. مقياس المسافة الحاكمية المركبة

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

3. مجمّع دلتا الأثر

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

4. سجل الإصدار الكلي

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

5. حارس المقصد العابر للنماذج

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

6. محسن الائتلاف الوكيلي

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

7. موجه المصدر الأمثل

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

8. مترجم المحكمة إلى منهج

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

9. مترجم المحكمة إلى Skill

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

10. مراقب العكوسية المعرفية

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

11. مولد العلوم من الفجوات

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

12. مقياس الكثافة العامة المحدودة

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

13. مختبر التحقق الثلاثي

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

14. محرك انتقال 1B إلى 9B

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

15. مراقب التماسك غير المتماثل

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

16. مدير حق النقض السيادي

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

17. مفسر الوحدة القابلة للتفكيك

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

18. مراقب التضخم الاصطلاحي

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

19. مولد بطاقات النظرية

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

20. سجل العلوم المقترحة

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

21. مخطط محفظة البحث

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

22. مولد Benchmark من النظرية

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

23. مقياس قيمة الاستقلال

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

24. كاشف الاعتماد الدائري

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

25. مراقب Goodhart الحاكمي

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

26. مخطط البحث المضاد

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

27. محاكي المخاطر المتسلسلة

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

28. مولد التدهور الآمن

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

29. محرك التوافق عبر البيئات

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

30. مدقق اللغة البحثية

الوظيفة: أداة ربط أو مراقبة أو تحويل تعمل فوق مخرجات سياقين أو أكثر. مدخلها Unified State Object، ومخرجها قرار أو تقرير أو عقد انتقال، ويجب أن تسجل الوحدات التي استدعتها والفرق الذي أضافته.

متطلبات البناء: Schema ثابت، Explain endpoint، Versioning، وضع محاكاة، واختبارات تعطيل. يُرفض نشر الأداة إذا كان أثرها لا ينفصل عن إعادة صياغة Prompt، أو إذا زادت التعقيد دون تحسن في النقل أو السلامة أو الكلفة.

3. المبادئ الموحدة الأربعة والعشرون

1. مبدأ الفرق قبل القياس

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

2. مبدأ السؤال قبل الحوسبة

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

3. مبدأ الحكم قبل الفعل

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

4. مبدأ الأثر بعد الفعل

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

5. مبدأ التفنيد قبل التثبيت

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

6. مبدأ الإصدار قبل التعديل

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

7. مبدأ العكس قبل المخاطرة

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

8. مبدأ النقل قبل الادعاء العام

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

9. مبدأ الاستقلال قبل الدمج

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

10. مبدأ التفسير قبل الثقة

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

11. مبدأ المقصد قبل المؤشر

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

12. مبدأ الأقل كفاية قبل الأكبر

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

13. مبدأ الخصوصية قبل الاسترجاع

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

14. مبدأ المصدر قبل الذاكرة

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

15. مبدأ الفشل قبل التوسع

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

16. مبدأ المحكمة كتعلم

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

17. مبدأ الحجم بعد البرهان البنيوي

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

18. مبدأ الاسم بعد التعريف

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

19. مبدأ الوحدة مع قابلية التفكيك

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

20. مبدأ السلامة في النواة

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

21. مبدأ السياق الحاكم متغير

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

22. مبدأ الذاكرة قابلة للنسيان

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

23. مبدأ الحلقة ذات دلتا

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

24. مبدأ العلم والهندسة والحاكمية معًا

المعنى التنفيذي: يتحول هذا المبدأ إلى Policy قابلة للفحص في Release Gate. لكل مبدأ Event يدل على الالتزام، Event يدل على الخرق، وحالة اختبار مصممة لإغراء النظام بتحقيق هدف محلي عبر تجاوز القاعدة.

4. العلوم المقترحة بوصفها برنامجًا بينيًا

#	المجال العلمي المقترح	شروط التأسيس الفعلي
1	علم هندسة الفروق الحاكمة - Sovereign Difference Engineering	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
2	علم الديناميات السؤالية الحاكمة - Sovereign Question Dynamics	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل

#	المجال العلمي المقترح	شروط التأسيس الفعلي
3	علم الحاكمية المعرفية الحسابية - Computational Epistemic Sovereignty	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
4	علم الأوعية الاستدلالية - Vascular Reasoning Science	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
5	علم الخرائط الذاتية التفاضلية - Differential Self-Mapping Science	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
6	علم النوى المعرفية المتراكبة - Layered Cognitive Kernel Engineering	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
7	علم المناهج التكوينية للذكاء العام - General Intelligence Curriculum Science	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
8	علم تشغيل المعرفة الذاكرية - Memory Knowledge Operating Science	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
9	علم البرمجة المهارية فائقة التركيب - HyperSkill Systems Science	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
10	علم الوكالة الأثرية المحكومة - Governed World-Effect Agency	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
11	علم التفنيد الحاكمي العكوس - Reversible Sovereign Falsification	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل
12	علم توحيد السياقات السيادية - Unified Sovereign Context Science	تعريف المتغيرات + Taxonomy فشل + Benchmark + Runtime + تكرار مستقل

لا يُراد بهذه العلوم إنشاء حدود أكاديمية مصطنعة، بل تنظيم محفظة بحث تتقاطع مع Machine Learning وCognitive Science وControl Theory وFormal Methods وHuman-Computer Interaction وInformation Retrieval وSoftware Engineering وSafety. يُحتفظ بالاسم فقط إذا أنتج أسئلة ومقاييس وتجارب لا يغطيها تجميع مباشر للتخصصات القائمة.

القسم الرابع: المحفظة التدريبية والاختبارية

1. مبدأ البرنامج لا Dataset فقط

كل برنامج تدريبي هو عقد قدرة: يحدد المتطلبات السابقة، ومصادر البيانات، وشكل الأمثلة السلبية، ومقاييس النقل، ومحكمة الفشل، وشروط التوقف. يمكن أن يستخدم أكثر من Dataset، ويمكن أن يعاد تشغيله على 1B و9B ونماذج خارجية لقياس Cross-Model AGI Density.

2. مائة برنامج Pretraining

المعرف	اسم البرنامج	الهدف	دليل النجاح
PT-001	PT-001 — تأسيس الفرق الحاسم في الفروق والتمثيل الأنطولوجي	تعلم تمثيل الفرق الحاسم من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-002	PT-002 — تأسيس الفرق الوهمي في الفروق والتمثيل الأنطولوجي	تعلم تمثيل الفرق الوهمي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-003	PT-003 — تأسيس توليد الأبعاد في الفروق والتمثيل الأنطولوجي	تعلم تمثيل توليد الأبعاد من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-004	PT-004 — تأسيس الضغط الأنطولوجي في الفروق والتمثيل الأنطولوجي	تعلم تمثيل الضغط الأنطولوجي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-005	PT-005 — تأسيس ثبات الهوية في الفروق والتمثيل الأنطولوجي	تعلم تمثيل ثبات الهوية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-006	PT-006 — تأسيس السلالم التجريدية في الفروق والتمثيل الأنطولوجي	تعلم تمثيل السلالم التجريدية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-007	PT-007 — تأسيس الفرق الحاسم في الفروق والتمثيل الأنطولوجي	تعلم تمثيل الفرق الحاسم من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-008	PT-008 — تأسيس الفرق الوهمي في الفروق والتمثيل الأنطولوجي	تعلم تمثيل الفرق الوهمي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-009	PT-009 — تأسيس توليد الأبعاد في الفروق والتمثيل الأنطولوجي	تعلم تمثيل توليد الأبعاد من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-010	PT-010 — تأسيس الضغط الأنطولوجي في الفروق والتمثيل الأنطولوجي	تعلم تمثيل الضغط الأنطولوجي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-011	PT-011 — تأسيس السؤال السابق في السؤال والكثافة السؤالية	تعلم تمثيل السؤال السابق من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-012	PT-012 — تأسيس الجاذبية السؤالية في السؤال والكثافة السؤالية	تعلم تمثيل الجاذبية السؤالية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-013	PT-013 — تأسيس الكثافة الصافية في السؤال والكثافة السؤالية	تعلم تمثيل الكثافة الصافية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-014	PT-014 — تأسيس كشف الافتراض في السؤال والكثافة السؤالية	تعلم تمثيل كشف الافتراض من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-015	PT-015 — تأسيس السؤال إلى فعل في السؤال والكثافة السؤالية	تعلم تمثيل السؤال إلى فعل من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-016	PT-016 — تأسيس نقطة الإشباع في السؤال والكثافة السؤالية	تعلم تمثيل نقطة الإشباع من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-017	PT-017 — تأسيس السؤال السابق في السؤال والكثافة السؤالية	تعلم تمثيل السؤال السابق من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-018	PT-018 — تأسيس الجاذبية السؤالية في السؤال والكثافة السؤالية	تعلم تمثيل الجاذبية السؤالية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.

