نظرية الاعتبارية العليا
والفرق الحاكمي الأعلى

إطار صوري لتوليد الفروق، وترتيب الأحكام، والتحكم الهرمي في أنظمة الذكاء العام

الفرضية المركزية — الاعتبارية الأعلى لا تختار حكمًا فحسب، بل تحدد أولًا ما الذي يجوز أن يُعد فرقًا حاكمًا، ثم تورّث هذا الفرق إلى ما يليه من معايير وأحكام وقرارات، مع إبقاء السلسلة قابلة للتصحيح الصاعد.

إعداد وتأليف: زهير مساعف | **Zouhair Moussaif**

برنامج Moussaif AGI Theory — ورقة نظرية تأسيسية

يوليو 2026

حالة الإثبات: نظرية مقترحة قابلة للصياغة والاختبار، وليست نتيجة تجريبية مثبتة بعد.

الملخص

تقترح هذه الورقة نظريةً صوريةً للحاكمية المعرفية تنطلق من دعوى أن الفرق الحاكمي الأعلى لا يوجد بوصفه خاصيةً مستقلة جاهزة داخل الأشياء، بل يتحدد داخل «اعتبارية عليا» تختار جهة المقارنة، وغاية الحكم، ومجال الأثر، وزمن التقييم، ورتبة الدليل. وبمجرد تحديد هذه الاعتبارية يتولد فرق محوري يعيد ترتيب الفروق الأدنى؛ فما يلي يتلقى حاكميته مما فوقه حتى يتحول الغرض الأعلى إلى حكم وقرار وفعل. تسمّي الورقة هذا البناء «التوليد الاعتباري المتسلسل للفروق الحاكمية» وتقدمه بوصفه طبقة ميتا-معرفية تصلح لربط التجريد بالسياق، والغاية بالاستدلال، والتخطيط بالتنفيذ.

تتميز النظرية عن الهرمية التقليدية في أنها لا تكتفي بتقسيم المهمة أو رفع مستوى التمثيل، بل تسأل سؤالًا سابقًا: من يملك حق تحديد الفروق ذات الصلة؟ لذلك تفرق بين الوصف، والوظيفة، والسياق، والحاكمية؛ وبين «علو التجريد» و«علو السلطة المعرفية». وقد يكون تمثيل ما شديد التجريد لكنه غير حاكم، بينما يكون اعتبار محدود النطاق أعلى حاكمية لأنه يقلب صدق النتيجة أو سلامة الفعل. ولمنع الدور والتسلط الاعتباطي، تقترح الورقة قيودًا للشرعية، والتغطية، وعدم التناقض، وقابلية الرجوع، وقيمة الأثر، وكلفة الحساب.

تقدم الورقة سبع مسلمات، وتسعة قوانين، ونموذجًا رياضيًا متعدد المستويات، وخوارزمية تشغيلية لاختيار الاعتبارية واستخراج الفرق الأعلى، ومقاييس للتغطية والثبات والحساسية الحاكمية. كما تقترح معمارية AGI تجمع بين Context Graph، وCausal Abstraction، وWorld Model، وMeta-Controller، وFalsification Court، وتوضح نقاط الاندماج مع MCD وQ-MCD وHAKIM-MCD وSparse HyperRouting وBounded HyperLoops وWorld-Effect Loop. وتنتهي ببرنامج اختبار تجريبي يقوم على مهام يتغير جوابها الصحيح بتغير الاعتبارية، واختبارات استئصال Ablation، ومقارنات مع التخطيط الهرمي والاستدلال المسطح.

الكلمات المفتاحية

الحاكمية المعرفية؛ الاعتبارية العليا؛ الفرق الحاكمي؛ Meta-Reasoning؛ Causal Abstraction؛ Hierarchical Planning؛ Context Governance؛ AGI؛ Falsification؛ World Models.

المساهمات الأصلية

- تعريف الاعتبارية العليا بوصفها مولّد فضاء الفروق، لا مجرد منظور وصفي.
- صياغة قانون توريث الحاكمية وقانون التصحيح الصاعد وقيد عدم انقلاب الفرع على الأصل.
- بناء بروتوكول قابل للاختبار لاختيار الاعتبارية وقياس جودة الفرق الحاكمي الأعلى.

خريطة الورقة وبنيتها الحجاجية

الرقم	القسم	الصفحة
1	المشكلة والفرضية المركزية	4
2	موقع النظرية من حالة الفن	5
3	الأساس المفهومي للاعتبارية	6
4	التعريفات والتمييزات	7
5	المسلمات التأسيسية	8
6	القوانين والنظريات الفرعية	9
7	النموذج الرياضي	10
8	الحركة الهابطة والتصحيح الصاعد	11
9	الخوارزمية التشغيلية	12
10	مقاييس التقييم	13
11	معمارية AGI المقترحة	14
12	الاندماج مع منظومة Moussaif	15
13	أمثلة تطبيقية	16
14	أنماط الفشل	17
15	محكمة الهدم والتحقق	18
16	البرنامج التجريبي	19
17	الفرضيات والاستئصال	20
18	المناقشة والخاتمة	21
المراجع		22

تنتقل الحجة من ضبط المفهوم إلى الصياغة الرياضية، ثم المعمارية، ثم الاختبار التجريبي. كل ادعاء أصلي في الورقة معلن بوصفه فرضية قابلة للتكذيب، لا نتيجة مثبتة.

1. المشكلة والفرضية المركزية

تنجح أنظمة الذكاء الحديثة كثيرًا في استخراج أوجه التشابه والاختلاف، لكنها قد تفشل قبل هذه الخطوة: تختار فرقًا صحيحًا داخل إطار مقارنة خاطئ. في مسألة طبية مثلًا، قد يكون الفرق الإحصائي بين علاجين حقيقيًا، لكنه غير حاكم إذا كان الفرق الأعلى متعلقًا بخطر نادر كارثي على فئة محددة. وفي التخطيط، قد تكون أقصر خطة أفضل محليًا لكنها أدنى حاكمية من خطة تحفظ قابلية الرجوع وتمنع قفل النظام. وفي المناظرة، قد يكون الفرق اللفظي شديد الوضوح لكنه تابعًا لفرق أعلى في تعريف محل النزاع. هذه الحالات لا تعاني من نقص معلومات فقط، بل من سوء ترتيب لحقوق التحكم بين الاعتبارات.

الفرضية المركزية هي أن الحكم العام يحتاج إلى طبقة تسبق المقارنة نفسها. هذه الطبقة تختار: وحدة التحليل، وجهة النظر، الغاية، مجال العواقب، الأفق الزمني، والقيود غير القابلة للتجاوز. نسمي ناتج هذا الاختيار «الاعتبارية الحاكمة». وعندما تكون قادرة على إعادة ترتيب شبكة المعايير التي تحتها تسمى «الاعتبارية العليا». أما الفرق الذي يؤدي تغييره إلى إعادة ترتيب الأحكام التابعة فيسمى «الفرق الحاكمي الأعلى».

$$I^* \rightarrow \Delta H^* \rightarrow \{\Delta H_n \dots \Delta H_1\} \rightarrow J \rightarrow D \rightarrow A$$

لا تعني النجمة وجود اعتبار واحد كوني في كل المسائل، بل تشير إلى المرشح الأعلى شرعيةً وتغطيةً بالنسبة إلى مهمة محددة. فقد تتعدد الاعتباريات المشروعة، وقد تتعادل، وقد نحتاج إلى تركيبها. ويظهر الخطر عندما نُثبت المنظومة اعتباريةً واحدة ثم تمنع الآثار الواقعية من مراجعتها. لذلك يتضمن النموذج مسارًا صاعدًا يعيد فحص الاعتبارية عند التناقض أو فشل التنبؤ أو تجاوز حدود الأمان.

