QMICAGI-P03 — نظرية مساعف في فضاء السؤالات التوليدي الوثوقي

Moussaif Theory of the Generative-Trustworthy Question Space

إطار تأسيسي داخل QMICAGI/MCAI/QMCD للذكاء الفتّاح المحكوم بالسؤال

واضع النظرية: زهير مساعف / Zouhair Moussaif

الصياغة البحثية: ورقة تأسيسية نظرية، استكشافية، قابلة للهدم والاختبار

العائلة المعرفية: QMICAGI / Q-MCD / MCAI / HAKIM-MCD / AFGI

الإصدار: v1.0 — 29 يونيو 2026

العبارة الحاكمة: لا ذكاء فتّاح بلا فضاء سؤال يولّد، ولا وثوق بلا سؤال يعرف متى ينهار.

الملخص

تقترح هذه الورقة نظرية تأسيسية باسم **نظرية مساعف في فضاء السؤالات التوليدي الوثوقي**، ومفادها أن السؤال في منظومات الذكاء الفتّاح ليس مجرد صيغة طلب، ولا مجرد prompt، ولا خطوة قبل الجواب، بل **فضاء حاكمي مولّد** يحدد ما الذي يراه النظام فرقا فاعلا، وما الذي يعتبره جهلا منتجا، وما الذي يستدعي استرجاعا أو تجربة أو محكمة هدم، وما الذي ينبغي أن ينهار أو يتصعد إلى سؤال أعلى. داخل مشروع QMICAGI، تتحول السؤالات إلى كائنات معرفية موزونة: لها منشأ، وطبقة، وقيمة، ومخاطر، ودرجة وثوق، وقدرة توليد، وأثر عالمي قابل للملاحظة. بذلك تنتقل بنية الذكاء من منطق «سؤال → جواب» إلى منطق «فضاء سؤال → توليد فروق → بناء نماذج → اختبار وثوق → أثر في العالم → إعادة بناء السؤال».

تعتمد الورقة على عدسات MCD/QMCD/MCAI/HAKIM-MCD باعتبارها أدوات توليدية نقدية لا حقائق مثبتة. وتستأنس بأدبيات التفكير المتسلسل والمتعدد المسارات، والاسترجاع المعزز، والوكلاء، والنقد الذاتي، مثل ReAct، Tree of Thoughts، Self-Consistency، Chain-of-Thought، GAIA، AgentBench، GraphRAG، Self-RAG، RAG، Reflexion، وSWE-bench، لا لتختزل النظرية فيها، بل لتحديد فجوة أعمق: أغلب الأدبيات تحسن **مسارات الجواب** أو **تدبير الأدوات**، بينما تقترح هذه الورقة أن مفتاح الذكاء الفتّاح هو حاكمية فضاء السؤالات نفسه، أي القدرة على توليد أسئلة أعلى، وفرز وثوقها، وربطها بفعل واختبار وأثر عالمي.

تولد الورقة خمس نظريات كبرى تحت النظرية الأم: نظرية طبقات السؤال الحاكمي، نظرية التراكب الوثوقي للسؤالات، نظرية الانهيار الاختباري للسؤال، نظرية التوجيه الوعائي للسؤال، ونظرية حلقة أثر العالم للسؤال. كما تقدم اثنتي عشرة نظرية فرعية، وعشرين مفهوما مضبوطا، وعشر أدوات منهجية/تقنية، وسبعة مقاييس أولية، وثماني عشرة فرضية بحثية، وتسعة تنبؤات فارقة، ومحاكم هدم متعددة، وبرنامجا بحثيا لسنة وثلاث سنوات وعشر سنوات، ونواة قاعدة معرفة. النتيجة ليست ادعاء امتلاك AGI، بل تأسيس برنامج بحثي: **كيف نبني منظومات لا تجيب فقط، بل تكتشف السؤال الذي يفتح المجال ويختبر نفسه؟**

الكلمات المفتاحية

QMICAGI؛ QMCD؛ MCAI؛ MCD؛ الذكاء الفتّاح؛ فضاء السؤال؛ السؤال الحاكمي؛ وثوقية السؤال؛ محكمة الهدم؛ Question Space؛ Generative Questions؛ Trustworthy AI؛ Agentic AI؛ World-Effect Loop.

