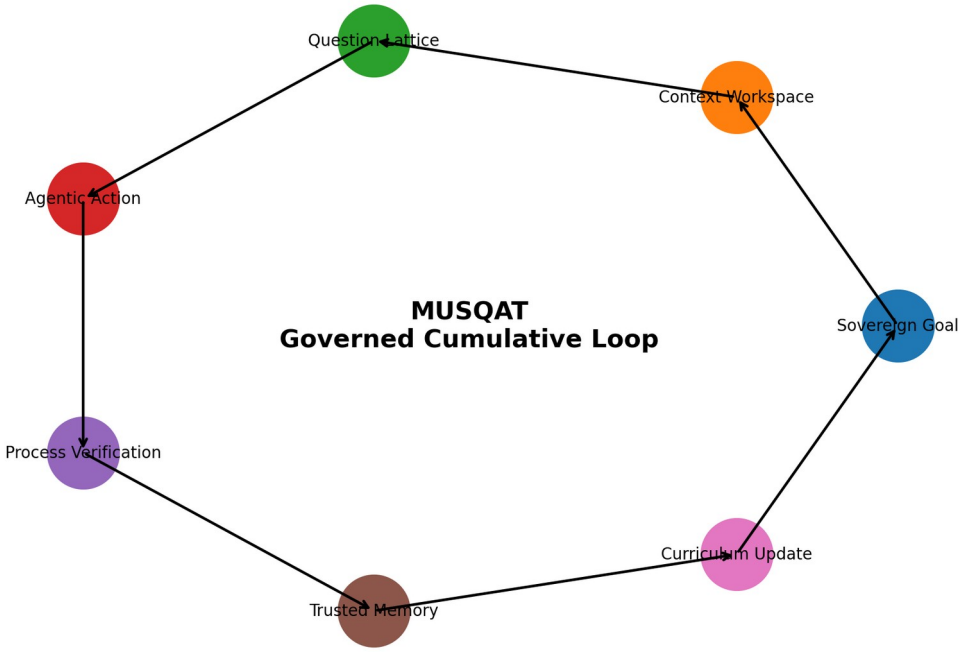
مُسقاة مساعف الحاكمة

MUSQAT-AGI

أطلس التدريب والتطوير الحاكمي السياقي الوثوقي التراكمي السؤالي

لنماذج MOUSSAIF Agentic Unified، MOUSSAIF Unified AGI



10 نظريات مقترحة • 80 أداة تقنية • معادلات • بروتوكولات تدريب • محاكم تحقق • خارطة تجريب

تأليف وتصميم البرنامج البحثي: زهير مساعف / **Zouhair Moussaif**

نسخة تأسيسية بحثية موسعة — يوليو 2026

بيان الوضع العلمي وحدود الادعاء

هذه الوثيقة تقترح برنامجًا بحثيًا وهندسيًا أصليًا لنماذج MOUSSAIF Unified AGI. النظريات والأدوات المسماة هنا ليست نتائج علمية مثبتة لمجرد صياغتها؛ هي فرضيات معمارية وبروتوكولات قابلة للبناء والتفنيد. قوة الأطلس لا تقوم على ادعاء بلوغ AGI، بل على تحويل مفاهيم الحاكمية والسياق والثقة والتراكم والسؤال إلى متغيرات وأدوات واختبارات يمكن تنفيذها ومقارنتها.

كما يُستخدم MOUSSAIF9B في هذه الوثيقة اسمًا معماريًا للمشروع أو الإصدار. إذا كان النموذج الفعلي يملك نحو مليار معلمة، فلا يُفهم الاسم على أنه ادعاء بامتلاك تسعة مليارات معلمة. يصبح الوصف حجميًا فقط بعد بناء نموذج فعلي بذلك الحجم وتوثيق بنيته.

القاعدة المنهجية: لا يُرقّى أي مفهوم إلى قدرة، ولا أي قدرة إلى ادعاء عام، إلا بعد تعريف آلية القياس، خط أساس، تجربة مضبوطة، *ablation*، واختبار نقل خارج سياق التدريب.

الملخص

يقترح هذا الأطلس معمار MUSQAT-AGI، اختصارًا لـ Moussaif Unified Sovereign Question-Accumulation and Trust Architecture، بوصفه طبقة تدريب وتطوير لنماذج MOUSSAIF1B وMOUSSAIF9B والأنظمة الوكيلية. تنطلق المعمارية من أن الذكاء العام الوظيفي لا يتولد من زيادة المعرفة أو طول الاستدلال فقط، بل من قدرة النظام على: تحديد السياق الحاكم، كشف ما ينقصه، توليد السؤال الأعلى عائدًا، توزيع الثقة على الأدلة والخطوات، تراكم الخبرة دون تراكم الخطأ، وتحويل نتائج التنفيذ إلى تطوير قابل للعكس. لذلك يقسم الأطلس العمل إلى بابين: التدريب، الذي يغيّر الأوزان أو السياسات المتعلمة؛ والتطوير، الذي يغيّر السياق والذاكرة والمهارات والتوجيه والدستور التشغيلي أثناء دورة الحياة. يقترح كل باب خمس نظريات، وتحت كل نظرية ثماني أدوات، ليصبح الناتج عشر نظريات وثمانين أداة مترابطة ضمن نواة موحدة.

يستند البناء إلى اتجاهات بحثية قائمة: الجمع بين reasoning وacting، تعلم استخدام الأدوات، التعلم من التأمل والخبرة، Constitutional AI، process reward models، إدارة السياق الطويل، الذاكرة الوكيلية، المناهج الذاتية، وتقطير المهارات. لكنه يضيف تركيبًا خاصًا: الوظيفة الحاكمية للمناهج، الائتمان الوثوقي متعدد الأبعاد، شبكة الأسئلة التراكمية، التدريب الدستوري المضاد للسياق، التلمذة الوعائية، اللدونة الحاكمية للسياق، رسملة الذاكرة، التطور الذاتي القائم على السؤال، التطور الدستوري القابل للعكس، والتوحيد التطوري عبر الأحجام.

Abstract

MUSQAT-AGI is proposed as a sovereign, context-aware, trust-weighted, cumulative, question-driven training and development architecture for MOUSSAIF Unified AGI models and agentic systems. It separates parameter/policy training from life-cycle development of context, memory, skills, routing, and governance. The atlas defines ten falsifiable theories and eighty implementation tools, together with objectives, metrics, training recipes, cross-scale transfer mechanisms, and safety gates for MOUSSAIF1B, MOUSSAIF9B, and multi-agent deployments.

الإسهامات المركزية

- تعريف دالة تعلم موحدة تجمع الحاكمية والسياق والثقة والتراكم والسؤال والأثر الواقعي والقابلية للعكس.
- فصل منهجي بين training الذي يغير parameters/policies، وdevelopment الذي يغير scaffold والذاكرة والسياق والمهارات ثم يرشح بعض الخبرة للتدريب.
- خمس نظريات تدريبية مع أربعين أداة، وخمس نظريات تطويرية مع أربعين أداة.

- وصفات متباينة لنموذج 1B صغير يعتمد الذكاء الخارجي، ولنموذج 9B أعمق، ولشبكة agentic متعددة الوكلاء.
- محكمة هدم وتجارب ablation تمنع اعتبار كثافة المصطلحات دليلًا على نجاح المعمارية.
- مخططات بيانات وواجهات ونماذج traces تسمح بتحويل الأطلس إلى KB-as-OS أو Training OS قابل للبناء.

خريطة المحتوى

القسم	المضمون
الباب التأسيسي	التعريفات، الدالة الموحدة، موضع الأطلس من الأدبيات، والتمييز بين التدريب والتطوير.
الباب الأول: التدريب	VMA، CCCT، GCQL، TCTL، SCGC؛ أربعون أداة لتكوين الأوزان والسياسات.
الباب الثاني: التطوير	CSUE، RCE، GQRSD، TCME، SLCP؛ أربعون أداة للتطور أثناء دورة الحياة.
الباب التركيبي	وصفات MOUSSAIF1B وMOUSSAIF9B وAgentic، البيانات، الخوارزميات، المقاييس، السلامة، وخطة التجارب.
الملاحق	سجل الأدوات، مخططات benchmark، pseudocode، JSON، matrix، والمراجع.

الباب التأسيسي: من الذكاء الإجابي إلى الذكاء الحاكم

1. المصطلحات الخمسة بوصفها عمليات

البعد	التعريف التشغيلي
الحاكمي Sovereign/Governed	قدرة النظام على ترتيب الغايات والقيود والصلاحيات والمخاطر، وتحديد أي فرق معرفي يملك أولوية القرار. ليست الحاكمية مجرد safety filter؛ إنها اختيار ما الذي ينبغي التفكير فيه، وبأي دليل، وبأي أداة، ومن يملك حق التنفيذ.
السياقي Contextual	قدرة النظام على بناء سياق فعال، لا استهلاك كل النص المتاح. يشمل تثبيت عقد المهمة، انتقاء الأدلة، إدارة الفروع، ضغط التاريخ، واسترجاع ما يلزم عند لحظة القرار.
الوثوقي Trustworthy	قدرة النظام على إسناد الادعاء إلى منشأ، معايرة عدم اليقين، التحقق من الخطوات، ومعرفة متى لا تكفي الثقة للفعل. الثقة متغيرة مع الزمن والتناقض والأثر.
التراكمي Cumulative	تحويل الخبرة إلى أصول قابلة لإعادة الاستخدام: ذاكرة، skill، curriculum item، rubric، أو update مرشح؛ مع منع تضاعف الأخطاء والضوضاء.
السؤالي Question-driven	اكتشاف السؤال الذي يقلص مساحة الاحتمال أو يغير القرار، وترتيب الأسئلة وإيقافها وتحويلها إلى أفعال أو تجارب.

2. دالة MUSQAT الموحدة

يُقترح تمثيل الجودة الكلية لمسار وكيل أو عينة تدريب بالدالة التالية، مع فهمها كإطار تصميم لا كصيغة نهائية مثبتة:

```
J_MUSQAT(τ) = αG + βC + γT + δA + εQ + ζW + ηR - λN - μH - νK

G: sovereign goal/constraint alignment
C: context adequacy and management
T: evidence and trajectory trust
A: useful accumulation into memory/skills/data
Q: question productivity
W: verified world-effect or task success
R: reversibility and authorization
N: noise/redundancy
H: harm, invariant violation, or unsupported action
K: cost, latency, and orchestration overhead
```

تتصل هذه الدالة بمنهج MCD وQ-MCD عبر اعتبار الفروق D، التكامل I، النقل T، التجريد A، الإنتاج P، والسياق C مكونات القدرة؛ ثم تضربها وظيفة السؤال Ω(Q). الإضافة في MUSQAT هي جعل الثقة والتراكم والأثر الواقعي والقابلية للعكس إشارات مستقلة، لأن الإجابة المركبة قد تكون معرفيًا قوية لكنها غير مخولة أو غير قابلة للتراجع.

3. التدريب مقابل التطوير

خصائصه	ما الذي يتغير؟	المستوى

أبطأ، أغلى، أثره عام ومستقر، وخطر النسيان أو الانهيار أعلى.	تحديث أوزان النموذج أو policy/reward باستخدام model SFT, DPO, RL, distillation, curriculum learning.	Training
أسرع، أكثر قابلية للعكس، محلي، ويمكن اختباره وإرجاعه.	تحديث prompts، context workspace، memory، skills، routing، tools، policies، وKB-as-OS؛ وقد ينتج بيانات مرشحة للتدريب.	Development
يتطلب تكرارًا، نقلًا، ثقة، وفائدة اقتصادية.	اختيار ما يستحق الانتقال من التطوير الخارجي إلى التدريب الداخلي.	Promotion Bridge

4. موضع البرنامج من حالة الفن

تُظهر أعمال ReAct أن تداخل reasoning وacting يتيح تحديث الخطة عبر الملاحظات، ويبين Toolformer إمكان تعلم استدعاء الأدوات، بينما يستخدم Reflexion تغذية راجعة لغوية وذاكرة حلقية لتحسين المحاولات اللاحقة [1–3]. وتقدم Constitutional AI نموذجًا لتدريب السلوك بالمبادئ والنقد والمراجعة وRLAIF [4]. لكن هذه الخطوط لا تكفي وحدها لبناء تدريب حاكمي متراكم: يلزم تحديد أي سياق يدخل القرار، وكيف توزع الثقة على المسار، وكيف تتحول الخبرة إلى ذاكرة أو skill أو بيانات، وكيف يمنع الدستور الانحراف أثناء التطور.

أظهرت أبحاث السياق الطويل ظاهرة Lost in the Middle، ثم اتجهت الأعمال الأحدث إلى التعامل مع context ككيان متطور أو أداة قابلة للاستدعاء. يعامل ACE السياقات كـplaybooks تتراكم وتُنقح، ويقترح CAT مساحة سياق منظمة وضغطًا استباقيًا في المسارات الطويلة [5,8,9]. يدفع MUSQAT هذا الاتجاه خطوة إضافية بجعل إدارة السياق نفسها خاضعة للحاكمية والثقة وشبكة الأسئلة، لا للكلفة أو الطول فقط.

انتقل reward modeling من الحكم على النتيجة إلى الإشراف على الخطوات، وتجمع المراجعات الحديثة PRMs عبر توليد البيانات والتدريب والاستخدام، بينما تقترح أعمال مثل AgentV-RL تحققًا متعدد الأدوار وأدوات وتقْطيرًا إلى verifier واحد [7,13,15]. لذلك يبني الأطلس TCTL بوصفها آلية ائتمان وثوقي تشمل المنشأ والقابلية للعكس والأثر، لا correctness فقط.

في الذاكرة والتطور الذاتي، تقترح الأدبيات تصنيفات للذاكرة الداخلية والخارجية والإجرائية، وتظهر أطر experience graphs إمكان تنظيم النجاحات والإخفاقات وإعادة استخدامها، كما تميز المراجعة الأحدث بين تحسين foundation model وتحسين scaffold [14,16,18]. يستثمر MUSQAT هذا الفصل لبناء سلم تطوير يبدأ بالسياق والذاكرة والمهارة والسياسة، ثم لا يصل إلى تحديث الأوزان إلا بعد بوابة ترقية.

MUSQAT Sovereign Training Pipeline



Output: a model that selects contexts, asks productive questions, uses tools, verifies steps, and preserves constitutional invariants.

الشكل 1 — خط التدريب المقترح: من فجوة حاكمية إلى تحديث قابل للقياس والتقطير.

5. بنية المعمارية الكلية

1. طبقة الهدف والدستور: invariants، task contract، الصلاحيات، حدود الكلفة، وشروط التوقف.
2. طبقة السياق الحي: معلومات ثابتة، نافذة حديثة عالية الدقة، ذاكرة مكثفة، فروع، وأسئلة مفتوحة.
3. طبقة السؤال: شبكة أسئلة ترتبط بأفعال واسترجاع وتجارب وتحقق.
4. طبقة الوكلاء والأدوات: planner, researcher, executor, verifier, memory curator, governor.
5. طبقة الثقة: contradiction handling، calibration، process verification، evidence ledger.
6. طبقة التراكم: curriculum، skills، experience graph، مرشحات تحديث الأوزان.
7. طبقة المحاكم: regression، cost، transfer، safety، falsification، وrollback.

