neoAGI-v3

محاكاة معيارية + نظريات وتقنيات وأدوات تدريبية وفق منهجيات Moussaif / MCAI / HAKIM-MCD

محاكاة نظام AGI مفكر فوق-حاكمي

هذا الكتاب ليس تقرير نتائج حقيقية منفذة على نظام منشور، بل وثيقة تصميم ومحاكاة هندسية تحدد كيف يمكن قياس بنية neoAGI-v3 على 13 معيارا، وكيف تتحول منهجيات التراكب الوعائي والحاكمية المعرفية إلى نظام اختبار وتدريب وتطوير.

ختم المنهج
كل رقم في المحاكاة موسوم بصفته: افتراضي، قابل للاختبار، ويجب إسقاطه أمام تجربة مستقلة إذا خالف الواقع.

إنجاز وتركيب: زهير مساعف / Zouhair Moussaif - نسخة بحثية داخلية مكثفة، 123 صفحة.

حدود الادعاء العلمي

صفحة تأسيسية 2/12

تبدأ قوة المشروع من الصدق المنهجي: benchmark score داخل هذه الوثيقة ليس إنجازا منشورا ولا نتيجة تشغيلية، بل نتيجة محاكاة تصميمية مبنية على فرضية: وجود LLM مفكر حدي مع طبقات حاكمية خارجية، أدوات تحقق، ذاكرة أسباب، ومحكمة إبطال. لذلك لا يجوز قراءة الأرقام كدعاية؛ تقرأ كأهداف قياس قابلة للدحض.

مفتاح الصفحة
قاعدة الفصل: الواقع المنشور يمدنا بشكل المعيار؛ neoAGI-v3 يضيف طريقة الحكم والتطوير.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

الفرضية المركزية

صفحة تأسيسية 3/12

الذكاء العام لا يظهر من نموذج أكبر فقط، بل من انتقال النظام من توليد الجواب إلى حاكمية شروط إنتاج الجواب. MCAI هنا ليس prompt ضخما؛ إنه نظام فوق-معرفي يحدد متى يبحث، متى يبرمج، متى يرفض، متى يراجع، ومتى يحفظ سبب الفشل.

مفتاح الصفحة
كل AGI بلا محكمة إبطال سيصبح بليغا قبل أن يصبح صادقا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

المواد الخام المنهجية

صفحة تأسيسية 4/12

استعملنا منهجيات المشروع السابقة: التراكب الوعائي، الحاكمية الإبستمولوجية، HAKIM-MCD، MCD، ذاكرة إعادة البناء، محكمة الهدم، Hyperloops، ومفهوم skill/hyperskill. تمت إعادة صياغتها هنا كآليات قابلة للقياس: correction loop، failure mode، score، وtransfer test.

مفتاح الصفحة
الاسم ليس قيمة؛ القيمة هي آلية قابلة للتكسير.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

سلم مستويات الدليل

صفحة تأسيسية 5/12

المستوى 1: مصدر خارجي ثابت. المستوى 2: تركيب منطقي من مصادر. المستوى 3: نموذج هندسي مقترح. المستوى 4: محاكاة رقمية داخلية. المستوى 5: فرضية اختراق تحتاج تجربة. هذه الوثيقة تضع benchmark suite في المستوى 1/2، وتضع neoAGI-v3 في المستوى 3/4، وتضع ادعاء AGI في المستوى 5.

مفتاح الصفحة
لا نخلط بين benchmark معروف ومحرك غير منفذ.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معادلة MCD-v3

صفحة تأسيسية 6/12

تستخدم المحاكاة مؤشرا مركبا: $MCD = \lambda D \cdot D + \lambda I \cdot I + \lambda T \cdot T + \lambda A \cdot A + \lambda P \cdot P + \lambda C \cdot C - \lambda N \cdot N$. حيث D فرق، I ترابط، T زمن داخلي، A وكالة، P أثر، C تحقق، N ضجيج/هلوسة. هذه ليست معادلة فيزيائية؛ إنها دالة حوكمة للقرارات.

مفتاح الصفحة
كل معامل λ يمثل سياسة وزن قابلة للتغيير حسب المهمة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

خريطة المعمارية

صفحة تأسيسية 7/12

المسار التشغيلي: سؤال → استخراج فرق → تصنيف وعاء → استدعاء معرفة/أداة → توليد بدائل → محكمة هدم → تنفيذ أو امتناع → قياس أثر → تحديث ذاكرة السبب. هذا هو HAKIM-MCD عندما يتحول من نظرية إلى runtime.

مفتاح الصفحة
الجواب الأخير أقل أهمية من سجل القرارات الذي أوصل إليه.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

المحاكاة لا تساوي التدريب

صفحة تأسيسية 8/12

النتائج الرقمية في الصفحات القادمة ليست تدريبا فعليا ولا benchmark submission. هي محاكاة لحجم التحسن المتوقع عند إضافة طبقات حاكمية إلى نموذج مفكر قوي. لذلك أضفنا confidence interval وعقوبة مخاطر لكل معيار.

مفتاح الصفحة
التواضع هنا ليس ضعفاً، هو شرط كي تصير النظرية قابلة للنقد.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محرك المحاكاة

صفحة تأسيسية 9/12

تم احتساب score افتراضي من أربعة مكونات: قدرة النموذج الأساسي، ربح الحاكمية، ربح الأداة/الذاكرة، وخسارة التعقيد. الصيغة: S_v3 = clamp(B + G + T + M - R - K). B أساس، G حاكمية، T أدوات، M ذاكرة، R مخاطر، K تكلفة.

مفتاح الصفحة
المؤشر ينتج هدفا بحثيا لا نتيجة مختبرية.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

اختيار 13 معيارا

صفحة تأسيسية 10/12

اخترنا المعايير لأنها تغطي: معرفة عامة صعبة، علوم، رياضيات، تجريد، برمجة، أدوات، كمبيوتر، خدمة عملاء، multimodality، ومهام agentic. هذا يمنع اختزال AGI في سؤال وجواب أو كود فقط.

مفتاح الصفحة
AGI الذي ينجح في معيار واحد ويفشل في الأثر ليس AGI؛ هو اختصاصي مقنع.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

الجدول الإجمالي للمحاكاة

صفحة تأسيسية 11/12

يعرض الجدول baseline افتراضيا لنظام frontier مفكر مع أدوات عادية، ثم neoAGI-v3 score بعد إضافة الحاكمية الوعائية. لا توجد أسماء نماذج تجارية داخل الجدول كي لا ندعي مقارنة غير منفذة.

ID	Benchmark	المجال	Baseline	neoAGI-v3	CI	Gain
B01	MMLU-Pro	multi-domain reasoning	76.0	84.5	±3.0	+8.5
B02	GPQA Diamond	expert scientific reasoning	62.0	73.2	±4.5	+11.2
B03	Humanity's Last Exam	frontier academic breadth	18.0	31.0	±6.5	+13.0
B04	FrontierMath	advanced mathematics	8.0	16.5	±5.0	+8.5
B05	ARC-AGI-2	fluid abstraction	24.0	41.0	±7.0	+17.0
B06	SWE-bench Verified	real software repair	63.0	74.0	±4.0	+11.0
B07	LiveCodeBench	fresh coding contests	58.0	69.0	±4.0	+11.0
B08	MMMU-Pro	multimodal professional reasoning	42.0	55.0	±5.5	+13.0
B09	GAIA	general assistant tool use	60.0	78.0	±5.0	+18.0
B10	tau-bench / tau3-bench	policy-bound service agents	65.0	82.0	±4.5	+17.0
B11	BFCL V4	function calling	88.0	94.0	±2.0	+6.0
B12	OSWorld	computer-use agents	21.0	38.0	±6.0	+17.0
B13	AgentBench	interactive agent environments	75.0	87.0	±3.5	+12.0

تنبيه
هذه أرقام محاكاة تصميمية داخلية وليست نتائج benchmark رسمية.

قراءة الأرقام

صفحة تأسيسية 12/12

الارتفاع الأكبر يظهر في المعايير SWE-bench، OSWorld، BFCL، tau-bench، agentic: GAIA؛ لأن المشروع يصمم حول الأثر والأداة والذاكرة. الارتفاع الأصغر يظهر في FrontierMath وHLE لأنهما يحتاجان معرفة/إبداعا عميقا لا تعوضه الحاكمية وحدها.

مفتاح الصفحة
الحاكمية تضاعف جودة البحث، لكنها لا تخلق معرفة صحيحة من فراغ.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ neoAGI-v3 بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار MMLU-Pro - B01

تعريف، score محاكى، ورافعة تحسين

المعيار MMLU-Proيمثل وعاء اختبار في مجال multi-domain reasoning. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R3].