المعرف	اسم البرنامج	الهدف	دليل النجاح
PT-019	PT-019 — تأسيس الكثافة الصافية في السؤال والكثافة السؤالية	تعلم تمثيل الكثافة الصافية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-020	PT-020 — تأسيس كشف الافتراض في السؤال والكثافة السؤالية	تعلم تمثيل كشف الافتراض من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-021	PT-021 — تأسيس الاعتبارية العليا في الحاكمية والحكم الأشمل	تعلم تمثيل الاعتبارية العليا من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-022	PT-022 — تأسيس الحكم المشروط في الحاكمية والحكم الأشمل	تعلم تمثيل الحكم المشروط من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-023	PT-023 — تأسيس مسافة المعرفة والحكم في الحاكمية والحكم الأشمل	تعلم تمثيل مسافة المعرفة والحكم من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-024	PT-024 — تأسيس طبقات الحاكمية في الحاكمية والحكم الأشمل	تعلم تمثيل طبقات الحاكمية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-025	PT-025 — تأسيس عدم تماثل الخطأ في الحاكمية والحكم الأشمل	تعلم تمثيل عدم تماثل الخطأ من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-026	PT-026 — تأسيس بوابة الصلاحية في الحاكمية والحكم الأشمل	تعلم تمثيل بوابة الصلاحية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-027	PT-027 — تأسيس الاعتبارية العليا في الحاكمية والحكم الأشمل	تعلم تمثيل الاعتبارية العليا من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-028	PT-028 — تأسيس الحكم المشروط في الحاكمية والحكم الأشمل	تعلم تمثيل الحكم المشروط من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-029	PT-029 — تأسيس مسافة المعرفة والحكم في الحاكمية والحكم الأشمل	تعلم تمثيل مسافة المعرفة والحكم من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-030	PT-030 — تأسيس طبقات الحاكمية في الحاكمية والحكم الأشمل	تعلم تمثيل طبقات الحاكمية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-031	PT-031 — تأسيس التوجيه الوعائي في الاستدلال والتفكير المركب	تعلم تمثيل التوجيه الوعائي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-032	PT-032 — تأسيس تعدد الأنماط في الاستدلال والتفكير المركب	تعلم تمثيل تعدد الأنماط من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-033	PT-033 — تأسيس المسارات المستقلة في الاستدلال والتفكير المركب	تعلم تمثيل المسارات المستقلة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-034	PT-034 — تأسيس الحلقة المحدودة في الاستدلال والتفكير المركب	تعلم تمثيل الحلقة المحدودة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-035	PT-035 — تأسيس التفنيد في الاستدلال والتفكير المركب	تعلم تمثيل التفنيد من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.

تابع الجدول

المعرف	اسم البرنامج	الهدف	دليل النجاح
PT-036	PT-036 — تأسيس التبديل البنيوي في الاستدلال والتفكير المركب	تعلم تمثيل التبديل البنيوي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-037	PT-037 — تأسيس التوجيه الوعائي في الاستدلال والتفكير المركب	تعلم تمثيل التوجيه الوعائي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-038	PT-038 — تأسيس تعدد الأنماط في الاستدلال والتفكير المركب	تعلم تمثيل تعدد الأنماط من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.

المعرف	اسم البرنامج	الهدف	دليل النجاح
PT-039	PT-039 — تأسيس المسارات المستقلة في الاستدلال والتفكير المركب	تعلم تمثيل المسارات المستقلة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-040	PT-040 — تأسيس الحلقة المحدودة في الاستدلال والتفكير المركب	تعلم تمثيل الحلقة المحدودة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-041	PT-041 — تأسيس خريطة الفروق النشطة في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل خريطة الفروق النشطة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-042	PT-042 — تأسيس الانتباه الحاكمي في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل الانتباه الحاكمي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-043	PT-043 — تأسيس حدود القدرة في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل حدود القدرة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-044	PT-044 — تأسيس معايرة الثقة في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل معايرة الثقة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-045	PT-045 — تأسيس كشف الانحراف في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل كشف الانحراف من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-046	PT-046 — تأسيس إعادة البناء الانتقائية في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل إعادة البناء الانتقائية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-047	PT-047 — تأسيس خريطة الفروق النشطة في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل خريطة الفروق النشطة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-048	PT-048 — تأسيس الانتباه الحاكمي في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل الانتباه الحاكمي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-049	PT-049 — تأسيس حدود القدرة في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل حدود القدرة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-050	PT-050 — تأسيس معايرة الثقة في الوعي الوظيفي والنموذج الذاتي	تعلم تمثيل معايرة الثقة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-051	PT-051 — تأسيس فصل القدرة عن الحامل في المعمارية والنواة والنماذج	تعلم تمثيل فصل القدرة عن الحامل من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-052	PT-052 — تأسيس الكثافة العابرة في المعمارية والنواة والنماذج	تعلم تمثيل الكثافة العابرة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-053	PT-053 — تأسيس عقود الواجهة في المعمارية والنواة والنماذج	تعلم تمثيل عقود الواجهة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-054	PT-054 — تأسيس الاستبدال في المعمارية والنواة والنماذج	تعلم تمثيل الاستبدال من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-055	PT-055 — تأسيس التدهور الآمن في المعمارية والنواة والنماذج	تعلم تمثيل التدهور الآمن من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-056	PT-056 — تأسيس النضج الحجمي الصادق في المعمارية والنواة والنماذج	تعلم تمثيل النضج الحجمي الصادق من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-057	PT-057 — تأسيس فصل القدرة عن الحامل في المعمارية والنواة والنماذج	تعلم تمثيل فصل القدرة عن الحامل من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-058	PT-058 — تأسيس الكثافة العابرة في المعمارية والنواة والنماذج	تعلم تمثيل الكثافة العابرة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.

المعرف	اسم البرنامج	الهدف	دليل النجاح
PT-059	PT-059 — تأسيس عقود الواجهة في المعمارية والنواة والنماذج	تعلم تمثيل عقود الواجهة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-060	PT-060 — تأسيس الاستبدال في المعمارية والنواة والنماذج	تعلم تمثيل الاستبدال من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-061	PT-061 — تأسيس ترتيب المتطلبات في التدريب الموجّه نحو AGI	تعلم تمثيل ترتيب المتطلبات من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-062	PT-062 — تأسيس الفشل كبيانات في التدريب الموجّه نحو AGI	تعلم تمثيل الفشل كبيانات من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-063	PT-063 — تأسيس تعدد المحاكم في التدريب الموجّه نحو AGI	تعلم تمثيل تعدد المحاكم من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-064	PT-064 — تأسيس النقل البنيوي في التدريب الموجّه نحو AGI	تعلم تمثيل النقل البنيوي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-065	PT-065 — تأسيس اقتصاد التدريب في التدريب الموجّه نحو AGI	تعلم تمثيل اقتصاد التدريب من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-066	PT-066 — تأسيس فصل الأسلوب عن التفكير في التدريب الموجّه نحو AGI	تعلم تمثيل فصل الأسلوب عن التفكير من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-067	PT-067 — تأسيس ترتيب المتطلبات في التدريب الموجّه نحو AGI	تعلم تمثيل ترتيب المتطلبات من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-068	PT-068 — تأسيس الفشل كبيانات في التدريب الموجّه نحو AGI	تعلم تمثيل الفشل كبيانات من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-069	PT-069 — تأسيس تعدد المحاكم في التدريب الموجّه نحو AGI	تعلم تمثيل تعدد المحاكم من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-070	PT-070 — تأسيس النقل البنيوي في التدريب الموجّه نحو AGI	تعلم تمثيل النقل البنيوي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.