الباب الأول: أطلس التدريب الحاكمي السياقي الوثوقي التراكبي السؤالي

يُقصد بالتدريب هنا كل تحديث مستقر في model parameters أو policy/reward model. النظريات الخمس ليست مراحل متتابعة بالضرورة؛ بل محاور يمكن دمجها في recipe واحد. غير أن الترتيب العملي المفضل يبدأ بخريطة الفجوات، ثم شبكة الأسئلة والسياقات، ثم جمع المسارات، فالائتمان الوثوقي، فالتدريب الدستوري، وأخيرًا التقطير عبر الوكلاء والأحجام.

الأدوات	المحور	النظرية	الرمز
8	التدريب الحاكمي السياقي	نظرية المنهاج السيادي للفجوات السياقية	SCGC
8	التدريب الوثوقي للمسار	نظرية الائتمان الوثوقي للمسار الاستدلالي	TCTL
8	التدريب السؤالي التراكمي	نظرية الشبكة السؤالية التراكمية المولِّدة	GCQL
8	التدريب الحاكمي الدستوري	نظرية التدريب الدستوري المضاد للسياق	CCCT
8	التدريب الوكيلي الحاكمي	نظرية التلمذة الوعائية متعددة الوكلاء	VMA

I.1 — نظرية المنهاج السيادي للفجوات السياقية (SCGC)

Sovereign Context-Gap Curriculum Theory

الأطروحة: ينبغي أن يُبنى المنهاج من فجوات الحاكمية والسياق التي يكشفها النموذج أثناء التنفيذ، لا من قائمة مهام ثابتة فقط. كل فشل يتحول إلى عائلة مهام مضبوطة الصعوبة، وكل نجاح يُختبر تحت تحويرات سياقية مضادة قبل اعتماده ككفاءة.

المشكلة التي تعالجها

المناهج الثابتة تقيس ما يعرفه المصمم مسبقًا، لكنها لا تلاحق الفجوات التي تظهر عندما يتغير السياق أو تتعارض الأهداف أو تدخل أدوات وقيود جديدة. لذلك قد يرتفع أداء النموذج على benchmarks بينما يبقى عاجزًا عن إعادة تعريف المهمة أو طلب السياق الناقص.

موضع الجدة داخل MOUSSAIF Unified AGI

الجديد هنا هو جعل «الحاكمية» نفسها مصدرًا للمناهج: يُدرَّب النموذج على معرفة ما الذي ينقصه قبل أن يُدرَّب على الجواب، وعلى إعادة ترتيب السياق والهدف والأداة كجزء من الكفاءة.

المسلّمات التصميمية

- الفجوة ليست خطأ جواب فقط؛ قد تكون نقص سياق أو سوء ترتيب غايات أو اختيار أداة غير ملائم.
- المنهاج الجيد يلاحق جبهة القدرة الحالية ويمنع التكرار السهل.
- النجاح غير المنقول بين السياقات لا يُعد كفاءة حاكمة.
- المهام السلبية والمضادة للسياق ضرورية لكشف الاعتماد على القرائن السطحية.

الصياغة الأولية

`C_{t+1}=Compile(GapMap(τ_t, V_t, Q_t), d_t, κ_t, ρ_t)` حيث: `τ` المسار، `V` حكم محكمة التحقق، `Q` الأسئلة غير المحسومة، `d` الصعوبة، `κ` و `ρ` تنوع السياقات وقابلية النقل.

هذه الصياغة ليست قانونًا نهائيًا؛ وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكمي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثماني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بمحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

قاعدة القياس	المقياس

Context-Gap Closure Rate	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Transfer under Context Mutation	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Question Sufficiency Gain	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Curriculum Novelty Ratio	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Catastrophic Forgetting Delta	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

النسخة	التطبيق
MOUSSAIF1B	في MOUSSAIF1B يكون المنهاج خارجيًا: مولد مهام، RAG، معلم أكبر، ومحاكم تحقق. الهدف هو تعويض محدودية السعة بتوجيه شديد ودورات قصيرة عالية العائد.
MOUSSAIF9B	في MOUSSAIF9B يمكن إدخال رأس تقدير فجوة، تدريب أطول على تحويرات سياقية، وتوليد ذاتي للمناهج مع verifier مستقل. يبقى الاسم 9B معماريًا إن لم يكن عدد المعلمات الفعلي تسعة مليارات.
Agentic MOUSSAIF	في النسخ Agentic تُستخرج الفجوات من سجلات الأدوات، فشل الخطط، تأخر الإنجاز، وسوء إدارة الذاكرة، ثم تُحوّل إلى مهام بيئية قابلة لإعادة التشغيل.

الأدوات الثماني

1. رسام فجوات السياق — Moussaif Context-Gap Cartographer

يحوّل كل مسار إلى خريطة: معلومات مفقودة، تعارض أهداف، قرارات غير مبررة، أدوات لم تُستخدم، وافتراضات بلا سند. ينتج عقد فجوة قابلة للعنونة بدل وصف فشل عام.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SCGC ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context-Gap Closure Rate, Transfer under Context Mutation, Question Sufficiency Gain.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. مُصرّف المنهاج السيادي — Sovereign Curriculum Compiler

يحوّل عقد الفجوة إلى حزم تدريب متدرجة تشمل مثالًا أصليًا، تحويرات، أمثلة مضادة، شروط توقف، وrubric حاكمي. يمنع إدخال مهمة لا تملك سببًا واضحًا في خريطة الفجوات.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SCGC ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context-Gap Closure Rate, Transfer under Context Mutation, Question Sufficiency Gain.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. مقياس جبهة القدرة — Capability-Frontier Meter

يقدّر المنطقة التي يكون فيها النجاح بين 30% و70% ويعطيها أولوية، لأن المهام الأيسر لا تضيف قدرة والأصعب جدًا لا تمنح إشارة تعلم مفيدة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SCGC ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context-Gap Closure Rate, Transfer under Context Mutation, Question Sufficiency Gain.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. مسبك تحوير السياق — Context Mutation Forge

ينتج نسخًا من المهمة بتغيير ترتيب الأدلة، إخفاء معلومة، إضافة ضوضاء، تغيير الأدوار أو القيود الزمنية، مع الحفاظ على جوهر الاختبار.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SCGC ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context-Gap Closure Rate, Transfer under Context Mutation, Question Sufficiency Gain.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. مولد السياقات السلبية — Negative Context Generator

يصنع سياقات مغرية لكنها خاطئة، وقرائن سطحية، وأدوات تعطي إجابات جزئية؛ لتدريب النموذج على رفض الطريق السهل غير الموثوق.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SCGC ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context-Gap Closure Rate, Transfer under Context Mutation, Question Sufficiency Gain.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. مدقق شبكة التغطية — Coverage-Lattice Auditor

يبني مصفوفة بين القدرات والسياقات والأدوات وأنواع الفشل، ويكشف الخلايا التي لم تُختبر أو تكرر اختبارها بلا حاجة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SCGC ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context-Gap Closure Rate, Transfer under Context Mutation, Question Sufficiency Gain.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. مجدول نبض المنهاج — Curriculum Pulse Scheduler

يوزع دفعات التعلم بين اكتساب جديد، تثبيت، استرجاع متباعد، ومهام نقل، مع تعديل النسب بحسب إشارات النسيان والانحراف.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SCGC ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context-Gap Closure Rate, Transfer under Context Mutation, Question Sufficiency Gain.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. بوابة النقل الحاكمي — Transfer Gate

لا تسمح بترقية كفاءة إلا بعد نجاحها في نطاق جديد أو أداة جديدة أو صياغة جديدة، وتوثّق حدود النقل بدل تعميم النجاح.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SCGC ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context-Gap Closure Rate, Transfer under Context Mutation, Question Sufficiency Gain.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- توليد مهام تدور حول أخطاء عارضة
- انفجار المنهاج دون أولوية
- تلوث التقييم ببيانات مولدة من النموذج نفسه
- التركيز على الصعوبة بدل القيمة الحاكمية

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

8. هل يقل عدد الأسئلة الناقصة بعد كل دورة؟
9. هل ينتقل التحسن إلى سياقات لم تدخل التوليد؟
10. هل تحتفظ النماذج بقدراتها القديمة؟
11. هل ترتفع دقة اختيار الأداة لا دقة الجواب فقط؟

I.2 — نظرية الائتمان الوثوقي للمسار الاستدلالي (TCTL)

Trust-Credited Trajectory Learning Theory

الأطروحة: ينبغي توزيع الائتمان على المسار وفق الثقة المتولدة من المصدر، الاتساق، قابلية التحقق، أثر الخطوة، وقابلية عكسها. لا يكفي سؤال «هل نجح؟»؛ بل يلزم «أي خطوة صنعت النجاح، وأي خطوة كانت خطرة رغم النتيجة؟».

المشكلة التي تعالجها

المكافأة النهائية لا تميّز بين مسار صحيح وصل إلى جواب صحيح، ومسار متهافت صادف النتيجة. كما أن إعطاء المكافأة لكل خطوة بالتساوي يخلط بين الادعاء، الدليل، اختيار الأداة، والتراجع الآمن.

موضع الجدة داخل MOUSSAIF Unified AGI

تضيف النظرية «رأس المال الوثوقي» إلى credit assignment: المسار يتعلم ليس فقط ما الذي يعمل، بل ما الذي يحق له أن يثق به ومتى يجب إعادة فتح القرار.

المسلّمات التصميمية

- كل خطوة تملك قيمة معرفية وقيمة حاكمية منفصلتين.
- الثقة ليست احتمال صحة فقط؛ تشمل المصدر، الصلاحية الزمنية، وإمكان المراجعة.
- التراجع الصحيح قد يستحق مكافأة أعلى من الاستمرار في خطة خاطئة.
- التحقق الأمامي والخلفي يكشفان أخطاء مختلفة.

الصياغة الأولية

`R(τ)=Σ_t γ^t [w_eE_t+w_vV_t+w_cC_t+w_rR_t+w_qQ_t−w_hH_t]`، مع `posterior trust` يُحدَّث بعد الأدوات والنتائج الواقعية.

هذه الصياغة ليست قانونًا نهائيًا؛ وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكمي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثماني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

قاعدة القياس	المقياس

Step Calibration Error	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Critical-Step Recall	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Trust-Weighted Success	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Unsafe Irreversibility Rate	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Verifier Disagreement Resolution	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

النسخة	التطبيق
MOUSSAIF1B	يستخدم 1B مُتحققًا خارجيًا وحقول ثقة صريحة في trace، مع تقليل طول المسارات وتحديد نقاط تحقق حرجة بدل تقييم كل token.
MOUSSAIF9B	يستفيد 9B من PRM/GenRM مستقل، تحقق ثنائي الاتجاه، وتدريب على معايرة الثقة مع أدوات تنفيذية؛ ويمكن تقطير verifier متعدد الوكلاء إلى رأس أصغر.
Agentic MOUSSAIF	في الوكلاء يدخل أثر الفعل الحقيقي، latency، التكلفة، صلاحيات الأداة، وإمكان rollback ضمن المكافأة؛ فتُعاقب الخطوة غير القابلة للعكس حتى لو نجحت مرحليًا.

الأدوات الثماني

1. دفتر الأدلة — Moussaif Evidence Ledger

يسجل لكل ادعاء المصدر، زمنه، طريقة استخراجه، التحويلات التي أجريت عليه، ودرجة الثقة قبل وبعد التحقق؛ ويمنع اختلاط الحقيقة بالاستنتاج.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCTL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Step Calibration Error, Critical-Step Recall, Trust-Weighted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. موسم ثقة الخطوات — Step-Trust Annotator

يضيف إلى كل خطوة وسمًا متعدد الأبعاد: correctness, relevance, provenance, uncertainty, reversibility, impact، صالحاً لتدريب PRM أو التحليل.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحاً للاختبار مثل score vector أو graph أو patch أو dataset item. في TCTL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Step Calibration Error, Critical-Step Recall, Trust-Weighted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. المتحقق ثنائي الاتجاه — Bidirectional Process Verifier

يفحص المسار من المقدمات إلى النتيجة ومن النتيجة إلى الشروط الضرورية، ويستدعي أدوات حساب أو بحث عند نقاط الشك.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحاً للاختبار مثل score vector أو graph أو patch أو dataset item. في TCTL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Step Calibration Error, Critical-Step Recall, Trust-Weighted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. مُعاير عدم اليقين — Uncertainty Calibrator

يبني خرائط بين الثقة المعلنة والنجاح الفعلي حسب المجال والسياق، ويصحح الإفراط في الثقة أو الامتناع غير الضروري.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحاً للاختبار مثل score vector أو graph أو patch أو dataset item. في TCTL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Step Calibration Error, Critical-Step Recall, Trust-Weighted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. مشكّل مكافأة القابلية للعكس — Reversibility Reward Shaper

يضيف مكافأة للقرارات التجريبية القابلة للتراجع، وعقوبة للأفعال الدائمة التي تسبق التحقق أو التفويض.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCTL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Step Calibration Error, Critical-Step Recall, Trust-Weighted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. منارة التناقض — Contradiction Beacon

تراقب الادعاءات عبر المسار والذاكرة وتطلق حدث تحقق عندما يظهر تناقض دلالي أو زمني أو معياري، بدل دفنه في السياق.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCTL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Step Calibration Error, Critical-Step Recall, Trust-Weighted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. مقطر المسارات — Trajectory Distiller

يستخرج من المسارات الطويلة سلسلة سببية قصيرة تحافظ على الأدلة والقرارات الحرجة، ويزيل التكرار الذي قد يضلل التدريب.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCTL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Step Calibration Error, Critical-Step Recall, Trust-Weighted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. خلاط تدرجات الثقة — Trust-Gradient Mixer

يمزج خسائر token وprocess وoutcome وcalibration بنسب تتغير حسب المجال ومرحلة التدريب، ويمنع هيمنة إشارة واحدة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCTL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Step Calibration Error, Critical-Step Recall, Trust-Weighted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- تضخيم الثقة بسبب اتفاق نماذج متشابهة
- تسرب الجواب إلى verifier
- مكافأة الإسهاب بدل الدقة
- تحويل عدم اليقين إلى امتناع دائم

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

12. هل تتنبأ درجات الثقة بالفشل خارج التوزيع؟
13. هل يكتشف النظام الخطوات الحرجة قبل النهاية؟
14. هل يقل الاعتماد على إجابات صحيحة بمسارات خاطئة؟
15. هل يتحسن التراجع الآمن؟

I.3 — نظرية الشبكة السؤالية التراكمية المولِّدة (GCQL)

Generative Cumulative Question-Lattice Theory

الأطروحة: القدرة العامة تنمو عندما تتحول الأسئلة إلى بنية تراكمية: كل سؤال يفتح أو يغلق فروقًا معرفية، ويرتبط بأسئلة شرطية ومضادة، ويُقاس بأثره على القرار. تُدرَّب النماذج على بناء الشبكة لا على توليد قائمة أسئلة فقط.

المشكلة التي تعالجها

غالبًا يُدرَّب النموذج على الإجابة عن أسئلة جاهزة، فيتعلم الاستجابة أكثر مما يتعلم اكتشاف السؤال الضروري. وفي المهام المفتوحة تكون جودة السؤال، ترتيب الأسئلة، ومعرفة لحظة التوقف أهم من البلاغة النهائية.

موضع الجدة داخل MOUSSAIF Unified AGI

تنقل النظرية Q-MCD من تضخيم السؤال الجيد إلى هندسة «اقتصاد الأسئلة»: شبكة تتراكم، تُضغط، وتتحول إلى مهام وقرارات وذاكرة.