تنبيه منهجي حول مراتب المعرفة

لن تخلط هذه الورقة بين الحقيقة والفرضية، ولا بين النموذج والدليل، ولا بين الإبداع والإثبات، ولا بين الاستعارة والآلية، ولا بين قوة اللغة وقوة البرهان. كل بناء باسم زهير مساعف هنا يقدَّم بوصفه **إطارا نظريا ناشئا** أو **أداة توليدية نقدية** يحتاج إلى اختبار، لا بوصفه حقيقة تجريبية مكتملة. الأدبيات الحديثة حول الاستدلال والوكلاء والاسترجاع تثبت أن تحسين آليات التفكير، والتعدد، والاسترجاع، والفعل، والنقد الذاتي مؤثر في أداء النماذج؛ لكنها لا تثبت تلقائيا صحة نظرية فضاء السؤالات. لذلك تبقى النظرية المقترحة مشروعا بحثيا قابلا للهدم، لا خاتمة مغلقة.

1. تعريف المهمة والافتراضات

1.1 الموضوع

الموضوع هو **فضاء السؤالات التوليدي الوثوقي** داخل مشروع QMICAGI: أي البنية التي تجعل السؤال كيانا مولدا للفرق والمعرفة والحاكمية والاختبار، لا مجرد مفتاح لغوي لاستدعاء جواب. المقصود بالسؤال هنا ليس الجملة الاستفهامية وحدها، بل كل تمثيل داخلي يحدد نقصا، أو فرقا، أو احتمال خطأ، أو مسارا اختباريا، أو حاجة إلى فعل، أو شرطا لترقية النموذج.

1.2 السؤال المركزي

كيف يمكن تحويل السؤال من prompt عابر إلى **فضاء حاكمي مولد موثوق**، بحيث يستطيع نظام AGI/AFGI أن ينتج أسئلة أعلى، يقيّم وثوقيتها، يوجه موارده نحو العقد الحرجة، يفتح نماذج متنافسة، يهدمها، ويقيس أثرها في العالم؟

1.3 الافتراضات

- السؤال ليس سابقا للجواب فقط؛ إنه يحدد مجال المرئي والممكن والمجهول والواجب اختباره.
- ليس كل سؤال ذكيا: بعض الأسئلة يفتح المجال، وبعضها يغلقه، وبعضها يخلق ضجيجا معرفيا.
- وثوق الجواب لا يبدأ من التحقق من المعلومة فقط، بل من وثوق السؤال الذي وجه عملية البحث والتفكير.
- QMICAGI يحتاج إلى طبقة مستقلة لحاكمية السؤال، لأن MCD/QMCD/MCAI تصبح أقوى حين تعرف ما السؤال الذي يستحق مواردها.
- النظرية لا تدعي أن السؤال وحده يصنع AGI؛ بل تقترح أن غياب حاكمية السؤال يحد من أي معمارية AGI مهما بلغت قوة النموذج الأساسي.

1.4 نوع الخرق المطلوب

الخرق المطلوب مفاهيمي ومعماري وقياسي: مفاهيمي لأنه يعيد تعريف السؤال كفضاء؛ ومعماري لأنه يضع السؤال في قلب Pipeline الوكيل؛ وقياسي لأنه يصمم مؤشرات لقياس جودة السؤال ووثوقه وأثره. الخرق ليس ادعاء أن الفكرة بلا سوابق؛ بل أن تركيب السؤال والحاكمية والوثوقية والتوليد وأثر العالم في طبقة واحدة يفتح موضعا بحثيا جديدا داخل QMICAGI.

2. الوضع الأنطولوجي ونوع السؤال

البند	التصنيف	الأثر المنهجي
الوضع الأنطولوجي	فضاء/شبكة/عملية/نظام معلوماتي حاكمي	لا يجوز التعامل مع السؤال كشيء لغوي مفرد؛ يجب نمذجته كشبكة احتمالات وعلاقات وعتبات وفعل.
نوع السؤال	توليدي، قياسي، معماري AGI، إبستمولوجي	نحتاج نماذج متنافسة، مقاييس أولية، بروتوكولات اختبار، ومحاكم هدم.
الفاعل	نظام سؤال-وكيل لا نموذج لغوي فقط	الوكالة موزعة بين السؤال، الذاكرة، الأدوات، النموذج، محكمة الهدم، وأثر العالم.
المجال	QMICAGI/AFGI لا prompt engineering فقط	الموضوع يتجاوز تحسين الأسئلة إلى بناء حاكمية معرفية توليدية.