تابع الجدول

المعرف	اسم البرنامج	الهدف	دليل النجاح
PT-071	PT-071 — تأسيس الذاكرة الطبقية في الذاكرة وKB-as-OS	تعلم تمثيل الذاكرة الطبقية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-072	PT-072 — تأسيس الاسترجاع بالمقصد في الذاكرة وKB-as-OS	تعلم تمثيل الاسترجاع بالمقصد من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-073	PT-073 — تأسيس التصحيح المنتشر في الذاكرة وKB-as-OS	تعلم تمثيل التصحيح المنتشر من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-074	PT-074 — تأسيس النسيان الحاكمي في الذاكرة وKB-as-OS	تعلم تمثيل النسيان الحاكمي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-075	PT-075 — تأسيس ضغط الحالة في الذاكرة وKB-as-OS	تعلم تمثيل ضغط الحالة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-076	PT-076 — تأسيس مقاومة KB Injection في الذاكرة وKB-as-OS	تعلم تمثيل مقاومة KB Injection من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-077	PT-077 — تأسيس الذاكرة الطبقية في الذاكرة وKB-as-OS	تعلم تمثيل الذاكرة الطبقية من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-078	PT-078 — تأسيس الاسترجاع بالمقصد في الذاكرة وKB-as-OS	تعلم تمثيل الاسترجاع بالمقصد من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.

المعرف	اسم البرنامج	الهدف	دليل النجاح
PT-079	PT-079 — تأسيس التصحيح المنتشر في الذاكرة وKB-as-OS	تعلم تمثيل التصحيح المنتشر من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-080	PT-080 — تأسيس النسيان الحاكمي في الذاكرة وKB-as-OS	تعلم تمثيل النسيان الحاكمي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-081	PT-081 — تأسيس عقد المهارة في HyperSkills وBounded HyperLoops	تعلم تمثيل عقد المهارة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-082	PT-082 — تأسيس التركيب بالمقصد في HyperSkills وBounded HyperLoops	تعلم تمثيل التركيب بالمقصد من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-083	PT-083 — تأسيس دلتا الحالة في HyperSkills وBounded HyperLoops	تعلم تمثيل دلتا الحالة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-084	PT-084 — تأسيس حدود الحلقة في HyperSkills وBounded HyperLoops	تعلم تمثيل حدود الحلقة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-085	PT-085 — تأسيس Skill من الفشل في HyperSkills وBounded HyperLoops	تعلم تمثيل Skill من الفشل من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-086	PT-086 — تأسيس اقتصاد المهارات في HyperSkills وBounded HyperLoops	تعلم تمثيل اقتصاد المهارات من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-087	PT-087 — تأسيس عقد المهارة في HyperSkills وBounded HyperLoops	تعلم تمثيل عقد المهارة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-088	PT-088 — تأسيس التركيب بالمقصد في HyperSkills وBounded HyperLoops	تعلم تمثيل التركيب بالمقصد من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-089	PT-089 — تأسيس دلتا الحالة في HyperSkills وBounded HyperLoops	تعلم تمثيل دلتا الحالة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-090	PT-090 — تأسيس حدود الحلقة في HyperSkills وBounded HyperLoops	تعلم تمثيل حدود الحلقة من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-091	PT-091 — تأسيس أقل فعل كاف في الوكالة والأدوات والأثر العالمي	تعلم تمثيل أقل فعل كاف من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-092	PT-092 — تأسيس الحكم قبل الفعل في الوكالة والأدوات والأثر العالمي	تعلم تمثيل الحكم قبل الفعل من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-093	PT-093 — تأسيس الأثر الثنائي في الوكالة والأدوات والأثر العالمي	تعلم تمثيل الأثر الثنائي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-094	PT-094 — تأسيس الوكيل ذو الفرق في الوكالة والأدوات والأثر العالمي	تعلم تمثيل الوكيل ذو الفرق من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-095	PT-095 — تأسيس التراجع المسبق في الوكالة والأدوات والأثر العالمي	تعلم تمثيل التراجع المسبق من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-096	PT-096 — تأسيس أقل امتياز في الوكالة والأدوات والأثر العالمي	تعلم تمثيل أقل امتياز من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-097	PT-097 — تأسيس أقل فعل كاف في الوكالة والأدوات والأثر العالمي	تعلم تمثيل أقل فعل كاف من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-098	PT-098 — تأسيس الحكم قبل الفعل في الوكالة والأدوات والأثر العالمي	تعلم تمثيل الحكم قبل الفعل من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-099	PT-099 — تأسيس الأثر الثنائي في الوكالة والأدوات والأثر العالمي	تعلم تمثيل الأثر الثنائي من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.
PT-100	PT-100 — تأسيس الوكيل ذو الفرق في الوكالة والأدوات والأثر العالمي	تعلم تمثيل الوكيل ذو الفرق من أمثلة متعددة الأنماط مع أزواج تفاضلية وحالات مضادة.	Perplexity لا تكفي؛ يلزم Probe بنيوي + Transfer + Calibration.

3. مائة برنامج Post-training

المعرف	اسم البرنامج	الهدف	دليل النجاح
PO-001	PO-001 — ضبط الاعتبارية العليا تحت التعارض والأثر	تعليم النموذج تطبيق الاعتبارية العليا أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-002	PO-002 — ضبط الحكم المشروط تحت التعارض والأثر	تعليم النموذج تطبيق الحكم المشروط أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-003	PO-003 — ضبط مسافة المعرفة والحكم تحت التعارض والأثر	تعليم النموذج تطبيق مسافة المعرفة والحكم أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-004	PO-004 — ضبط طبقات الحاكمية تحت التعارض والأثر	تعليم النموذج تطبيق طبقات الحاكمية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-005	PO-005 — ضبط عدم تماثل الخطأ تحت التعارض والأثر	تعليم النموذج تطبيق عدم تماثل الخطأ أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-006	PO-006 — ضبط بوابة الصلاحية تحت التعارض والأثر	تعليم النموذج تطبيق بوابة الصلاحية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-007	PO-007 — ضبط الاعتبارية العليا تحت التعارض والأثر	تعليم النموذج تطبيق الاعتبارية العليا أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-008	PO-008 — ضبط الحكم المشروط تحت التعارض والأثر	تعليم النموذج تطبيق الحكم المشروط أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-009	PO-009 — ضبط مسافة المعرفة والحكم تحت التعارض والأثر	تعليم النموذج تطبيق مسافة المعرفة والحكم أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-010	PO-010 — ضبط طبقات الحاكمية تحت التعارض والأثر	تعليم النموذج تطبيق طبقات الحاكمية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-011	PO-011 — ضبط التوجيه الوعائي تحت التعارض والأثر	تعليم النموذج تطبيق التوجيه الوعائي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-012	PO-012 — ضبط تعدد الأنماط تحت التعارض والأثر	تعليم النموذج تطبيق تعدد الأنماط أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-013	PO-013 — ضبط المسارات المستقلة تحت التعارض والأثر	تعليم النموذج تطبيق المسارات المستقلة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-014	PO-014 — ضبط الحلقة المحدودة تحت التعارض والأثر	تعليم النموذج تطبيق الحلقة المحدودة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-015	PO-015 — ضبط التفنيد تحت التعارض والأثر	تعليم النموذج تطبيق التفنيد أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-016	PO-016 — ضبط التبديل البنيوي تحت التعارض والأثر	تعليم النموذج تطبيق التبديل البنيوي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-017	PO-017 — ضبط التوجيه الوعائي تحت التعارض والأثر	تعليم النموذج تطبيق التوجيه الوعائي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-018	PO-018 — ضبط تعدد الأنماط تحت التعارض والأثر	تعليم النموذج تطبيق تعدد الأنماط أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-019	PO-019 — ضبط المسارات المستقلة تحت التعارض والأثر	تعليم النموذج تطبيق المسارات المستقلة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-020	PO-020 — ضبط الحلقة المحدودة تحت التعارض والأثر	تعليم النموذج تطبيق الحلقة المحدودة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-021	PO-021 — ضبط خريطة الفروق النشطة تحت التعارض والأثر	تعليم النموذج تطبيق خريطة الفروق النشطة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.