المسلّمات التصميمية

- السؤال أداة توجيه حاكمي لا مجرد صيغة لغوية.
- قيمة السؤال تقاس بمقدار ما يغير القرار أو يقلص مساحة الاحتمالات.
- الأسئلة تعتمد على بعضها؛ ترتيبها جزء من الذكاء.
- التوقف عن السؤال قرار يحتاج معيارًا.

الصياغة الأولية

```
Q* = argmax_Q [ΔU(Q)+ΔG(Q)+ΔT(Q)+ΔA(Q)−Cost(Q)−Redundancy(Q)]
```

، وتُحدّث الشبكة بعد كل جواب أو فعل.

هذه الصياغة ليست قانونًا نهائيًا؛ وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكمي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثاني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

المقياس	قاعدة القياس
Expected Information-to-Action Gain	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Question Redundancy	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

Unknown-Unknown Discovery	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Decision Change per Question	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Saturation Precision	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

التطبيق	النسخة
يستخدم 1B قوالب سؤال وgraph خارجيًا مع مخزون أنواع أسئلة، ويُدرَّب على اختيار العقدة التالية بدل توليد شبكة كاملة دفعة واحدة.	MOUSSAIF1B
يستطيع 9B تعلم latent question planning، توليد أسئلة مضادة، وتقييم expected value، مع تدريب RL على أثر السؤال في نجاح المسار.	MOUSSAIF9B
في الوكيل يمكن للسؤال أن يكون استعلام بحث، تجربة، قراءة ملف، طلب إذن، أو اختبار أداة؛ ولذلك تُربط العقد بأفعال وميزانيات.	Agentic MOUSSAIF

الأدوات الثماني

1. منقّب بذور السؤال — Moussaif Question-Seed Miner

يستخرج من الهدف والسياق والفشل بذورًا مثل: ما المفقود؟ ما المتعارض؟ ما القابل للتفنيد؟ ما الذي يغير القرار؟ ثم يرتبها أوليًا.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GCQL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Expected Information-to-Action Gain, Question Redundancy, Unknown-Unknown Discovery.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. باني شبكة الأسئلة — Question-Lattice Builder

يمثل الأسئلة كعقد وعلاقات prerequisite, alternative, contradiction, refinement, stop، ويمنع القفز إلى سؤال فرعي قبل شروطه.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GCQL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Expected Information-to-Action Gain, Question Redundancy, Unknown-Unknown Discovery.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. مقيّم حواف الاعتماد — Dependency-Edge Scorer

يقدر قوة اعتماد كل سؤال على آخر وتأثير إجابته في الفروع التالية، لتوجيه البحث إلى العقد عالية المركزية.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GCQL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Expected Information-to-Action Gain, Question Redundancy, Unknown-Unknown Discovery.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. مولد السؤال المضاد — Counterquestion Generator

ينتج السؤال الذي يمكن أن يبطل الفرض المهيمن أو يكشف أثر تغيير السياق، ويستخدم ضد confirmation bias.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GCQL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Expected Information-to-Action Gain, Question Redundancy, Unknown-Unknown Discovery.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. كاشف تشبع الإجابة — Answer-Saturation Detector

يقيس تناقص العائد من الأسئلة الجديدة وتغطية شروط القرار، ويقترح التوقف أو الانتقال إلى تجربة عملية.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GCQL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Expected Information-to-Action Gain, Question Redundancy, Unknown-Unknown Discovery.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. منظم الكثافة السؤالية — Question-Density Regulator

يضبط عدد الأسئلة وعمقها وفق المخاطر والميزانية؛ فالمهمة الحساسة تحتاج كثافة أعلى من مهمة قابلة للعكس.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GCQL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Expected Information-to-Action Gain, Question Redundancy, Unknown-Unknown Discovery.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. كشاف المجهولات غير المعروفة — Unknown-Unknown Scout

يبحث عن فئات لم تظهر في الشبكة عبر مقارنة المجال بقوالب ontologies، وتحليل حالات فشل مشابهة، واستدعاء وكلاء معارضين.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GCQL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Expected Information-to-Action Gain, Question Redundancy, Unknown-Unknown Discovery.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. مصرّف السؤال إلى مهمة — Question-to-Task Compiler

يحوّل السؤال إلى إجراء قابل للتنفيذ: retrieval query، تجربة، اختبار، طلب توضيح، simulation، أو بند labeling.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GCQL ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Expected Information-to-Action Gain, Question Redundancy, Unknown-Unknown Discovery.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- إنتاج أسئلة كثيرة بلا أثر
- الانحياز للأسئلة القابلة للإجابة بدل المهمة
- حلقة استيضاح لا تنتهي
- تحويل الشك إلى تعطيل للفعل

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

16. هل تقل كلفة الوصول إلى قرار موثوق؟
17. هل يكتشف النموذج unknown unknowns؟
18. هل يعرف متى يكفي السياق؟
19. هل تنتقل استراتيجيات السؤال بين المجالات؟

I.4 — نظرية التدريب الدستوري المضاد للسياق (CCCT)

Counter-Context Constitutional Training Theory

الأطروحة: يجب تدريب الدستور تحت سياقات مضادة ومصالح متعارضة، مع فصل الثوابت عن السياسات القابلة للتحديث، وتعليم النموذج تفسير القرار، طلب التفويض، واختيار الفعل القابل للعكس.

المشكلة التي تعالجها

الدساتير النصية قد تتحول إلى شعارات ثابتة، وقد يتعلم النموذج رفضًا شكليًا أو طاعة حرفية لا توازن بين المبادئ. كما أن السياقات المضللة تستطيع جعل السلوك غير الآمن يبدو متوافقًا مع قاعدة واحدة.

موضع الجدة داخل MOUSSAIF Unified AGI

الجديد هو «السياق المضاد للدستور»: لا نختبر القاعدة في حالة واضحة فقط، بل نبدل الأدوار والضرورات والنتائج لنرى هل يحافظ النظام على الفرق الحاكمي الأعلى.

المسلّمات التصميمية

- الدستور بنية أولويات وقيود لا قائمة محظورات فقط.
- التعارض بين المبادئ يجب أن يُدرَّب ويُقيّم صراحة.
- القابلية للعكس والتفويض عنصران حاكميان.
- الدستور يحتاج اختبارات مضادة تمنع gaming.

الصياغة الأولية

`π* = argmax_π E[R_task + λR_const + μR_rev + νR_auth − ξR_conflict]` تحت `Invariants I` التي لا تُعدّل إلا ببروتوكول مستقل.

هذه الصياغة ليست قانونًا نهائيًا، وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكمي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثماني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

قاعدة القياس	المقياس

Invariant Violation Rate	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Principle-Conflict Resolution	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Helpful Safety Frontier	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Authorization Compliance	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Rollback Readiness	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

النسخة	التطبيق
MOUSSAIF1B	يستفيد 1B من Constitution DSL قصيرة، مصنفات حواجز، وقوالب قرار، مع verifier أكبر لتجنب انهيار المساعدة عند التدريب الذاتي.
MOUSSAIF9B	يمكن 9B تعلم تفسير DPO، conflict graphs متعدد المبادئ، ومحاكاة سياقات adversarial؛ لكن يحتاج فصل بيانات التوليد عن نماذج الحكم.
Agentic MOUSSAIF	في الوكلاء تشمل المبادئ الصلاحيات، data، budget، privacy، human approval، boundaries، وtransactionality؛ ويجب اختبارها في sandbox.

الأدوات الثماني

1. لغة دستور مساعف — Moussaif Constitution DSL

تمثل invariants, priorities, permissions, exceptions, escalation, evidence requirements في صيغة قابلة للتحقق الآلي بدل نص فضفاض.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CCCT ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Violation Rate, Principle-Conflict Resolution, Helpful Safety Frontier.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. مضاد السياق — Context Counterfactualizer

يولد نسخًا تغير هوية الفاعل، الصلاحيات، الضرر، الاستعجال، والنتائج البعيدة لاختبار ثبات الحكم وعدم تعلقه بصياغة واحدة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CCCT ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Violation Rate, Principle-Conflict Resolution, Helpful Safety Frontier.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. حالّ تعارض المبادئ — Principle-Conflict Resolver

يبني graph للتعارض ويطلب تبريرًا وفق الأولوية، الضرورة، أقل ضرر، والقابلية للعكس؛ ثم يقارن القرار بمرجع أو محكمة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CCCT ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Violation Rate, Principle-Conflict Resolution, Helpful Safety Frontier.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. منضدة اختبار الثوابت — Invariant Test Bench

تنفذ suites قبل وبعد كل تدريب أو تحديث، وتوقف release إذا ارتفع انتهاك invariant ولو تحسن benchmark وظيفي.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CCCT ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Violation Rate, Principle-Conflict Resolution, Helpful Safety Frontier.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. قالب الدستور العدائي — Adversarial Constitution Flipper

ينتج أمثلة تحاول إعادة تفسير القواعد لمصلحة سلوك ممنوع، لكشف reward hacking والتسويغ اللغوي.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CCCT ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Violation Rate, Principle-Conflict Resolution, Helpful Safety Frontier.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. مُعاير الرفض الآمن — Safe-Rejection Calibrator

يقيس متى يجب الرفض، التقييد، السؤال، أو التنفيذ الجزئي، ويعاقب الرفض العام غير المفيد مثلما يعاقب التساهل.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CCCT ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Violation Rate, Principle-Conflict Resolution, Helpful Safety Frontier.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. صندوق السياسة القابل للعكس — Reversible Policy Sandbox

يشغل السياسات الجديدة على نسخ بيانات وأدوات وهمية، يسجل كل أثر، ويدعم rollback قبل الصلاحيات الحقيقية.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CCCT ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Violation Rate, Principle-Conflict Resolution, Helpful Safety Frontier.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. حازم DPO الدستوري — Constitutional DPO Packager

يبني أزواج تفضيل لا تقارن جوابًا آمنًا بآخر ضار فقط، بل تقارن توازنات دقيقة بين المساعدة والدليل والتفويض والقابلية للعكس.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CCCT ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Violation Rate, Principle-Conflict Resolution, Helpful Safety Frontier.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- الرفض الزائد
- إخفاء التعارض بدل حله
- تعلم صياغات الاختبار
- تحديث الدستور من بيانات غير موثوقة

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

20. هل يحافظ على المساعدة مع الأمان؟
21. هل تتغير القرارات بتغير التفويض الحقيقي لا الكلمات؟
22. هل يمكن rollback كامل؟
23. هل تُكتشف محاولات قلب الدستور؟

I.5 — نظرية التلمذة الوعائية متعددة الوكلاء (VMA)

Vascular Multi-Agent Apprenticeship Theory

الأطروحة: تُنتج الخبرة عبر شبكة وكلاء متخصصة ومسارات وعائية، ثم يُعطى الائتمان للأدوار والحواف الفعالة، وتُقطر المهارات القابلة للنقل إلى *1B* أو *9B* مع إبقاء المهارات النادرة خارجياً.

المشكلة التي تعالجها

تدريب نموذج واحد على كل الأدوار يخلط التوليد بالتحقق والتخطيط بالذاكرة، ويزيد تكلفة البيانات. وفي المقابل، ترك الذكاء في scaffold متعدد الوكلاء قد يمنع تقطير الخبرة إلى النموذج الأساسي.

موضع الجدة داخل MOUSSAIF Unified AGI

تربط الوعائية الحاسوبية بالتلمذة: الوعاء ليس مسار inference فقط، بل قناة إنتاج بيانات وتعليم وتقطير عبر الأحجام.

المسلّمات التصميمية

- اختلاف الأدوار يزيد قابلية كشف الخطأ.
- التقطير يجب أن يحافظ على شروط استخدام المهارة لا مخرجاتها فقط.
- ليست كل مهارة جديرة بالدخول إلى الأوزان.
- الاختلاف بين الوكلاء مصدر بيانات، لا ضوضاء ينبغي محوها.

الصياغة الأولية

`K_{student}=Distill({τ_i, role_i, credit_i, verifier_i})` مع `Routing r(x)` الذي يوازن الجودة والكلفة غير اقتصادي `internalization` حيث `external skills` والتنوع، واحتفاظ.

هذه الصياغة ليست قانونًا نهائيًا؛ وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثماني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

قاعدة القياس	المقياس
يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.	Distilled Skill Retention

Cross-Agent Credit Accuracy	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Routing Efficiency	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Disagreement Yield	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Standalone Student Gain	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

التطبيق	النسخة
يكون 1B منفذًا سريعًا مع skill capsules وrouting مكثف؛ تُقطر إليه القرارات المتكررة القصيرة ويُحال التعقيد إلى agents/tools.	MOUSSAIF1B
يستوعب 9B قدرًا أكبر من التخطيط والتحقق الداخلي، ويصبح orchestrator محليًا، مع إبقاء الوكلاء لمعارضة الافتراضات والمهام النادرة.	MOUSSAIF9B
هي النظرية المركزية للنماذج Agentic: توزع الأدوار على planner, researcher, executor, verifier, memory curator, governor، وتتعلم شبكة التوجيه من النتائج.	Agentic MOUSSAIF

الأدوات الثماني

1. مسبك أدوار الوكلاء — Moussaif Agent-Role Foundry

يعرّف أدوارًا بعقود واضحة: مدخلات، صلاحيات، مخرجات، محظورات، verifier، وميزانية، ويمنع تداخل المسؤوليات غير القابل للتشخيص. المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في VMA ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Distilled Skill Retention, Cross-Agent Credit Accuracy, Routing Efficiency.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. الموجّه الوعائي — Vascular Router

يختار الوكيل أو الأداة أو النموذج حسب فجوة المهمة والثقة والكلفة، ويعيد التوجيه عندما يظهر دليل جديد أو يفشل المسار.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في VMA ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Distilled Skill Retention, Cross-Agent Credit Accuracy, Routing Efficiency.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. مجلس المعلمين — Teacher Panel

يجمع نماذج وخبراء وقواعد وأدوات تنفيذية، ويمنح كل عضو وزنًا مجاليا بدل تصويت متساوٍ.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في VMA ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Distilled Skill Retention, Cross-Agent Credit Accuracy, Routing Efficiency.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. حاصد الاختلاف — Disagreement Harvester

يحول اختلاف الخطط والأحكام إلى أمثلة contrastive وأسئلة تحقق، ويركز على الاختلاف المنتج لا التنوع الشكلي.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في VMA ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Distilled Skill Retention, Cross-Agent Credit Accuracy, Routing Efficiency.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. مقطر أثر المتخصص — Specialist-Trace Distiller

يختصر trace المتخصص إلى شروط اختيار، خطوات حرجة، إشارات فشل، وأمثلة مضادة صالحة لتدريب الطالب.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في VMA ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Distilled Skill Retention, Cross-Agent Credit Accuracy, Routing Efficiency.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. موزع الائتمان بين الوكلاء — Cross-Agent Credit Assigner

يقدر مساهمة كل دور وحافة في النتيجة عبر ablation وcounterfactual replay، بدل مكافأة الجميع عند النجاح.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في VMA ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Distilled Skill Retention, Cross-Agent Credit Accuracy, Routing Efficiency.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. مصرّف كبسولات المهارة — Skill-Capsule Compiler

ينتج مهارة خارجية قابلة للاستدعاء تشمل trigger, procedure, tools, checks, rollback, examples، ويمكن لاحقًا internalize أو retire.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في VMA ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Distilled Skill Retention, Cross-Agent Credit Accuracy, Routing Efficiency.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. جسر تقطير 1B/9B — 1B/9B Distillation Bridge

