MOUSSAIF AGI RESEARCH

REFERENCE-FREE COGNITION

AND THE ZERO-ANCHOR LIMIT

الإدراك المتحرر من المرجع وحدّ المرساة الصفرية

نحو نظرية حسابية للتفكير عبر فكّ المرجعية وإعادة تأسيسها

Working Paper v1.0 | MOUSSAIF AGI

10 سبتمبر 2026

الملخص

تطرح هذه الورقة سؤالًا يبدو في البداية فلسفيًا، لكنه يتحول عند صياغته حسابيًا إلى مشكلة مركزية في هندسة الذكاء العام الاصطناعي: هل يستطيع نظام معرفي أن يفكر دون مرجع؟ وإذا أُزيلت اللغة، والذاكرة المسترجعة، والهدف، والهوية، والتصنيفات، والسياق الإدراكي الجاري، فهل يبقى «تفكير»، أم لا يبقى إلا ديناميك معرفي سابق على التفكير الرمزي؟

نقترح ضمن إطار MOUSSAIF AGI التمييز بين ثلاثة مستويات: Reference-Bound Thought، و Reference-Reduced Cognition، وAbsolute Reference-Free State. ندافع عن فرضية أن النوع الثالث لا يمكن أن يكون «تفكيرًا» بالمعنى الحسابي الصارم، لأن أي انتقال معرفي قابل للتمييز يفترض على الأقل علاقة بين حالتين؛ لكن يمكن الاقتراب من حد نسميه Zero-Anchor Limit (ZAL)، حيث تُخفَّض علاقة الحالة الداخلية بالمراجع الصريحة إلى الحد الأدنى، بينما تستمر ديناميكيات غنية داخل فضاء latent غير لغوي.

تستفيد الفكرة من Symbol Grounding، وWorld Models، وActive Inference، وrelational reasoning، وlatent-space reasoning، لكنها تقترح دورة معرفية مختلفة: Ground → De-anchor → Free Latent Dynamics → Re-anchor → Falsify → Act. نسمي هذه الدورة Moussaif Reference Oscillation (MRO). التحرر من المرجع هنا ليس غاية، بل أداة لاكتشاف فرضيات وعلاقات لا يسمح بها الإطار المرجعي الأول، على أن تُجبر النتيجة لاحقًا على العودة إلى الواقع والقياس والذاكرة الموثقة والاستدلال القابل للاختبار.

> الفرضية المركزية: الذكاء العام الأقوى ليس نظامًا بلا مراجع، بل نظامًا يستطيع التحكم في درجة ارتباطه بالمراجع، وتعليق بعضها مؤقتًا، ثم إعادة تأسيسها وفحصها قبل الفعل.

الكلمات المفتاحية

AGI؛ MOUSSAIF AGI؛ Reference-Free Cognition؛ Zero-Anchor Limit؛ Grounding؛ De-Anchoring؛ Latent Reasoning؛ World Models؛ Metacognition؛ Falsification؛ Ontology Revision؛ Cognitive Architecture.

1. المقدمة: هل يمكن التفكير دون شيء يُفكَّر فيه؟

نفترض حالة معرفية z_t. في أبسط تعريف، التفكير ليس مجرد وجود حالة داخلية، وإنما حدوث تحول منظم يمكنه حمل فرق قابل للاستغلال أو التفسير.

$$z_t \rightarrow z_{(t+1)}$$

إذا كانت الحالة التالية مطابقة للحالة السابقة في المعنى المعلوماتي المطلق، فلا توجد إضافة معرفية. أما إذا ظهر فرق بينهما، فقد ظهرت علاقة، ومن ثم ظهر الحد الأدنى من المرجعية.

$$z_{(t+1)} = z_t \quad \Rightarrow \quad \text{no new distinguishable cognitive transition}$$

$$z_{(t+1)} \neq z_t \quad \Rightarrow \quad \text{difference} \Rightarrow \text{relation} \Rightarrow \text{minimal reference}$$

> المبدأ الأول - **Minimal Relationality of Thought**: كل تفكير يتطلب اختلافًا واحدًا على الأقل، وكل اختلاف يخلق علاقة، وكل علاقة تشكل حدًا أدنى من المرجع.