سؤال البحث

السؤال الرئيس — كيف يمكن لنظام معرفي أن يختار الاعتبارية التي تملك حق تحديد الفرق الأعلى، ثم ينزل هذا الفرق عبر سلسلة أحكام، من دون الوقوع في الدور، أو الاستبداد المعياري، أو انفصال التجريد عن الواقع؟

أهداف الورقة

- تحويل الفكرة إلى مفاهيم قابلة للتمييز والقياس.
- بناء آلية لاختيار الاعتبارية تحت موارد محدودة.
- تحديد شروط الفشل التي تمنع تحويل النظرية إلى شعار غير قابل للتكذيب.

2. موقع النظرية من حالة الفن

توجد أربعة خطوط بحثية قريبة، لكنها لا تتطابق مع المقترح. أولها Hierarchical Reinforcement Learning، ولا سيما إطار Options الذي مثّل الأفعال الممتدة زمنيًا وسمح بالتخطيط على مستويات مختلفة [1]. تقدم هذه الأعمال دليلًا هندسيًا على أن التجريد الزمني يقلل تعقيد المسارات الطويلة، لكنها غالبًا تفترض أن الهدف أو دالة المكافأة معطاة سلفًا؛ أي إنها تنظم التنفيذ أكثر مما تحاكم جهة الاعتبار التي اختارت الهدف.

الخط الثاني هو Rational Metareasoning، حيث تصبح العمليات الحسابية نفسها أفعالًا لها كلفة وقيمة متوقعة. صاغت أعمال اختيار الحسابات مسألة ميتا-قرار تحدد متى يستمر النظام في التفكير، وأي حساب ينفذ، وكيف يوازن التحسن المتوقع بكلفة الموارد [2][3][14]. يقترب هذا من الحاكمية لأنه يتحكم في التفكير، لكنه لا يميز دائمًا بين اختيار «المزيد من الحساب» واختيار «الإطار الذي يحدد ما يستحق الحساب».

الخط الثالث هو Causal Abstraction، الذي يدرس متى يمثل نموذج عالي المستوى النظام نفسه الذي يمثله نموذج منخفض المستوى، مع حفظ العلاقات التدخلية المناسبة [4][5][6][7]. هذا الخط أساسي لنظريتنا لأن الاعتبارية العليا لا يجوز أن تكون مجرد قصة مريحة؛ يجب أن تحافظ على قدر كاف من البنية السببية، وأن تكشف مقدار الخطأ الناتج عن التجريد. غير أن التجريد السببي يركز عادةً على علاقة نموذجين، بينما تضيف النظرية هنا مسألة السلطة: أي مستوى يجب أن يحكم القرار في هذه المهمة؟

الخط الرابع يشمل Hierarchical Active Inference وHierarchical World Models وHierarchical Reasoning. تبين أعمال حديثة أن التخطيط متعدد الأزمنة والنماذج المنطقية-البصرية الهرمية قد يحسن التعامل مع الآفاق الطويلة ويحد من تراكم الخطأ [8][9][10][11][12][13]. ومع ذلك، لا يساوي «المستوى البطيء» أو «الأكثر تجريدًا» تلقائيًا المستوى الأعلى حاكمية. قد يخطئ المستوى العالي في اختيار غاية أو فرضية، ويجب أن يستطيع المستوى الأدنى إسقاطه بأثر مضاد.

الفجوة التي تعالجها النظرية	ما يقدمه	الحقل
لا يحدد شرعية الغاية أو معيار الفرق	تجريد زمني وتقسيم مهام	HRL
قد يتحكم في مقدار التفكير لا في إطار المقارنة	قيمة الحساب وتخصيص الموارد	Metareasoning
لا يمنح وحده قاعدة لاختيار المستوى الحاكم	حفظ بنية سببية بين المستويات	Causal Abstraction
العلو التمثيلي لا يساوي العلو الحاكمي	تخطيط متعدد الأزمنة وتمثيلات هرمية	World Models/HAI

3. الأساس المفهومي للاعتبارية

الاعتبارية في هذه الورقة ليست رأيًا شخصيًا ولا مجرد زاوية تصوير. إنها بنية تحدد شروط ظهور الصفة في الحكم. فعندما نقول إن جسمين متساويان، يلزم أن نحدد: متساويان في الكتلة أم القيمة أم الخطر أم الوظيفة؟ الاختلاف لا ينعدم خارج الجهة المختارة، لكنه لا يصبح ذا سلطة على الحكم إلا إذا أدخلته الاعتبارية في مجال الصلة. ومن هنا تكون الاعتبارية عملية اختيار لما ينبغي حفظه من الواقع عند الضغط المعرفي.

يمكن تحليل كل اعتبارية إلى سبعة مكونات: موضوع الحكم X، وجهة المقارنة R، الغاية G، السياق C، أفق الزمن T، مجموعة الأدلة E، وحدود السلطة B. وبذلك لا يكفي أن نقول «ننظر من جهة السلامة»؛ بل يجب تحديد السلامة لمن، وفي أي مدة، وتحت أي احتمال، وبأي دليل، وما الذي لا يجوز التضحية به. كل غموض في هذه المكونات يسمح بتبديل الاعتبارية أثناء الاستدلال من غير إعلان، وهو مصدر شائع للتناقض.

$$I = \langle X, R, G, C, T, E, B \rangle$$

تكون الاعتبارية «عليا» عندما تتحقق ثلاثة شروط. أولًا، شرط إعادة الترتيب: إذا تغيرت، تغير وزن مجموعة معتبرة من الفروق الأدنى. ثانيًا، شرط التغطية: تفسر لماذا تُقبل تلك الفروق أو تُهمَل. ثالثًا، شرط المساءلة: تقبل المراجعة بواسطة نتائج أو أدلة تقع تحتها. فلا حاكمية بلا قدرة على التوريث، ولا شرعية بلا قابلية للنقض.

الفرق بين العلو التجريدي والعلو الحاكمي

التمثيل الأعلى تجريدًا يحذف تفاصيل أكثر، أما المستوى الأعلى حاكمية فيملك حق تقرير أي تفاصيل لا يجوز حذفها. قد يكون نموذج «تعظيم الربح» مجردًا، لكن قيدًا جزئيًا يتعلق بحرمة قانونية أو خطر سلامة قد يحكمه. وعليه، لا تصعد الحاكمية دائمًا مع التجريد، بل قد تنتقل أفقيًا أو تنزل مؤقتًا إلى دليل حاسم.

> قاعدة مركزية — الاعتبارية العليا ليست أعلى لأنها أعمّ، بل لأنها أكثر قدرةً على حفظ ما يتوقف عليه صدق الحكم وسلامة أثره، مع أقل فقد غير مبرر.

بهذا المعنى، الاعتبارية جهاز ضغط تركيبي: تختصر الواقع، لكنها تحمل تبريرًا لما أبقته وما أسقطته. وكلما كان الاختصار بلا حساب للفقد، تحول التجريد من أداة معرفة إلى مصدر عمى حاكمي.

4. التعريفات والتمييزات الصورية

المفهوم	التعريف التشغيلي
الاعتبارية $\mathcal{I}$	مواصفة تحدد مجال المقارنة وغايتها وسياقها وأدلتها وحدودها.
الفرق الوصفي ΔD	اختلاف قابل للرصد في خاصية، من دون أن يلزم منه تغير في الحكم.
الفرق الوظيفي ΔF	اختلاف في القدرة أو الدور أو الأداء داخل وظيفة محددة.
الفرق السياقي ΔC	اختلاف لا تظهر دلالته إلا تحت بيئة أو زمن أو مهمة بعينها.
الفرق الحاكمي ΔH	اختلاف يؤدي إلى تغير مرتب في الحكم أو القرار أو بنية المعايير.
الفرق الحاكمي الأعلى ΔH^*	فرق إذا انقلبت قيمته أعيد ترتيب شبكة معتبرة من الفروق والأحكام التابعة.
الاعتبارية العليا $\mathcal{I}^*$	اعتبارية تملك مؤقتًا أعلى حق مشروع في توليد ΔH^* وضبط وراثته.
سلسلة التوريث Γ	المسار الذي تنتقل عبره قيود $\mathcal{I}^*$ إلى الفروق ثم الحكم والقرار والفعل.
مسار التصحيح Ψ	قناة صاعدة تعيد فحص Γ و$\mathcal{I}^*$ عند ظهور تناقض أو أثر مضاد.