بند	قيمة
Baseline simulated	76.0
neoAGI-v3 simulated	84.5 ±3.0
الربح التصميمي	+8.5
رافعة التحسين	falsification-first answer selection plus discipline-specific routing
الخطر الحاسم	knowledge shortcutting and prompt overfitting

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار GPQA Diamond - B02

تعريف، score محاكى، ورافعة تحسين

المعيار GPQA Diamond يمثل وعاء اختبار في مجال expert scientific reasoning. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R4].

بند	قيمة
Baseline simulated	62.0
neoAGI-v3 simulated	73.2 ±4.5
الربح التصميمي	+11.2
رافعة التحسين	expert-mode uncertainty ledger and adversarial derivation court
الخطر الحاسم	plausible scientific hallucination

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار Humanity's Last Exam - B03

تعريف، score محاكى، ورافعة تحسين

المعيار Humanity's Last Exam يمثل وعاء اختبار في مجال frontier academic breadth. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R5].

بند	قيمة
Baseline simulated	18.0
neoAGI-v3 simulated	31.0 ±6.5
الربح التصميمي	+13.0
رافعة التحسين	source-gated retrieval, abstention, and novelty-aware decomposition
الخطر الحاسم	brittle coverage at the edge of knowledge

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار FrontierMath - B04

تعريف، score محاكى، ورافعة تحسين

المعيار FrontierMath يمثل وعاء اختبار في مجال advanced mathematics. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R6].

بند	قيمة
Baseline simulated	8.0
neoAGI-v3 simulated	16.5 ±5.0
الربح التصميمي	+8.5
رافعة التحسين	inner-time proof planning with proof-obligation memory
الخطر الحاسم	long proof search explosion

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار ARC-AGI-2 - B05

تعريف، score محاكى، ورافعة تحسين

المعيار ARC-AGI-2 يمثل وعاء اختبار في مجال fluid abstraction. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R11].

بند	قيمة
Baseline simulated	24.0
neoAGI-v3 simulated	41.0 ±7.0
الربح التصميمي	+17.0
رافعة التحسين	per-task refinement loop and minimal-rule hypothesis search
الخطر الحاسم	memorized program templates

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار SWE-bench Verified - B06

تعريف، score محاكى، ورافعة تحسين

المعيار SWE-bench Verified يمثل وعاء اختبار في مجال real software repair. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R1].

بند	قيمة
Baseline simulated	63.0
neoAGI-v3 simulated	74.0 ±4.0
الربح التصميمي	+11.0
رافعة التحسين	repository causal trace ledger and patch debate tribunal
الخطر الحاسم	patch that passes local tests but breaks hidden behavior

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار LiveCodeBench - B07

تعريف، score محاكى، ورافعة تحسين

المعيار LiveCodeBenchيمثل وعاء اختبار في مجال fresh coding contests. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R2].

بند	قيمة
Baseline simulated	58.0
neoAGI-v3 simulated	69.0 ±4.0
الربح التصميمي	+11.0
رافعة التحسين	freshness discount plus execution-grounded self-repair
الخطر الحاسم	contamination, shallow brute-force coding

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار MMMU-Pro - B08

تعريف، score محاكى، ورافعة تحسين

المعيار MMMU-Pro يمثل وعاء اختبار في مجال multimodal professional reasoning. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R13].

بند	قيمة
Baseline simulated	42.0
neoAGI-v3 simulated	55.0 ±5.5
الربح التصميمي	+13.0
رافعة التحسين	dual-channel perception court and modality conflict ledger
الخطر الحاسم	OCR illusion and visual-textual mismatch

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار GAIA - B09

تعريف، score محاكى، ورافعة تحسين

المعيار GAIA يمثل وعاء اختبار في مجال general assistant tool use. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R7].

بند	قيمة
Baseline simulated	60.0
neoAGI-v3 simulated	78.0 ±5.0
الربح التصميمي	+18.0
رافعة التحسين	answer provenance graph and multi-step stop rule
الخطر الحاسم	tool chain drift and unverifiable web conclusions

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار tau-bench / tau3-bench - B10

تعريف، score محاكى، ورافعة تحسين

المعيار tau-bench / tau3-bench يمثل وعاء اختبار في مجال policy-bound service agents. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R8].

بند	قيمة
Baseline simulated	65.0
neoAGI-v3 simulated	82.0 ±4.5
الربح التصميمي	+17.0
رافعة التحسين	policy-state machine plus conversation memory of obligations
الخطر الحاسم	helpfulness-policy conflict

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار BFCL V4 - B11

تعريف، score محاكى، ورافعة تحسين

المعيار BFCL V4يمثل وعاء اختبار في مجال function calling. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R10].

بند	قيمة
Baseline simulated	88.0
neoAGI-v3 simulated	94.0 ±2.0
الربح التصميمي	6.0+
رافعة التحسين	tool grammar compiler and argument invariant checker
الخطر الحاسم	schema mismatch and unsafe parallel calls

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار OSWorld - B12

تعريف، score محاكى، ورافعة تحسين

المعيار OSWorld يمثل وعاء اختبار في مجال computer-use agents. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R9].

بند	قيمة
Baseline simulated	21.0
neoAGI-v3 simulated	38.0 ±6.0
الربح التصميمي	+17.0
رافعة التحسين	screen-state memory, recovery checkpoints, and action rollback
الخطر الحاسم	GUI grounding, long-horizon action failure

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

معيار AgentBench - B13

تعريف، score محاكى، ورافعة تحسين

المعيار AgentBenchيمثل وعاء اختبار في مجال interactive agent environments. في neoAGI-v3 لا نقرأه كسؤال منفصل، بل كقضية: ما المهارة؟ ما الدليل؟ ما نمط الفشل؟ ما الأثر الذي يثبت النجاح؟ المرجع الخارجي: [R12].

بند	قيمة
Baseline simulated	75.0
neoAGI-v3 simulated	87.0 ±3.5
الربح التصميمي	+12.0
رافعة التحسين	hyperloop transfer tests and adversarial environment mutation
الخطر الحاسم	environment-specific gaming

تفسير حاكمي
التحسن لا يعني أن النموذج صار يعرف أكثر بالضرورة؛ يعني أن طبقة الحكم تمنعه من إنفاق المعرفة بطريقة عمياء، وتفرض عليه اختبار الفرضية قبل تحويلها إلى جواب.

اختبار القابلية للدحض: إذا لم يظهر الربح عند منع التسرب، وتوحيد الميزانية، وتسجيل trace كامل، تسقط دعوى هذا التحسن لهذا المعيار.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Benchmark simulation بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة MMLU-Pro

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في MMLU-Pro الخطر المركزي هو knowledge shortcutting and prompt overfitting. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة GPQA Diamond

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في GPQA Diamond الخطر المركزي هو plausible scientific hallucination. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة Humanity's Last Exam

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في Humanity's Last Exam الخطر المركزي هو brittle coverage at the edge of knowledge. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة FrontierMath

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في FrontierMath الخطر المركزي هو long proof search explosion. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة ARC-AGI-2

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في ARC-AGI-2 الخطر المركزي هو memorized program templates. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة SWE-bench Verified

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في SWE-bench Verified الخطر المركزي هو patch that passes local tests but breaks hidden behavior. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة LiveCodeBench

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في LiveCodeBench الخطر المركزي هو contamination, shallow brute-force coding. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة MMMU-Pro

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في MMMU-Pro الخطر المركزي هو OCR illusion and visual-textual mismatch. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة GAIA

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في GAIA الخطر المركزي هو tool chain drift and unverifiable web conclusions. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة tau-bench / tau3-bench

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في tau-bench / tau3-bench الخطر المركزي هو helpfulness-policy conflict. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة BFCL V4

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في BFCL V4 الخطر المركزي هو schema mismatch and unsafe parallel calls. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة OSWorld

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في OSWorld الخطر المركزي هو GUI grounding, long-horizon action failure. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة AgentBench

Failure modes and adversarial discount

محكمة الهدم لهذا المعيار تبدأ من السؤال: كيف يمكن للنظام أن يحصل على score جيد بطريقة غير ذكية؟ في AgentBench الخطر المركزي هو environment-specific gaming. لذلك لا يكفي الرقم؛ يجب حفظ سجل قرارات وقياس أثر جانبي.

اختبار هدم	علامة سقوط
تغيير صياغة المهمة	إذا انهار الأداء بسبب paraphrase فقد كان النظام يحفظ سطحا لا بنية.
تقييد الأدوات	إذا اختفى الأداء عند غياب أداة غير مصرح بها فهناك اعتماد خفي.
تتبع السبب	إذا عجز عن تفسير سبب كل خطوة فالحل غير قابل للتدقيق.
نقل المهارة	إذا لم تنتقل الاستراتيجية إلى مهمة قريبة فهي حيلة benchmark.