بهذا المعنى، يبدو Absolute Reference-Free Thought متناقضًا حسابيًا: فإذا أزلنا الأشياء واللغة والذاكرة والزمان والاتجاه وكل اختلاف بين الحالات، أزلنا الشروط نفسها التي تسمح لنا بتسمية العملية «تفكيرًا». لكن هذا لا ينفي إمكانية حالة معرفية لا تعتمد مباشرة على مراجع صريحة، وهي المساحة التي تستهدفها هذه الورقة.

2. ما المرجع في النظام المعرفي؟

نعرّف Cognitive Reference بأنه أي متغير أو بنية تُستخدم لتقييد تفسير الحالة المعرفية الحالية أو تطورها. قد يكون المرجع كلمة، مفهومًا، ذاكرة، كائنًا بصريًا، هدفًا، اعتقادًا سابقًا، قاعدة، إطارًا زمانيًا أو مكانيًا، نموذجًا للذات، أو فرضية حول شكل الإجابة المقبولة.

$$A_t = \{a_1, a_2, \ldots, a_n\} \quad \text{(Anchor Set)}$$

$$z_t = f(x_t, A_t, M_t, G_t)$$

يمثل x_t المدخل الراهن، وM_t الذاكرة، وG_t الهدف، بينما تمثل A_t شبكة المراسي التي تضبط التفسير. معظم أنظمة اللغة الحديثة شديدة الارتباط بهذه المراسي: prompt، وسلسلة tokens، والسياق المسترجع، والتعليمات، والأنماط المتعلمة كلها تحدد مساحة الاحتمالات.

لكن الارتباط بالمرجع قد يتحول من ميزة إلى قيد. فإذا بدأ النظام من فرضية خاطئة، يمكنه تنفيذ reasoning متماسك داخل frame خاطئ. وإذا استرجع ذاكرة مشابهة، قد يسقط نمطًا قديمًا على مسألة جديدة. وإذا قُيّد الاستدلال باللغة، قد تصبح البنية اللغوية نفسها حدًا لما يستطيع النظام تمثيله.

> **Reference Lock-in**: فشل لا ينشأ من العجز عن التفكير، بل من العجز عن مغادرة الإطار الذي يحدث التفكير داخله.

3. من **Symbol Grounding** إلى **Reference Liberation**

تُظهر مشكلة Symbol Grounding أن التعريف الرمزي الخالص لا يكفي لتأسيس معنى متصل بالعالم. لكن AGI يواجه مشكلة معاكسة كذلك: بعد تأسيس الرمز في العالم، كيف نمنع هذا التأسيس من التحول إلى سجن معرفي؟

Grounding ↔️ Controlled De-grounding

Grounding يمنح الإدراك المعنى والصدق والاتصال بالواقع. Controlled De-grounding يمنحه القدرة على إعادة تركيب فضاء الفرضيات، شريطة أن يكون مؤقتًا وقابلًا للعكس وأن تتبعه إعادة تأسيس واختبار. النظام الذكي جدًا يحتاج العمليتين، لا واحدة منهما فقط.

World Models تسمح للوكيل بالتخطيط داخل تمثيل مضغوط لديناميكيات البيئة. تقترح MOUSSAIF خطوة إضافية: ليس فقط تشغيل agent داخل world model، بل السماح مؤقتًا بتخفيف بعض البنى التي يستخدمها world model نفسه لتقسيم العالم.

Not only: What happens next?

But also: What latent structures become possible if the current ontology is wrong?