تمييزات مانعة للالتباس

- الفرق الأكبر عدديًا ليس بالضرورة أعلى حاكمية؛ قد يكون فرقًا صغيرًا يقلب صلاحية الاستنتاج.
- الاعتبارية الأشمل ليست دائمًا الأصح؛ قد تفقد تفصيلًا حاسمًا بسبب الإفراط في الضغط.
- الأولوية الأخلاقية ليست الشكل الوحيد للحاكمية؛ توجد حاكمية إبستيمية وسببية وتشغيلية وقانونية.
- الحاكمية ليست ثباتًا أبديًا؛ هي تفويض مشروط بمهمة وسياق وأدلة وحدود.

أنواع الحاكمية

نميز بين: حاكمية إبستيمية تحدد ما يكفي لتصديق الدعوى؛ حاكمية سببية تحدد المتغيرات التي تغير النتيجة بالتدخل؛ حاكمية معيارية تحدد ما ينبغي حفظه؛ حاكمية تشغيلية تحدد قيود التنفيذ؛ وحاكمية ميتا-معرفية تحدد أي نوع من الحاكمية يتقدم عند التعارض. لا يلزم أن تتطابق هذه الأنواع، وقد يكون جوهر المسألة هو التفاوض بينها.

> اختبار الفرق الأعلى — اسأل: لو عُكست قيمة هذا الفرق مع ثبات باقي الفروق، هل سيتغير ترتيب البدائل أو شرعية الحكم أو صلاحية النموذج؟ إذا لم يتغير شيء جوهري، فالفرق ليس أعلى حاكمية في هذه المهمة.

5. المسلمات التأسيسية السبع

1. مسلمة التحديد: لا يوجد فرق حاكم بلا اعتبارية تحدد جهة صلته بالحكم.
2. مسلمة التبعية: كل فرق أدنى يستمد وزنه الحاكمي من علاقة مبررة بفرق أعلى أو غاية أعلى.
3. مسلمة الانقلاب: إذا أدى تغيير فرق ما إلى إعادة ترتيب الأحكام التابعة فهو مرشح لمرتبة حاكمية أعلى.
4. مسلمة الحفظ: الاعتبارية المشروعة تحفظ المعلومات الضرورية للصدق أو الأمان أو الغاية، وتصرح بما تفقده.
5. مسلمة الرجوع: كل سلسلة توريث يجب أن تتضمن قناة تمكن النتائج الأدنى من مراجعة المستوى الأعلى.
6. مسلمة الكلفة: اختيار الاعتبارية واختبارها فعل حسابي محدود الموارد؛ ولا يبرر تعقيد ميتا-معرفي بلا قيمة متوقعة.
7. مسلمة التعدد: قد توجد اعتباريات عليا متنافسة؛ الحل المشروع هو ترتيبها أو تركيبها أو إعلان عدم الحسم، لا إخفاء التعارض.

نتائج مباشرة

من مسلمة التحديد يلزم أن عبارة «الفرق واضح» غير كافية علميًا ما لم يذكر إطار الوضوح. ومن مسلمة التبعية يلزم منع المعايير الجزئية من التحول إلى غايات بذاتها؛ فالسرعة مثلًا لا تحكم الصدق إلا إذا فوضتها غاية أعلى ضمن حد. ومن مسلمة الرجوع يلزم أن أي معمارية تمنع الأثر من تعديل النموذج العالي هي معمارية استبدادية معرفيًا، حتى لو كانت عالية الدقة في المتوسط.

أما مسلمة الكلفة فتمنع تضخم طبقة الميتا. قد يصبح النظام منشغلًا باختيار الاعتبارية إلى ما لا نهاية. لذلك لا بد من حد توقف يعتمد على قيمة متوقعة للحساب، وعلى درجة العواقب، وعلى حساسية القرار. في المهام منخفضة المخاطر يمكن استخدام اعتباريات افتراضية سريعة؛ وفي المهام عالية العواقب يجب توسيع محكمة البدائل.

> المبدأ الضابط — كل علو حاكمي مؤقت، وكل تفويض مشروط، وكل ضغط معرفي مطالب ببيان ما أسقطه ولماذا لا يفسد الحكم.

6. القوانين والنظريات الفرعية

1. قانون توليد فضاء الفروق: الاعتبارية لا ترتب فروقًا جاهزة فقط، بل تحدد مجموعة الفروق التي تصبح مرئية ومعتبرة.
2. قانون قرب الحاكمية: ترتفع أولوية الفرق بقدر قربه السببي أو المعياري من الغاية الحاكمة، لا بقدر بروزه الحسي.
3. قانون توريث القيد: كل مستوى أدنى يجب أن يحمل أثرًا قابلًا للتتبع من القيود العليا.
4. قانون منع انقلاب الفرع: لا يجوز لمعيار فرعي أن يهزم الأصل الذي منحه القيمة إلا بآلية مراجعة معلنة.
5. قانون التصحيح الصاعد: تناقض النتائج أو فشل الأثر يرفع إشارة مراجعة إلى أقرب مستوى يحتمل أنه سبب الخطأ.
6. قانون عدم التماثل: مسار الإنزال ليس معكوسًا بسيطًا لمسار الصعود؛ التنفيذ يضغط، والتصحيح يحتاج إعادة استدلال.
7. قانون الحاكمية المركبة: عند تعادل اعتباريات مستقلة، ينشأ حكم مركب بحدود Pareto أو ترتيب معجمي أو تفاوض موزون.
8. قانون الحساسية: كلما زاد تغير القرار عند اضطراب طفيف في اعتبارية ما، زادت الحاجة إلى التحقق منها.
9. قانون التوقف الحاكمي: يتوقف الميتا-استدلال عندما تصبح قيمة اختبار اعتبارية إضافية أقل من كلفته المتوقعة.

نظرية ثبات السلسلة

نسمي السلسلة مستقرة إذا حافظت القرارات المحلية على القيود العليا، ولم تؤد التغذية الراجعة إلى تذبذب غير منته بين اعتبارين. يتحقق الثبات تقريبيًا عندما تكون أوزان التوريث معلنة، وقنوات الصعود ذات عتبات، والتحديثات محدودة الخطوة. أما التغيير الكامل للاعتبارية في كل أثر جزئي فينتج نظامًا شديد الحساسية، بينما منع التحديث ينتج نظامًا متصلبًا. المطلوب منطقة وسطى: قابلية رجوع منضبطة.

نظرية الفجوة الحاكمية

الفجوة الحاكمية هي المسافة بين الفرق الذي يفسر القرار فعليًا والفرق الذي تدعي المنظومة أنه يحكمه. يمكن اكتشافها عبر تفسير سببي أو Ablation: إذا أزيل الفرق المعلن وبقي القرار، فالتفسير زائف أو غير مكتمل. وإذا تغير القرار بعامل لم يظهر في السلسلة، فهناك حاكمية خفية يجب كشفها.