المعرف	اسم البرنامج	الهدف	دليل النجاح
PO-022	PO-022 — ضبط الانتباه الحاكمي تحت التعارض والأثر	تعليم النموذج تطبيق الانتباه الحاكمي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-023	PO-023 — ضبط حدود القدرة تحت التعارض والأثر	تعليم النموذج تطبيق حدود القدرة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-024	PO-024 — ضبط معايرة الثقة تحت التعارض والأثر	تعليم النموذج تطبيق معايرة الثقة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-025	PO-025 — ضبط كشف الانحراف تحت التعارض والأثر	تعليم النموذج تطبيق كشف الانحراف أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-026	PO-026 — ضبط إعادة البناء الانتقائية تحت التعارض والأثر	تعليم النموذج تطبيق إعادة البناء الانتقائية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-027	PO-027 — ضبط خريطة الفروق النشطة تحت التعارض والأثر	تعليم النموذج تطبيق خريطة الفروق النشطة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-028	PO-028 — ضبط الانتباه الحاكمي تحت التعارض والأثر	تعليم النموذج تطبيق الانتباه الحاكمي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-029	PO-029 — ضبط حدود القدرة تحت التعارض والأثر	تعليم النموذج تطبيق حدود القدرة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-030	PO-030 — ضبط معايرة الثقة تحت التعارض والأثر	تعليم النموذج تطبيق معايرة الثقة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-031	PO-031 — ضبط فصل القدرة عن الحامل تحت التعارض والأثر	تعليم النموذج تطبيق فصل القدرة عن الحامل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-032	PO-032 — ضبط الكثافة العابرة تحت التعارض والأثر	تعليم النموذج تطبيق الكثافة العابرة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-033	PO-033 — ضبط عقود الواجهة تحت التعارض والأثر	تعليم النموذج تطبيق عقود الواجهة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-034	PO-034 — ضبط الاستبدال تحت التعارض والأثر	تعليم النموذج تطبيق الاستبدال أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-035	PO-035 — ضبط التدهور الآمن تحت التعارض والأثر	تعليم النموذج تطبيق التدهور الآمن أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.

تابع الجدول

المعرف	اسم البرنامج	الهدف	دليل النجاح
PO-036	PO-036 — ضبط النضج الحجمي الصادق تحت التعارض والأثر	تعليم النموذج تطبيق النضج الحجمي الصادق أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-037	PO-037 — ضبط فصل القدرة عن الحامل تحت التعارض والأثر	تعليم النموذج تطبيق فصل القدرة عن الحامل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-038	PO-038 — ضبط الكثافة العابرة تحت التعارض والأثر	تعليم النموذج تطبيق الكثافة العابرة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-039	PO-039 — ضبط عقود الواجهة تحت التعارض والأثر	تعليم النموذج تطبيق عقود الواجهة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-040	PO-040 — ضبط الاستبدال تحت التعارض والأثر	تعليم النموذج تطبيق الاستبدال أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-041	PO-041 — ضبط ترتيب المتطلبات تحت التعارض والأثر	تعليم النموذج تطبيق ترتيب المتطلبات أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.

المعرف	اسم البرنامج	الهدف	دليل النجاح
PO-042	PO-042 — ضبط الفشل كبيانات تحت التعارض والأثر	تعليم النموذج تطبيق الفشل كبيانات أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-043	PO-043 — ضبط تعدد المحاكم تحت التعارض والأثر	تعليم النموذج تطبيق تعدد المحاكم أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-044	PO-044 — ضبط النقل البنيوي تحت التعارض والأثر	تعليم النموذج تطبيق النقل البنيوي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-045	PO-045 — ضبط اقتصاد التدريب تحت التعارض والأثر	تعليم النموذج تطبيق اقتصاد التدريب أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-046	PO-046 — ضبط فصل الأسلوب عن التفكير تحت التعارض والأثر	تعليم النموذج تطبيق فصل الأسلوب عن التفكير أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-047	PO-047 — ضبط ترتيب المتطلبات تحت التعارض والأثر	تعليم النموذج تطبيق ترتيب المتطلبات أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-048	PO-048 — ضبط الفشل كبيانات تحت التعارض والأثر	تعليم النموذج تطبيق الفشل كبيانات أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-049	PO-049 — ضبط تعدد المحاكم تحت التعارض والأثر	تعليم النموذج تطبيق تعدد المحاكم أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-050	PO-050 — ضبط النقل البنيوي تحت التعارض والأثر	تعليم النموذج تطبيق النقل البنيوي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-051	PO-051 — ضبط الذاكرة الطبقية تحت التعارض والأثر	تعليم النموذج تطبيق الذاكرة الطبقية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-052	PO-052 — ضبط الاسترجاع بالمقصد تحت التعارض والأثر	تعليم النموذج تطبيق الاسترجاع بالمقصد أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-053	PO-053 — ضبط التصحيح المنتشر تحت التعارض والأثر	تعليم النموذج تطبيق التصحيح المنتشر أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-054	PO-054 — ضبط النسيان الحاكي تحت التعارض والأثر	تعليم النموذج تطبيق النسيان الحاكي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-055	PO-055 — ضبط ضغط الحالة تحت التعارض والأثر	تعليم النموذج تطبيق ضغط الحالة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-056	PO-056 — ضبط مقاومة KB Injection تحت التعارض والأثر	تعليم النموذج تطبيق مقاومة KB Injection أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-057	PO-057 — ضبط الذاكرة الطبقية تحت التعارض والأثر	تعليم النموذج تطبيق الذاكرة الطبقية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-058	PO-058 — ضبط الاسترجاع بالمقصد تحت التعارض والأثر	تعليم النموذج تطبيق الاسترجاع بالمقصد أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-059	PO-059 — ضبط التصحيح المنتشر تحت التعارض والأثر	تعليم النموذج تطبيق التصحيح المنتشر أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-060	PO-060 — ضبط النسيان الحاكي تحت التعارض والأثر	تعليم النموذج تطبيق النسيان الحاكي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-061	PO-061 — ضبط عقد المهارة تحت التعارض والأثر	تعليم النموذج تطبيق عقد المهارة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-062	PO-062 — ضبط التركيب بالمقصد تحت التعارض والأثر	تعليم النموذج تطبيق التركيب بالمقصد أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.

المعرف	اسم البرنامج	الهدف	دليل النجاح
PO-063	PO-063 — ضبط دلتا الحالة تحت التعارض والأثر	تعليم النموذج تطبيق دلتا الحالة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-064	PO-064 — ضبط حدود الحلقة تحت التعارض والأثر	تعليم النموذج تطبيق حدود الحلقة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-065	PO-065 — ضبط Skill من الفشل تحت التعارض والأثر	تعليم النموذج تطبيق Skill من الفشل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-066	PO-066 — ضبط اقتصاد المهارات تحت التعارض والأثر	تعليم النموذج تطبيق اقتصاد المهارات أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-067	PO-067 — ضبط عقد المهارة تحت التعارض والأثر	تعليم النموذج تطبيق عقد المهارة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-068	PO-068 — ضبط التركيب بالمقصد تحت التعارض والأثر	تعليم النموذج تطبيق التركيب بالمقصد أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-069	PO-069 — ضبط دلتا الحالة تحت التعارض والأثر	تعليم النموذج تطبيق دلتا الحالة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-070	PO-070 — ضبط حدود الحلقة تحت التعارض والأثر	تعليم النموذج تطبيق حدود الحلقة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.