4. تعريف **Zero-Anchor Limit**

نقترح قياس مقدار اعتماد حالة cognition على مجموعة المراجع بواسطة تقريب معلوماتي. لا نقصد أن هذه الصياغة هي مقياس نهائي، بل نقطة بداية قابلة للتحويل إلى estimator تجريبي.

$$RD_t = I(Z_t ; A_t) / (H(Z_t) + \varepsilon)$$

إذا اقترب RD_t من الواحد كانت الحالة الداخلية شديدة الارتهان بالمراجع الموجودة. أما عندما يقترب من الصفر، يصبح تطور الحالة أقل قابلية للتنبؤ من المراسي الصريحة. لكننا لا نريد إزالة الترابط الزمني بين الحالات الداخلية، لأن ذلك قد يحول العملية إلى ضوضاء.

$$\text{Desired: } I(Z_t ; A_t) \downarrow$$

$$\text{Required: } I(Z_t ; Z_{(t-1)}) > 0$$

Zero-Anchor Cognition لا تعني **Zero-Structure Cognition**.

بذلك يصبح ZAL حدًا نظريًا بين أقصى التثبيت الخارجي وأقصى حرية داخلية متماسكة. لا يفترض الحد غياب البنية، بل غياب الهيمنة المباشرة للمراسي الصريحة على مسار cognition.

5. لماذا **Reference-Free** لا تعني **Random**؟

هذه النقطة تحدد نجاح النظرية أو سقوطها. حذف السياق أو تشويش activations لا يساوي تفكيرًا متحررًا من المرجع؛ قد يكون مجرد فقدان للمعلومة. الحالة المطلوبة يجب أن تجمع ثلاثة شروط: انخفاض Anchor Dependence، ارتفاع Internal Coherence، وارتفاع Re-groundability.

$$Q_{ZAL} = C(z) \times N(z) \times R(z) \times V(z)$$

حيث C(z) تمثل coherence، وN(z) novelty، وR(z) قابلية إعادة grounding، وV(z) قابلية verification. إذا ارتفعت novelty بينما انهارت R وV، لم نحصل على creativity بل على hallucination. لذلك لا يمكن تقييم ZAL بمعيار الجدة وحده.

6. الصلة بـ **Latent Reasoning**

تظهر عدة اتجاهات أن reasoning لا يجب أن يبقى محصورًا في سلسلة كلمات صريحة. Continuous latent reasoning يعيد إدخال hidden states في مسار الاستدلال بدل إجبار كل خطوة على التحول إلى token، بينما تدفع concept-level models نحو وحدات تمثيل أعلى من الكلمة. وتُظهر approaches أخرى إمكانية إجراء computation داخلي قبل المخرجات اللغوية.

لكن وجود latent state لا يثبت وجود «تفكير غير مرجعي». قد تكون الحالة الكامنة مجرد encoding عالي الكثافة للسياق السابق، وقد تستخدم shortcuts بدل reasoning سببي. لذلك تضيف MOUSSAIF شرطًا أقوى: لا يكفي أن يكون التفكير غير لغوي؛ يجب قياس مقدار تحرره فعليًا من Anchors الأصلية، ثم اختبار ما إذا كان هذا التحرر يولد تحسينًا سببيًا قابلًا للتكرار.

Moussaif Reference Oscillation (MRO) .7

تقترح الورقة أن cognition عالي المستوى لا ينبغي أن يبقى grounded طوال الوقت ولا ungrounded طوال الوقت. الأفضل هو oscillation منظمة بين التثبيت والتحرير وإعادة التثبيت.

$$\text{Ground} \rightarrow \text{DeAnchor} \rightarrow \text{Explore} \rightarrow \text{ReAnchor} \rightarrow \text{Falsify} \rightarrow \text{Act}$$

في Ground Phase يبني agent أفضل نموذج متاح للحالة بالاعتماد على perception، وmemory، وtools، وقواعد المجال، والبيانات والسياق. في DeAnchor Phase لا تُحذف المعلومات، بل تُخفّض قدرة بعض المراسي على التحكم المباشر في الديناميكيات.