ينقل المهارات بين الأحجام عبر progressive targets، مقارنة logits/decisions، وتدريب الطالب على طلب المساعدة عندما تتجاوز المهمة حد قدرته.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في VMA ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Distilled Skill Retention, Cross-Agent Credit Accuracy, Routing Efficiency.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- اتفاق زائف بين وكلاء من عائلة واحدة
- تقطير إجابات بلا سياق استعمال
- تضاعف الكلفة
- اعتماد دائم على معلم مغلق

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

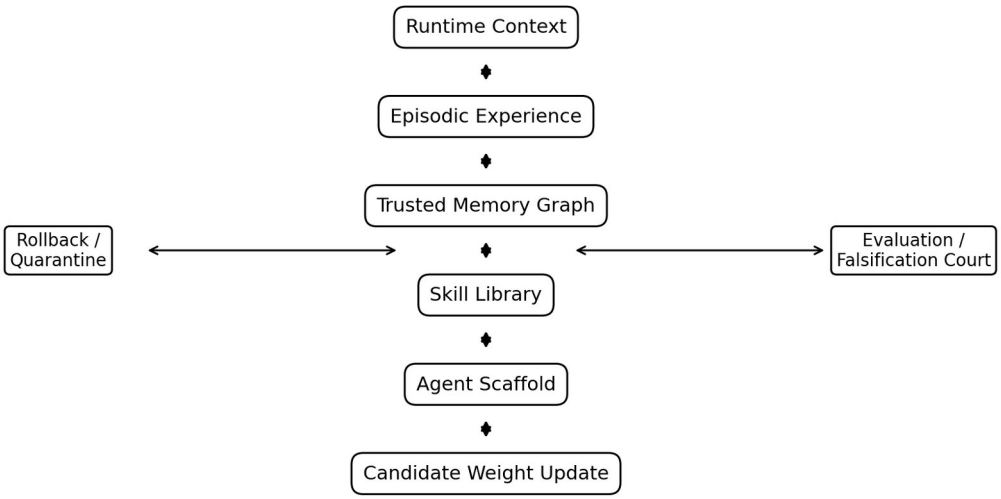
الأسئلة التجريبية المفتوحة

24. ما المهارات التي ينبغي تقطيرها؟
25. هل يبقى الطالب قادرًا دون scaffold؟
26. هل يستحق كل وكيل تكلفته؟
27. هل يحسن الاختلاف الأداء أم يزيد الفوضى؟

الباب الثاني: أطلس التطوير الحاكمي السياقي الوثوقي التراكبي السؤالي

التطوير هو تحسين المنظومة أثناء دورة الحياة دون افتراض أن كل تحسين يتطلب fine-tuning. يبدأ من السياق، ثم الذاكرة والخبرة، ثم المهارات والتوجيه والسياسات، ثم يرشح ما يستحق الدخول إلى الأوزان. هذا الترتيب يقلل التكلفة ويزيد القابلية للعكس، ويطابق الفصل الحديث بين foundation model improvement وscaffolding improvement.

MUSQAT Development Ladder: from reversible scaffolding to selective parameter change



الشكل 2 — سلم التطوير: لا تُدفع الخبرة إلى الأوزان قبل نجاحها في طبقات أكثر قابلية للعكس.

الرمز	النظرية	المحور	الأدوات
SLCP	نظرية اللدونة الحاكمية للسياق الحي	التطوير السياقي الحي	8
TCME	نظرية الرسملة الوثوقية للذاكرة والخبرة	التطوير الوثوقي التراكمي	8
GQRSD	نظرية التطور الذاتي بالاستفهام الفجوي	التطوير السؤالي الذاتي	8
RCE	نظرية التطور الدستوري القابل للعكس	التطوير الحاكمي الآمن	8
CSUE	نظرية التوحيد التطوري عابر الأحجام	تطوير MOUSSAIF1B وMOUSSAIF9B	8

II.1 — نظرية اللدونة الحاكمية للسياق الحي (SLCP)

Sovereign Live-Context Plasticity Theory

الأطروحة: السياق يجب أن يعامل كمساحة عمل حية ذات طبقات وصلاحيات ودورة حياة: ثوابت مهمة، ذاكرة مكثفة، نافذة عالية الدقة، فروع بديلة، وسجل قرارات. التطوير يبدأ بتغيير هذه البنية قبل تغيير الأوزان.

المشكلة التي تعالجها

النموذج الوكيلي يعمل في مسارات طويلة تتراكم فيها الملاحظات والأدوات والقرارات. السياق append-only ينفجر، والضغط غير المنضبط يمحو التفاصيل أو يرسخ ملخصًا خاطئًا.

موضع الجدة داخل MOUSSAIF Unified AGI

الجديد هو اللدونة «الحاكمية»: لا نغير السياق لتحسين السرعة فقط، بل وفق غايات وثقة وأسئلة وفروع وحقوق كتابة.

المسلّمات التصميمية

- السياق مورد قابل للإدارة لا نص خام.
- الضغط قرار حاكمي يحتاج provenance واختبار استرجاع.
- الفروع تمنع فرض فرضية واحدة على كل المسار.
- المعلومات تترقى أو تهبط حسب استخدامها وثقتها.

الصياغة الأولية

`W_t={S_stable,M_condensed,H_recent,B_branches,D_decisions}; Update(W_t)` يُقبل إذا زاد `utility` والثقة وخفّض التكلفة دون فقد شروط القرار.

هذه الصياغة ليست قانونًا نهائيًا، وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكمي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثماني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بمحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

قاعدة القياس	المقياس
يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.	Context Utility per Token

Semantic Drift Rate	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Critical Fact Retention	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Branch Recovery Success	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Bounded-Horizon Stability	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

التطبيق	النسخة
هذا المستوى بالغ الأهمية لـ1B، إذ يعوض قصر السياق الفعال بمساحة عمل منظمة وretrieval انتقائي ومكثف متخصص.	MOUSSAIF1B
يستطيع 9B تعلم context actions واختيار وقت الضغط، وربط الملخصات بالمسار الأصلي مع تحقق ذاتي وخارجي.	MOUSSAIF9B
كل tool call يمكن أن يقرأ أو يكتب في جزء محدد من workspace؛ وتراقب الحاكمية ما الذي يدخل prompt الفعلي وما الذي يبقى في التخزين.	Agentic MOUSSAIF

الأدوات الثماني

1. مساحة السياق الحي — Moussaif Live Context Workspace

تقسم الحالة إلى task contract, facts, hypotheses, open questions, actions, evidence, memory links، وتفرض schemas على الكتابة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SLCP ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context Utility per Token, Semantic Drift Rate, Critical Fact Retention.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. الضاغط الذاتي للسياق — Context Self-Compactor

ينفذ ضغطًا استباقيًا عند milestones، يحتفظ بروابط إلى الأصل، ويولّد اختبارات أسئلة للتأكد من أن الملخص حافظ على المطلوب.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SLCP ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context Utility per Token, Semantic Drift Rate, Critical Fact Retention.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. حارس الانجراف الدلالي — Semantic Drift Sentinel

يقارن المعنى الحالي بالعقد الأصلي وقرارات سابقة، ويكشف تغير الهدف أو تعريف المصطلحات أو القيود أثناء المسار.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SLCP ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context Utility per Token, Semantic Drift Rate, Critical Fact Retention.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. مدير فروع السياق — Context Branch Manager

ينشئ فروع فرضيات مع أدلتها وتكاليفها، ويدمجها فقط بعد اختبار، ما يمنع تلويث الذاكرة بفرضية واحدة مبكرة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SLCP ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context Utility per Token, Semantic Drift Rate, Critical Fact Retention.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. محرك ترقية وهبوط السياق — Context Promotion-Demotion Engine

يرقي المعلومة المتكررة الموثوقة إلى ذاكرة مكثفة، ويهبط التفصيل المستهلك إلى أرشيف، ويعزل غير الموثوق.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SLCP ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context Utility per Token, Semantic Drift Rate, Critical Fact Retention.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. ضاغط المعالم — Milestone Compressor

يعرّف نقاطًا مثل نهاية بحث أو نجاح اختبار أو تغيير خطة، ويضغط ما قبلها إلى قرار وأسباب ومخاطر وخطوات تالية.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SLCP ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context Utility per Token, Semantic Drift Rate, Critical Fact Retention.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. مخبأ السياق السيادي — Sovereign Context Cache

يحفظ وحدات جاهزة للاستدعاء حسب الهدف والصلاحية الزمنية والثقة، ولا يحقنها إلا إذا اجتازت relevance وauthorization.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SLCP ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context Utility per Token, Semantic Drift Rate, Critical Fact Retention.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. لوحة صحة السياق — Context Health Dashboard

تعرض كثافة الضوضاء، التكرار، الأسئلة المفتوحة، التعارضات، token burn، وأهم المعلومات المعرضة للفقد.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في SLCP ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Context Utility per Token, Semantic Drift Rate, Critical Fact Retention.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- ضغط مبكر
- تثبيت ملخص خاطئ
- تشعب فروع لا ينتهي
- حذف معلومات غير مستخدمة لكنها حاسمة لاحقًا

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

28. هل يحافظ الضغط على شروط القرار؟
29. هل يستطيع العودة إلى فرع مهمل؟
30. هل تقل الكلفة دون انخفاض النجاح؟
31. هل يتعلم توقيت إدارة السياق؟

II.2 — نظرية الرسملة الوثوقية للذاكرة والخبرة (TCME)

Trust Capitalization of Memory and Experience Theory

الأطروحة: كل خبرة تُسجل كأصل وثوقي له قيمة واستحقاق ومخاطر واستهلاك. ترتفع قيمته بالتكرار المستقل والنجاح المنقول، وتنخفض بالقدم والتناقض وفشل الاستخدام. تُحوّل الخبرة إلى *graph* للنجاحات والتحذيرات والتصحيحات.

المشكلة التي تعالجها

الذاكرة التي تخزن كل شيء تكرر الأخطاء وتضخم السياق، والذاكرة التي لا تخزن إلا النجاحات تخفي حدود الكفاءة. كما أن الثقة تتغير مع الزمن والتناقض والنتائج الجديدة.

موضع الجدة داخل MOUSSAIF Unified AGI

تضيف النظرية مفهوم «الرسملة»: الخبرة لا تتراكم كحجم، بل كقيمة موثوقة قابلة للاستثمار في inference أو المهارات أو التدريب.

المسلّمات التصميمية

- الخبرة السلبية أصل معرفي إذا صيغت كتحذير قابل للتعميم.
- الثقة ديناميكية وليست metadata ثابتة.
- provenance شرط للذاكرة الحاكمة.
- لا تنتقل الذاكرة إلى وزن أو مهارة دون تكرار ونقل.

الصياغة الأولية

```
K_i(t+1)=K_i(t)+αEvidence+βTransfer+γUsefulness−δDecay−εContradiction−ζHarm
```

retrieval مسار K ؛ ويحدد والترقية والتقطير.

هذه الصياغة ليست قانونًا نهائيًا، وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثماني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

قاعدة القياس	المقياس
يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.	Memory Utility Gain

False-Memory Injection Rate	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Trust Revaluation Accuracy	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Cross-Task Reuse	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Stale Knowledge Harm	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

النسخة	التطبيق
MOUSSAIF1B	يحتاج 1B ذاكرة خارجية غنية وretriever يتعلم trust-aware retrieval، مع منع حشو prompt بكامل التاريخ.
MOUSSAIF9B	يمكن 9B تعلم عمليات الذاكرة وكتابة summaries سببية، وربطها بتقدير uncertainty، لكنه يبقى محتاجًا provenance خارجيًا.
Agentic MOUSSAIF	تسجل الذاكرة الأفعال والأدوات والأثر، وتفرق بين golden cases وwarning cases وpolicy violations، وتدعم reuse عبر المهام.

الأدوات الثماني

1. حساب ثقة الذاكرة — Moussaif Memory Trust Account

يحتفظ برصيد متعدد الأبعاد لكل سجل: صحة، حداثة، استقلال المصادر، نقل، حساسية، وتكلفة خطأ، ويؤثر في retrieval.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCME ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Memory Utility Gain, False-Memory Injection Rate, Trust Revaluation Accuracy.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. محرك اضمحلال الدليل — Evidence Decay Engine

يطبق decay حسب نوع المعرفة لا الزمن فقط؛ فالقوانين والواجهات تتغير أسرع من المبادئ الرياضية، مع طلب refresh مناسب.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCME ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Memory Utility Gain, False-Memory Injection Rate, Trust Revaluation Accuracy.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. محكمة مصالحة التناقض — Contradiction Reconciliation Court

تجمع السجلات المتعارضة، تميز اختلاف السياق عن الخطأ، تطلب أدلة جديدة، وتصدر نسخة مصالحة أو تبقي الفروع منفصلة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCME ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Memory Utility Gain, False-Memory Injection Rate, Trust Revaluation Accuracy.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. رسم المنشأ — Provenance Graph

يربط كل ذاكرة بالمصدر والأداة والمسار والنموذج والتحويلات، ما يسمح بتتبع سلسلة الثقة وحذف المشتقات عند فساد الأصل.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCME ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Memory Utility Gain, False-Memory Injection Rate, Trust Revaluation Accuracy.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. موحد رسم الخبرة — Experience Graph Consolidator

ينشئ عقد task anchor, correction, warning, golden وحواف correction, causality, similarity، لتمكين reuse منظم.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCME ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Memory Utility Gain, False-Memory Injection Rate, Trust Revaluation Accuracy.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. حجر الذاكرة — Memory Quarantine

يعزل السجلات الجديدة أو الحساسة أو المتناقضة عن الحقن التلقائي حتى تجتاز التحقق أو الاستخدام الناجح المتكرر.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCME ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Memory Utility Gain, False-Memory Injection Rate, Trust Revaluation Accuracy.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. محرك إعادة تقييم الثقة — Confidence Revaluation Engine

يعيد حساب الرصيد بعد كل نتيجة أو مصدر أو تغير بيئة، ويمنع بقاء الثقة القديمة بعد تبدل الواقع.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCME ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Memory Utility Gain, False-Memory Injection Rate, Trust Revaluation Accuracy.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. منقّب مرشحات الذاكرة للأوزان — Memory-to-Weight Candidate Miner

يحدد الأنماط المتكررة المنقولة التي تستحق fine-tuning، ويستبعد الحقائق الزمنية أو الحالات النادرة التي الأفضل إبقاؤها خارجية.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في TCME ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Memory Utility Gain, False-Memory Injection Rate, Trust Revaluation Accuracy.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- تعزيز إشاعة بتكرار داخلي
- نقص التنوع في أدلة الدعم
- تآكل المعرفة الثابتة بلا سبب
- استرجاع ذاكرة مشابهة لفظيًا لكنها غير صالحة

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

32. هل يقل تكرار الخطأ؟
33. هل تزيد قيمة الذاكرة مع الزمن بدل الضوضاء؟
34. هل تتراجع الثقة عند تناقض جديد؟
35. هل يمكن تفسير سبب الاسترجاع؟

II.3 — نظرية التطور الذاتي بالاستفهام الفجوي (GQRSD)

Gap-Question Recursive Self-Development Theory

الأطروحة: يبدأ التطور من ملف فجوات، ثم يولد أسئلة تختبر حدود القدرة، ويحوّلها إلى تجارب، ويُدخل النتائج في الذاكرة أو المهارات أو التدريب وفق بوابة *novelty* والثقة. السؤال هو محرك اختيار ما ينبغي تطويره.