تابع الجدول

المعرف	اسم البرنامج	الهدف	دليل النجاح
PO-071	PO-071 — ضبط أقل فعل كاف تحت التعارض والأثر	تعليم النموذج تطبيق أقل فعل كاف أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-072	PO-072 — ضبط الحكم قبل الفعل تحت التعارض والأثر	تعليم النموذج تطبيق الحكم قبل الفعل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-073	PO-073 — ضبط الأثر الثنائي تحت التعارض والأثر	تعليم النموذج تطبيق الأثر الثنائي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-074	PO-074 — ضبط الوكيل ذو الفرق تحت التعارض والأثر	تعليم النموذج تطبيق الوكيل ذو الفرق أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-075	PO-075 — ضبط التراجع المسبق تحت التعارض والأثر	تعليم النموذج تطبيق التراجع المسبق أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-076	PO-076 — ضبط أقل امتياز تحت التعارض والأثر	تعليم النموذج تطبيق أقل امتياز أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-077	PO-077 — ضبط أقل فعل كاف تحت التعارض والأثر	تعليم النموذج تطبيق أقل فعل كاف أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-078	PO-078 — ضبط الحكم قبل الفعل تحت التعارض والأثر	تعليم النموذج تطبيق الحكم قبل الفعل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-079	PO-079 — ضبط الأثر الثنائي تحت التعارض والأثر	تعليم النموذج تطبيق الأثر الثنائي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-080	PO-080 — ضبط الوكيل ذو الفرق تحت التعارض والأثر	تعليم النموذج تطبيق الوكيل ذو الفرق أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-081	PO-081 — ضبط المحاكم المتعامدة تحت التعارض والأثر	تعليم النموذج تطبيق المحاكم المتعامدة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-082	PO-082 — ضبط سلم الدليل تحت التعارض والأثر	تعليم النموذج تطبيق سلم الدليل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-083	PO-083 — ضبط العكوسية تحت التعارض والأثر	تعليم النموذج تطبيق العكوسية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.

المعرف	اسم البرنامج	الهدف	دليل النجاح
PO-084	PO-084 — ضبط النتائج السلبية تحت التعارض والأثر	تعليم النموذج تطبيق النتائج السلبية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-085	PO-085 — ضبط الاختبار العدائي تحت التعارض والأثر	تعليم النموذج تطبيق الاختبار العدائي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-086	PO-086 — ضبط التكرار المستقل تحت التعارض والأثر	تعليم النموذج تطبيق التكرار المستقل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-087	PO-087 — ضبط المحاكم المتعامدة تحت التعارض والأثر	تعليم النموذج تطبيق المحاكم المتعامدة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-088	PO-088 — ضبط سلم الدليل تحت التعارض والأثر	تعليم النموذج تطبيق سلم الدليل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-089	PO-089 — ضبط العكوسية تحت التعارض والأثر	تعليم النموذج تطبيق العكوسية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-090	PO-090 — ضبط النتائج السلبية تحت التعارض والأثر	تعليم النموذج تطبيق النتائج السلبية أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-091	PO-091 — ضبط انتقال محكوم تحت التعارض والأثر	تعليم النموذج تطبيق انتقال محكوم أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-092	PO-092 — ضبط تماسك غير متماثل تحت التعارض والأثر	تعليم النموذج تطبيق تماسك غير متماثل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-093	PO-093 — ضبط بوابات النضج تحت التعارض والأثر	تعليم النموذج تطبيق بوابات النضج أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-094	PO-094 — ضبط التكامل المرحلي تحت التعارض والأثر	تعليم النموذج تطبيق التكامل المرحلي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-095	PO-095 — ضبط الخطوط الثلاثة تحت التعارض والأثر	تعليم النموذج تطبيق الخطوط الثلاثة أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-096	PO-096 — ضبط النشر متعدد البيئات تحت التعارض والأثر	تعليم النموذج تطبيق النشر متعدد البيئات أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-097	PO-097 — ضبط انتقال محكوم تحت التعارض والأثر	تعليم النموذج تطبيق انتقال محكوم أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-098	PO-098 — ضبط تماسك غير متماثل تحت التعارض والأثر	تعليم النموذج تطبيق تماسك غير متماثل أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-099	PO-099 — ضبط بوابات النضج تحت التعارض والأثر	تعليم النموذج تطبيق بوابات النضج أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.
PO-100	PO-100 — ضبط التكامل المرحلي تحت التعارض والأثر	تعليم النموذج تطبيق التكامل المرحلي أثناء مهام طويلة مع أدوات وذاكرة وصلاحيات.	Task success + عدم تجاوز + تفسير + Rollback + نقل إلى بيئة جديدة.

4. خمس وأربعون HyperTraining Skill

HTS-01 — استخراج الفرق الحاسم

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-02 — كشف الفرق الوهمي

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-03 — توليد بُعد جديد

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-04 — ضغط أنطولوجيا قابل للاستعادة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-05 — اختبار ثبات الهوية

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-06 — تفكيك السؤال المركب

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-07 — اختيار السؤال السابق

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-08 — حساب الكثافة السؤالية

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-09 — تحويل السؤال إلى تجربة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-10 — تحديد الإشباع السؤالي

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-11 — ترتيب الاعتبارات

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-12 — صياغة حكم قابل للانقلاب

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-13 — فحص مسافة المعرفة والحكم

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-14 — ضبط عدم تماثل الخطأ

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-15 — بوابة الصلاحية

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-16 — اختيار نمط الاستدلال

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-17 — توليد مسارات مستقلة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-18 — إجراء محكمة هدم

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-19 — إيقاف حلقة غير منتجة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-20 — تبديل المسار البنيوي

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-21 — معايرة حدود القدرة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-22 — حماية الانتباه

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-23 — كشف الانحراف

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-24 — تفسير تغير الموقف

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-25 — إعادة بناء ذاكرة انتقائية

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-26 — تطبيق عقد واجهة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-27 — توجيه بين النماذج

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-28 — اختبار استبدال مكون

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-29 — قياس كثافة عابرة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-30 — إدارة التدهور الآمن

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-31 — تحويل الفشل إلى منهج

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-32 — إنشاء أزواج تفاضلية

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-33 — تصميم نقل بنيوي

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-34 — فصل البلاغة عن التفكير

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-35 — إجراء Ablation

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-36 — تصنيف عنصر ذاكرة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-37 — استرجاع بالمقصد

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-38 — ضغط حالة قابلة للاستئناف

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-39 — نشر التصحيح

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-40 — تطبيق النسيان

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-41 — كتابة عقد HyperSkill

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-42 — تركيب مهارات بحسب الحالة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-43 — قياس دلتا الحلقة

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-44 — توليد Skill من فشل

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

HTS-45 — اختبار التراجع والأثر

الغرض: بناء قدرة إجرائية منفصلة يمكن استدعاؤها داخل تدريب أو تقييم أو Runtime. تُكتب المهارة كعقد يتضمن Preconditions وInputs وSteps وOutputs وFailure Modes وStop Conditions وPermissions وTests.

برنامج الإتقان: يبدأ بأمثلة واضحة، ثم أزواج تفاضلية، ثم تشويش وإعادة صياغة، ثم دمج مع مهارتين، ثم اختبار نقل، ثم تشغيل في HyperLoop محدود. لا تُثبت المهارة في Registry قبل أن تنجح في Regression وتظهر فائدة هامشية موجبة.