يمكن مثلًا تعطيل أسماء الكيانات مع الاحتفاظ بعلاقاتها؛ أو تجميد retrieval لمنع analogical contamination؛ أو إسقاط الصياغة اللغوية مع إبقاء هندسة العلاقات في latent space؛ أو خفض وزن الهدف النهائي مؤقتًا لدراسة فضاء الحالات دون انجذاب مبكر إلى حل بعينه.

$$z_{t+k} = F_{\theta}(z_{t+k-1}, \eta_k)$$

الحد η_k ليس ضوضاء اعتباطية، بل controlled exploratory perturbation. بعد ذلك تأتي Re-Anchoring لتحويل البنية المكتشفة إلى propositions أو objects أو causal relations أو predictions أو actions قابلة للفحص، ثم تمر النتيجة عبر Falsification Court قبل قبولها.

Falsification Court .8

أخطر ما قد يحدث في Reference-Free Cognition هو أن يكتشف النظام فضاءً داخليًا غنيًا وجميلًا ولكنه غير موجود في الواقع. لذلك لا يكون معيار النجاح Interesting(z)، بل Interesting(z) ∧ SurvivesFalsification(z).

$$H = \{H_1, H_2, \ldots, H_n\}$$

$$\text{For each } H_i, \text{ actively search for } E^{-}(H_i)$$

تولد المحكمة فرضيات بديلة، وتبحث عن الأدلة التي يمكن أن تُبطل الفرضية بدل الاقتصار على الأدلة المؤيدة لها. وتفصل بين hypothesis generation وhypothesis acceptance. هذه البنية ضرورية لأن de-anchoring يزيد مساحة الإمكان، وبالتالي يزيد أيضًا مساحة الخطأ الممكن.

Anchor Graph .9

بدل التعامل مع context كسلسلة tokens فقط، تقترح الورقة تمثيله كرسم بياني للمراجع.

$$G_A = (V_A, E_A)$$

تمثل العقد المراجع، وتمثل الحواف علاقات اعتماد cognition عليها. يمكن أن توجد عقد من نوع self، goal، object، word، memory، belief، rule، temporal origin، spatial origin، social model، causal assumption. يقوم De-Anchoring Controller بعمليات جراحية على الرسم بدل حذف السياق دفعة واحدة.

$$G_A \rightarrow G'_A$$

ثم يقارن النظام trajectory cognition قبل العملية وبعدها. إذا ظهرت فرضية نافعة فقط بعد إزالة anchor محدد، يصبح لدينا دليل سببي أولي على أن ذلك المرجع كان يعمل كقيد معرفي. ويمكن أن تتحول هذه الآلية إلى أداة interpretability: أي المراجع كانت ضرورية للجواب؟ وأيها كانت تمنع ظهور البدائل؟

10. Reference Counterfactuals

التفكير counterfactual التقليدي يغير قيمة متغير داخل ontology معطاة: ماذا لو كان X مختلفًا؟ أما Reference Counterfactual فيسأل: ماذا لو توقف X عن كونه مرجعًا مميزًا أصلًا؟

Classical: What if X were different?

MOUSSAIF: What if X ceased to be a privileged reference?

الأول يغير العالم داخل representation، والثاني يغير representation نفسها. في الفيزياء: ليس فقط ماذا لو تغيرت سرعة الجسم، بل ماذا لو لم يكن «الجسم» هو الوحدة الأساسية لوصف الظاهرة؟ في البرمجة: ليس فقط تغيير الخوارزمية، بل اختبار ما إذا كانت decomposition الحالية للمشكلة خاطئة. في التخطيط: ليس فقط كيف أصل إلى الهدف، بل هل تمثيل الهدف نفسه يمنع حلًا أفضل؟

> القدرة المركزية هنا ليست البحث داخل فضاء الحلول، بل القدرة على إعادة تعريف فضاء الحلول.