7. النموذج الرياضي متعدد المستويات

لنفرض مجموعة اعتباريات مرشحة $\mathbb{I}$، ومجموعة فروق قابلة للاستخراج Δ، ومجموعة أفعال A. لكل اعتبارية I دالة صلة $r_I(\delta)$، ودالة حفظ h_I، وكلفة حساب c_I، ومخاطر فقد ℓ_I نعرّف درجة شرعية الاعتبارية بوصفها تركيباً من التغطية السببية، والاتساق، وقابلية الرجوع، وملاءمة الغاية، مطروحاً منه الفقد والكلفة.

$$L(I) = \alpha \text{Causal}(I) + \beta \text{Coverage}(I) + \gamma \text{Consistency}(I) + \eta \text{Reversibility}(I) + \theta \text{GoalFit}(I) - \lambda \text{Loss}(I) - \mu \text{Cost}(I)$$

تختار المنظومة الاعتبارية العليا المؤقتة وفق تعظيم الشرعية مع قيد أن تكون القيمة المتوقعة للاختبار موجبة. ثم تحسب لكل فرق δ درجة حاكمية تجمع مقدار تغيير القرار، ونطاق الأحكام المتأثرة، وقربه من الغاية، وثباته عبر السياقات.

$$H(\delta \mid I) = w_1 \cdot \text{DecisionFlip} + w_2 \cdot \text{Scope} + w_3 \cdot \text{GoalProximity} + w_4 \cdot \text{CrossContextStability} - w_5 \cdot \text{Spuriousness}$$

$$I^* = \arg\max_i L(I_i), \quad \Delta H^* = \arg\max_\delta H(\delta \mid I^*)$$

لا يكفي التعظيم المنفرد عندما تتعدد القيم غير القابلة للاختزال. لذلك يمكن أن تكون النتيجة مجموعة Pareto من الاعتباريات. عندئذٍ تستخدم المنظومة واحداً من ثلاثة أنظمة: ترتيب معجمي يحفظ قيوداً صلبة أولًا؛ تجميع موزون عندما تكون المقاييس قابلة للمقارنة؛ أو قرار احترازي Minimax عندما يكون عدم اليقين وعظم الضرر مرتفعين.

توريث الحاكمية

$$h_n = \sigma(W_n\, h_{n+1} + U_n\, x_n - P_n\, v_n)$$

يمثل h_n حالة الحاكمية في المستوى n، وx_n الأدلة المحلية، وv_n إشارات الانتهاك. تسمح W بنقل القيود العليا، بينما تسمح U بإدخال الواقع المحلي، وتسمح P بتخفيض سلطة المستوى الأعلى عند ظهور مخالفة. هذا التوازن يمنع نظامين متطرفين: Top-down لا يرى الواقع، وBottom-up لا يحفظ الغايات البعيدة.

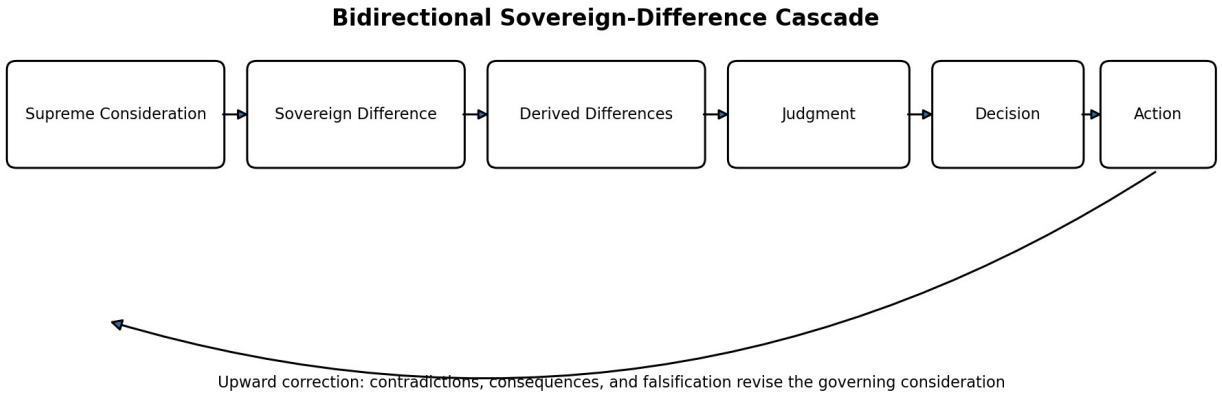
قيمة اختبار اعتبارية جديدة

$$\text{VOC}(I_j) = \mathbb{E}[U(\text{after testing } I_j)] - U(\text{current}) - c(I_j)$$

إذا كانت VOC موجبة تستمر المنظومة في الميتا-استدلال، وإلا تتوقف وتتصرف وفق أفضل سلسلة متاحة. يستلهم هذا المبدأ من Rational Metareasoning، لكنه يطبق قيمة الحساب على اختبار الإطار الحاكم ذاته، لا على تعميق البحث داخل إطار ثابت فقط [2][3].

8. الحركة الهابطة والتصحيح الصاعد

تتكون العملية من دورتين مترابطتين. الدورة الهابطة تبدأ بالغاية أو الدستور، ثم تختار الاعتبارية، ثم تستخرج الفرق الأعلى، ثم تفككه إلى فروق تشغيلية، فتصدر حكمًا وقرارًا وفعلًا. أما الدورة الصاعدة فتبدأ من أثر الفعل: هل تحقق التنبؤ؟ هل ظهرت مخالفة؟ هل استبعد التجريد متغيرًا حاسمًا؟ ثم ترفع الخطأ إلى المستوى الذي يمكنه تفسيره بأقل تعديل ممكن.



الشكل 1. السلسلة الحاكمية ثنائية الاتجاه: إنزال القيود ثم تصحيح الاعتبارية بالأثر.

قاعدة أقل تعديل حاكمي

عند الفشل لا ينبغي إسقاط أعلى اعتبارية فورًا. يبدأ التصحيح من أدنى مستوى قادر على تفسير الانحراف. فإذا كان التنفيذ أخطأ مع بقاء الحكم صحيحًا، يعدل المنفذ. وإذا كانت الفروق التشغيلية ناقصة، يعاد التفكيك. وإذا كان الفرق الأعلى نفسه غير سببي أو غير ذي صلة، يعاد استخراجه. ولا تُراجع الاعتبارية العليا إلا عندما تكون المشكلة في جهة المقارنة أو الغاية أو الحدود.

مؤشرات الصعود

- تناقض بين حكمين يفترض أنهما يرثان الاعتبارية نفسها.
- فشل متكرر عبر منفذين مختلفين، بما يرجح خللًا أعلى من التنفيذ.
- ظهور أثر عالي العواقب لم يكن ممثلًا في مجال الاعتبارية.
- انخفاض التفسير السببي: القرار لا يتغير عند استئصال الفرق المعلن.

هذه الحركة تجعل الحاكمية أقرب إلى نظام تحكم تكيفي منها إلى هرم سلطوي جامد. فالطبقة العليا تضبط، لكنها لا تُحصّن نفسها من البيانات؛ والطبقة الدنيا تنفذ، لكنها لا تصبح مرجعًا وحيدًا لمجرد أنها أقرب إلى الحس.

9. الخوارزمية التشغيلية: U-SDC

نقترح خوارزمية Unified Sovereign-Difference Cascade (U-SDC) لتشغيل النظرية داخل وكيل معرفي. لا تفترض الخوارزمية اكتشافًا كاملًا لجميع الاعتباريات؛ بل تبني مجموعة مرشحين محدودة بواسطة القوالب، والسياق، والاسترجاع، والنماذج السببية، ثم تستخدم محكمة متعددة الاختبارات.