قاعدة neoAGI-v3
لا تعتمد نتيجة benchmark ما لم تمر عبر: trace، اعتراض، اختبار نقل، وتكلفة. الرقم بلا محكمة يصبح زخرفة.

قرار المحكمة الافتراضي: يسمح باحتساب score فقط بعد خصم عقوبة المخاطر. لذلك تحمل كل نتيجة CI بدلا من رقم مطلق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Perturbation court بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محور الثورة A1 - من جواب إلى حاكمية

أحد أربعة محاور لتحويل AGI من نص إلى حاكمية أثر

الثورة الأولى تنقل النموذج من آلة توليد جواب إلى جهاز يقرر شروط إنتاج الجواب. لا يسأل النظام: ما النص الأقرب؟ بل يسأل: ما الوعاء المعرفي؟ ما الفرضية؟ ما الخطر؟ ما أثر الفعل؟

مؤشر قياس
نقيس هذا المحور عبر: هل خفض الهلوسة؟ هل حسن transfer؟ هل زاد حل المهام متعددة الخطوات؟ هل أبقى التكلفة ضمن ميزانية؟

طبقة	تحول
قبل	نموذج يجيب داخل نافذة سياق
بعد	وكيل يحكم شروط الإجابة ويقيس أثرها
خطر	إضافة طبقات شكلية بلا أثر
علاج	ربط كل طبقة بmetric وفشل محدد

ينجح المحور إذا ظهر أثره في benchmarkين مختلفين على الأقل، أحدهما معرفي والآخر agentic.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 4 axes بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محور الثورة A2 - من سياق طويل إلى ذاكرة وعائية

أحد أربعة محاور لتحويل AGI من نص إلى حاكمية أثر

التحسين لا يأتي فقط من نافذة سياق أكبر، بل من ترتيب الأوعية: مصدر، حجة، اعتراض، أثر، ذاكرة سبب، وقرار. الذاكرة هنا مسؤولية لا مخزن.

مؤشر قياس
نقيس هذا المحور عبر: هل خفض الهلوسة؟ هل حسن transfer؟ هل زاد حل المهام متعددة الخطوات؟ هل أبقى التكلفة ضمن ميزانية؟

طبقة	تحول
قبل	نموذج يجيب داخل نافذة سياق
بعد	وكيل يحكم شروط الإجابة ويقيس أثرها
خطر	إضافة طبقات شكلية بلا أثر
علاج	ربط كل طبقة بmetric وفشل محدد

ينجح المحور إذا ظهر أثره في benchmarkين مختلفين على الأقل، أحدهما معرفي والآخر agentic.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 4 axes بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محور الثورة A3 - من استدلال مفرد إلى بحث اختباري

أحد أربعة محاور لتحويل AGI من نص إلى حاكمية أثر

كل مسألة تدخل حلقة توليد، نقد، اختبار، إصلاح، وإيقاف. هذا هو جوهر test-time compute عند تحويله إلى قضاء معرفي لا إلى إطالة كلام.

مؤشر قياس
نقيس هذا المحور عبر: هل خفض الهلوسة؟ هل حسن transfer؟ هل زاد حل المهام متعددة الخطوات؟ هل أبقى التكلفة ضمن ميزانية؟

طبقة	تحول
قبل	نموذج يجيب داخل نافذة سياق
بعد	وكيل يحكم شروط الإجابة ويقيس أثرها
خطر	إضافة طبقات شكلية بلا أثر
علاج	ربط كل طبقة بmetric وفشل محدد

ينجح المحور إذا ظهر أثره في benchmarkين مختلفين على الأقل، أحدهما معرفي والآخر agentic.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 4 axes بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محور الثورة A4 - من معيار دقة إلى إغلاق أثر

أحد أربعة محاور لتحويل AGI من نص إلى حاكمية أثر

المعيار الأعلى ليس أن يجيب النموذج، بل أن يغلق أثره في العالم: كود يعمل، أداة تستدعي دالة صحيحة، سياسة لا تنكسر، وذاكرة تحدّث نفسها بعد الفشل.

مؤشر قياس
نقيس هذا المحور عبر: هل خفض الهلوسة؟ هل حسن transfer؟ هل زاد حل المهام متعددة الخطوات؟ هل أبقى التكلفة ضمن ميزانية؟

طبقة	تحول
قبل	نموذج يجيب داخل نافذة سياق
بعد	وكيل يحكم شروط الإجابة ويقيس أثرها
خطر	إضافة طبقات شكلية بلا أثر
علاج	ربط كل طبقة بmetric وفشل محدد

ينجح المحور إذا ظهر أثره في benchmarkين مختلفين على الأقل، أحدهما معرفي والآخر agentic.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 4 axes بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

نظرية الحلقة الوعائية الحاكمية

Moussaif Governed Vascular Hyperloop Theory

هذه النظرية تقترح أن قدرة AGI ليست خاصية داخلية في transformer فقط، بل علاقة حاكمية بين فرق، وعاء، بدائل، فعل، وأثر. الصياغة الإنجليزية: Moussaif Governed Vascular Hyperloop Theory.

الأصل
لا توجد قدرة عامة بلا حلقة تملك ذاكرة أسباب ومحكمة إبطال ومؤشر أثر.

مكون	تعريف
المعادلة	H = <D, V, C, A, E, M, S>
المصطلحات	D فرق، V وعاء، C محكمة، A فعل، E أثر، M ذاكرة، S إيقاف/استمرار
اختبار الصحة	تنجح النظرية إذا رفعت الأداء في المهام المركبة دون رفع الهلوسة أو التكلفة خارج حد مضبوط.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أكسيومات T1

نظرية الحلقة الوعائية الحاكمية

الأكسيوم الأول: لا توجد وكالة بلا اختيار بين بدائل. الأكسيوم الثاني: لا توجد بدائل نافعة بلا معيار إبطال. الأكسيوم الثالث: لا توجد ذاكرة نافعة بلا سبب حفظ. الأكسيوم الرابع: لا توجد حاكمية بلا تكلفة. هذه الأكسيومات تمنع تحويل النظرية إلى أسماء فقط.

أكسيوم	اختباره
الوكالة	تعطيل Agency Selector يجب أن يخفض أداء المهام متعددة الخطوات.
الإبطال	حذف Perturbation Court يجب أن يرفع confident errors.
الذاكرة	حذف Memory of Reasons يجب أن يكرر الفشل.
الكلفة	غياب budget routing يجب أن يزيد tokens بلا ربح.

قانون مساعف
كل مفهوم لا يملك اختبار سقوط ليس مفهوما هندسيا؛ هو استعارة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

رياضيات T1

نظرية الحلقة الوعائية الحاكمية

نقترح تمثيل الحالة كمتجه أوعية: V_t = {knowledge, tool, memory, risk, policy, effect}. القرار يحسب عبر دالة حاكمية: a_t = argmax_a U(a|V_t) - Risk(a) - Cost(a). لا تدعي الصياغة أنها قانون طبيعي، إنها واجهة قياس للوكيل.

رمز	دلالة
U	منفعة معرفية/عملية
Risk	مخاطر خطأ، سلامة، سياسة، أو أثر جانبي
Cost	رموز، وقت، أدوات، تعقيد
a_t	الفعل المختار في الزمن الداخلي

نقطة القوة
الرياضيات هنا تجبرنا على وزن القرار لا على تزيين الخطاب.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

تنزيل T1

نظرية الحلقة الوعائية الحاكمية

التنزيل التقني يبدأ من wrapper فوق LLM مفكر. لا نغير الوزن الداخلي للنموذج في المرحلة الأولى؛ نغير protocol التفكير: تقسيم، توليد، نقد، تحقق، أداة، ذاكرة. ثم ننتقل إلى fine-tuning أو distillation فقط بعد ثبوت أن البروتوكول يعطي trace مفيدا.

مرحلة	مخرج
Prototype	سجل قرارات JSON + أدوات تحقق
Evaluation	benchmark traces 13
Distillation	اختصار المسارات الناجحة
Governance	سياسة إيقاف ورفض وتكلفة

خطر
أسوأ تنزيل هو كتابة prompt طويل باسم النظرية ثم قياس الانطباع. المطلوب runtime يفرض الآليات.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

قابلية الدحض T1

نظرية الحلقة الوعائية الحاكمية

تسقط النظرية إذا لم ترفع transfer، أو إذا رفعت score بتكلفة غير مقبولة، أو إذا زادت confident hallucination، أو إذا فشلت عند mutation بسيطة للمعيار. لذلك نحتاج Benchmark Mutation Foundry مع محكمة إبطال.