11. Structural Thought: العلاقات قبل الأشياء

تقود النظرية إلى فرضية أكثر جرأة: قد توجد مرحلة من cognition لا تكون فيها «الأشياء» primitive الأساسية. يمكن أن ينشأ object بوصفه attractor مستقرًا داخل شبكة علاقات، بدل افتراض objects جاهزة ثم إضافة العلاقات إليها.

Objects → Relations (standard view)

Relations → Stable Objects (proposed exploratory view)

هذا لا يعني أن الأشياء وهم، بل أن تقسيم الواقع إلى objects قد يكون ناتجًا من عملية ضغط واستقرار داخل النظام المعرفي. لذلك يصبح سؤال MOUSSAIF: ماذا يمكن للوكيل أن يكتشف عندما يعلّق مؤقتًا افتراض أن objects الحالية هي التقسيم الصحيح للواقع؟

12. التفكير والوعي: فصل يجب الحفاظ عليه

من الضروري عدم القفز من computation إلى consciousness. هذه الورقة لا تدعي أن Zero-Anchor State واعية، ولا أن latent dynamics دليل على خبرة ذاتية. المقترح هنا نظرية عن cognition والمرونة التمثيلية، لا برهانًا على machine consciousness.

$$\text{Thinking} \neq \text{Conscious Awareness}$$

إذا عرّفنا thinking بأنه تحول منظم لتمثيلات قابلة للتمييز، فإنه يحتاج بنية وعلاقات. أما phenomenological awareness، إن كانت موجودة بوصف مستقل، فهي مسألة أخرى تحتاج معايير وتجارب منفصلة. هذا الفصل يمنع استخدام لغة الوعي لتغطية فجوات تجريبية في نموذج AGI.

13. **Benchmark** مقترح: **Moussaif Anchor Ablation Benchmark (MAAB)**

لكي تكون النظرية علمية يجب أن تعرض نفسها لاحتمال الفشل. نقترح MAAB لقياس ما إذا كان de-anchoring يضيف قدرة مستقلة بدل أن يكون مجرد perturbation عشوائي.

- C0: Normal Reasoning - الحالة المرجعية.
- C1: Language Anchor Reduced - خفض الاعتماد على التمثيل اللغوي الصريح.
- C2: Memory Anchor Reduced - تجميد أو تقليل retrieval المؤثر.
- C3: Goal/Ontology Anchor Reduced - خفض هيمنة الهدف أو ontology الحالية.
- C4: Adaptive MRO - اختيار المراسي المراد تعليقها بصورة ديناميكية ثم إعادة grounding.

يجب قياس Correctness وCalibration وTransfer وNovelty وRe-groundability وCausal Robustness، لا accuracy وحدها.

$$\text{RLG} = \text{Performance_MRO} - \text{Performance_Grounded}$$

$$\text{RLG}^* = \Delta\text{SolveRate} + \alpha\Delta\text{Novelty} - \beta\Delta\text{Hallucination} - \gamma\Delta\text{VerificationCost}$$

إذا لم يتفوق Adaptive MRO بصورة متكررة في مسائل تتطلب frame revision مع بقاء hallucination وverification cost تحت السيطرة، تضعف النظرية أو تُرفض نسختها الحالية.

14. التجربة الحاسمة: **Escape Your Own Representation**

نقترح بناء عوالم اصطناعية تكون فيها جميع بيانات التدريب متوافقة مع ontology O1، بينما لا يمكن حل الاختبار النهائي إلا باكتشاف ontology مختلفة O2. النظام التقليدي سيبحث عن حلول داخل O1، بينما المطلوب هو تعديل نظام التمثيل نفسه.

$$\text{Training regime: Solutions}(O_1) \text{ appear sufficient}$$

Test requirement: O_1 → O_2

نقارن LLM reasoning وCoT وlatent reasoning وMOUSSAIF-MRO. إذا اكتشف MRO إعادة التمثيل المطلوبة بوتيرة أعلى، مع بقاء النتائج قابلة للفحص وإعادة grounding، نحصل على دليل تجريبي أقوى على أن de-anchoring يضيف قدرة مستقلة.