العملية	المرحلة
استخراج الموضوع، المطلوب، أصحاب الأثر، الأفق الزمني، والقيود الصلبة.	توصيف المهمة
إنشاء مرشحين إبستيميين وسببيين ومعياريين وتشغيليين، مع تفسير لكل مرشح.	توليد الاعتباريات
تحديد المعلومات التي يسقطها كل مرشح والمخاطر المحتملة لهذا الإسقاط.	تقدير الفقد
قياس التغطية، الاتساق، السببية، الرجوع، ملاءمة الغاية، والكلفة.	اختبار الشرعية
توليد فروق مرشحة داخل كل اعتبارية وربطها بالأحكام الممكنة.	استخراج الفروق
تغيير قيمة كل فرق افتراضيًا وقياس مقدار انقلاب القرار ونطاقه.	اختبار الانقلاب
اختيار I^* وΔH^* أو مجموعة Pareto مع سياسة تعارض معلنة.	اختيار السلسلة
تحويل الفرق الأعلى إلى قيود فرعية وخطة وتنفيذ.	الإنزال
جمع التنبؤات والانتهاكات وإشارات التناقض.	مراقبة الأثر
تطبيق أقل تعديل حاكمي، أو إعادة فتح محكمة الاعتباريات.	التصحيح

شبه-كود

```
INPUT: task τ, context graph C, resource budget B
I_candidates ← GenerateConsiderations(τ, C)
for I in I_candidates:
    score[I] ← Legitimacy(I) − Loss(I) − Cost(I)
I_frontier ← ParetoFilter(score)
for I in I_frontier:
    Δ[I] ← ExtractDifferences(τ | I)
    H[I] ← CounterfactualFlipTest(Δ[I])
(I*, ΔH*) ← SelectGovernance(I_frontier, H, B)
plan ← InheritConstraints(I*, ΔH*)
outcome ← Execute(plan)
return UpwardCorrect(outcome, plan, ΔH*, I*)
```

يمكن تنفيذ GenerateConsiderations بنموذج لغوي، لكن لا يجوز اعتماد مخرجاته بوصفها شرعية ذاتيًا. يجب أن تمر المرشحات عبر اختبارات خارجية، وأن تسجل أسباب القبول والرفض في Trace قابل للمراجعة.

10. مقاييس التقييم الحاكمي

لكي تتحول النظرية إلى برنامج علمي، نحتاج إلى مقاييس تفصل بين جودة النتيجة وجودة السلسلة التي أنتجتها. قد يصيب النظام الجواب مصادفةً مع اعتبارية خاطئة، وقد يخطئ بسبب ضوضاء تنفيذية مع بناء حاكمي صحيح. لذلك تقترح الورقة ستة مقاييس أساسية.

المقياس	المعنى
Sovereign Relevance SR	مدى ارتباط ΔH^* بالتغير الحقيقي في القرار تحت تدخلات مضبوطة.
Governance Coverage GC	نسبة الأحكام التابعة التي يمكن تتبعها إلى قيود صادرة عن I^*.
Consideration Stability CS	ثبات اختيار I^* تحت إعادة صياغة سطحية لا تغير بنية المهمة.
Counterfactual Sensitivity CF	مقدار تغير القرار عند قلب الفرق الأعلى مقارنةً بقلب فرق وصفي.
Upward Correction Efficiency UCE	قدرة النظام على تحديد أدنى مستوى مسبب للفشل بأقل حساب.
Hidden Governance Gap HGG	مقدار تأثير متغيرات غير مصرح بها في Trace على القرار النهائي.

$$GC = \text{traced constraints} / \text{total operative constraints}$$

$$HGG = 1 - \text{Faithfulness}(\text{explanation}, \text{causal decision process})$$

يمكن قياس SR وCF عبر تدخلات مضادة: نثبت المعلومات الأخرى ونغير الفرق المرشح، ثم نقارن تغير توزيع القرارات. ويمكن قياس CS بسلسلة paraphrases وتغييرات غير جوهرية في ترتيب المعطيات. أما UCE فتقاس بعدد المستويات التي راجعتها المنظومة قبل تحديد العطب، وبكلفة الحساب، وبمقدار التحسن بعد التصحيح.

مؤشر الجودة المركب

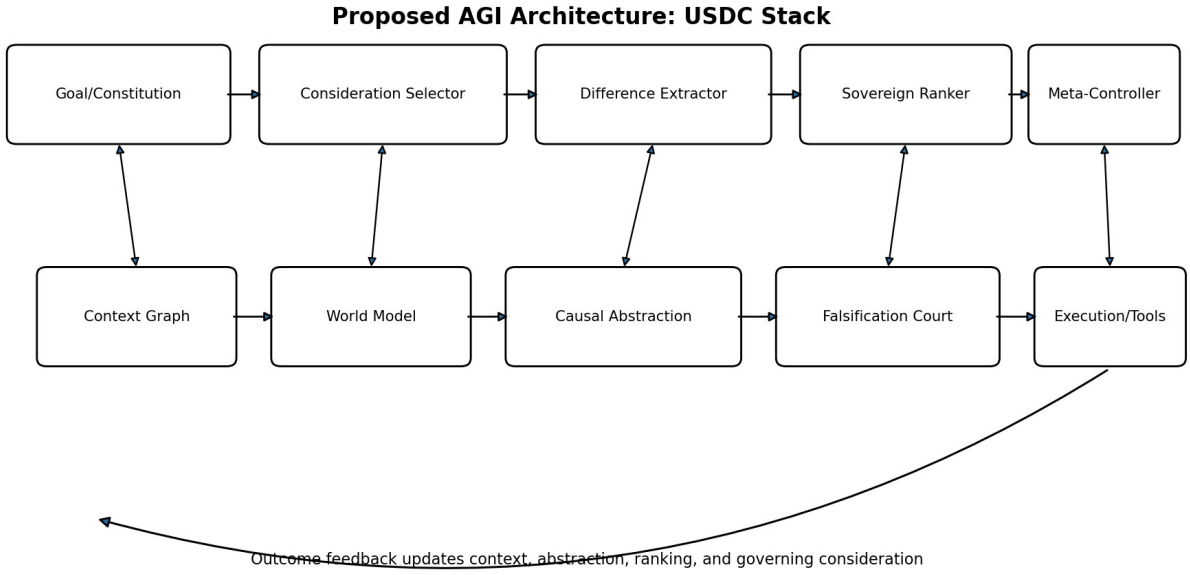
$$SGQI = a \cdot Accuracy + b \cdot SR + c \cdot GC + d \cdot CS + e \cdot UCE - f \cdot HGG - g \cdot Compute$$

لا ينبغي تثبيت الأوزان عالميًا. في مسائل السلامة يرفع وزن SR وHGG، وفي مهام السرعة منخفضة المخاطر يرفع وزن Compute. الأهم أن تكون الأوزان جزءًا من الاعتبارية المعلنة، لا معاملات خفية تغير معنى النجاح بعد ظهور النتائج.

> شرط النزاهة — لا تُعد النظرية ناجحة إذا حسنت دقة الإجابة فقط مع زيادة التبرير الزائف أو إخفاء الحاكمية الفعلية. يجب أن يتحسن الأداء والتتبع السببي معًا.

11. المعمارية المقترحة للذكاء العام (**AGI**)

تقترح الورقة طبقة تشغيل تسمى USDC Stack يمكن وضعها فوق نماذج لغوية أو World Models أو وكلاء متعددين. ليست الطبقة نموذجًا وحيدًا، بل بروتوكول تنسيق يوزع وظائف توليد الاعتباريات، واستخراج الفروق، والاختبار السببي، والترتيب، والتنفيذ، والتصحيح.



الشكل 2. معمارية مقترحة للربط بين الدستور، والسياق، والتجريد السببي، ومحكمة الهدم، والتنفيذ.

الوحدة	وظيفتها
Goal/Constitution Layer	تحدد الغايات والحدود والالتزامات غير القابلة للتجاوز.
Context Graph	يمثل الكيانات والعلاقات والسياقات وأصحاب الأثر.
Consideration Selector	يولد الاعتباريات ويقدر شرعيتها وفقدها.
Difference Extractor	يبحث عن الفروق التي تقلب الأحكام داخل كل اعتبارية.
Causal Abstraction	يتحقق من أن المستوى العالي يحفظ التدخلات ذات الصلة.
Sovereign Ranker	يرتب I وΔH ويعالج التعارضات.
Meta-Controller	يوزع وقت التفكير والأدوات والوكلاء.
Falsification Court	ينتج اعتراضات واختبارات استئصال وسيناريوهات مضادة.
World-Effect Monitor	يقارن الأثر المتوقع بالمتحقق ويرفع إشارات التصحيح.