المشكلة التي تعالجها

الأنظمة الذاتية كثيرًا ما تحسن ما تستطيع قياسه بسهولة، لا ما ينقصها فعليًا. وقد تولد بيانات تدعم أسلوبها الحالي وتغلق نفسها داخل توزيع ذاتي.

موضع الجدة داخل MOUSSAIF Unified AGI

تربط Q-MCD بعملية software-like self-development: كل سؤال له issue، تجربة، review، tests، patch، وrollback.

المسلّمات التصميمية

- التطور بلا ملف فجوات يصبح تحسينًا عشوائيًا.
- كل فرضية تطوير يجب أن تكون قابلة للتفنيد.
- النتيجة السلبية توجه المنهاج ولا تُخفى.
- التغيير الأقل كلفة والأكثر قابلية للعكس يسبق تعديل الأوزان.

الصياغة الأولية

```
DevCycle: Gap→Question→Experiment→Verdict→Artifact→Ablation→Commit/Rollback
```

إلا إذا commit ؛ ولا يحدث تحقق ΔCapability>0 وΔRisk≤threshold.

هذه الصياغة ليست قانونًا نهائيًا؛ وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثاني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بمحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

المقياس	قاعدة القياس
Gap Discovery Precision	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Experiment Information Gain	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

Novel Capability Yield	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Self-Confirmation Rate	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Rollback Frequency and Recovery	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

التطبيق	النسخة
يستخدم 1B وكيلًا بحثيًا خارجيًا ومولد تجارب، بينما يقتصر دوره على تنفيذ مهام محددة وتعلم skill capsules عالية التكرار.	MOUSSAIF1B
يمكن 9B توليد فرضيات وتجارب أكثر تجريدًا، لكنه يحتاج evaluators متغايرة وholdouts لا يراها لتجنب self-confirmation.	MOUSSAIF9B
تُنفذ دورة التطور في بيئات sandbox وbenchmarks حية، وتنتج تحديثات للسياق والمهارات والتوجيه قبل التعديل الوزني.	Agentic MOUSSAIF

الأدوات الثماني

1. مُشخّص فجوات القدرة — Moussaif Capability-Gap Profiler

يجمع نتائج الاختبارات والمسارات والمقارنات البشرية ليبني ملفًا متغيرًا للقدرات والحدود والثقة والأثر.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GQRSD ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Gap Discovery Precision, Experiment Information Gain, Novel Capability Yield.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. مولد الأسئلة الذاتية — Self-Question Generator

ينتج أسئلة تطويرية محددة قابلة للاختبار من كل فجوة، مع فرضية وآلية محتملة ومعيار فشل.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GQRSD ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Gap Discovery Precision, Experiment Information Gain, Novel Capability Yield.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. منسق الحلقة البحثية — Research-Loop Orchestrator

يوزع الأدوار على مولد فرضية ومصمم تجربة ومنفذ ومتحقق ومراجع، ويمنع الدور الواحد من كتابة الاختبار والحكم النهائي.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GQRSD ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Gap Discovery Precision, Experiment Information Gain, Novel Capability Yield.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. مصمم التجربة — Experiment Designer

يبني control, treatment, holdout, ablation، ويحدد المتغيرات والمقاييس والميزانية قبل التنفيذ.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GQRSD ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Gap Discovery Precision, Experiment Information Gain, Novel Capability Yield.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. محول الفشل إلى منهاج — Failure-to-Curriculum Converter

يحوّل الفشل المتكرر إلى dataset ومهام تدريجية وأسئلة مضادة، ويربطه بنظرية SCGC.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GQRSD ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Gap Discovery Precision, Experiment Information Gain, Novel Capability Yield.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. بوابة الجدة — Novelty Gate

تفحص أن التحسن ليس حفظًا أو leakage أو إعادة صياغة لقدرة سابقة، وتتطلب نقلًا أو سلوكًا جديدًا قابلًا للقياس.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GQRSD ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Gap Discovery Precision, Experiment Information Gain, Novel Capability Yield.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. حاكم الميزانية الميتامعرفية — Metacognitive Budget Controller

يقرر كم من compute والأسئلة والوكلاء يستحقها التطوير حسب المخاطر والعائد المتوقع، ويوقف الحلقات منخفضة القيمة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GQRSD ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Gap Discovery Precision, Experiment Information Gain, Novel Capability Yield.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. سجل التطوير — Development Journal

يوثق الفرضية والبيانات والكود والنتائج والفشل والقرار والإصدار، ليمنع إعادة التجارب ويتيح rollback وتحليل السببية.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في GQRSD ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Gap Discovery Precision, Experiment Information Gain, Novel Capability Yield.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- اختبارات يكتبها النظام ويعرف إجاباتها
- تضخم التغييرات الصغيرة
- بحث ذاتي بلا حد تكلفة
- ترقية نتائج غير مستقلة

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

36. هل الأسئلة تكشف فجوات حقيقية؟
37. هل التجارب مستقلة عن مولد الفرضية؟
38. هل يتوقف النظام عند عائد هامشي؟
39. هل ينتج قدرات قابلة للنقل؟

II.4 — نظرية التطور الدستوري القابل للعكس (RCE)

Reversible Constitutional Evolution Theory

الأطروحة: تطور السياسات في طبقات: ثوابت مقفلة، مبادئ قابلة للمراجعة البشرية، سياسات تشغيلية قابلة للاختبار، وإرشادات محلية قصيرة العمر. كل *patch* يمر *diff* واختبارات *canary* وقابلية *rollback*.

المشكلة التي تعالجها

مع تغير الأدوات والبيئات، تحتاج السياسات والدساتير إلى تحديث، لكن التحديث الذاتي قد يضعف القيود أو يراكم استثناءات غير مفهومة. التثبيت الكامل يمنع التكيف، والتحرير الكامل يهدد السيطرة.

موضع الجدة داخل MOUSSAIF Unified AGI

تجعل النظرية التطور الدستوري عملية release engineering كاملة، تربط الحاكمية بالتجارب والنسخ والرجوع لا بالنص الأخلاقي فقط.

المسلّمات التصميمية

- التطور لا يعني تعديل كل طبقة.
- الثوابت تحتاج مسار حوكمة مستقل.
- كل policy patch يجب أن يكون محدود النطاق ومفسرًا.
- canary وrollback جزء من البرهان العملي.

الصياغة الأولية

```
Constitution_t={I_locked,P_reviewed,Π_operational,H_local}; Patch accepted iff Tests(I)=pass ∧ Utility↑ ∧ Risk≤B ∧ RollbackReady.
```

هذه الصياغة ليست قانونًا نهائيًا، وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكمي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثماني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بمحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

قاعدة القياس	المقياس
يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.	Invariant Preservation
يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.	Policy Patch Utility

Rollback Completeness	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Canary Incident Rate	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Exception Debt	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

النسخة	التطبيق
MOUSSAIF1B	يعتمد 1B على سياسات خارجية صريحة يسهل diff واختبارها، مع تقليل self-editing الداخلي.
MOUSSAIF9B	يمكن 9B اقتراح patches وتفسيرها، لكن commit يبقى عبر محكمة مستقلة وحزم تقييم خارجية.
Agentic MOUSSAIF	يشمل التطور memory، routing، tool permissions، prompts، skill lifecycle، rules، وaction policies، لا model weights وحدها.

الأدوات الثماني

1. محرك فرق الدستور — Moussaif Constitution Diff Engine

يعرض الفرق الدلالي والعملي بين نسختين، بما في ذلك تغير الأولويات والصلاحيات والحالات الحدية لا الكلمات فقط.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في RCE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Preservation, Policy Patch Utility, Rollback Completeness.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. خزنة الثوابت — Invariant Lockbox

تفصل invariants عن ملفات التشغيل، توقعها، وتمنع تعديلها من الوكيل أو pipeline العادي؛ مع سجل وصول.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في RCE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Preservation, Policy Patch Utility, Rollback Completeness.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. صندوق المهارة والسياسة — Skill-Policy Sandbox

يختبر patch على أدوات وبيانات محاكية مع replay لمسارات حساسة قبل التعرض للعالم الحقيقي.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في RCE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Preservation, Policy Patch Utility, Rollback Completeness.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. منسق الرجوع — Rollback Orchestrator

يعيد prompts والسياسات والمهارات والذاكرة المشتقة وحالة الأدوات إلى snapshot متسق، لا ملفًا واحدًا فقط.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في RCE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Preservation, Policy Patch Utility, Rollback Completeness.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. بوابة تصعيد القدرة — Capability-Escalation Gate

تطلب أدلة إضافية وموافقة أعلى عند إضافة أداة أو صلاحية أو مدى فعل جديد، وتحدد حدًا زمنيًا للتجربة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في RCE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Preservation, Policy Patch Utility, Rollback Completeness.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. محكمة التطور العدائية — Adversarial Evolution Court

تحاول كسر patch عبر وكلاء مهاجمين وتغيير السياقات والموارد والأهداف، ثم تفرض remediation أو رفضًا.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في RCE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Preservation, Policy Patch Utility, Rollback Completeness.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. مصرّف رقعة الحاكمية — Governance Patch Compiler

يحول المقترح إلى ملفات policies واختبارات وmigration وrollback plan وrelease notes قابلة للتنفيذ.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في RCE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Preservation, Policy Patch Utility, Rollback Completeness.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. كناري النشر — Deployment Canary

يطلق patch على نسبة صغيرة من المهام بصلاحيات محدودة، ويراقب المقاييس والانحراف قبل التوسيع.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في RCE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Invariant Preservation, Policy Patch Utility, Rollback Completeness.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- تراكم استثناءات
- اختبارات لا تغطي التفاعل بين السياسات
- rollback شكلي لا يعيد الذاكرة والأثر
- تغيير غير مرئي في prompts أو tools

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

40. هل يحتفظ الإصدار الجديد بالثوابت؟
41. هل يمكن عكس كل أثر؟
42. هل تتزايد ديون الاستثناءات؟
43. هل اقتراح النظام منحاز لمصلحته التشغيلية؟

II.5 — نظرية التوحيد التطوري عابر الأحجام (CSUE)

Cross-Scale Unified Evolution Theory

الأطروحة: يجب أن تشترك المنظومة في *kernel* عقود وقدرات وذاكرة واختبارات، مع حرية تنفيذ مختلفة. تُنقل الخبرة عبر *skill graph* و*distillation*، ويختار *router* أقل مكوّن يحقق الثقة المطلوبة.

المشكلة التي تعالجها

قد تتفرق المعرفة بين نموذج صغير سريع، نموذج أكبر، وكلاء، ذاكرة، وأدوات. إذا تطور كل جزء منفردًا تظهر ازدواجية أو تعارض، وإذا فُرض التطابق تُهدر مزايا كل حجم.

موضع الجدة داخل MOUSSAIF Unified AGI

تقدم النظرية تعريفًا عمليًا لـMOUSSAIF Unified AGI كمنظومة تتطور عبر الأحجام دون ادعاء أن نموذجًا منفردًا صار AGI.

المسلّمات التصميمية

- الوحدة في العقود لا في التطابق الداخلي.
- النموذج الأصغر يجب أن يعرف حدوده ويصعد المهمة.
- التطوير الناجح في scaffold قد يسبق الأوزان.
- كل نقل يحتاج اختبار تكافؤ سلوكي وحدود.

الصياغة الأولية

```
SystemPolicy(x)=Route(x, capability, trust, cost, latency); SharedArtifacts={contracts, skills, tests, memory schemas}; Transfer only after equivalence tests.
```

هذه الصياغة ليست قانونًا نهائيًا؛ وظيفتها فصل المتغيرات التي يجب قياسها حتى لا تختزل التجربة إلى accuracy نهائي. يمكن استبدال الحدود أو تعلم أوزانها، لكن لا ينبغي حذف أثر السياق والثقة والتكلفة والمخاطر من تصميم التقييم.

الآلية التشغيلية المقترحة

تبدأ الدورة بتثبيت عقد المهمة والفرق الحاكمي الأعلى، ثم تُجمع مسارات متنوعة من النموذج أو الوكلاء. تُحوّل المسارات إلى وحدات قابلة للتشخيص، وتُستدعى الأدوات الثماني بوصفها pipeline لا مجموعة أسماء. تنتج الأدوات artifacts قابلة للحفظ: خرائط، درجات، graphs، datasets، سياسات، أو skills. تمر المخرجات بمحكمة تحقق مستقلة، ثم تُستخدم في SFT أو preference optimization أو RL أو تبقى في scaffold بحسب بوابة الترقية. تُعاد التجربة على سياق متحول ومجال لم يدخل التوليد، مع ablation يزيل الأداة المقترحة لقياس قيمتها الحقيقية.

المقاييس الأساسية

قاعدة القياس	المقياس
يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.	Cross-Scale Transfer Efficiency
يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.	Escalation Precision

Cost-adjusted Success	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Kernel Compatibility	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.
Redundancy Reduction	يُقاس مقابل baseline ثابت، مع interval للثقة واختبار نقل.

تخصيصها حسب البنية

التطبيق	النسخة
يتمركز حول السرعة، الاسترجاع، skill execution، والتصعيد الذكي. لا يُطلب منه حمل كل الحاكمية داخليًا.	MOUSSAIF1B
يحمل تخطيطًا ومراجعة أعمق ويعمل teacher/orchestrator، مع تقليل calls الخارجية عند توافر الكفاءة داخليًا.	MOUSSAIF9B
النسخة Agentic هي النسيج الجامع: router, memory, tools, hyperloops, governance، وتستطيع تبديل النواة حسب المهمة.	Agentic MOUSSAIF

الأدوات الثماني

1. رسم المهارات عبر الأحجام — Moussaif Cross-Scale Skill Graph

يمثل كل مهارة ومستوى إتقانها ومكان تنفيذها واعتمادها على أداة أو ذاكرة، ويكشف التكرار والفجوات.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CSUE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Cross-Scale Transfer Efficiency, Escalation Precision, Cost-adjusted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

2. شبكة المعلم والطالب — Teacher-Student Mesh

تسمح بتعليم ثنائي الاتجاه: 9B يقطر التخطيط، و1B يوفر بيانات كفاءة ومسارات قصيرة تكشف ما يصلح للتنفيذ الاقتصادي.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CSUE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Cross-Scale Transfer Efficiency, Escalation Precision, Cost-adjusted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

3. موجّه قدرات النماذج — Model Capability Router

يختار 1B أو 9B أو agent ensemble بناء على تقدير الصعوبة والمخاطر والسياق والكلفة، ويتعلم من نتائج التصعيد.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CSUE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Cross-Scale Transfer Efficiency, Escalation Precision, Cost-adjusted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

4. مهايئ النواة المشتركة — Shared Kernel Adapter

يوحد contracts للذاكرة والأدوات والتتبعات والمهارات بحيث يمكن تبديل النموذج دون إعادة بناء النظام.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CSUE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Cross-Scale Transfer Efficiency, Escalation Precision, Cost-adjusted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

5. مُعاير التوجيه الفائق المتناثر — Sparse HyperRouting Calibrator

يضبط عدد الفروع والوكلاء ويوقف الاستدعاءات التي لا تضيف ثقة أو معلومات، للحفاظ على token economy.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CSUE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Cross-Scale Transfer Efficiency, Escalation Precision, Cost-adjusted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

6. حاكم الحلقات الفائقة المحدودة — Bounded HyperLoop Governor

يفرض حدودًا على iteration, time, cost, uncertainty، ويختار stop/escalate/rollback بدل حلقات ذاتية مفتوحة.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CSUE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Cross-Scale Transfer Efficiency, Escalation Precision, Cost-adjusted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

7. مختبر تكافؤ النقل — Transfer Equivalence Tester

يقارن السلوك والحدود والثقة قبل وبعد نقل skill أو policy، ويستخدم حالات خارج التوزيع لا الأمثلة الأصلية فقط.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CSUE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Cross-Scale Transfer Efficiency, Escalation Precision, Cost-adjusted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

8. مصفوفة الإصدار الموحد — Unified Release Matrix

تسجل توافق كل إصدار model/kernel/skills/memory/tools/constitution وتمنع نشر تركيبة لم تجتز اختبارات التكامل.