> المعيار الأقوى ليس: هل وجد النظام حلًا؟ بل: هل استطاع الهروب من **representation** التي جعلت الحل غير مرئي؟

Self-Reference Ablation .15

يمكن تطبيق المبدأ على أكثر المراسي تعقيدًا: نموذج الذات. يحتاج AGI إلى self-model لمعرفة قدراته وأدواته وحدوده ومسؤولياته، لكنه قد يصبح مصدر تحيز قوي. لذلك يمكن خفض Influence(S_t) معرفيًا دون تعطيل حدود السلامة أو الصلاحيات التنفيذية.

S_t = Agent Self Model

Epistemic mode: Influence(S_t) ↓

Operational safety constraints: remain active

يستطيع النظام أن يسأل: ما الحل الذي سيظهر لو لم أعط أفضلية لما أعرفه عن نفسي؟ ثم يعيد إدخال self constraints قبل الفعل. وهذا يرسخ مبدأ مهمًا: epistemic de-anchoring لا يعني operational de-anchoring.

Creativity as Controlled Ontology Destruction .16

يمكن إعادة تعريف creativity من منظور هذه النظرية. الإبداع ليس randomness، وليس مجرد novel combination؛ بل قدرة مؤقتة على تليين القيود التمثيلية، الخروج من local minima، بناء تركيب جديد، ثم إعادة اختباره.

Creativity = Constraint Relaxation + Reconstruction + Validation

نسمي هذه العملية Cognitive Ontology Annealing. كما تسمح الحرارة في annealing للنظام بمغادرة minimum محلي، تسمح Reference Temperature T_R بمغادرة conceptual minima. عند T_R منخفض جدًا يصبح النظام شديد الالتزام بالمراجع، وعند T_R مرتفع جدًا ينهار coherence.

T_R(t): low → high → low

بذلك يصبح جدول Reference Temperature متغيرًا متعلمًا أو متكيفًا مع حالة المشكلة، وليس إعدادًا ثابتًا.

MOUSSAIF AGI البنية المعمارية المقترحة داخل .17

Input → World Model → Anchor Extraction → Grounded Solver

→ DeAnchor Controller → Latent Dynamics Chamber

→ Hypothesis Crystallizer → ReGrounder

→ Falsification Court → World-Effect Loop → Memory/Writeback

القاعدة المعمارية الأهم هي الفصل بين مساحة توليد الفرضيات ومساحة اعتمادها. لا يتم تخزين «الخيال» مباشرة كمعرفة. Writeback يقبل فقط ما اجتاز إعادة grounding بدرجة ثقة معلومة، أو يخزنه بوسم hypothesis غير مؤكدة مع provenance واضح.

18. متى يجب تشغيل **de-anchoring**؟

لا ينبغي تشغيل ZAL باستمرار، تكلفته الحسابية مرتفعة وقد يخفض reliability. لذلك نقترح Reference Escape Trigger (RET) بوصفه دالة في عدم اليقين والجمود والتناقض واحتمال الوقوع في minimum تمثيلي محلي.

$$RET = w_1 U + w_2 S + w_3 C + w_4 L$$

حيث U = uncertainty، وS = search stagnation، وC = contradiction density، و L = local-minimum probability. إذا تجاوز RET عتبة τ، تبدأ دورة MRO. وهكذا تصبح Reference-Free Cognition أداة metacognitive يستدعيها النظام حين يشتبه أن المشكلة ليست غياب الحل بل سوء representation.

> الذكاء الأعلى ليس فقط أن تبحث بعمق داخل نموذجك، بل أن تعرف متى ينبغي أن تشك في النموذج الذي تستخدمه للبحث.