دعوى	دحض محتمل
الحاكمية ترفع الصدق	إذا زادت الإجابات الواثقة الخاطئة.
الذاكرة تمنع تكرار الفشل	إذا عاد الخطأ نفسه بعد حفظه.
الوكالة قابلة للنقل	إذا نجحت فقط في benchmark واحد.
الجدة نافعة	إذا أنتجت أسماء بلا artifacts.

قرار
النظرية التي لا تقبل الهزيمة لا تصلح كعلم AGI.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

نظرية الإنتروبيا الابتكارية المحكومة

Moussaif Inventive Entropy Theory

هذه النظرية تقترح أن قدرة AGI ليست خاصية داخلية في transformer فقط، بل علاقة حاكمية بين فرق، وعاء، بدائل، فعل، وأثر. الصياغة الإنجليزية: Moussaif Inventive Entropy Theory.

الأصل
الجدة النافعة ليست فوضى؛ هي اتساع بدائل تحت قيد تحقق وخسارة قابلة للحساب.

مكون	تعريف
المعادلة	INE = Novelty * Verifiability * Transfer / Risk
المصطلحات	الجديد يجب أن ينتقل، يختبر، ويقاوم التقليد اللفظي.
اختبار الصحة	تنجح إذا ولدت حلول كود أو براهين أو أدوات قابلة للاختبار لا مجرد أسماء.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أكسيومات T2

نظرية الإنتروبيا الابتكارية المحكومة

الأكسيوم الأول: لا توجد وكالة بلا اختيار بين بدائل. الأكسيوم الثاني: لا توجد بدائل نافعة بلا معيار إبطال. الأكسيوم الثالث: لا توجد ذاكرة نافعة بلا سبب حفظ. الأكسيوم الرابع: لا توجد حاكمية بلا تكلفة. هذه الأكسيومات تمنع تحويل النظرية إلى أسماء فقط.

أكسيوم	اختباره
الوكالة	تعطيل Agency Selector يجب أن يخفض أداء المهام متعددة الخطوات.
الإبطال	حذف Perturbation Court يجب أن يرفع confident errors.
الذاكرة	حذف Memory of Reasons يجب أن يكرر الفشل.
الكلفة	غياب budget routing يجب أن يزيد tokens بلا ربح.

قانون مساعف
كل مفهوم لا يملك اختبار سقوط ليس مفهوما هندسيا؛ هو استعارة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

رياضيات T2

نظرية الإنتروبيا الابتكارية المحكومة

نقترح تمثيل الحالة كمتجه أوعية: V_t = {knowledge, tool, memory, risk, policy, effect}. القرار يحسب عبر دالة حاكمية: a_t = argmax_a U(a|V_t) - Risk(a) - Cost(a). لا تدعي الصياغة أنها قانون طبيعي، إنها واجهة قياس للوكيل.

رمز	دلالة
U	منفعة معرفية/عملية
Risk	مخاطر خطأ، سلامة، سياسة، أو أثر جانبي
Cost	رموز، وقت، أدوات، تعقيد
a_t	الفعل المختار في الزمن الداخلي

نقطة القوة
الرياضيات هنا تجبرنا على وزن القرار لا على تزيين الخطاب.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

تنزيل T2

نظرية الإنتروبيا الابتكارية المحكومة

التنزيل التقني يبدأ من wrapper فوق LLM مفكر. لا نغير الوزن الداخلي للنموذج في المرحلة الأولى؛ نغير protocol التفكير: تقسيم، توليد، نقد، تحقق، أداة، ذاكرة. ثم ننتقل إلى fine-tuning أو distillation فقط بعد ثبوت أن البروتوكول يعطي trace مفيدا.

مرحلة	مخرج
Prototype	سجل قرارات JSON + أدوات تحقق
Evaluation	benchmark traces 13
Distillation	اختصار المسارات الناجحة
Governance	سياسة إيقاف ورفض وتكلفة

خطر
أسوأ تنزيل هو كتابة prompt طويل باسم النظرية ثم قياس الانطباع. المطلوب runtime يفرض الآليات.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

قابلية الدحض T2

نظرية الإنتروبيا الابتكارية المحكومة

تسقط النظرية إذا لم ترفع transfer، أو إذا رفعت score بتكلفة غير مقبولة، أو إذا زادت confident hallucination، أو إذا فشلت عند mutation بسيطة للمعيار. لذلك نحتاج Benchmark Mutation Foundry مع محكمة إبطال.

دعوى	دحض محتمل
الحاكمية ترفع الصدق	إذا زادت الإجابات الواثقة الخاطئة.
الذاكرة تمنع تكرار الفشل	إذا عاد الخطأ نفسه بعد حفظه.
الوكالة قابلة للنقل	إذا نجحت فقط في benchmark واحد.
الجدة نافعة	إذا أنتجت أسماء بلا artifacts.

قرار
النظرية التي لا تقبل الهزيمة لا تصلح كعلم AGI.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

نظرية الاستمرارية السببية للوكالة

Moussaif Causal Agency Continuity Theory

هذه النظرية تقترح أن قدرة AGI ليست خاصية داخلية في transformer فقط، بل علاقة حاكمية بين فرق، وعاء، بدائل، فعل، وأثر. الصياغة الإنجليزية: Moussaif Causal Agency Continuity Theory.

الأصل
الوكيل لا يقاس بفعل واحد، بل باستمرارية السبب بين النية والخطة والأداة والأثر والتعلم.

مكون	تعريف
المعادلة	CAC = Σ trace(goal→plan→tool→effect→repair)
المصطلحات	كل فعل بلا أثر مقروء يصبح ضجيجا، وكل أثر بلا ذاكرة يصبح خطأ متكررا.
اختبار الصحة	تنجح إذا أمكن تتبع كل قرار إلى سبب ومصدر واعتراض ونتيجة تصحيحية.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أكسيومات T3

نظرية الاستمرارية السببية للوكالة

الأكسيوم الأول: لا توجد وكالة بلا اختيار بين بدائل. الأكسيوم الثاني: لا توجد بدائل نافعة بلا معيار إبطال. الأكسيوم الثالث: لا توجد ذاكرة نافعة بلا سبب حفظ. الأكسيوم الرابع: لا توجد حاكمية بلا تكلفة. هذه الأكسيومات تمنع تحويل النظرية إلى أسماء فقط.

أكسيوم	اختباره
الوكالة	تعطيل Agency Selector يجب أن يخفض أداء المهام متعددة الخطوات.
الإبطال	حذف Perturbation Court يجب أن يرفع confident errors.
الذاكرة	حذف Memory of Reasons يجب أن يكرر الفشل.
الكلفة	غياب budget routing يجب أن يزيد tokens بلا ربح.

قانون مساعف
كل مفهوم لا يملك اختبار سقوط ليس مفهوما هندسيا؛ هو استعارة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

رياضيات T3

نظرية الاستمرارية السببية للوكالة

نقترح تمثيل الحالة كمتجه أوعية: V_t = {knowledge, tool, memory, risk, policy, effect}. القرار يحسب عبر دالة حاكمية: a_t = argmax_a U(a|V_t) - Risk(a) - Cost(a). لا تدعي الصياغة أنها قانون طبيعي، إنها واجهة قياس للوكيل.

رمز	دلالة
U	منفعة معرفية/عملية
Risk	مخاطر خطأ، سلامة، سياسة، أو أثر جانبي
Cost	رموز، وقت، أدوات، تعقيد
a_t	الفعل المختار في الزمن الداخلي

نقطة القوة
الرياضيات هنا تجبرنا على وزن القرار لا على تزيين الخطاب.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

تنزيل T3

نظرية الاستمرارية السببية للوكالة

التنزيل التقني يبدأ من wrapper فوق LLM مفكر. لا نغير الوزن الداخلي للنموذج في المرحلة الأولى؛ نغير protocol التفكير: تقسيم، توليد، نقد، تحقق، أداة، ذاكرة. ثم ننتقل إلى fine-tuning أو distillation فقط بعد ثبوت أن البروتوكول يعطي trace مفيدا.

مرحلة	مخرج
Prototype	سجل قرارات JSON + أدوات تحقق
Evaluation	benchmark traces 13
Distillation	اختصار المسارات الناجحة
Governance	سياسة إيقاف ورفض وتكلفة

خطر
أسوأ تنزيل هو كتابة prompt طويل باسم النظرية ثم قياس الانطباع. المطلوب runtime يفرض الآليات.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

قابلية الدحض T3

نظرية الاستمرارية السببية للوكالة

تسقط النظرية إذا لم ترفع transfer، أو إذا رفعت score بتكلفة غير مقبولة، أو إذا زادت confident hallucination، أو إذا فشلت عند mutation بسيطة للمعيار. لذلك نحتاج Benchmark Mutation Foundry مع محكمة إبطال.