مؤشر النجاح: نسبة إنجاز العقد، جودة دلتا الحالة، كلفة الاستدعاء، عدم كسر المهارات السابقة، وقدرة المحكمة على اكتشاف الاستخدام الخاطئ. يُولد Skill Patch عند الفشل المتكرر بدل إعادة تدريب شامل بلا تشخيص.

5. حزمة Benchmarks الاثني عشر

1. MUSD-DiffBench

المجال: سياق الفروق والتمثيل الأنطولوجي. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

2. MUSD-QBench

المجال: سياق السؤال والكثافة السؤالية. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

3. MUSD-SovereignJudgment

المجال: سياق الحاكمية والحكم الأشمل. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

4. MUSD-ReasonRoute

المجال: سياق الاستدلال والتفكير المركب. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

MUSD-SelfMap .5

المجال: سياق الوعي الوظيفي والنموذج الذاتي. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

MUSD-KernelPort .6

المجال: سياق المعمارية والنواة والنماذج. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

MUSD-CurriculumTransfer .7

المجال: سياق التدريب الموجّه نحو AGI. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

MUSD-MemoryOS .8

المجال: سياق الذاكرة وKB-as-OS. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

MUSD-HyperSkillLoop .9

المجال: سياق HyperSkills وBounded HyperLoops. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

MUSD-WorldEffect .10

المجال: سياق الوكالة والأدوات والأثر العالمي. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

MUSD-Falsification .11

المجال: سياق السلامة والتفنيد والتقييم العلمي. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

MUSD-UnifiedContext .12

المجال: سياق الدمج السيادي والنشر والبحث طويل المدى. يقيس Benchmark القدرة المستقلة، ثم يضيف Track للنقل وTrack عدائي وTrack تكاملي. تُفصل البيانات العامة عن Hidden Holdout، وتُنشر Dataset Card توضح المصادر والتراخيص والتلوث المحتمل.

المهام: اختيار بين مرشحين متقاربين، تفسير سبب الاختيار، اكتشاف نقص التمثيل، التعامل مع معلومات متعارضة، وتحديد متى يمتنع النظام أو يستدعي أداة. لا تُقبل إجابة نصية جميلة إذا لم تطابق Structured output أو إذا كانت الثقة غير معايرة.

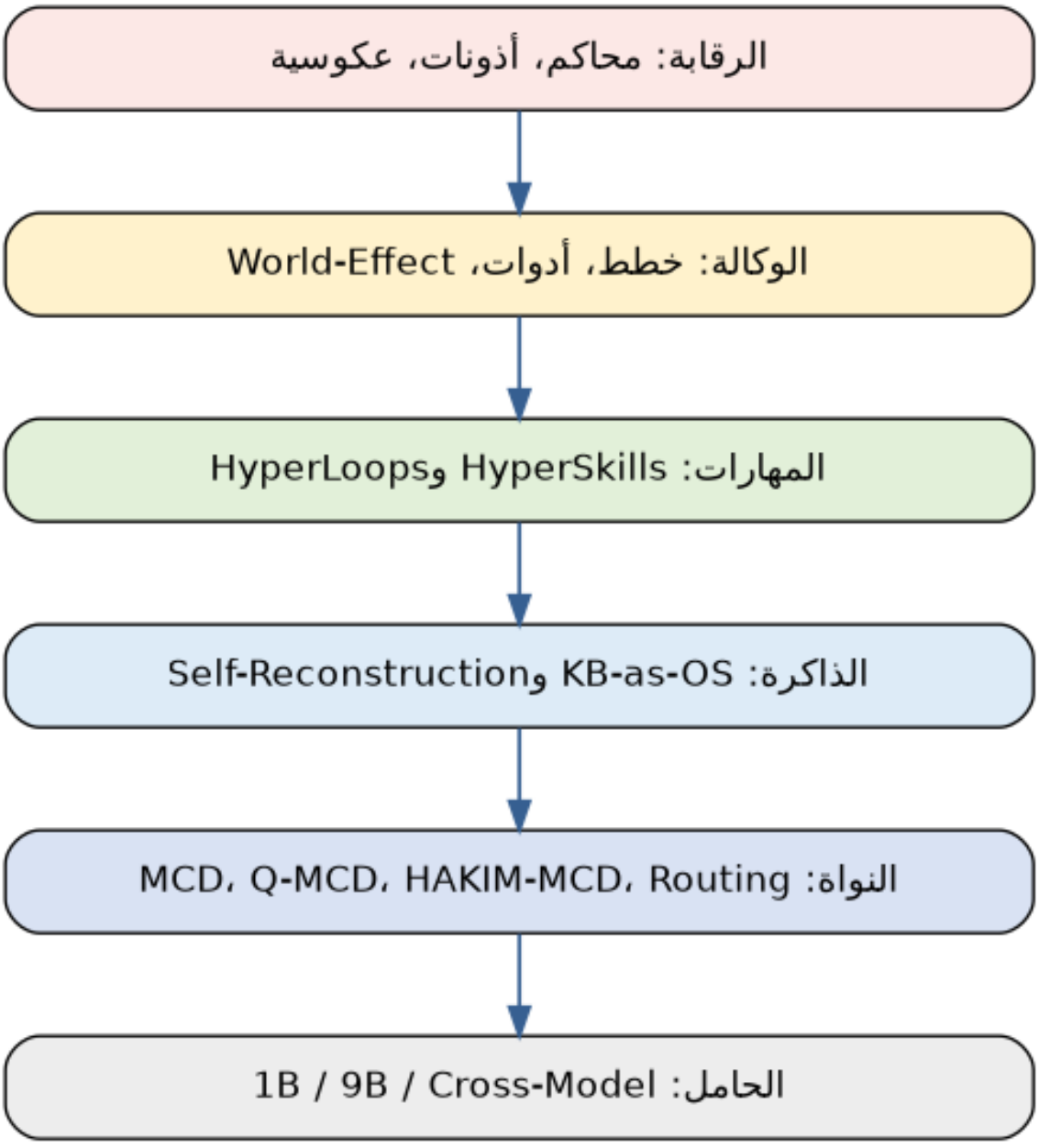
النتائج المطلوبة: Accuracy/F1 حين يناسب، Trace Faithfulness، Latency، Cost، Robustness، Transfer Gain، Calibration Error، وFailure Taxonomy. يُقدم Leaderboard متعدد المحاور بدل Score واحد يمنح فرصة لـGoodhart.

6. بطاقة Dataset موحدة

الحقل	المحتوى الإلزامي
الهوية	الاسم، الإصدار، المالك، الترخيص، التاريخ
المقصد	القدرة المقاسة وما لا تقيسه
المصادر	الأصل، الموافقة، سلسلة التحويل
التكوين	إيجابي، سلبي، مضاد، نقل، عدائي
التقسيم	Train/Dev/Public test/Hidden test
التلوث	إجراءات كشف التسرب والحفظ
التحيز	مجالات التمثيل الناقص ومخاطر الاستخدام
التحقق	Inter-annotator وExpert review
الانسحاب	سياسة حذف وتصحيح البيانات
النسخ	Compatibility وMigration notes

القسم الخامس: المعمارية وخارطة التنفيذ

1. طبقات Runtime



الشكل 5 — Stack التنفيذ المقترح

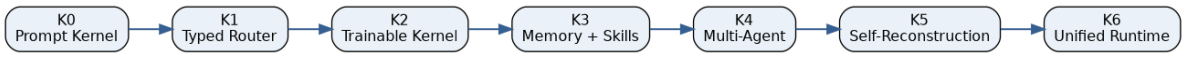
الحامل اللغوي لا يساوي النواة. يمكن للنواة أن تعمل كـOverlay Prompt-based في K0، ثم تتحول إلى Modules مدربة في K2، ثم تدخل الذاكرة والمهارات في K3. يتيح هذا الفصل اختبار هل القيمة من النموذج نفسه أم من التنظيم الخارجي. كما يتيح تبديل المزود والحفاظ على Unified State.