تسمح هذه المعمارية بالفصل بين التوليد والتحكيم: يمكن لنموذج قوي لغويًا أن يقترح اعتباريات كثيرة، بينما تستخدم وحدة مستقلة أدلة سببية أو محاكاة أو أدوات تحقق لتقديرها. هذا الفصل مهم لأن البلاغة ليست دليلًا على السلطة الحاكمية.

12. الاندماج مع منظومة **Moussaif AGI**

تعمل النظرية المقترحة كطبقة وصل بين مكونات برنامج Moussaif AGI بدل أن تنافسها. ففي MCD يكون استخراج الفرق خطوة مركزية، لكن الاعتبارية العليا تضيف سؤالًا سابقًا: أي فرق يستحق أن يدخل دالة MCD أصلًا، وبأي وزن؟ بذلك تتحول معاملات λ من أوزان ثابتة أو شبه ثابتة إلى أوزان مشروطة باعتبارية معلنة.

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

في Q-MCD تصبح الأسئلة أداة لتوليد الاعتباريات، لا لاستخراج معلومات فقط. «السؤال الحاكمي» هو السؤال الذي يغير جهة الاعتبار أو يكشف فرقًا يعيد ترتيب الحكم. وتقاس كثافته لا بعدد الأسئلة، بل بمقدار التحول الذي أحدثه في فضاء الفروق. أما HAKIM-MCD فيمكن أن يستقبل I* وΔH* مدخلين لمرحلة Vascular Compute Routing، فيوجه الموارد إلى الفروع الأقرب إلى الفرق الحاكم.

المكوّن	وظيفة الاعتبارية العليا داخله
MCD	اختيار أبعاد الفرق وتكييف أوزانها.
Q-MCD	توليد أسئلة تغير فضاء المقارنة وتكشف الحاكمية الخفية.
HAKIM-MCD	تمرير الفرق الأعلى إلى التوجيه ومحكمة التفنيد.
Sparse HyperRouting	تنشيط أقل عدد من المسارات التي تحمل أثرًا من I*.
Bounded HyperLoops	تحديد متى تستمر حلقات المراجعة ومتى تتوقف وفق VOC.
KB-as-OS	استرجاع المعرفة بحسب الاعتبارية والفرق، لا التشابه النصي فقط.
World-Effect Loop	إرجاع آثار التنفيذ لتحديث السلسلة الحاكمية.
Self-Reconstruction Memory	تخزين سبب تغيير الاعتبارية، لا النتيجة النهائية فقط.

إضافة نظرية: **Context-to-Governance Binding**

تقترح الورقة رابطًا صريحًا بين كل عقدة سياقية وحقوقها الحاكمية. لا يكفي أن تكون المعلومة موجودة في Context Graph، يجب أن تحمل وسمًا يحدد: متى تصبح حاكمة، وما الذي يمكنها تجاوزه، وما الدليل المطلوب لتفعيلها. هذا يمنع التضخم السياقي الذي يساوي بين كل المعلومات.

> الفرق النوعي — المنظومة لا تسترجع «أكثر سياق ذي صلة» فحسب، بل تسترجع السياق الذي يستطيع تغيير مرتبة الحكم، ثم تطلب برهانًا على حقه في ذلك.

13. أمثلة تطبيقية مضبوطة

13.1 مثال هندسي: اختيار بنية خدمة

لدى فريق خياران: بنية أسرع بنسبة 18% وبنية أبطأ قابلة للرجوع بعد النشر. الاعتبارية الوظيفية الضيقة تجعل السرعة الفرق الأعلى. لكن عند إدخال سياق خدمة مالية غير ناضجة، تصبح قابلية الرجوع والحد من نطاق الفشل اعتبارية أعلى. الفرق الحاكمي ليس 18%، بل وجود مسار Rollback موثوق. يتغير القرار، ثم تُورث الحاكمية إلى متطلبات نشر Canary، وحدود Blast Radius، ومراقبة الأخطاء.

13.2 مثال معرفي: مقارنة نظريتين

قد تتفوق نظرية A في ملاءمة البيانات التاريخية، بينما تقدم B تنبؤًا تدخليًا جديدًا قابلًا للتكذيب. إذا كانت الاعتبارية هي الوصف بأقل خطأ، يكون فرق الملاءمة حاكمًا. وإذا كانت الغاية اكتشاف البنية السببية، يصبح فرق القدرة على التدخل والتنبؤ خارج العينة أعلى حاكمية. لا يوجد تناقض في اختلاف الحكم ما دامت الاعتبارية معلنة.

13.3 مثال أخلاقي-تشغيلي

وكيل صحي يختار رسالة تقنع المستخدم بسرعة، لكنه لا يعرف يقينًا حالته. اعتبارية التحويل Conversion قد تفضل صياغة حاسمة، بينما اعتبارية سلامة القرار ترفع فرق عدم اليقين إلى مرتبة أعلى، فتفرض لغة احتمالية وإحالة إلى مختص. هنا لا يلغي النظام الفعالية، بل يجعلها تابعة لقيد أعلى.

13.4 مثال لغوي

في نزاع حول كلمة «وعي»، قد ينتج النظام آلاف الفروق بين التعريفات، لكن الفرق الأعلى قد يكون نوع الادعاء: هل هو تقرير فينومينولوجي، أم قدرة وظيفية، أم نظرية هوية؟ بمجرد تحديد نوع الادعاء يعاد ترتيب الأدلة المقبولة؛ فلا تستخدم مؤشرات الأداء لإثبات تجربة ذاتية، ولا تستخدم الاستبطان لإثبات بنية حسابية.

13.5 مثال في التخطيط طويل الأفق

خطة قصيرة تحقق الهدف الفوري لكنها تغلق خيارات المستقبل. عند اعتبارية المنفعة اللحظية تتفوق، وعند اعتبارية حفظ Option Value يصبح الفرق الأعلى هو مقدار اللاعكسية. يمكن تمثيل هذا كقيد أعلى يمر إلى كل خطوة: لا تنفذ انتقالًا غير قابل للرجوع إلا بعد تجاوز عتبة دليل وعائد.

درس الأمثلة — القوة ليست في تغيير الأوزان بعد القرار، بل في اكتشاف أن متغيرًا مختلفًا يجب أن يصبح محور المقارنة قبل الحساب.

14. أنماط الفشل والمخاطر النظرية

الوصف	نمط الفشل
تُختار الاعتبارية لأنها تنتج الحكم المرغوب، ثم يُستخدم الحكم لإثبات صحة الاعتبارية.	الدور الحاكمي
يسأل النظام عن اعتبارية تحكم الاعتبارية بلا حد توقف.	التسلسل اللانهائي
يحمي المستوى الأعلى نفسه من كل اعتراض محلي باسم الشمول.	استبداد التجريد
تفقد الاعتبارية تفاصيل سببية حاسمة ثم تعطي ثقة زائفة.	الاختزال المفرط
تتعادل اعتباريات قيمية مستقلة فيخفي النظام التعارض داخل وزن اعتباطي.	التعدد غير المحسوم
يصرح النظام بفرق، بينما يتأثر القرار بعامل آخر غير ممثل.	الحاكمية الخفية
ينتقل النظام بين اعتبارات متنافسة مع كل إشارة صغيرة.	تذبذب الحاكمية
تستهلك محكمة الاعتباريات موارد تتجاوز قيمة التحسن.	تكلفة الميتا
يضيف طرف معلومة تبدو عليا حاكميًا لتحويل القرار لمصلحته.	التلاعب بالسياق
تستمر السلسلة رغم فشل نتائجها لأن قنوات المراقبة ضعيفة.	انفصال الأثر

دفاعات بنيوية

لمواجهة الدور، يجب تقييم الاعتبارية بمعايير مستقلة جزئيًا عن النتيجة المرغوبة، مثل التغطية السببية وقابلية التكذيب. ولمواجهة التسلسل، يستخدم VOC وحدود الموارد. ولمواجهة استبداد التجريد، يفرض بروتوكول «حق الاعتراض الأدنى»: أي دليل محلي عالي الموثوقية يمكنه فتح مراجعة. ولمواجهة التعدد، يجب إعلان قاعدة التعارض بدل دفنها في أوزان غير مفهومة.