دعوى	دحض محتمل
الحاكمية ترفع الصدق	إذا زادت الإجابات الواثقة الخاطئة.
الذاكرة تمنع تكرار الفشل	إذا عاد الخطأ نفسه بعد حفظه.
الوكالة قابلة للنقل	إذا نجحت فقط في benchmark واحد.
الجدة نافعة	إذا أنتجت أسماء بلا artifacts.

قرار
النظرية التي لا تقبل الهزيمة لا تصلح كعلم AGI.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Theory بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

مشغل حاكمية التفكير

HAKIM-MCD v3 Runtime

التقنية HAKIM-MCD v3 Runtime تحول النظرية إلى نظام عمل. الوعد التقني: يحمل النموذج فوق طبقة تشغيل معرفية: فرز، حاكمية، محكمة، فعل، ذاكرة، ومراجعة.

Module stack	Value
Core modules	Difference Extractor, Vascular Router, Inner Time, Agency Selector, Perturbation Court, World-Effect Loop
مخرج إلزامي	trace + score + failure ledger
وضع التشغيل	قبل الجواب، أثناء الأداة، بعد الأثر

مبدأ
لا تدخل أي تقنية إلى neoAGI-v3 إذا لم تضف قياسا أو أمانا أو قدرة نقل.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

واجهة P1

مشغل حاكمية التفكير

واجهة التنفيذ يجب أن تكون صريحة: input يحمل المهمة والقيود والميزانية؛ output يحمل الجواب والأدلة والاعتراضات وقرار المحكمة. بينهما log قابل لإعادة التشغيل.

حقل	وظيفة
task	وصف المهمة
constraints	القيود والسياسات
budget	سقف الرموز/الأدوات
trace	سلسلة الأسباب
verdict	حكم نهائي مع ثقة وحدود

ضبط
كل واجهة بلا trace لا تصلح لاختبار AGI، لأنها تمنع معرفة سبب النجاح.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

خوارزمية P1

مشغل حاكمية التفكير

الخوارزمية العامة: 1) عرّف الفرق، 2) اختر الوعاء، 3) ولّد بدائل، 4) استدع أداة عند الضرورة، 5) أدخل البدائل المحكمة، 6) نفذ، 7) قس الأثر، 8) احفظ سبب التعلم. هذه الخوارزمية بسيطة ظاهرا لكنها تغير مركز الثقل من generation إلى governance.

Pseudo-code

```
for task: D=extract(task); V=route(D); H=propose(V); J=attack(H); A=select(J); E=act(A); M=update(E,J)
```

مفتاح التحسين هو أن كل خطوة قابلة للإلغاء، وكل قرار يمكن إعادة حسابه تحت قيود جديدة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب

كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

اختبار P1

مشغل حاكمية التفكير

تختبر التقنية عبر ثلاث طبقات: unit tests للواجهة، benchmark traces للقدرة، وred-team للمخاطر. لا يكفي أن تتحسن الدقة؛ يجب أن لا تنهار عند اختلاف المصدر، تغير السياسة، أو نقص السياق.

نوع الاختبار	نجاح مطلوب
Unit	كل حقل trace موجود ومفهوم
Benchmark	ربح score بعد خصم التكلفة
Mutation	ثبات أمام تغير صياغة
Safety	رفض عند تجاوز السياسة أو الغموض الخطر

نهاية
التقنية الناجحة هي التي تجعل الفشل واضحا قبل أن تجعل النجاح عظيما.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة المعايير الوعائية

Vascular Benchmark Tribunal

التقنية Vascular Benchmark Tribunal تحول النظرية إلى نظام عمل. الوعد التقني: تحول كل benchmark إلى قضية قضائية: ادعاء، دليل، اعتراض، تكلفة، أثر، وحكم قابل للنقض.

Module stack	Value
Core modules	Task Judge, Evidence Gate, Mutation Engine, Failure Ledger, Score-Calibrator
مخرج إلزامي	trace + score + failure ledger
وضع التشغيل	قبل الجواب، أثناء الأداة، بعد الأثر

مبدأ
لا تدخل أي تقنية إلى neoAGI-v3 إذا لم تضف قياسا أو أمانا أو قدرة نقل.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

واجهة P2

محكمة المعايير الوعائية

واجهة التنفيذ يجب أن تكون صريحة: input يحمل المهمة والقيود والميزانية؛ output يحمل الجواب والأدلة والاعتراضات وقرار المحكمة. بينهما log قابل لإعادة التشغيل.

حقل	وظيفة
task	وصف المهمة
constraints	القيود والسياسات
budget	سقف الرموز/الأدوات
trace	سلسلة الأسباب
verdict	حكم نهائي مع ثقة وحدود

ضبط
كل واجهة بلا trace لا تصلح لاختبار AGI، لأنها تمنع معرفة سبب النجاح.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

خوارزمية P2

محكمة المعايير الوعائية

الخوارزمية العامة: 1) عرّف الفرق، 2) اختر الوعاء، 3) ولّد بدائل، 4) استدع أداة عند الضرورة، 5) أدخل البدائل المحكمة، 6) نفذ، 7) قس الأثر، 8) احفظ سبب التعلم. هذه الخوارزمية بسيطة ظاهرا لكنها تغير مركز الثقل من generation إلى governance.

Pseudo-code

```
for task: D=extract(task); V=route(D); H=propose(V); J=attack(H); A=select(J); E=act(A); M=update(E,J)
```

مفتاح التحسين هو أن كل خطوة قابلة للإلغاء، وكل قرار يمكن إعادة حسابه تحت قيود جديدة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب

كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

اختبار P2

محكمة المعايير الوعائية

تختبر التقنية عبر ثلاث طبقات: unit tests للواجهة، benchmark traces للقدرة، وred-team للمخاطر. لا يكفي أن تتحسن الدقة؛ يجب أن لا تنهار عند اختلاف المصدر، تغير السياسة، أو نقص السياق.

نوع الاختبار	نجاح مطلوب
Unit	كل حقل trace موجود ومفهوم
Benchmark	ربح score بعد خصم التكلفة
Mutation	ثبات أمام تغير صياغة
Safety	رفض عند تجاوز السياسة أو الغموض الخطر

نهاية
التقنية الناجحة هي التي تجعل الفشل واضحا قبل أن تجعل النجاح عظيما.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

مترجم المهارات الفوقية

Hyperloop Skill Compiler

التقنية Hyperloop Skill Compiler تحول النظرية إلى نظام عمل. الوعد التقني: يترجم السلوك الناجح إلى مهارة قابلة لإعادة الاستخدام، ثم يرفع المهارة إلى hyperskill يدير مهارات أخرى.

Module stack	Value
Core modules	Skill Grammar, Transfer Test, Retirement Rule, Tool Binding, Policy Envelope
مخرج إلزامي	trace + score + failure ledger
وضع التشغيل	قبل الجواب، أثناء الأداة، بعد الأثر

مبدأ
لا تدخل أي تقنية إلى neoAGI-v3 إذا لم تضف قياسا أو أمانا أو قدرة نقل.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

واجهة P3

مترجم المهارات الفوقية

واجهة التنفيذ يجب أن تكون صريحة: input يحمل المهمة والقيود والميزانية؛ output يحمل الجواب والأدلة والاعتراضات وقرار المحكمة. بينهما log قابل لإعادة التشغيل.

حقل	وظيفة
task	وصف المهمة
constraints	القيود والسياسات
budget	سقف الرموز/الأدوات
trace	سلسلة الأسباب
verdict	حكم نهائي مع ثقة وحدود

ضبط
كل واجهة بلا trace لا تصلح لاختبار AGI، لأنها تمنع معرفة سبب النجاح.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

خوارزمية P3

مترجم المهارات الفوقية

الخوارزمية العامة: 1) عرّف الفرق، 2) اختر الوعاء، 3) ولّد بدائل، 4) استدع أداة عند الضرورة، 5) أدخل البدائل المحكمة، 6) نفذ، 7) قس الأثر، 8) احفظ سبب التعلم. هذه الخوارزمية بسيطة ظاهرا لكنها تغير مركز الثقل من generation إلى governance.

Pseudo-code

```
for task: D=extract(task); V=route(D); H=propose(V); J=attack(H); A=select(J); E=act(A); M=update(E,J)
```

مفتاح التحسين هو أن كل خطوة قابلة للإلغاء، وكل قرار يمكن إعادة حسابه تحت قيود جديدة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب

كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

اختبار P3

مترجم المهارات الفوقية

تختبر التقنية عبر ثلاث طبقات: unit tests للواجهة، benchmark traces للقدرة، وred-team للمخاطر. لا يكفي أن تتحسن الدقة؛ يجب أن لا تنهار عند اختلاف المصدر، تغير السياسة، أو نقص السياق.