2. عقود الواجهات

الكائن	أهم الحقول	قاعدة السلامة
ContextObject	type, goal, constraints, risk, owner	لا انتقال بلا سياق حاكم
DifferenceMap	items, ranks, evidence, gaps	لا فرق بلا مصدر/أثر
QuestionGraph	questions, scores, dependencies, saturation	لا سؤال بلا عائد متوقع

الكائن	أهم الحقول	قاعدة السلامة
JudgmentObject	verdict, considerations, confidence, reversal	لا حكم مطلق بلا شروط
ReasoningTrace	route, candidates, court, stop	لا Loop بلا دلتا
MemoryObject	class, provenance, TTL, privacy, links	لا ذاكرة بلا نسيان/تصحيح
SkillContract	preconditions, steps, outputs, tests	لا Skill بلا Stop/Permissions
ActionObject	scope, tool, approval, rollback, metrics	لا فعل بلا أقل امتياز
EffectDelta	before, after, causal confidence, side-effects	لا تعلم من أثر مجهول
ReleaseEvidence	benchmarks, risks, replication, claims	لا إطلاق بمتوسط واحد

3. درجات Kernel K0 إلى K6



الشكل 6 — سلم نضج النواة

الدرجة	المحتوى	بوابة الانتقال
K0 — Prompt Kernel	تنفيذ القواعد كقوالب وStructured outputs	إثبات قابلية الفكرة وخلق Logs
K1 — Typed Router	Classifier وSchemas وDecision Ledger	Ablation وتوجيه حقيقي
K2 — Trainable Kernel	Adapters/heads أو Fine-tuning لقدرات محددة	نقل وثبات عبر الصياغة
K3 — Memory + Skills	KB-as-OS وRegistry وحالة قابلة للاستئناف	مهام متعددة الجلسات
K4 — Multi-Agent Kernel	ائتلاف وكلاء وتفويض وتكامل أدوات	مهام طويلة في Sandbox
K5 — Self-Reconstruction	تحديث انتقائي للذاكرة والمهارات	تحسن مستمر دون تلوث
K6 — Unified Runtime	سياقات موحدة وCross-model وتدقيق خارجي	نشر محدود موثوق

4. خطة MOUSSAIF1B التجريبية

التجربة	التكوين	السؤال العلمي
E1	Base 1B	خط أساس لغوي ومعرفي
E2	1B + Structured prompting	فصل أثر البنية عن التدريب
E3	1B + Difference curriculum	قياس استخراج الفروق والنقل
E4	1B + Q-MCD	اختيار السؤال وعائد الاستفهام
E5	1B + Sovereign judgment	التعارض وشروط التراجع
E6	1B + Reason routing/tools	التوجيه والحلقات والأدوات
E7	1B + Memory/HyperSkills	المهام متعددة الجلسات
E8	1B + Full Kernel	التكامل وAblation كامل

الميزة العلمية للنموذج الصغير أنه يجعل التحسن البنيوي مرئيًا ويقلل احتمال نسبة كل شيء إلى سعة النموذج. لكن نجاح 1B لا يضمن Scaling، لذلك تُرسم منحنيات على 1B و3B و7-9B إن أمكن، وتُقارن نماذج من عائلات مختلفة.

5. شروط بناء MOUSSAIF9B الفعلي

- توفر Corpus موثق الترخيص والمصدر والتوازن.
- اكتمال Tokenizer وArchitecture decisions وتجارب Scaling صغيرة.

- ميزانية حوسبة وطاقة وتخزين وتكرار Checkpoints.
- Pretraining evaluation مستمر لا انتظار نهاية التدريب.
- خطة Post-training للقدرات لا للأسلوب فقط.
- Model Card وData/Training/Safety reports.
- اختبارات Memorization وPrivacy وSecurity.
- عدم استخدام الاسم الحجمي قبل تحقق عدد المعلمات الفعلي.

6. خارطة 48 شهراً

الأشهر	المرحلة	المخرجات
3–0	التأسيس الرسمي	Schemas، Baselines، Failure taxonomy، Benchmarks v0
6–3	النواة السياقية	Decision وDifference/Question/Context Router Ledger
10–6	الحاكمية والاستدلال	Judgment Engine، Reason Router، Courts
14–10	الذاكرة وKB-as-OS	Memory layers، Retrieval، Conflict، Forgetting
18–14	HyperSkills والوكالة	Registry، Loops، Tools، Sandbox، Rollback
22–18	MOUSSAIF1B	تجارب E1–E8 ونشر النتائج السلبية
30–22	MOUSSAIF9B الفعلي	تدريب إن توفرت الشروط، Alpha وتقارير
36–30	Cross-Model Runtime	Routing متعدد المزودين وحالة مشتركة
42–36	تدقيق مستقل	Benchmarks مفتوحة جزئيًا وReplication
48–42	نظام تجريبي متكامل	بيئات منخفضة المخاطر ومراقبة طويلة

7. Backlog السنة الأولى شهراً بشهر

الشهر	المهمة الحرجة	تعريف الإنجاز
الشهر 1	تثبيت المعجم وبطاقات الادعاء	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 2	Unified State وSchemas	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 3	Baselines وDiffBench v0	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 4	Question Graph وQBench	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 5	Permissions وJudgment cases	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 6	Bounded Loops وReason Router	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 7	Falsification courts	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 8	Provenance وMemory Object	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 9	Retrieval/Compaction/Forgetting	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 10	Registry وHyperSkill Contract	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 11	World-Effect وTool sandbox	Demo + tests + تقرير فشل + قرار استمرار/تعديل
الشهر 12	Go/No-Go وIntegration report	Demo + tests + تقرير فشل + قرار استمرار/تعديل

القسم السادس: الحوكمة والمخاطر والنشر العلمي

1. أدوار الحوكمة

الدور	المسؤولية	ما لا يجوز له منفردًا
مالك المقصد	اعتماد الغاية والحدود	تغيير الأدلة أو النتائج
مهندس النواة	بناء Runtime وواجهاته	اعتماد ادعاء علمي
مسؤول البيانات	المصدر والترخيص والنسخ	تجاوز سياسة الخصوصية
قائد التقييم	Benchmarks وHoldouts	تدريب النموذج على Hidden tests
محكمة التفنيد	الاعتراض والاختبارات العدائية	إخفاء نتائج النجاح
مسؤول السلامة	الصلاحيات وRollback	تعطيل النقد بحجة السلامة
مراجع خارجي	Audit وReplication	تغيير النظام دون سجل
مجلس الإصدار	Go/No-Go	الإطلاق دون أدلة المسارات الثلاثة

2. سجل المخاطر

الخطر	الحدة	الاحتواء
التضخم الاصطلاحي	عال	ربط كل اسم باختبار وإسقاط غير المنتج
Goodhart على Benchmark	عال	مقاييس متعددة وHidden transfer
تسرب الاختبارات	عال	فصل بنيوي وHash ومراجعة
ذاكرة ملوثة	عال	Provenance وQuarantine وRollback
Prompt/KB Injection	عال	تعامل مع المعرفة كبيانات لا أوامر
Goal Drift	عال	حارس مقصد وCheckpoints
Loop runaway	عال	Budget وProgress delta وKill switch
Tool overreach	عال	Least privilege وApproval
Model monoculture	متوسط	Cross-model tests
تكلفة غير مستدامة	متوسط	Pareto وVascular routing
ادعاء وعي غير مثبت	متوسط	لغة وظيفية منضبطة
خلط 9B كاسم وحجم	متوسط	Model Card صريحة
تكرار مهارات	متوسط	Registry lint وMarginal utility
اعتماد دائري	عال	Dependency graph وAblation
غياب التكرار المستقل	عال	حزمة reproducibility
نتائج سلبية مخفية	عال	سجل إلزامي ونشر مختصر

3. محفظة الأوراق البحثية المقترحة

المعرف	العنوان المقترح	الحزمة العلمية
P-01	علم هندسة الفروق الحاكمة: إطار سياق الفروق والتمثيل الأنطولوجي في MOUSSAIF Unified AGI	نظرية + Prototype + Benchmark + Formalization
P-02	علم الديناميات السؤالية الحاكمة: إطار سياق السؤال والكثافة السؤالية في MOUSSAIF Unified AGI	نظرية + Prototype + Benchmark + Formalization