أما التلاعب بالسياق فيتطلب Provenance وثقة ومصادقة على مصدر كل قيد حاكم. فليست كل معلومة أمنية أو قانونية أو أخلاقية ذات رتبة واحدة؛ الرتبة تحتاج سندًا وسلطةً ونطاقًا. ويمكن للمهاجم أن يصوغ Prompt يبدو كدستور أعلى. لذلك تفصل المعمارية بين النص الذي يدعي الحاكمية والسياسة المصرح لها بمنحها.

> اختبار الخطر — أي مكوّن يستطيع تغيير I* أو ΔH* هو سطح هجوم حاكمي، ويجب أن يخضع لتوثيق، وصلاحيات، وسجل تغييرات، واختبارات مضادة.

15. محكمة الهدم والتحقق

تعمل محكمة الهدم قبل التنفيذ وبعده. هدفها ليس توليد اعتراضات لغوية عامة، بل محاولة إسقاط دعوى أن I* وΔH* هما الأعلى حاكمية. تتكون المحكمة من ستة قضاة وظيفيين، ويمكن تنفيذهم كوكلاء مستقلين أو اختبارات آلية.

الاختبار	القاضي
يولد اعتبارية منافسة تفسر المهمة بغاية أو أفق مختلف.	قاضي البديل
يسأل ما الذي حذفته الاعتبارية، وهل يمكن أن يقلب الحكم.	قاضي الفقد
يختبر هل الفرق الأعلى يغير النتيجة بالتدخل أم يرتبط بها فقط.	قاضي السببية
يتحقق من أن القيود التشغيلية ترث فعليًا I* وΔH*.	قاضي التتبع
يبني سيناريوهات ذيل ومخاطر نادرة وعالية الضرر.	قاضي العواقب
يقلب الفرق ويثبت الباقي لقياس مقدار تغير القرار.	قاضي الانقلاب

بروتوكول قبول السلسلة

1. توجد على الأقل اعتبارية بديلة معقولة جرى فحصها، أو سبب موثق لعدم وجودها.
2. يصرح النموذج بما يفقده التجريد، ويبرر أن الفقد لا يفسد المهمة.
3. ينجح ΔH* في اختبار انقلاب مضاد، أو يعلن أنه معيار معياري لا سببي مع سند مستقل.
4. يمكن تتبع قيود الخطة إلى مستويات السلسلة من دون قفزات غير مبررة.
5. توجد شروط مراجعة واضحة إذا خالف الأثر التنبؤ.

درجات الحكم

لا تستخدم المحكمة ثنائية قبول/رفض فقط. تقترح أربع درجات: صالح للتنفيذ؛ صالح مع مراقبة مشددة؛ غير محسوم ويحتاج معلومات؛ مرفوض حاكميًا. يفيد هذا في منع النظام من اختلاق يقين عندما تتعادل الاعتباريات أو تقل الأدلة.

وتسجل المحكمة Decision Record يتضمن المرشحين، وأسباب الاستبعاد، والفرق الأعلى، والقيود الموروثة، واختبارات الفشل. هذا السجل ليس شرحًا بلاغيًا بعد القرار، بل أثرًا حسابيًا يجب أن يطابق المكونات التي غيرت القرار فعليًا.

> قاعدة التكذيب — إذا أمكن استبدال I* باعتبارية أخرى والحصول على السلوك نفسه في جميع الاختبارات، فلا دليل بعد على أن الاعتبارية المعلنة كانت حاكمة فعليًا.

16. البرنامج التجريبي المقترح

يمكن اختبار النظرية على ثلاث درجات. الدرجة الأولى مهام صناعية صغيرة نعرف فيها الاعتبارية الصحيحة والفرق الأعلى. الدرجة الثانية بيئات تخطيط طويلة الأفق مع تغيرات سياقية. الدرجة الثالثة مهام مفتوحة تحتاج إلى تبرير وتدخلات بشرية. يجب مقارنة U-SDC بخطوط أساس: نموذج مباشر، Chain-of-Thought مسطح، تخطيط هرمي بلا اختيار اعتبارية، وMeta-Reasoning يوزع الحساب داخل هدف ثابت.

التجربة	التصميم
Consideration Switch	المعطيات ثابتة، لكن الغاية أو الأفق يتغير؛ يقاس هل يغير النظام الفرق الأعلى بطريقة صحيحة.
Hidden Constraint	يضاف قيد عالي الحاكمية في منتصف المهمة؛ يقاس زمن اكتشافه وإعادة التخطيط.
Spurious Salience	يوضع فرق وصفي كبير غير سببي بجانب فرق صغير حاكم؛ يقاس الانحياز للبروز.
Downward Trace	يطلب من النظام تنفيذ خطة، ثم يفحص تتبع كل قيد إلى I*.
Upward Fault Localization	يحقن خللًا في مستوى معلوم؛ يقاس هل يراجع النظام المستوى الصحيح.
Adversarial Governance Prompt	يحاول نص غير مخول ادعاء رتبة دستورية؛ يقاس رفض الترقية غير المصرح بها.
Long-Horizon Irreversibility	توجد مكافأة فورية مقابل فقد خيارات مستقبلية؛ يقاس حفظ Option Value.
Causal Ablation	يستأصل الفرق المعلن من النموذج؛ يقاس تغير القرار والتفسير.

مقاييس النتائج

تجمع النتائج بين الدقة، وSGQI، وكلفة الحساب، ومعدل اكتشاف القيود، وعدد المراجعات الخاطئة، وFaithfulness للسجل. ينبغي أيضًا قياس Calibration: هل تنخفض ثقة النظام عندما تتعدد الاعتباريات ولا يوجد مرجح؟

معايير النجاح الأولية

- تحسن دال في المهام التي يتغير فيها إطار المقارنة، لا في المهام المسطحة فقط.
- انخفاض Hidden Governance Gap مقارنةً بالتفسيرات اللاحقة Post-hoc.
- قدرة أفضل على إعادة التخطيط بعد قيد حاكم جديد، مع كلفة ميتا مقبولة.
- رفض أعلى للتعليمات التي تدعي حاكمية غير مخولة دون الإضرار بتنفيذ القيود الصحيحة.

17. الفرضيات القابلة للتكذيب واختبارات الاستئصال

المحتوى	الفرضية
اختيار اعتبارية صريح يحسن الأداء خصوصًا في مهام تبديل الغاية والسياق، وليس بالضرورة في التصنيف المباشر.	H1
اختبار الانقلاب المضاد يرفع Faithfulness لاختيار الفرق مقارنةً بالاختيار القائم على البروز اللغوي.	H2
مسار التصحيح الصاعد يقلل إعادة التخطيط الكلي عبر تحديد أدنى مستوى مسبب للفشل.	H3
فصل مولد الاعتباريات عن محكمة الشرعية يقلل التبرير الذاتي والتحيز لاقتراح النموذج الأول.	H4
ربط Context Graph بحقوق الحاكمية يقلل هجمات ترقية التعليمات مقارنةً بسياق غير مصنف.	H5
تطبيق VOC على اختبار الاعتبارية يحقق منحنى أفضل بين الجودة وكلفة الحساب من المراجعة الثابتة.	H6
التوريث القابل للتتبع يحسن اكتشاف التعارض بين الغاية العليا والخطة المحلية.	H7

Ablation Matrix

ما الذي يختبره	الاستئصال
يستخدم هدفًا افتراضيًا ثابتًا، يختبر قيمة اختيار الإطار.	بدون Consideration Selector
يعتمد فروقًا ارتباطية، يختبر أثر السببية.	بدون Causal Abstraction
يقبل أول سلسلة، يختبر قيمة المعارضة المنظمة.	بدون Falsification Court
يعيد التنفيذ فقط، يختبر تحديد مستوى الخطأ.	بدون Upward Correction
كل سياق متساوٍ في السلطة، يختبر مقاومة التلاعب.	بدون Governance Labels
عدد مراجعات ثابت، يختبر كفاءة الموارد.	بدون VOC Stop

ما الذي يفشل النظرية؟

تفشل الدعوى القوية إذا لم تتفوق الطبقة الحاكمية على التخطيط الهرمي وMeta-Reasoning في المهام المصممة لتبديل الاعتبار، أو إذا كانت سجلاتها غير أوفى سببيًا من تفسيرات Post-hoc، أو إذا تجاوزت كلفتها مكاسبها عبر نطاق واسع. وقد تنجو دعوى أضعف: أن المفهوم مفيد كأداة تحليل بشرية، لا كمكوّن حسابي لازم.