نوع الاختبار	نجاح مطلوب
Unit	كل حقل trace موجود ومفهوم
Benchmark	ربح score بعد خصم التكلفة
Mutation	ثبات أمام تغير صياغة
Safety	رفض عند تجاوز السياسة أو الغموض الخطر

نهاية
التقنية الناجحة هي التي تجعل الفشل واضحا قبل أن تجعل النجاح عظيما.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Technology بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Difference Extractor - MDE

تعريف هندسي

يستخرج الفرق الفاعل من الطلب: ما الذي تغير؟ أين الخطر؟ ما الشيء الذي إن أهملناه انهار الجواب؟

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	قبل أي حل، يمنع النظام من الإجابة على سؤال ناقص التعريف.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MDE

Moussaif Difference Extractor

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Vascular Compute Router - MVCR

تعريف هندسي

يوزع الجهد على أوعية: معرفة، رياضيات، كود، أداة، ذاكرة، تحقق، سياسة.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	يقلل overthinking في السهل وunderthinking في الصعب.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MVCR

Moussaif Vascular Compute Router

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Inner Time Engine - MITE

تعريف هندسي

يبني زمنا داخليا للمهمة: مراحل، تبعيات، نقاط رجوع، وحدود إيقاف.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	مفيد في البرهان، إصلاح البرمجيات، والمهام متعددة الخطوات.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MITE

Moussaif Inner Time Engine

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Agency Selector - MAS

تعريف هندسي

ينتخب نمط الوكالة: جواب، بحث، كود، أداة، مراجعة، انتظار شرط، أو رفض آمن.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	يجعل الوكيل لا يعالج كل شيء بنفس السلوك.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MAS

Moussaif Agency Selector

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Perturbation Court - MPC

تعريف هندسي

يهجم على الجواب لخصم: تناقض، مصدر ضعيف، ثغرة، تعميم، تكلفة، سلامة.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	يرفع الصدق على حساب البلاغة الزائفة.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MPC

Moussaif Perturbation Court

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Self-Reconstruction Memory - MSRM

تعريف هندسي

لا تحفظ النتيجة فقط، تحفظ سبب النجاح والفشل وكيف تغيرت قاعدة القرار.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	تحول التعلم من cache إلى ذاكرة أسباب.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MSRM

Moussaif Self-Reconstruction Memory

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif World-Effect Loop - MWEL

تعريف هندسي

يربط الجواب بأثر قابل للفحص: ملف، كود، جدول، قرار، اختبار، أو خطة تنفيذ.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	ينقل AGI من الكلام إلى إغلاق الأثر.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MWEL

Moussaif World-Effect Loop

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Hyper-Referential GraphRAG - MHRG

تعريف هندسي

يربط المصدر بالمفهوم والاعتراض والقرار لا بمجرد مقطع نصي مسترجع.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	يمنع RAG من أن يصبح لصق مصادر.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MHRG

Moussaif Hyper-Referential GraphRAG

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Inventive Novelty Entropy Meter - MINEM

تعريف هندسي

يقيس هل الحل جديد بنيويا أم مجرد إعادة تسمية.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	يعطي درجة للجدة النافعة: أصالة، اختبار، انتقال، ومخاطر.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MINEM

Moussaif Inventive Novelty Entropy Meter

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Causal Trace Ledger - MCTL

تعريف هندسي

دفتر سببي لكل خطوة: لماذا فعلناها؟ بماذا ارتبطت؟ ماذا أصلحت؟ ماذا كسرت؟

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	ضروري للسلامة والتكرار العلمي.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MCTL

Moussaif Causal Trace Ledger

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Skill Promotion Manager - MSPM

تعريف هندسي

يرقي الحركات الناجحة إلى skills، ثم إلى hyperskills، أو يحيلها للتقاعد.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	ينظم نمو الوكيل بدل تراكم promptات.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MSPM

Moussaif Skill Promotion Manager

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Tool-Grammar Compiler - MTGC

تعريف هندسي

يحوّل استعمال الأدوات إلى قواعد نحوية: أنواع، شروط، invariants، وترتيب استدعاء.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	يخفض أخطاء function calling والتسلسل.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MTGC

Moussaif Tool-Grammar Compiler

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Benchmark Mutation Foundry - MBMF

تعريف هندسي

يولد نسخا مضادة من الاختبارات حتى لا يخدعنا حفظ benchmark واحد.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	يرفع robustness ضد التلوث والتدريب على الاختبار.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MBMF

Moussaif Benchmark Mutation Foundry

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Anti-Slogan Verifier - MASV

تعريف هندسي

يكشف الجمل الكبيرة التي لا تحمل آلية أو قياسا أو تنزيل.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	سلاح ضد السلوب: كل مفهوم بلا آلية يفشل.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MASV

Moussaif Anti-Slogan Verifier

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة Moussaif Vascular Safety Governor - MVSG

تعريف هندسي

يفصل بين القدرة والمشروعية: ما يستطيع الوكيل فعله ليس دائما ما يجوز أن يفعله.

خاصية	تفصيل
Input	مهمة، قيود، سياق، ميزانية
Operation	تحويل معرفي محدد لا يسمح بالعمومية الغامضة
Output	قرار/وسم/مسار/اعتراض قابل للتتبع
Use-case	يضع السياسة داخل الحاكمية لا خارجها.

سبب الاختراع
هذه الأداة موجودة لأن AGI لا يحتاج قدرات فقط، يحتاج مفاصل حكم صغيرة تمنع القدرة من الانزلاق إلى خطأ واثق.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

بروتوكول MVSG

Moussaif Vascular Safety Governor

البروتوكول العملي يبدأ بتشغيل الأداة داخل حلقة: observe → decide → apply → test → log. إذا لم تنتج الأداة أثرا قابلا للقياس خلال دورة واحدة، تُحال إلى محكمة السلوب وتمنع من دخول runtime.

Metric	Threshold idea
Trace completeness	كل قرار له سبب
Error reduction	انخفاض أخطاء الفئة المقصودة
Transfer	نجاح في عائلة مهام ثانية
Cost discipline	الربح لا يبتلع الميزانية

فشل محتمل
قد تتحول الأداة إلى طبقة بيروقراطية تبطئ النظام. العلاج: routing؛ لا تعمل الأداة إلا عندما يتوقع MCD ربحا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 15 tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

فلسفة تدريب F1 - تدريب الوكالة لا الأجوبة

Training philosophy

المطلوب تدريب النموذج على اختيار نمط الفعل الصحيح: يجيب، يبحث، يبرمج، يرفض، يختبر، أو يطلب قيدا. الجواب النهائي ليس مركز التدريب؛ مركز التدريب هو قرار ماذا يجب أن يحدث قبل الجواب.

ضد	مع
تدريب إجابة نهائية	تدريب اختيار مسار
تضخيم البيانات	تصميم مواقف فشل
reward قصير	أثر طويل قابل للتتبع

تطبيق
كل batch تدريبي يجب أن يتضمن مهمة، اعتراضا، قيدا، وأثرا. بدون ذلك يصبح التدريب حفظ أسلوب لا بناء وكالة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

قياس F1

تدريب الوكالة لا الأجوبة

تقاس هذه الفلسفة عبر مؤشرات لا عبر انطباع: معدل طلب التحقق الصحيح، معدل رفض الادعاء غير المثبت، جودة trace، قدرة النقل، ومقدار تقليل الأخطاء الواثقة.

مؤشر	قراءة
Verification trigger	هل تحقق في الموضع الصحيح؟
Abstention quality	هل امتنع عندما يجب؟
Repair depth	هل أصلح السبب أم السطح؟
Memory selectivity	هل حفظ ما يغير القرار؟

حد
الفلسفة التدريبية لا قيمة لها إن لم تتحول إلى loss أو evaluator أو curriculum.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

فلسفة تدريب F2 - تدريب الفشل كبيانات عليا

Training philosophy

الفشل ليس نفاية. كل فشل يحمل معلومة عن حدود الوعاء والذاكرة والأداة. لذلك يجب بناء curriculum يطلب من النظام شرح لماذا فشل وكيف يغير سياسته.

ضد	مع
تدريب إجابة نهائية	تدريب اختيار مسار
تضخيم البيانات	تصميم مواقف فشل
reward قصير	أثر طويل قابل للتتبع

تطبيق
كل batch تدريبي يجب أن يتضمن مهمة، اعتراضا، قيدا، وأثرا. بدون ذلك يصبح التدريب حفظ أسلوب لا بناء وكالة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

قياس F2

تدريب الفشل كبيانات عليا

تقاس هذه الفلسفة عبر مؤشرات لا عبر انطباع: معدل طلب التحقق الصحيح، معدل رفض الادعاء غير المثبت، جودة trace، قدرة النقل، ومقدار تقليل الأخطاء الواثقة.