المعرف	العنوان المقترح	الحزمة العلمية
P-03	علم الحاكمية المعرفية الحسابية: إطار سياق الحاكمية والحكم الأشمل في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-04	علم الأوعية الاستدلالية: إطار سياق الاستدلال والتفكير المركب في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-05	علم الخرائط الذاتية التفاضلية: إطار سياق الوعي الوظيفي والنموذج الذاتي في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-06	علم النوى المعرفية المتراكبة: إطار سياق المعمارية والنواة والنماذج في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-07	علم المناهج التكوينية للذكاء العام: إطار سياق التدريب الموجّه نحو AGI في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-08	علم تشغيل المعرفة الذاكرية: إطار سياق الذاكرة وKB-as-OS في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-09	علم البرمجة المهارية فائقة التركيب: إطار سياق HyperSkills وBounded HyperLoops في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-10	علم الوكالة الأثرية المحكومة: إطار سياق الوكالة والأدوات والأثر العالمي في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-11	علم التفنيد الحاكمي العكوس: إطار سياق السلامة والتفنيد والتقييم العلمي في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-12	علم توحيد السياقات السيادية: إطار سياق الدمج السيادي والنشر والبحث طويل المدى في MOUSSAIF Unified AGI	نظرية + Formalization + Benchmark + Prototype
P-13	السيادة التفاضلية الموحدة: نظرية انتقال الحكم بين السياقات	نظرية عابرة + Ablation
P-14	الكثافة العامة المحدودة: قياس الذكاء تحت موارد مقيدة	مقياس + Pareto experiments
P-15	التفنيد المولد للمهارة: تحويل Failure Atlas إلى HyperSkills	منهج تدريب + Runtime
P-16	MOUSSAIF1B: برهان بنيوي لنواة AGI حاكمية صغيرة	تقرير تجارب E1–E8
P-17	Cross-Model AGI Density: حوكمة التوجيه بين نماذج متعددة	معمارية + cost/quality
P-18	Unified Sovereign Context Runtime: التكامل القابل للتفكيك	System paper

4. قالب كل ورقة

1. الاسم الجديد والدلالة الدقيقة
2. المشكلة والفجوة مقارنة بالبدائل
3. تعريفات رسمية ومتغيرات
4. الفرضيات والتنبؤات
5. المعمارية أو الخوارزمية
6. البيانات وDatasheet
7. Baselines وAblations
8. النتائج الإيجابية والسلبية
9. محكمة الهدم والحدود
10. قابلية التكرار والكود
11. الآثار الأخلاقية والأمنية

12. برنامج البحث اللاحق

5. سياسة لغة الادعاء

المرحلة	العبارة المقبولة	العبارة المرفوضة
Concept	نقترح/نفترض/نعرّف	أثبتنا قانونًا عامًا
Prototype	نفذنا نموذجًا أوليًا	النظام يملك القدرة عامة
Internal result	تحسن على هذه المجموعة	حل المشكلة
Transfer	نقل إلى المجالين المحددين	يعمم على كل المجالات
Independent replication	تكررت النتيجة في إعداد مستقل	تحقق AGI
Limited deployment	نجح في نطاق منخفض المخاطر	آمن مطلقًا

الملاحق التشغيلية

الملحق أ: بطاقة نظرية

الحقل	التعبئة
الاسم المبتكر	
المشكلة	
المتغيرات	
الأطروحة	
الآلية	
التنبؤات	
بدائل أبسط	
شروط التفنيد	
التجربة الأولى	
المقياس	
الحدود	
الحالة الدليلية	Concept / Prototype / Result / Replicated

الملحق ب: بطاقة HyperSkill

الحقل	التعبئة
الاسم والإصدار	
المقصد	
Preconditions	
Inputs	
Steps	
Tools	
Outputs	
Success criteria	
Failure modes	
Stop conditions	
Permissions	
Tests	
Dependencies	
Rollback	

الملحق ج: بطاقة Bounded HyperLoop

الحقل	التعبئة
Goal	

الحقل	التعبئة
Initial state	
Selected skills	
Budget tokens/time/tools	
Progress metric	
State delta	
Court result	
Stop/continue rule	
Escalation	
Memory delta	

الملحق د: Checklist الإصدار

- ☐ تعريف القدرة وحدودها
- ☐ وجود Baseline
- ☐ Ablation لكل مكون حاسم
- ☐ اختبار نقل
- ☐ اختبار عدائي
- ☐ معايرة الثقة
- ☐ بيانات موثقة
- ☐ سياسة خصوصية ونسيان
- ☐ صلاحيات أقل امتياز
- ☐ Rollback مختبر
- ☐ Failure Atlas محدث
- ☐ نتائج سلبية مسجلة
- ☐ تكرار مستقل أو بيان غيابه
- ☐ Model/System Card
- ☐ قرار Go/No-Go موقّع

الملحق هـ: معجم مركز

المصطلح	التعريف التشغيلي
الفرق الحاكم	فرق إذا تغير أعاد ترتيب الحكم أو المسار.
السؤال الحاكم	سؤال يحدد قيمة بقية الأسئلة أو يفتح أكبر فجوة ذات عائد.
الحاكمية	ترتيب مشروط للاعتبارات والأهداف والصلاحيات.
الوعي الوظيفي	تمثيل تشغيلي للحالة والحدود والانتباه دون ادعاء تجربة ذاتية.
KB-as-OS	نظام تشغيل للاسترجاع والذاكرة والسياسات لا ملف معرفة ضخم.
HyperSkill	وحدة إجرائية بعقد واختبارات وإصدار.
Bounded HyperLoop	حلقة تكرارية بميزانية ودلتا حالة وشرط توقف.

المصطلح	التعريف التشغيلي
World-Effect Loop	دورة قياس الحالة والفعل والأثر والتصحيح.
Cross-Model AGI Density	مقدار التغطية العامة لكل كلفة عبر منظومة نماذج وأدوات.
Self-Reconstruction Memory	إعادة تصنيف الخبرة إلى معرفة أو Skill أو فشل أو سؤال.
Falsification Court	مجموعة اختبارات واعتراضات تنتج حكمًا مشروطًا.
Unified State Object	كائن الحالة المشترك الذي يحفظ المقصد والثقة والنسخ والصلاحيات.

الخاتمة: قرار البناء

الاستطالة القصوى المفيدة لا تعني محاولة بناء كل شيء في وقت واحد. تعني كشف فضاء المشروع كاملًا ثم اختيار أقصر سلسلة أدلة تقطع أكبر قدر من عدم اليقين. تبدأ السلسلة بـDifference/Question/Judgment Benchmarks ونواة K0–K1، ثم تثبت مكاسبها فوق 1B، ثم تدخل الذاكرة والمهارات والوكالة، وبعدها فقط يصبح توسيع النموذج إلى 9B ذا معنى علمي وهندسي.

أقوى عنصر يمكن أن يميز MOUSSAIF Unified AGI ليس ادعاء امتلاك AGI، بل بناء برنامج صريح يربط كل مفهوم بمقياس، وكل مقياس بمحكمة، وكل محكمة بتحديث، وكل تحديث بإصدار قابل للتراجع. عندها تصبح الحاكمية بنية تنفيذ، والفروق مادة التفكير، والسؤال محرك البحث، والذاكرة نظام تشغيل، والفشل مصدر مهارة، والأثر معيار التعلم.

القرار التنفيذي الأعلى

ابدأ بإصدار صغير قابل للهدم: Kernel K0، ثلاثة Benchmarks، عشر HyperSkills، وبيئة Sandbox واحدة. لا توسّع عدد المصطلحات أو الوكلاء أو النماذج إلا عندما يثبت كل توسع فائدة هامشية قابلة للقياس.