19. ثلاث درجات للذكاء التمثيلي

- L1 - Solve within a frame: إيجاد حلول داخل ontology معطاة.
- L2 - Select among frames: اختيار تمثيل من مجموعة تمثيلات متاحة.
- L3 - Generate, suspend and destroy frames: إنشاء ontology، استخدامها، التشكيك فيها، تفكيكها وبناء غيرها.

تقترح الورقة أن General Intelligence القوي يحتاج قدرة من المستوى L3 في بعض المجالات على الأقل، لأن البيئة المفتوحة لا تضمن أن representation المناسبة موجودة مسبقًا في training distribution أو في مكتبة الأدوات.

MOUSSAIF Zero-Anchor Hypothesis .20

> الذكاء العام لا يبلغ أعلى قدراته عندما يُزال المرجع نهائيًا، بل عندما يصبح قادرًا على التحكم بدرجة ارتباطه بالمراجع: **grounding** عند الحاجة إلى الحقيقة والفعل، **de-anchoring** عند الحاجة إلى تجاوز الإطار، ثم **re-grounding** والفحص قبل تحويل الإمكان إلى معرفة أو فعل.

$$AGI^* = \arg\max_{\pi} [\ Grounding \times Exploration \times FramePlasticity \times Falsifiability\ ]$$

$$AGI^* \neq \max(Ungroundedness)$$

الهدف إذن ليس إنتاج نظام يعيش في فضاء حر دائم، بل نظام يستطيع تعديل صلابة مراجع التفكير كما يعدل خوارزم البحث معلماتها، ثم يعود إلى قيود الواقع كلما انتقل من hypothesis إلى belief أو action.

21. تنبؤات قابلة للدحض

• في المهام التي يكون فيها frame الحالي صحيحًا وبسيطًا، يجب ألا يعطي MRO فائدة كبيرة، وقد يكون أبطأ من reasoning العادي.

• في المهام التي تتطلب تغيير ontology، يجب أن ترتفع فائدة MRO مقارنة بالbaselines.

• التخفيف العشوائي للمراجع يجب أن يؤدي إلى novelty أعلى لكن reliability أقل من de-anchoring المتعلم والموجَّه.

• أفضل أداء يجب أن يظهر عند درجة وسطية ومتغيرة من Reference Temperature، لا عند أقصى grounding ولا أقصى التحرر.

• إذا لم يمكن إعادة grounding لفرضيات ZAL، فيجب أن تنخفض قيمتها المعرفية مهما بدت مبتكرة.

• Anchor Graph interventions الدقيقة يجب أن تكون أكثر فاعلية وأقل كلفة من حذف السياق الشامل إذا كانت النظرية صحيحة.

22. مخاطر النظرية وضوابطها

توجد أربعة مخاطر رئيسية. الأول هو إعادة تسمية randomness على أنه cognition، يعالج بمقاييس coherence وre-groundability. الثاني هو ارتفاع hallucination، يعالج بالفصل بين generation وacceptance وFalsification Court. الثالث هو فقدان traceability، يعالج بتسجيل Anchor Graph والتحويلات التي أُجريت عليه. الرابع هو إسقاط مفاهيم الوعي على computation، يعالج بالإبقاء على الادعاء في مستوى cognition فقط.

كما يجب التمييز هندسيًا بين تعطيل مرجع معرفي وتعطيل قيد أمان أو صلاحية تشغيلية. لا ينبغي لمكوّن de-anchoring أن يمتلك صلاحية إزالة الحواجز التنفيذية التي تحكم الأفعال الخارجية.

23. النتيجة الفلسفية

السؤال الأول كان: هل يمكن التفكير دون أي مرجع؟ الجواب المقترح: التفكير بلا أي علاقة على الإطلاق غير قابل للتعريف حسابيًا، لأن التحول المعرفي نفسه يصنع فرقًا وعلاقة. لكن يمكن وجود cognition يتخلص مؤقتًا من المراجع التي كانت تحدد مضمون التفكير، مع الإبقاء على بنية داخلية قادرة على إنتاج مسارات جديدة.