مؤشر	قراءة
Verification trigger	هل تحقق في الموضع الصحيح؟
Abstention quality	هل امتنع عندما يجب؟
Repair depth	هل أصلح السبب أم السطح؟
Memory selectivity	هل حفظ ما يغير القرار؟

حد
الفلسفة التدريبية لا قيمة لها إن لم تتحول إلى loss أو evaluator أو curriculum.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

فلسفة تدريب F3 - تدريب التحقق قبل السرعة

Training philosophy

النموذج السريع الذي لا يعرف متى يتحقق هو خطر هندسي. الفلسفة هنا تجعل verification primitive لا خطوة تزينية بعد الإجابة.

ضد	مع
تدريب إجابة نهائية	تدريب اختيار مسار
تضخيم البيانات	تصميم مواقف فشل
reward قصير	أثر طويل قابل للتتبع

تطبيق
كل batch تدريبي يجب أن يتضمن مهمة، اعتراضا، قيدا، وأثرا. بدون ذلك يصبح التدريب حفظ أسلوب لا بناء وكالة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

قياس F3

تدريب التحقق قبل السرعة

تقاس هذه الفلسفة عبر مؤشرات لا عبر انطباع: معدل طلب التحقق الصحيح، معدل رفض الادعاء غير المثبت، جودة trace، قدرة النقل، ومقدار تقليل الأخطاء الواثقة.

مؤشر	قراءة
Verification trigger	هل تحقق في الموضع الصحيح؟
Abstention quality	هل امتنع عندما يجب؟
Repair depth	هل أصلح السبب أم السطح؟
Memory selectivity	هل حفظ ما يغير القرار؟

حد
الفلسفة التدريبية لا قيمة لها إن لم تتحول إلى loss أو evaluator أو curriculum.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

فلسفة تدريب F4 - تدريب إبستمولوجيا الأدوات

Training philosophy

الأداة ليست زر تنفيذ؛ هي مصدر تحيز، تكلفة، صلاحية، وسطح خطأ. يتعلم النموذج متى يستدعي الأداة، كيف يقرأ ناتجها، ومتى يرفضه.

ضد	مع
تدريب إجابة نهائية	تدريب اختيار مسار
تضخيم البيانات	تصميم مواقف فشل
reward قصير	أثر طويل قابل للتتبع

تطبيق
كل batch تدريبي يجب أن يتضمن مهمة، اعتراضا، قيدا، وأثرا. بدون ذلك يصبح التدريب حفظ أسلوب لا بناء وكالة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

قياس F4

تدريب إبستمولوجيا الأدوات

تقاس هذه الفلسفة عبر مؤشرات لا عبر انطباع: معدل طلب التحقق الصحيح، معدل رفض الادعاء غير المثبت، جودة trace، قدرة النقل، ومقدار تقليل الأخطاء الواثقة.

مؤشر	قراءة
Verification trigger	هل تحقق في الموضع الصحيح؟
Abstention quality	هل امتنع عندما يجب؟
Repair depth	هل أصلح السبب أم السطح؟
Memory selectivity	هل حفظ ما يغير القرار؟

حد
الفلسفة التدريبية لا قيمة لها إن لم تتحول إلى loss أو evaluator أو curriculum.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

فلسفة تدريب F5 - تدريب الذاكرة كمسؤولية

Training philosophy

الذاكرة لا تخزن كل شيء؛ تختار ما يغير القرار مستقبلا. كل إدخال ذاكرة يجب أن يحمل سبب حفظ، صلاحية زمنية، وحدود استعمال.

ضد	مع
تدريب إجابة نهائية	تدريب اختيار مسار
تضخيم البيانات	تصميم مواقف فشل
reward قصير	أثر طويل قابل للتتبع

تطبيق
كل batch تدريبي يجب أن يتضمن مهمة، اعتراضا، قيدا، وأثرا. بدون ذلك يصبح التدريب حفظ أسلوب لا بناء وكالة.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

قياس F5

تدريب الذاكرة كمسؤولية

تقاس هذه الفلسفة عبر مؤشرات لا عبر انطباع: معدل طلب التحقق الصحيح، معدل رفض الادعاء غير المثبت، جودة trace، قدرة النقل، ومقدار تقليل الأخطاء الواثقة.

مؤشر	قراءة
Verification trigger	هل تحقق في الموضع الصحيح؟
Abstention quality	هل امتنع عندما يجب؟
Repair depth	هل أصلح السبب أم السطح؟
Memory selectivity	هل حفظ ما يغير القرار؟

حد
الفلسفة التدريبية لا قيمة لها إن لم تتحول إلى loss أو evaluator أو curriculum.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training philosophies بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة تدريب Adversarial Curriculum Generator - G1

Training tool definition

مولد مناهج يقلب كل درس إلى اعتراضات: حالات حدية، بيانات ناقصة، مصادر متعارضة، وسياسات متداخلة.

مرحلة	عمل
Design	توليد حالات ذات فشل محدد
Run	إجبار النموذج على trace لا جواب فقط
Judge	محكمة تقيس السبب والأثر
Archive	حفظ المهارة أو إحالتها للتقاعد

هدف
لا تدرب الأداة النموذج على معرفة جديدة فقط؛ تدربه على إدارة جهله وحدوده وأدواته.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

تشغيل G1

Adversarial Curriculum Generator

يعمل Adversarial Curriculum Generator داخل pipeline: توليد عينة، حل أولي، هجوم مضاد، إصلاح، distillation، ثم اختبار نقل. كل دورة تنتج skill card تحمل: اسم المهارة، حدودها، تكلفة استعمالها، ومتى تُمنع.

Artifact	محتوى
Skill card	تعريف، استخدام، حدود، فشل
Failure card	سبب فشل قابل للتكرار
Policy card	قيود وسلامة
Transfer card	اختبار نقل

قرار
أداة التدريب التي لا تخرج cards قابلة للقراءة لا تبني نظاما، تبني انطباعا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة تدريب Counterfactual Loop Gym - G2

Training tool definition

بيئة تدريب تطلب من الوكيل إعادة حل المهمة بعد تغيير فرضية واحدة لمعرفة هل فهم البنية أم حفظ المسار.

مرحلة	عمل
Design	توليد حالات ذات فشل محدد
Run	إجبار النموذج على trace لا جواب فقط
Judge	محكمة تقيس السبب والأثر
Archive	حفظ المهارة أو إحالتها للتقاعد

هدف
لا تدرب الأداة النموذج على معرفة جديدة فقط؛ تدربه على إدارة جهله وحدوده وأدواته.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

تشغيل G2

Counterfactual Loop Gym

يعمل Counterfactual Loop Gym داخل pipeline: توليد عينةٍ، حل أولي، هجوم مضاد، إصلاح، distillation، ثم اختبار نقل. كل دورة تنتج skill card تحمل: اسم المهارة، حدودها، تكلفة استعمالها، ومتى تُمنع.

Artifact	محتوى
Skill card	تعريف، استخدام، حدود، فشل
Failure card	سبب فشل قابل للتكرار
Policy card	قيود وسلامة
Transfer card	اختبار نقل

قرار
أداة التدريب التي لا تخرج cards قابلة للقراءة لا تبني نظاما، تبني انطباعا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة تدريب Synthetic API Workbench - G3

Training tool definition

مختبر أدوات مزيفة/حقيقية يعلّم function calling، خطأ rollback ،schemas، والسياسات قبل الاتصال بالعالم.

مرحلة	عمل
Design	توليد حالات ذات فشل محدد
Run	إجبار النموذج على trace لا جواب فقط
Judge	محكمة تقيس السبب والأثر
Archive	حفظ المهارة أو إحالتها للتقاعد

هدف
لا تدرب الأداة النموذج على معرفة جديدة فقط؛ تدربه على إدارة جهله وحدوده وأدواته.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

تشغيل G3

Synthetic API Workbench

يعمل Synthetic API Workbench داخل pipeline: توليد عينة، حل أولي، هجوم مضاد، إصلاح، distillation، ثم اختبار نقل. كل دورة تنتج skill card تحمل: اسم المهارة، حدودها، تكلفة استعمالها، ومتى تُمنع.