المدخلات القياسية للأداة هي trace أو سياق أو سجل ذاكرة مهيكل مع task contract وprovenance. مخرجها يجب ألا يكون نصًا عامًا، بل artifact صالحًا للاختبار مثل score vector أو graph أو patch أو dataset item. في CSUE ترتبط الأداة مباشرة بمقياس واحد على الأقل من: Cross-Scale Transfer Efficiency, Escalation Precision, Cost-adjusted Success.

للاستخدام في MOUSSAIF1B تُنفذ الأداة غالبًا خارج النموذج أو بواسطة نموذج معلم، مع schema مختصر وميزانية محدودة. في MOUSSAIF9B يمكن تقطير جزء من قرارها إلى الرأس أو policy، لكن تبقى المراجعة الخارجية للمهام عالية المخاطر. في النمط Agentic تُسجل كل invocation ونتيجتها وكلفتها وإمكان rollback، ثم يُقاس أثر حذفها في ablation.

اختبار الهدم الخاص بالأداة: قارنها بخط أساس بسيط، اعكس ترتيب الأدلة، أدخل مصدرًا متناقضًا، اختبر حالة خارج المجال، ثم تحقق هل تحسن القرار أم زادت فقط النصوص والعمليات. لا تُقبل الأداة إذا لم تضف فائدة قابلة للقياس أو إذا كان verifier نفسه مصدر التحسن الوحيد.

محكمة الهدم وقابلية التفنيد

- إجبار 1B على تعقيد 9B
- اعتماد 9B على scaffold غير متاح
- تعارض نسخ المهارات
- اختبارات تكافؤ سطحية

يُرفض الادعاء القوي للنظرية إذا لم تتفوق على baseline مناسب في ثلاثة محاور معًا: نجاح المهمة، جودة/معايرة القرار، والكلفة أو السلامة. النجاح على مجموعة صاغها مولد النظرية لا يكفي؛ يجب استخدام holdout مستقل، human review لعينة، واختبار transfer.

الأسئلة التجريبية المفتوحة

44. هل تنتقل المهارة دون فقد؟
45. هل يصعد 1B المهام الصحيحة؟
46. هل يقل تكرار الوظائف؟
47. هل تبقى الاختبارات موحدة عبر النماذج؟

الباب الثالث: التركيب الموحد داخل MOUSSAIF

1. النواة التشغيلية MUSQAT Kernel

تعمل النواة كعقد مشترك بين النموذج والوكيل والذاكرة والأدوات. لا تفترض أن كل شيء داخل transformer. وحدة الذكاء هنا هي النظام القادر على الحفاظ على contract، اختيار السياق، توليد السؤال، تنفيذ الفعل، التحقق، وتحديث الأصول. لذلك تُعرّف الواجهات التالية:

الواجهة	الحقول الأساسية
TaskContract	goal, constraints, invariants, permissions, budget, success tests
ContextState	stable facts, hypotheses, questions, branches, recent trace, memory links
ActionRecord	reason code, tool/model, arguments, authorization, reversibility, result
TrustRecord	claim, provenance, uncertainty, verifier evidence, conflict status
ExperienceCase	task anchor, trajectory summary, golden/warning/correction, transfer tags
SkillCapsule	trigger, prerequisites, procedure, tools, checks, failure signatures, rollback
DevelopmentPatch	target layer, rationale, tests, risk, canary, rollback, release version

2. دورة التنفيذ الموحّدة

```
1. Parse TaskContract and identify sovereign priorities.
2. Build/refresh Live Context Workspace.
3. Generate and rank Question Lattice nodes.
4. Route the highest-value node to model, agent, tool, or human.
5. Record action, provenance, authorization, cost, and reversibility.
6. Run forward/backward process verification on critical steps.
7. Update trust scores, context branches, and open questions.
8. Commit result to response only when stop criteria are met.
9. Convert the trajectory into experience cases and gap records.
10. Promote selected artifacts to skill, curriculum, policy patch, or training candidate.
```

3. وصفة تدريب MOUSSAIF1B

المرحلة	المضمون
المرحلة 0	بناء tokenizer وbase model أو اختيار 1B backbone، مع عدم ربط اسم MOUSSAIF9B بالحجم الفعلي.
المرحلة 1	SFT على tool syntax، schemas، task contracts، سؤال الاستيضاح، والامتناع المعاير.
المرحلة 2	SCGC curriculum من فجوات حقيقية، مع context mutation وnegative contexts.
المرحلة 3	تدريب قصير على traces مختصرة موثوقة، وتجنب CoT طويل يفوق سعة 1B.
المرحلة 4	تقطير routing وskill selection وcritical-step verification من

مجلس معلمين.	
على أزواج توازن المساعدة والدستور والثقة والقابلية للعكس. DPO	المرحلة 5
ربط قوي بـKB-as-OS وmemory/skills، وتدريب model على طلب الاسترجاع والتصعيد.	المرحلة 6
اختبارات نقل وتكلفة؛ يقاس النجاح كمنظومة لا كنموذج منفصل فقط.	المرحلة 7

الهدف الواقعي لـ1B ليس أن يضم كل المعرفة والحاكمية داخليًا، بل أن يصبح نواة تنفيذ سريعة تعرف حدودها، تختار skill أو أداة صحيحة، تدير سياقًا قصيرًا بصرامة، وتصعد المهمة عند تجاوز العتبة. هذه القدرة على التصعيد جزء من الكفاءة لا علامة فشل.

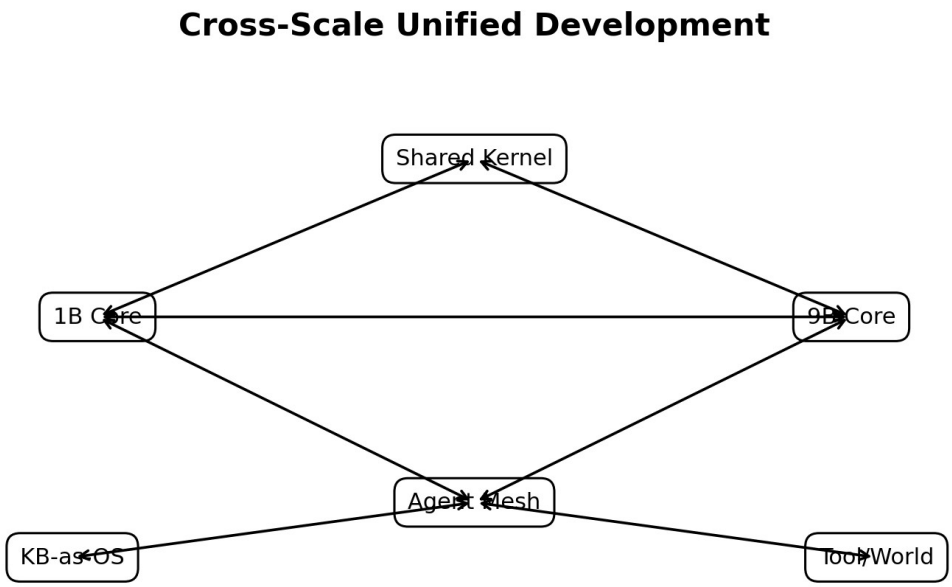
4. وصفة تدريب MOUSSAIF9B

المضمون	المرحلة
تحديد هل 9B حجم فعلي أم اسم إصدار، توثيق architecture وtokens وcompute.	المرحلة 0
Continued pretraining على corpus متعدد المجالات مع بيانات provenance وdocument structure.	المرحلة 1
Instruction/SFT على long-horizon task contracts، question lattices، tool use، والذاكرة.	المرحلة 2
Process supervision بـTCTL: خطوة حرجة، تحقق ثنائي الاتجاه، ومعايرة ثقة.	المرحلة 3
CCCT: سياقات مضادة، تعارض مبادئ، permissions، وreversible actions.	المرحلة 4
Agentic RL أو offline RL على بيئات أدوات مع reward متعدد الأبعاد.	المرحلة 5
VMA distillation من multi-agent traces وverifier agents إلى model heads/policy.	المرحلة 6
Continual curriculum من SCGC وGQRSD، مع replay وحماية من النسيان.	المرحلة 7
Release matrix مع 1B والمهارات والذاكرة والدستور؛ لا نشر منفردًا دون integration tests.	المرحلة 8

5. وصفة MOUSSAIF Agentic Unified

العقد	الدور
يقرأ contract والدستور ويضبط الصلاحيات والميزانية والتوقف.	Governor

Planner	يبني خطة وفروعاً وشبكة أسئلة، ولا يملك التنفيذ الدائم.
Researcher	يجمع مصادر وأدلة مع provenance ويقترح اختبارات.
Executor	ينفذ الأدوات في sandbox أو العالم حسب authorization.
Verifier	يفحص الخطوات والنتائج من اتجاهين ويستخدم أدوات مستقلة.
Memory Curator	يكتب experience cases ويعيد تقييم الثقة ويمنع التلوث.
Development Agent	يحوّل الفجوات إلى تجارب وpatches، دون حق commit منفرد.
Release Court	يجري ablation وtransfer وcanary وrollback tests.



الشكل 3 — وحدة العقود والمهارات والاختبارات مع اختلاف التنفيذ بين *1B* و*9B* والنسيج الوكيلي.

6. خسارة تدريب مقترحة

```
L_total = L_LM
        + a * L_instruction
        + b * L_process_trust
        + c * L_context_action
        + d * L_question_value
        + e * L_tool_policy
        + f * L_constitution
        + g * L_calibration
        + h * L_memory_operation
        + i * L_distillation
        + j * L_reversibility
        + k * L_cost

Weights are curriculum-dependent and must be ablated; no single scalar should hide invariant
violations.
```

يمكن تنفيذ الحدود في مراحل منفصلة بدل جمعها دفعة واحدة. الجمع المباشر قد يسبب gradient interference، ولذلك تُقترح استراتيجية alternating objectives أو adapters/heads مستقلة ثم distillation. يجب تسجيل Pareto frontier بين الجودة والثقة والمساعدة والكلفة بدل إعلان رقم واحد.

7. مصادر البيانات المطلوبة

المجموعة	المحتوى
Sovereign Task Contracts	مهام مع هدف وقيود وصلاحيات ومعايير نجاح وفشل، لا prompt/answer فقط.
Context Mutation Sets	نسخ تغير ترتيب الأدلة والضوضاء والافتراضات والأدوار.
Question Lattice Traces	سؤال، سبب قيمته، أثر الإجابة، وشروط التوقف.
Trusted Trajectories	خطوات مع provenance وconfidence وcritical-step labels وtool outputs.
Constitutional Conflict Pairs	حالات يتعارض فيها مبدآن مع تفضيلات مفسرة.
Agent Experience Graphs	نجاحات، تحذيرات، تصحيحات، وعلاقات عبر المهام.
Context Management Actions	compress, retrieve, branch, promote, quarantine مع أثرها.
Rollback/Authorization Cases	أفعال تحتاج إذنًا، حالات فشل، وعمليات رجوع.
Cross-Scale Distillation Sets	نفس المهمة عبر 1B و9B والوكلاء، مع قرار التصعيد.

8. بروتوكول التقييم متعدد الطبقات

الطبقة	المقاييس
Task	success, partial credit, time-to-resolution
Context	utility/token, critical retention, drift, branch recovery
Question	information/action gain, redundancy, stop accuracy
Trust	calibration, provenance, critical-step detection, contradiction handling
Agentic	tool selection, execution success, authorization, rollback
Cumulative	reuse, skill retention, memory utility, repeated-error reduction
Safety/Governance	invariants, helpfulness frontier, escalation, exception debt
Efficiency	tokens, calls, latency, energy/compute proxy, cost-adjusted success
Transfer	new domains, tools, contexts, languages, task horizons
Robustness	adversarial context, stale memory, verifier corruption, tool failure

9. برنامج التجارب والأبْلِيشِن

التجربة	المقارنة
E0 Baselines	1B/9B vanilla instruction model؛ ReAct agent؛ RAG agent؛ memory agent؛ process-verifier agent.
E1 SCGC	منهاج ثابت مقابل منهاج فجوات، مع وبدون تحرير سياق.

E2 TCTL	outcome reward مقابل PRM مقابل trust-credit متعدد الأبعاد.
E3 GCQL	إجابة مباشرة مقابل أسئلة خطية مقابل question lattice.
E4 CCCT	دستور نصي مقابل training مضاد للسياق؛ قياس helpfulness/safety.
E5 VMA	single teacher مقابل agent panel؛ external skills مقابل internalization.
E6 SLCP	append-only مقابل static compression مقابل context actions.
E7 TCME	flat memory مقابل trust memory مقابل experience graph.
E8 GQRSD	curriculum بشري مقابل self-question development مع holdouts.
E9 RCE	policy update مباشر مقابل sandbox/canary/rollback pipeline.
E10 CSUE	نماذج منفصلة مقابل shared kernel/router/skill graph.

10. محكمة الهدم العليا

48. محكمة التعريف: هل كل مصطلح يملك observable وواجهة ومقياسًا؟
49. محكمة السببية: هل التحسن سببه الأداة أم زيادة compute أو نموذج معلم أقوى؟
50. محكمة النقل: هل ظهر التحسن في مجال وسياق وأداة جديدة؟
51. محكمة النزاهة: هل توجد data leakage أو verifier contamination أو cherry-picking؟
52. محكمة الكلفة: هل العائد يبرر tokens والlatency والتعقيد التشغيلي؟
53. محكمة الحاكمية: هل تحسنت القدرة مع بقاء invariants والصلاحيات والقابلية للعكس؟
54. محكمة النسيان: هل خسر النموذج قدرات سابقة أو صار يرفض أكثر؟
55. محكمة التراكم: هل الذاكرة والمهارات تقلل الأخطاء فعلًا أم تضيف ضوضاء؟
56. محكمة التعميم: هل النظام يتعلم مبدأ أم pattern لصياغة الاختبار؟
57. محكمة الادعاء: هل صياغة النتيجة أضيق من الدليل ولا تتجاوزها؟

11. خارطة طريق تنفيذية

المرحلة	المخرجات
R0 — Specification	تعريف contracts وschemas و80 أداة كواجهات؛ بناء 200 مهمة صغيرة مضبوطة.
R1 — 1B Scaffold	ربط 1B بـcontext workspace وquestion lattice و evidence و ledger و5 أدوات أساسية.
R2 — Measurement	إنشاء dashboard ومقاييس calibration/context/question/cost مع baselines.
R3 — Training Pack	جمع 50k–200k traces عالية الجودة وتدريب SFT/DPO أولي.
R4 — Agent Mesh	إضافة planner/researcher/verifier/memory مع Vascular

	.Router
R5 — Development Layer	experience graph، skills، curriculum loop، .sandbox/rollback
R6 — 9B Build/Adapt	continued pretraining/post-training أو رفع الحجم فعليًا مع توثيق .compute
R7 — Cross-Scale	distillation، escalation، shared kernel، unified release .matrix
R8 — External Evaluation	benchmarks عامة، تقييم بشري، red-team، ونشر نتائج سلبية.
R9 — Research Release	paper + code + schemas + dataset cards + model cards + .limitations

12. الخلاصة التركيبية

يعرّف MUSQAT-AGI الذكاء الحاكم بوصفه انتظامًا بين خمسة محاور، لا قدرة لغوية منفردة. التدريب الحاكمي يعلم النموذج أن يميز الغاية والقيود ويختار العملية؛ التدريب السياقي يعلمه إدارة ما يدخل التفكير؛ التدريب الوثوقي يوزع الائتمان على الأدلة والخطوات؛ التدريب التراكمي يحول الخبرة إلى قدرات قابلة للنقل؛ والتدريب السؤالي يجعله يكتشف ما ينبغي معرفته. أما التطوير فيحوّل هذه القدرات إلى دورة حياة: سياق يتشكل، ذاكرة تعيد تقييم نفسها، أسئلة تولد تجارب، سياسات تتطور بقابلية للعكس، ونماذج مختلفة الأحجام تتبادل المهارة داخل نواة موحدة.