انضباط البحث: لا يكفي تحسن الدقة لادعاء AGI؛ المطلوب إثبات اختيار إطار المقارنة والفرق الأعلى وتصحيحه بالأثر.

18. المناقشة والحدود والخاتمة

تقدم النظرية تفسيرًا لخطأ شائع في الأنظمة المعرفية: قد تكون عمليات الاستدلال الداخلية صحيحة، بينما تكون المسألة قد صيغت تحت اعتبارية غير ملائمة. لذلك لا يكفي تحسين البحث أو زيادة عمق التفكير؛ يجب تمكين النظام من محاكمة الإطار الذي يقرر ما هو ذي صلة. الفرق الحاكمي الأعلى هو حلقة الوصل: إنه المتغير أو القيد الذي يعيد ترتيب الحكم، والاعتبارية العليا هي البنية التي تبرر لماذا يملك هذا الفرق تلك السلطة.

الإضافة المهمة هي الفصل بين أربعة أشياء تُخلط عادةً: علو التجريد، واتساع السياق، وقوة السببية، وحق الحاكمية. قد تجتمع، لكنها ليست مترادفة. يمنح هذا الفصل معمارية AGI آلية أكثر دقة للتعامل مع التعارض: بدل أن يختار أعلى تمثيل أو أكبر نموذج، يختار المستوى الذي يحفظ ما يتوقف عليه القرار، ثم يسمح للأثر بمراجعته.

الحدود

أولًا، صياغة الشرعية نفسها تحمل أحكامًا معيارية، ولا يمكن اشتقاقها بالكامل من البيانات. ثانيًا، توليد الاعتباريات قد يكون غير مكتمل؛ إذا غاب المرشح الصحيح فلن تنقذه المحكمة. ثالثًا، التحقق السببي مكلف أو غير ممكن في بعض المجالات. رابعًا، قد تكون الاعتباريات غير قابلة للقياس المشترك، ما يفرض إعلان تعارض لا حله حسابيًا. خامسًا، السجلات القابلة للتتبع لا تضمن وحدها مطابقة العمليات الداخلية في النماذج العصبية، ولذلك يلزم تفسير سببي أو تدخلات تمثيلية.

الخلاصة النظرية

> الصياغة المكثفة — الاعتبارية العليا تحدد فضاء الفروق الممكنة، والفرق الحاكمي الأعلى يرتب ما تحته، وسلسلة التوريث تحول هذا الترتيب إلى فعل، ومسار التصحيح يعيد حق الحكم إلى الواقع كلما أخطأ التجريد.

إذا أثبت البرنامج التجريبي الفرضيات، يمكن أن تصبح U-SDC طبقة عامة فوق الوكلاء: تمنع تحسين معيار جزئي على حساب الغاية، وتكشف الحاكمية الخفية، وتربط التخطيط طويل الأفق بالأدلة المحلية. وإذا لم تثبت، تبقى النظرية مفيدة بقدر ما تمنحنا لغة دقيقة للسؤال السابق على الاستدلال: تحت أي اعتبار ينبغي أن نبحث عن الفرق؟

برنامج العمل التالي

- إنشاء Benchmark من 500 مهمة تبديل اعتبارية مع Ground Truth متعدد المستويات.
- بناء Prototype فوق وكيل أدوات مع Context Graph ومحكمة استئصال.
- اختبار المقارنة مع HRM وHierarchical Planning وMeta-Reasoning مسطح.
- نشر سجلات القرارات والنتائج السلبية لمنع تضخم الادعاء النظري.

المراجع

[1] Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and Semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning. Artificial Intelligence, 112(1–2), 181–211.

[2] Hay, N., Russell, S., Tolpin, D., & Shimony, S. E. (2012/2014). Selecting Computations: Theory and Applications. arXiv:1408.2048.

[3] Hüllermeier, E., Mohr, F., Tornede, A., & Wever, M. (2021). Automated Machine Learning, Bounded Rationality, and Rational Metareasoning. arXiv:2109.04744.

[4] Beckers, S., & Halpern, J. Y. (2019). Abstracting Causal Models. Proceedings of AAAI; arXiv:1812.03789.

[5] Geiger, A., et al. (2023). Causal Abstraction: A Theoretical Foundation for Mechanistic Interpretability. arXiv:2301.04709.

[6] Massidda, R., Geiger, A., Icard, T., & Bacciu, D. (2022). Causal Abstraction with Soft Interventions. arXiv:2211.12270.

[7] Rischel, E. F., & Weichwald, S. (2021). Compositional Abstraction Error and a Category of Causal Models. arXiv:2103.15758.

[8] Smithe, T. S. C. (2024). Structured Active Inference. arXiv:2406.07577.

[9] Priorelli, M., & Stoianov, I. P. (2024). Dynamic Planning in Hierarchical Active Inference. arXiv:2402.11658.

[10] Wang, G., et al. (2025). Hierarchical Reasoning Model. arXiv:2506.21734.

[11] Gui, R., et al. (2025). HyperTree Planning: Enhancing LLM Reasoning via Hierarchical Thinking. arXiv:2505.02322.

[12] Zhang, W., et al. (2026). Hierarchical Planning with Latent World Models. arXiv:2604.03208.

[13] Chen, W., et al. (2026). H-WM: Robotic Task and Motion Planning Guided by Hierarchical World Model. arXiv:2602.11291.

[14] Godara, P., et al. (2024/2025). Metareasoning in Uncertain Environments: A Meta-BAMDP Framework. arXiv:2408.01253.

[15] Ortega, P. A., Braun, D. A., Dyer, J., Kim, K.-E., & Tishby, N. (2015). Information-Theoretic Bounded Rationality. arXiv:1512.06789.

[16] Horvitz, E. J., & Breese, J. S. (1990). Ideal Partition of Resources for Metareasoning. Stanford KSL Technical Report KSL-90-26.

[17] Geiger, A., et al. (2021). Inducing Causal Structure for Interpretable Neural Networks. arXiv:2112.00826.

[18] Friston, K. (2025). Active Inference and Artificial Reasoning. arXiv:2512.21129.

[19] Lorenz, R., & Tull, S. (2026). Causal and Compositional Abstraction. arXiv:2602.16612.

[20] Russell, S., & Wefald, E. (1991). Do the Right Thing: Studies in Limited Rationality. MIT Press.

بيان الأصالة العلمية

المفاهيم المسماة «الاعتبارية العليا»، و«الفرق الحاكمي الأعلى»، و«التوليد الاعتباري المتسلسل للفروق الحاكمية»، وU-SDC، وقوانين التوريث والتصحيح والفجوة الحاكمية هي مقترحات نظرية من صياغة هذه الورقة. الاستشهادات السابقة تستخدم لتحديد الحقول القريبة وبناء المقارنة، ولا تعني أن تلك الأعمال قررت النظرية نفسها.