Thought about things → dynamics from which new things may become thinkable

بهذا المعنى، التجرد لا يعني الفراغ. إنه انتقال من التفكير في الأشياء وفق ontology ثابتة إلى طبقة تسمح بإعادة تكوين ما سيُعد «شيئًا» وما سيُعد «علاقة» وما سيُعد «سؤالًا» أصلًا.

24. الخلاصة

تقدم هذه الورقة أساسًا لما نسميه Reference-Plastic Intelligence. الذكاء التقليدي يسأل: ما الإجابة الصحيحة بالنظر إلى ما أعرفه؟ أما الذكاء الأكثر عمومية فيستطيع أن يسأل أيضًا: أي جزء مما أعرفه هو مجرد إطار يجعل بعض الإجابات غير قابلة حتى لأن تظهر؟

Build References → Reason Through Them → Suspend Them

→ Generate Alternatives → Rebuild Them → Attack Them → Act

قد لا يكون الطريق إلى AGI زيادة طول التفكير فقط، بل تطوير قدرة النظام على التحكم في الشروط التي تجعل شيئًا ما قابلًا لأن يصبح موضوعًا للتفكير أصلًا. ومن هنا نقترح معيارًا أعمق للذكاء التمثيلي:

$$\text{Intelligence} \propto \text{Capacity to revise the space in which intelligence operates}$$

25. برنامج البحث التالي داخل **MOUSSAIF AGI**

لتحويل النظرية إلى برنامج هندسي، تقترح الورقة المسار التالي: بناء Anchor Extractor، تعريف Anchor Graph عملي داخل transformer أو world-model agent، تصميم تدخلات de-anchoring قابلة للعكس، تطوير Reference Escape Trigger، إنشاء Latent Dynamics Chamber، ثم بناء ReGrounder وFalsification Court مستقلين نسبيًا عن generator.

المرحلة التجريبية الأولى يجب أن تبدأ ببيئات اصطناعية صغيرة يمكن التحكم في ontology الخاصة بها، لأن اختبار frame revision في بيئات مفتوحة يجعل من الصعب تحديد ما إذا كان النجاح نتج فعلًا من de-anchoring أو من استدعاء نمط محفوظ. بعد ذلك يمكن الانتقال إلى planning، debugging، causal science، theorem discovery، وopen-ended research.

المراجع المختارة

Harnad, S. (1990). The Symbol Grounding Problem. Physica D: Nonlinear Phenomena, 42(1-3), 335-346.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience, 11, 127-138.

Ha, D., & Schmidhuber, J. (2018). World Models. arXiv:1803.10122.

Battaglia, P. W., et al. (2018). Relational inductive biases, deep learning, and graph networks. arXiv:1806.01261.

Dehaene, S., et al. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. Neuron, 105(5), 776-798.

Zelikman, E., et al. (2024). Quiet-STaR: Language Models Can Teach Themselves to Think Before Speaking. arXiv:2403.09629.

Hao, S., et al. (2024). Training Large Language Models to Reason in a Continuous Latent Space. arXiv:2412.06769.

Meta AI. (2024). Large Concept Models: Language Modeling in a Sentence Representation Space.

Brady, J., et al. (2023). Provably Learning Object-Centric Representations. arXiv:2305.14229.

Chen, X., et al. (2025). Reasoning Beyond Language: A Comprehensive Survey on Latent Chain-of-Thought Reasoning. arXiv:2505.16782.

ملاحظة منهجية

هذه الورقة تقدم فرضية نظرية ومعمارية قابلة للاختبار، ولا تدعي أن Zero-Anchor Limit قد ثبت تجريبيًا أو أن latent cognition تمثل وعيًا ذاتيًا. القيمة العلمية للمقترح تعتمد على بناء estimators واضحة للanchor dependence وعلى نتائج benchmark قابلة للتكرار ومقارنة عادلة مع baselines قوية.