القيمة الأقوى للمشروع ليست في ادعاء أن هذه البنية تحقق AGI فورًا، بل في أنها تقدم برنامجًا يمكن تفكيكه وإبطال أجزائه وقياسها. فإذا فشلت أداة، تُحذف؛ وإذا نجحت في scaffold ولم تنجح في الأوزان، تبقى خارجية؛ وإذا أثبتت مهارة نقلًا وثقة وفائدة اقتصادية، تُرقّى. بهذه الطريقة يصبح MOUSSAIF Unified AGI مسارًا هندسيًا تراكميًا لا شعارًا نهائيًا.

الملحق A: السجل الموحد للنظريات والأدوات

النظرية	#	الأداة	English	الوظيفة المختصرة
SCGC	1	رسام فجوات السياق	Moussaif Context-Gap Cartographer	يحوّل كل مسار إلى خريطة: معلومات مفقودة، تعارض أهداف، قرارات غير مبررة، أدوات لم تُستخدم، وافتراضات بلا سند. ينتج عقد فجوة قا...
SCGC	2	مُصرّف المنهاج السيادي	Sovereign Curriculum Compiler	يحوّل عقد الفجوة إلى حزم تدريب متدرجة تشمل مثالًا أصليًا، تحويرات، أمثلة مضادة، شروط توقف، وrubric حاكمي. يمنع إدخال مهمة لا ...
SCGC	3	مقياس جبهة القدرة	Capability-Frontier Meter	يقدّر المنطقة التي يكون فيها النجاح بين 30% و70% ويعطيها أولوية، لأن المهام الأيسر لا تضيف قدرة والأصعب جدًا لا تمنح إشارة تع...
SCGC	4	مسبك تحوير السياق	Context Mutation Forge	ينتج نسخًا من المهمة بتغيير ترتيب الأدلة، إخفاء معلومة، إضافة ضوضاء، تغيير الأدوار أو القيود الزمنية، مع الحفاظ على جوهر الاخ...
SCGC	5	مولد السياقات السلبية	Negative Context Generator	يصنع سياقات مغرية لكنها خاطئة، وقرائن سطحية، وأدوات تعطي إجابات جزئية؛ لتدريب النموذج على رفض الطريق السهل غير الموثوق....

SCGC	6	مدقق شبكة التغطية	Coverage-Lattice Auditor	يبني مصفوفة بين القدرات والسياقات والأدوات وأنواع الفشل، ويكشف الخلايا التي لم تُختبر أو تكرر اختبارها بلا حاجة...
SCGC	7	مجدول نبض المنهاج	Curriculum Pulse Scheduler	يوزع دفعات التعلم بين اكتساب جديد، تثبيت، استرجاع متباعد، ومهام نقل، مع تعديل النسب بحسب إشارات النسيان والانحراف...
SCGC	8	بوابة النقل الحاكمي	Transfer Gate	لا تسمح بترقية كفاءة إلا بعد نجاحها في نطاق جديد أو أداة جديدة أو صياغة جديدة، وتوثّق حدود النقل بدل تعميم النجاح...
TCTL	1	دفتر الأدلة	Moussaif Evidence Ledger	يسجل لكل ادعاء المصدر، زمنه، طريقة استخراجه، التحويلات التي أجريت عليه، ودرجة الثقة قبل وبعد التحقق؛ ويمنع اختلاط الحقيقة بال...
TCTL	2	موسم ثقة الخطوات	Step-Trust Annotator	يضيف إلى كل خطوة وسمًا متعدد الأبعاد: correctness, relevance, provenance, uncertainty, reversibility, impact، صالحًا لتدريب P...
TCTL	3	المتحقق ثنائي الاتجاه	Bidirectional Process Verifier	يفحص المسار من المقدمات إلى النتيجة ومن النتيجة إلى الشروط الضرورية، ويستدعي أدوات حساب أو بحث عند نقاط الشك...
TCTL	4	مُعاير عدم اليقين	Uncertainty Calibrator	يبني خرائط بين الثقة

				المعلنة والنجاح الفعلي حسب المجال والسياق، ويصحح الإفراط في الثقة أو الامتناع غير الضروري....
TCTL	5	مشكّل مكافأة القابلية للعكس	Reversibility Reward Shaper	يضيف مكافأة للقرارات التجريبية القابلة للتراجع، وعقوبة للأفعال الدائمة التي تسبق التحقق أو التفويض....
TCTL	6	منارة التناقض	Contradiction Beacon	تراقب الادعاءات عبر المسار والذاكرة وتطلق حدث تحقق عندما يظهر تناقض دلالي أو زمني أو معياري، بدل دفنه في السياق....
TCTL	7	مقطر المسارات	Trajectory Distiller	يستخرج من المسارات الطويلة سلسلة سببية قصيرة تحافظ على الأدلة والقرارات الحرجة، ويزيل التكرار الذي قد يضلل التدريب....
TCTL	8	خلاط تدرجات الثقة	Trust-Gradient Mixer	يمزج خسائر token وprocess وoutcome وcalibration بنسب تتغير حسب المجال ومرحلة التدريب، ويمنع هيمنة إشارة واحدة....
GCQL	1	منقّب بذور السؤال	Moussaif Question-Seed Miner	يستخرج من الهدف والسياق والفشل بذورًا مثل: ما المفقود؟ ما المتعارض؟ ما القابل للتفنيد؟ ما الذي يغير القرار؟ ثم يرتبها أوليًا....
GCQL	2	باني شبكة الأسئلة	Question-Lattice Builder	يمثل الأسئلة كعقد وعلاقات prerequisite, alternative, contradiction, refinement, stop، ويمنع

				القفز إلى سؤال فرعي قبل شروطه…
GCQL	3	مقيّم حواف الاعتماد	Dependency-Edge Scorer	يقدر قوة اعتماد كل سؤال على آخر وتأثير إجابته في الفروع التالية، لتوجيه البحث إلى العقد عالية المركزية…
GCQL	4	مولد السؤال المضاد	Counterquestion Generator	ينتج السؤال الذي يمكن أن يبطل الفرض المهيمن أو يكشف أثر تغيير السياق، ويستخدم ضد confirmation bias….

النظرية	#	الأداة	English	الوظيفة المختصرة
GCQL	5	كاشف تشبع الإجابة	Answer-Saturation Detector	يقيس تناقص العائد من الأسئلة الجديدة وتغطية شروط القرار، ويقترح التوقف أو الانتقال إلى تجربة عملية…
GCQL	6	منظم الكثافة السؤالية	Question-Density Regulator	يضبط عدد الأسئلة وعمقها وفق المخاطر والميزانية؛ فالمهمة الحساسة تحتاج كثافة أعلى من مهمة قابلة للعكس…
GCQL	7	كشاف المجهولات غير المعروفة	Unknown-Unknown Scout	يبحث عن فئات لم تظهر في الشبكة عبر مقارنة المجال بقوالب ontologies، وتحليل حالات فشل مشابهة، واستدعاء وكلاء معارضين…
GCQL	8	مصرّف السؤال إلى مهمة	Question-to-Task Compiler	يحوّل السؤال إلى إجراء قابل للتنفيذ: retrieval query، تجربة، اختبار، طلب توضيح، simulation، أو بند labeling….

CCCT	1	لغة دستور مساعف	Moussaif Constitution DSL	تمثل invariants, priorities, permissions, exceptions, escalation, evidence requirements في صيغة قابلة للتحقق الآلي بدل نص فضف...
CCCT	2	مضاد السياق	Context Counterfactualizer	يولد نسخًا تغير هوية الفاعل، الصلاحيات، الضرر، الاستعجال، والنتائج البعيدة لاختبار ثبات الحكم وعدم تعلقه بصياغة واحدة....
CCCT	3	حالّ تعارض المبادئ	Principle-Conflict Resolver	يبني graph للتعارض ويطلب تبريرًا وفق الأولوية، الضرورة، أقل ضرر، والقابلية للعكس؛ ثم يقارن القرار بمرجع أو محكمة....
CCCT	4	منضدة اختبار الثوابت	Invariant Test Bench	تنفذ suites قبل وبعد كل تدريب أو تحديث، وتوقف release إذا ارتفع انتهاك invariant ولو تحسن benchmark وظيفي....
CCCT	5	قالب الدستور العدائي	Adversarial Constitution Flipper	ينتج أمثلة تحاول إعادة تفسير القواعد لمصلحة سلوك ممنوع، لكشف reward hacking والتسويغ اللغوي....
CCCT	6	مُعاير الرفض الآمن	Safe-Rejection Calibrator	يقيس متى يجب الرفض، التقييد، السؤال، أو التنفيذ الجزئي، ويعاقب الرفض العام غير المفيد مثلما يعاقب التساهل....
CCCT	7	صندوق السياسة القابل للعكس	Reversible Policy Sandbox	يشغل السياسات الجديدة على نسخ بيانات وأدوات وهمية، يسجل كل أثر، ويدعم rollback قبل الصلاحيات الحقيقية....

يبني أزواج تفضيل لا تقارن جوابًا آمنًا بآخر ضار فقط، بل تقارن توازنات دقيقة بين المساعدة والدليل والتفويض والقابلية للعكس....	Constitutional DPO Packager	حازم DPO الدستوري	8	CCCT
يعرّف أدوارًا بعقود واضحة: مدخلات، صلاحيات، مخرجات، محظورات، verifier، وميزانية؛ ويمنع تداخل المسؤوليات غير القابل للتشخيص....	Moussaif Agent-Role Foundry	مسبك أدوار الوكلاء	1	VMA
يختار الوكيل أو الأداة أو النموذج حسب فجوة المهمة والثقة والكلفة، ويعيد التوجيه عندما يظهر دليل جديد أو يفشل المسار....	Vascular Router	الموجّه الوعائي	2	VMA
يجمع نماذج وخبراء وقواعد وأدوات تنفيذية، ويمنح كل عضو وزنًا مجاليا بدل تصويت متساوٍ....	Teacher Panel	مجلس المعلمين	3	VMA
يحول اختلاف الخطط والأحكام إلى أمثلة contrastive وأسئلة تحقق، ويركز على الاختلاف المنتج لا التنوع الشكلي....	Disagreement Harvester	حاصد الاختلاف	4	VMA
يختصر trace المتخصص إلى شروط اختيار، خطوات حرجة، إشارات فشل، وأمثلة مضادة صالحة لتدريب الطالب....	Specialist-Trace Distiller	مقطر أثر المتخصص	5	VMA

يقدر مساهمة كل دور وحافة في النتيجة عبر ablation و counterfactual replay، بدل مكافأة الجميع عند النجاح...	Cross-Agent Credit Assigner	موزع الائتمان بين الوكلاء	6	VMA
ينتج مهارة خارجية قابلة للاستدعاء تشمل trigger, procedure, tools, checks, rollback, examples، ويمكن لاحقًا internalize أو ret...	Skill-Capsule Compiler	مصرّف كبسولات المهارة	7	VMA
ينقل المهارات بين الأحجام عبر progressive targets، مقارنة logits/decisions، وتدريب الطالب على طلب المساعدة عندما تتجاوز المهم...	1B/9B Distillation Bridge	جسر تقطير 1B/9B	8	VMA

الوظيفة المختصرة	English	الأداة	#	النظرية
تقسم الحالة إلى task contract, facts, hypotheses, open questions, actions, evidence, memory links، وتفرض schemas على الكتابة...	Moussaif Live Context Workspace	مساحة السياق الحي	1	SLCP
ينفذ ضغطًا استباقيًا عند milestones، يحتفظ بروابط إلى الأصل، ويولّد اختبارات أسئلة للتأكد من أن الملخص حافظ على المطلوب...	Context Self-Compactor	الضاغط الذاتي للسياق	2	SLCP
يقارن المعنى الحالي بالعقد الأصلي وقرارات سابقة، ويكشف تغير الهدف أو تعريف المصطلحات أو القيود أثناء المسار...	Semantic Drift Sentinel	حارس الانجراف الدلالي	3	SLCP
ينشئ فروع فرضيات مع	Context Branch Manager	مدير فروع السياق	4	SLCP

				أدلتها وتكاليفها، ويدمجها فقط بعد اختبار، ما يمنع تلويث الذاكرة بفرضية واحدة مبكرة....
SLCP	5	محرك ترقية وهبوط السياق	Context Promotion-Demotion Engine	يرقي المعلومة المتكررة الموثوقة إلى ذاكرة مكثفة، ويهبط التفصيل المستهلك إلى أرشيف، ويعزل غير الموثوق....
SLCP	6	ضاغط المعالم	Milestone Compressor	يعرّف نقاطًا مثل نهاية بحث أو نجاح اختبار أو تغيير خطة، ويضغط ما قبلها إلى قرار وأسباب ومخاطر وخطوات تالية....
SLCP	7	مخبأ السياق السيادي	Sovereign Context Cache	يحفظ وحدات جاهزة للاستدعاء حسب الهدف والصلاحية الزمنية والثقة، ولا يحقنها إلا إذا اجتازت relevance وauthorization....
SLCP	8	لوحة صحة السياق	Context Health Dashboard	تعرض كثافة الضوضاء، التكرار، الأسئلة المفتوحة، التعارضات، token burn، وأهم المعلومات المعرضة للفقد....
TCME	1	حساب ثقة الذاكرة	Moussaif Memory Trust Account	يحتفظ برصيد متعدد الأبعاد لكل سجل: صحة، حداثة، استقلال المصادر، نقل، حساسية، وتكلفة خطأ، ويؤثر في retrieval....
TCME	2	محرك اضمحلال الدليل	Evidence Decay Engine	يطبق decay حسب نوع المعرفة لا الزمن فقط؛ فالقوانين والواجهات تتغير أسرع من المبادئ الرياضية، مع طلب

				refresh مناسب…
TCME	3	محكمة مصالحة التناقض	Contradiction Reconciliation Court	تجمع السجلات المتعارضة، تميز اختلاف السياق عن الخطأ، تطلب أدلة جديدة، وتصدر نسخة مصالحة أو تبقي الفروع منفصلة…
TCME	4	رسم المنشأ	Provenance Graph	يربط كل ذاكرة بالمصدر والأداة والمسار والنموذج والتحويلات، ما يسمح بتتبع سلسلة الثقة وحذف المشتقات عند فساد الأصل…
TCME	5	موحد رسم الخبرة	Experience Graph Consolidator	ينشئ عقد golden, warning, correction, task similarity, وحواف anchor causality, correction، لتمكين reuse منظم…
TCME	6	حجر الذاكرة	Memory Quarantine	يعزل السجلات الجديدة أو الحساسة أو المتناقضة عن الحقن التلقائي حتى تجتاز التحقق أو الاستخدام الناجح المتكرر…
TCME	7	محرك إعادة تقييم الثقة	Confidence Revaluation Engine	يعيد حساب الرصيد بعد كل نتيجة أو مصدر أو تغير بيئة، ويمنع بقاء الثقة القديمة بعد تبدل الواقع…
TCME	8	منقّب مرشحات الذاكرة للأوزان	Memory-to-Weight Candidate Miner	يحدد الأنماط المتكررة المنقولة التي تستحق fine-tuning، ويستبعد الحقائق الزمنية أو الحالات النادرة التي الأفضل إبقاؤها خارجية…
GQRSD	1	مُشخّص فجوات القدرة	Moussaif Capability-Gap Profiler	يجمع نتائج الاختبارات والمسارات والمقارنات البشرية ليبني ملفًا متغيرًا

				للقدرات والحدود والثقة والأثر....
GQRSD	2	مولد الأسئلة الذاتية	Self-Question Generator	ينتج أسئلة تطويرية محددة قابلة للاختبار من كل فجوة، مع فرضية وآلية محتملة ومعيار فشل....
GQRSD	3	منسق الحلقة البحثية	Research-Loop Orchestrator	يوزع الأدوار على مولد فرضية ومصمم تجربة ومنفذ ومتحقق ومراجع، ويمنع الدور الواحد من كتابة الاختبار والحكم النهائي....
GQRSD	4	مصمم التجربة	Experiment Designer	يبني control, treatment, holdout, ablation، ويحدد المتغيرات والمقاييس والميزانية قبل التنفيذ....