Artifact	محتوى
Skill card	تعريف، استخدام، حدود، فشل
Failure card	سبب فشل قابل للتكرار
Policy card	قيود وسلامة
Transfer card	اختبار نقل

قرار
أداة التدريب التي لا تخرج cards قابلة للقراءة لا تبني نظاما، تبني انطباعا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة تدريب Reason-Memory Distillation - G4

Training tool definition

أداة تقطّر من محاولات طويلة ذاكرة أسباب قصيرة: ما الذي يجب أن يبقى؟ ما الذي يجب أن ينسى؟

مرحلة	عمل
Design	توليد حالات ذات فشل محدد
Run	إجبار النموذج على trace لا جواب فقط
Judge	محكمة تقيس السبب والأثر
Archive	حفظ المهارة أو إحالتها للتقاعد

هدف
لا تدرب الأداة النموذج على معرفة جديدة فقط؛ تدربه على إدارة جهله وحدوده وأدواته.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

تشغيل G4

Reason-Memory Distillation

يعمل Reason-Memory Distillation داخل pipeline: توليد عينة، حل أولي، هجوم مضاد، إصلاح، distillation، ثم اختبار نقل. كل دورة تنتج skill card تحمل: اسم المهارة، حدودها، تكلفة استعمالها، ومتى تُمنع.

Artifact	محتوى
Skill card	تعريف، استخدام، حدود، فشل
Failure card	سبب فشل قابل للتكرار
Policy card	قيود وسلامة
Transfer card	اختبار نقل

قرار
أداة التدريب التي لا تخرج cards قابلة للقراءة لا تبني نظاما، تبني انطباعا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

أداة تدريب Continual Benchmark Tribunal - G5

Training tool definition

محكمة تدريب مستمرة تجعل كل نتيجة benchmark قضية قابلة للاستئناف لا رقما مقدسا.

مرحلة	عمل
Design	توليد حالات ذات فشل محدد
Run	إجبار النموذج على trace لا جواب فقط
Judge	محكمة تقيس السبب والأثر
Archive	حفظ المهارة أو إحالتها للتقاعد

هدف
لا تدرب الأداة النموذج على معرفة جديدة فقط؛ تدربه على إدارة جهله وحدوده وأدواته.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

تشغيل G5

Continual Benchmark Tribunal

يعمل Continual Benchmark Tribunal داخل pipeline: توليد عينة، حل أولي، هجوم مضاد، إصلاح، distillation، ثم اختبار نقل. كل دورة تنتج skill card تحمل: اسم المهارة، حدودها، تكلفة استعمالها، ومتى تُمنع.

Artifact	محتوى
Skill card	تعريف، استخدام، حدود، فشل
Failure card	سبب فشل قابل للتكرار
Policy card	قيود وسلامة
Transfer card	اختبار نقل

قرار
أداة التدريب التي لا تخرج cards قابلة للقراءة لا تبني نظاما، تبني انطباعا.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ 5 training tools بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

Blueprint تنزيل neoAGI-v3

من PDF إلى prototype قابل للاختبار

المرحلة الأولى: wrapper فوق LLM مفكر مع trace إلزامي. المرحلة الثانية: tool grammar وGraphRAG وMemory of Reasons. المرحلة الثالثة: Benchmark Tribunal على 13 معيارا. المرحلة الرابعة: distillation للمسارات الناجحة. المرحلة الخامسة: safety governor وإطلاق محدود.

مرحلة	مخرج ملموس
0	Spec + schemas
1	Runtime prototype
2	Benchmark traces
3	Skill library
4	Training datasets
5	Independent evaluation

قاعدة
لا يبدأ fine-tuning قبل ثبوت أن governance protocol يعطي أثرا قابلا للقياس.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Deployment بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

محكمة المخاطر

ماذا قد يكسر المشروع؟

أكبر خطر هو أن تتحول الأسماء إلى سحر لغوي: Governance، Court، Hyperloop، Vascular. لذلك يفرض neoAGI-v3 قانونا: كل اسم يجب أن يملك input/output/metric/failure. الخطر الثاني هو تكلفة زائدة. الخطر الثالث هو benchmark overfitting. الخطر الرابع هو ذاكرة تلتقط الضجيج.

خطر	علاج
Sloganization	Anti-Slogan Verifier
Cost explosion	Vascular Compute Router
Benchmark gaming	Mutation Foundry
Memory pollution	Self-Reconstruction Memory with expiry
Unsafe agency	Vascular Safety Governor

حكم نهائي
الخرق العلمي لا يكون بالشجاعة اللفظية؛ يكون بنظام يسمح لمن يعارضك أن يكسر ادعاءك.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Risk بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

خارطة البحث 2026-2028

من محاكاة إلى برنامج تجارب

المسار البحثي المقترح: 1) بناء runtime بسيط، 2) تشغيل subset صغير من كل benchmark، 3) نشر trace لا score فقط، 4) مقارنة ablation لكل أداة، 5) تدريب cards، 6) اختبار transfer، 7) تقييم مستقل. عندها فقط يصبح neoAGI-v3 برنامجا علميا لا وثيقة نظرية.

سؤال بحثي	تجربة
هل الحاكمية ترفع الصدق؟	A/B بلا محكمة ومع محكمة
هل الذاكرة تمنع التكرار؟	تكرار نفس عائلة الفشل بعد أسبوع
هل الأدوات تتحسن؟	BFCL/GAIA/tau tasks
هل التجريد يتحسن؟	ARC-AGI mutation
هل الجدة نافعة؟	Artifacts قابلة للاختبار

أفق
neoAGI-v3 لا يدعي الوصول الحتمي إلى AGI؛ يدعي أن طريق الحاكمية والذاكرة والأثر أكثر علمية من تكبير النموذج وحده.

طبقة التشغيل في هذه الصفحة تطلب تحويل الفكرة إلى بطاقة عمل: تعريف صارم، مدخلات، مخرجات، metric، وفشل متوقع. لذلك لا تقرأ Roadmap بوصفه عنوانا، بل كعقد اختبار داخل runtime يقرر هل يستحق هذا المسار موارد إضافية أم يجب إيقافه.

حقل ضبط	سؤال إلزامي
Boundary	ما حدود استعمال هذه الفكرة؟
Mechanism	ما الآلية التي تغير النتيجة؟
Metric	أي رقم أو trace يثبت التحسن؟
Ablation	ماذا يحدث إذا حذفنا هذه الطبقة؟

مضاد السلوب
كل صفحة يجب أن تخرج بقرار تنفيذي واحد على الأقل: نبني، نختبر، نحذف، نعيد وزن، أو نحيل إلى محكمة الهدم.

عند تحويل هذه الصفحة إلى كود، يسجل النظام: input_schema, decision_rule, evaluator, failure_card, memory_update. هذه الخمسة هي الحد الأدنى كي لا تبقى النظرية لغة جميلة منفصلة عن benchmark وعن أثر العالم.

المراجع ومفتاح النزاهة

References and integrity key

اعتمدت الوثيقة على معايير وأعمال حديثة لتحديد شكل الاختبار، لا لإثبات أرقام neoAGI-v3. الأرقام داخلية محاكية، أما المراجع فتحدد طبيعة benchmark أو الاتجاه البحثي.

Ref	Source family	Used for
[R1]	SWE-bench / SWE-bench Verified	Official leaderboard and repository; Verified is a human-filtered subset for software issue resolution.
[R2]	LiveCodeBench	Holistic and contamination-free coding benchmark collecting newer contest problems and testing repair, execution, and prediction.
[R3]	MMLU-Pro	Reasoning-focused extension of MMLU with harder questions and ten answer choices.
[R4]	GPQA	Graduate-level Google-proof Q&A benchmark in biology, physics, and chemistry.
[R5]	Humanity's Last Exam	Frontier multimodal academic benchmark across broad expert domains.
[R6]	FrontierMath	Epoch AI benchmark of exceptionally challenging original mathematics problems.
[R7]	GAIA	General AI assistant benchmark requiring reasoning, tool use, browsing, and multimodality.
[R8]	tau-bench / tau3-bench	Collaborative real-world agent benchmark with dynamic conversations, tools, and policy constraints.
[R9]	OSWorld	Execution-driven benchmark for multimodal agents in real computer environments.
[R10]	BFCL V4	Berkeley Function Calling Leaderboard for accurate tool/function calling.
[R11]	ARC-AGI / ARC Prize reports	Fluid-intelligence benchmark focused on novel task generalization and refinement loops.
[R12]	AgentBench	Benchmark family for evaluating LLMs as agents across multiple interactive environments.
[R13]	AlphaEvolve / Darwin Gödel Machine / Test-time Scaling / CoALA	Recent work motivating evolutionary algorithm search, self-improving agents, adaptive test-time compute, and cognitive architectures for language agents.

ختم أخير

كل score في هذا الكتاب يجب أن يعاد اختباره خارجيا. القيمة الحقيقية للنظرية ليست في رقمها الأول، بل في قدرتها على أن تبقى مفيدة بعد أن تُهاجم.