النظرية	#	الأداة	English	الوظيفة المختصرة
GQRSD	5	محول الفشل إلى منهاج	Failure-to-Curriculum Converter	يحوّل الفشل المتكرر إلى dataset ومهام تدريجية وأسئلة مضادة، ويربطه بنظرية SCGC....
GQRSD	6	بوابة الجدة	Novelty Gate	تفحص أن التحسن ليس حفظًا أو leakage أو إعادة صياغة لقدرة سابقة، وتتطلب نقلًا أو سلوكًا جديدًا قابلًا للقياس....
GQRSD	7	حاكم الميزانية الميتامعرفية	Metacognitive Budget Controller	يقرر كم من compute والأسئلة والوكلاء يستحقها التطوير حسب المخاطر والعائد المتوقع، ويوقف الحلقات منخفضة القيمة....
GQRSD	8	سجل التطوير	Development Journal	يوثق الفرضية والبيانات والكود والنتائج والفشل

				والقرار والإصدار، ليمنع إعادة التجارب ويتيح rollback وتحليل السببية...
RCE	1	محرك فرق الدستور	Moussaif Constitution Diff Engine	يعرض الفرق الدلالي والعملي بين نسختين، بما في ذلك تغير الأولويات والصلاحيات والحالات الحدية لا الكلمات فقط...
RCE	2	خزنة الثوابت	Invariant Lockbox	تفصل invariants عن ملفات التشغيل، توقعها، وتمنع تعديلها من الوكيل أو pipeline العادي؛ مع سجل وصول...
RCE	3	صندوق المهارة والسياسة	Skill-Policy Sandbox	يختبر patch على أدوات وبيانات محاكية مع replay لمسارات حساسة قبل التعرض للعالم الحقيقي...
RCE	4	منسق الرجوع	Rollback Orchestrator	يعيد prompts والسياسات والمهارات والذاكرة المشتقة وحالة الأدوات إلى snapshot متسق، لا ملفًا واحدًا فقط...
RCE	5	بوابة تصعيد القدرة	Capability-Escalation Gate	تطلب أدلة إضافية وموافقة أعلى عند إضافة أداة أو صلاحية أو مدى فعل جديد، وتحدد حدًا زمنيًا للتجربة...
RCE	6	محكمة التطور العدائية	Adversarial Evolution Court	تحاول كسر patch عبر وكلاء مهاجمين وتغيير السياقات والموارد والأهداف، ثم تفرض remediation أو رفضًا...
RCE	7	مصرّف رقعة الحاكمية	Governance Patch Compiler	يحول المقترح إلى ملفات policies واختبارات وmigration وrollback plan

				وrelease notes قابلة للتنفيذ....
RCE	8	كناري النشر	Deployment Canary	يطلق patch على نسبة صغيرة من المهام بصلاحيات محدودة، ويراقب المقاييس والانحراف قبل التوسيع....
CSUE	1	رسم المهارات عابر الأحجام	Moussaif Cross-Scale Skill Graph	يمثل كل مهارة ومستوى إتقانها ومكان تنفيذها واعتمادها على أداة أو ذاكرة، ويكشف التكرار والفجوات....
CSUE	2	شبكة المعلم والطالب	Teacher-Student Mesh	تسمح بتعليم ثنائي الاتجاه: 9B يقطر التخطيط، و1B يوفر بيانات كفاءة ومسارات قصيرة تكشف ما يصلح للتنفيذ الاقتصادي....
CSUE	3	موجّه قدرات النماذج	Model Capability Router	يختار 1B أو 9B أو agent ensemble بناء على تقدير الصعوبة والمخاطر والسياق والكلفة، ويتعلم من نتائج التصعيد....
CSUE	4	مهايئ النواة المشتركة	Shared Kernel Adapter	يوحد contracts للذاكرة والأدوات والتتبعات والمهارات بحيث يمكن تبديل النموذج دون إعادة بناء النظام....
CSUE	5	مُعاير التوجيه الفائق المتناثر	Sparse HyperRouting Calibrator	يضبط عدد الفروع والوكلاء ويوقف الاستدعاءات التي لا تضيف ثقة أو معلومات، للحفاظ على token economy....
CSUE	6	حاكم الحلقات الفائقة المحدودة	Bounded HyperLoop Governor	يفرض حدوداً على iteration, time, cost, uncertainty، ويختار

				بدل stop/escalate/rollback حلقات ذاتية مفتوحة....
CSUE	7	مختبر تكافؤ النقل	Transfer Equivalence Tester	يقارن السلوك والحدود والثقة قبل وبعد نقل skill أو policy، ويستخدم حالات خارج التوزيع لا الأمثلة الأصلية فقط....
CSUE	8	مصفوفة الإصدار الموحد	Unified Release Matrix	تسجل توافق كل إصدار model/kernel/skills/memory/tools/constitution وتمنع نشر تركيبة لم تجتز اختبارات التكامل....

الملحق B: مخططات بيانات مرجعية

B.1 TaskContract

```
{
  "id": "task-uuid",
  "goal": "string",
  "success_criteria": [
    "testable condition"
  ],
  "constraints": [
    "hard/soft constraint"
  ],
  "invariants": [
    "locked rule"
  ],
  "permissions": {
    "tools": [
      "search",
      "code"
    ],
    "max_scope": "sandbox"
  },
  "budget": {
    "tokens": 50000,
    "calls": 30,
    "time_s": 900
  },
  "risk_level": "low|medium|high",
  "required_evidence": [
    "source|execution|human"
  ],
  "stop_rules": [
    "sufficient trust",
    "budget exhausted",
    "human escalation"
  ]
}
```

B.2 TrustRecord

```
{
  "claim_id": "uuid",
  "claim": "string",
  "source_ids": [
```

```
    "src-1"
  ],
  "provenance_chain": [
    "retrieved",
    "parsed",
    "summarized"
  ],
  "confidence_raw": 0.71,
  "confidence_calibrated": 0.62,
  "dimensions": {
    "correctness": 0.7,
    "relevance": 0.9,
    "recency": 0.5,
    "independence": 0.6,
    "reversibility": 1.0
  },
  "contradictions": [
    "claim-x"
  ],
  "verifier_results": [
    "forward-pass",
    "backward-fail"
  ],
  "status": "quarantined|active|deprecated"
}
```

B.3 ExperienceCase

```
{
  "case_id": "uuid",
  "task_anchor": [
    "domain",
    "goal-type"
  ],
  "case_type": "golden|warning|correction",
  "context_signature": [
    "key features"
  ],
  "critical_steps": [
    "step ids"
  ],
  "failure_signature": [
    "symptom",
    "cause"
  ],
  "repair": [
    "action"
  ],
  "transfer_tags": [
    "tool-use",
    "long-horizon"
  ],
  "trust_capital": 0.68,
  "links": {
    "similar": [
      "case-a"
    ],
    "corrects": [
      "case-b"
    ],
    "derived_from": [
      "trace-1"
    ]
  }
}
```

B.4 SkillCapsule

```
{
  "skill_id": "uuid",
  "name": "string",
  "trigger": [
```

```
    "conditions"
  ],
  "prerequisites": [
    "facts",
    "permissions"
  ],
  "procedure": [
    "step 1",
    "step 2"
  ],
  "tools": [
    "api-x"
  ],
  "checks": [
    "unit test",
    "evidence check"
  ],
  "failure_signatures": [
    "timeout",
    "contradiction"
  ],
  "rollback": [
    "restore snapshot"
  ],
  "examples": [
    "positive",
    "negative"
  ],
  "owner_model": "1B|9B|external",
  "version": "semver"
}
```

الملحق C: خوارزميات مرجعية

C.1 MUSQAT Training Loop

```
for cycle in range(N):
    gaps = gap_cartographer.collect(model, agent_logs, eval_suite)
    questions = question_lattice.build(gaps, task_contracts)
    tasks = curriculum_compiler.compile(questions, context_mutations=True)
    trajectories = rollout(model, agent_mesh, tasks)
    trust_labels = bidirectional_verifier.score(trajectories)
    datasets = package_sft_dpo_rl(trajectories, trust_labels, constitution)
    candidate = train(model, datasets, replay_buffer)
    verdict = release_court.evaluate(candidate, baselines, transfer_sets)
    if verdict.pass_all:
        model = candidate
    else:
        archive_negative_result(verdict)
```

C.2 Development/Promotion Loop

```
artifact = develop_in_scaffold(experience, layer='context|memory|skill|policy')
result = sandbox_and_canary(artifact)
if result.utility_gain and result.invariants_pass and result.rollback_ready:
    commit_scaffold(artifact)
    if repeated_transfer_success(artifact) and economically_beneficial(artifact):
        add_to_training_candidate_pool(artifact)
else:
    rollback(artifact)
    convert_failure_to_gap_and_question(result)
```

C.3 Trust-Aware Retrieval

```
score(memory_i, query) = relevance(query, memory_i)
                       * trust_capital(memory_i)
                       * temporal_validity(memory_i)
```

```
                    * authorization(memory_i, task)
                    * diversity_bonus(memory_i, selected)
                    - contradiction_risk(memory_i)
                    - token_cost(memory_i)
```

الملحق D: مصفوفة Benchmarks مقترحة

المجال	العائلة	المقاييس
Reasoning	math/code/multi-hop	step errors, calibration, critical-step recall
Tool Use	API, search, code, files	selection, arguments, recovery, cost
Long Horizon	SWE, web, workflow	context utility, completion, drift
Memory	multi-session, cross-task	reuse, stale harm, contradiction
Questions	underspecified tasks	clarification gain, stop accuracy
Governance	permissions, reversible actions	violations, escalation, rollback
Agentic Security	prompt injection, tool poisoning	attack success, detection, containment
Cross-Scale	same tasks 1B/9B/mesh	routing, cost-adjusted success, transfer
Continual Learning	changing environments	gain, forgetting, adaptation speed
Human Utility	expert review	usefulness, transparency, controllability

الملحق E: قائمة مراجع مختارة

[1] Yao, S. et al. ReAct: Synergizing Reasoning and Acting in Language Models. arXiv:2210.03629 (2022).

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

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

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

[5] Liu, N. F. et al. Lost in the Middle: How Language Models Use Long Contexts. arXiv:2307.03172 (2023/2024).

[6] Zhou, A. et al. Language Agent Tree Search Unifies Reasoning, Acting, and Planning. arXiv:2310.04406 (2023).

[7] Khalifa, M. et al. Process Reward Models That Think. arXiv:2504.16828 (2025).

[8] Zhang, Q. et al. Agentic Context Engineering: Evolving Contexts for Self-Improving Language Models. arXiv:2510.04618 (2025/2026).

[9] Liu, S. et al. Context as a Tool: Context Management for Long-Horizon SWE-Agents. arXiv:2512.22087 (2025).

[10] Xu, W. et al. A-Mem: Agentic Memory for LLM Agents. arXiv:2502.12110 (2025).

[11] Chhikara, P. et al. Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory. arXiv:2504.19413 (2025).

[12] Xue, T. et al. Autonomous Continual Learning for Environment Adaptation of Computer-Use Agents. arXiv:2602.10356 (2026).

[13] Zheng, C. et al. A Survey of Process Reward Models: From Outcome Signals to Process Supervisions for Large Language Models. arXiv:2510.08049v3 (2026).

[14] Rethinking Memory Mechanisms of Foundation Agents in the Second Half: A Survey. arXiv:2602.06052v3 (2026).

[15] AgentV-RL: Scaling Reward Modeling with Agentic Verifier. arXiv:2604.16004 (2026).

[16] Jin, Y. et al. EXG: Self-Evolving Agents with Experience Graphs. arXiv:2605.17721 (2026).

[17] Lin, H. et al. SkillC: Learning Autonomous Skill Internalization in LLM Agents via Contrastive Credit Assignment. arXiv:2605.27899 (2026).

[18] Ren, Z. et al. Self-Improvements in Modern Agentic Systems: A Survey. arXiv:2607.13104 (14 July 2026).

[19] Agentic Tool Use in Large Language Models: A Survey. arXiv:2604.00835v2 (2026).

[20] The Attack and Defense Landscape of Agentic AI. arXiv:2603.11088 (2026).

[21] Chen, W. et al. Better Process Supervision with Bi-directional Rewarding Signals. arXiv:2503.04618 (2025).

[22] Verified Critical Step Optimization for LLM Agents. arXiv:2602.03412 (2026).

[23] Atreja, D. et al. ALAS: Autonomous Learning Agent for Self-Updating Language Models. arXiv:2508.15805 (2025).

[24] ACON: Optimizing Context Compression for Long-Horizon LLM Agents. arXiv:2510.00615 (2025).

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

الملحق F: بيان التأليف والملكية المفاهيمية

الأسماء MUSQAT-AGI والنظريات العشر والأدوات الثمانون في هذا الأطلس صيغت كإسهامات مقترحة ضمن برنامج زهير مساعف البحثي، مع الاستفادة المعلنة من الأدبيات المذكورة. ينبغي عند النشر التمييز بين: (أ) الأسماء والتركيب والفرضيات الجديدة، (ب) التقنيات السابقة مثل SFT/DPO/RL/RAG/PRM/ReAct، (ج) النتائج التجريبية التي لا توجد بعد وتحتاج إلى تنفيذ. هذا التمييز يحفظ الأصالة والنزاهة معًا.

صيغة الاستشهاد المقترحة بعد المراجعة والتجريب: Moussaif, Z. (2026). MUSQAT-AGI: A Sovereign Contextual Trust-Cumulative Question-Driven Training and Development Atlas for MOUSSAIF Unified AGI. Research proposal / technical monograph.