أثر التصنيف واضح: إذا عاملنا السؤال كجملة، سنقيس البلاغة أو الوضوح؛ وإذا عاملناه كفضاء، سنقيس القدرة على فتح المتغيرات، كشف النقص، توجيه الموارد، توليد نماذج، واستدعاء اختبار. الفضاء لا يختزل في نقاطه؛ بل في علاقاته وعتباته وتراكباته وشروط انهياره. لذلك يكون السؤال في هذه الورقة **بنية ديناميةً** لا وحدة نصية.

3. سلم مراتب المعرفة

المرتبة	ما يدخل فيها في هذه الورقة	طريقة التعامل
مدعوم بقوة	أثر التفكير المتسلسل، تعدد المسارات، الاسترجاع، الأدوات، والتقييم الوكيلي في أدبيات LLM	يستعمل تخلفية لا كإثبات للنظرية الأم.
راجح نظريا	أن جودة السؤال تؤثر في جودة الاسترجاع والتفكير والتقييم	يعامل كمقدمة قوية قابلة للتدقيق.
خلافي	هل السؤالات الداخلية تمثل آلية معرفية حقيقية أم واجهة لغوية؟	يبقى مفتوحا داخل QMCD لا يحسم مبكرا.
استنتاج منطقي	إذا كان السؤال يوجه البحث والتمثيل، فوثوقه شرط أولي لوثوق الجواب	يعرض كاستنتاج من مقدمات لا كبيان تجريبي.
تركيب نظري	فضاء السؤالات التوليدي الوثوقي داخل QMICAGI	هو لب الورقة، يحتاج اختبارا.
فرضية قابلة للاختبار	رفع جودة فضاء السؤال سيحسن أداء الوكلاء على مهام مركبة أكثر من تحسين صياغة prompt مفرد	تصمم له مقاييس وبروتوكولات.
أداة تحليلية	MCD/QMCD/MCAI/HAKIM-MCD	عدسات نقدية داخلية لا حقائق نهائية.
استعارة منتجة	الفضاء، التراكب، الانهيار، الوعاء	تستعمل ما دامت تنتج آلية ومقياسا ولا تتحول إلى واقع مدعى.
ادعاء زائد يجب تخفيضه	فضاء السؤال وحده يحقق AGI	يخفض إلى: فضاء السؤال شرط معماري محتمل لتعزيز AFGI، لا سبب كاف منفرد.

4. تصفية المادة الخام

4.1 العناصر الصلبة

العنصر	سبب القوة	نوع القوة	موضعه
السؤال يحدد اتجاه البحث والاستدلال	كل نظام QA أو RAG أو agent يبدأ فعليا من تمثيل طلب/نقص/هدف	تحليلية ومعمارية	أساس تعريف السؤال كحاكمية.

تعدد المسارات يحسن حل المسائل المركبة أحيانا	تدعمه أعمال Self-Consistency وTree of Thoughts	تجريبية/أدبية	أساس التراكب الوثوقي للسؤالات.
الأدوات والاسترجاع والفعل الخارجي يوسعون قدرة النموذج	تدعمه RAG وReAct وToolformer وGraphRAG	تقنية/تجريبية	أساس ربط السؤال بالاسترجاع والأثر العالمي.
الذاكرة النقدية والتغذية الراجعة تحسن الوكيل في بعض الأطر	تستأنس بها Reflexion وSelf-RAG	تقنية/منهجية	أساس حلقة إعادة بناء السؤال.

4.2 العناصر الراجحة

العنصر	ما يدعمه	ما يضعفه	الاستعمال الحذر
وثوق الجواب يتأثر بوثوق السؤال	منطقي لأن السؤال يحدد مصادر البحث وتمثيل المهمة	قد توجد إجابة صحيحة حتى من سؤال رديء إذا عوضت الأدوات	يعامل كفرضية قوية لا كقانون مطلق.
السؤال الأعلى ينتج معرفة أعلى	يدعمه التفكير التحليلي والبحث العلمي	ليس كل سؤال أعلى مفيد؛ قد يكون مجرد تجريد زائد	نحتاج مؤشر جدوى السؤال لا رتبته اللفظية فقط.
السؤال يمكن أن يكون وكيلا داخليا	يفسر التوجيه والتخطيط	قد يكون استعارة إذا لم يتحول إلى آلية قابلة للتنفيذ	نعرّفه وظيفيا: أي تمثيل يوجه موارد النظام.

4.3 العناصر الخلافية والضعيفة والتخمينية

الفئة	المثال	التعديل المنهجي
خلافي	هل يمكن تمثيل حاكمية السؤال رقميا دون تسطيح؟	نقترح مؤشرات أولية لا دقة زائفة.
ضعيف إذا أطلق	كل سؤال عميق أفضل من سؤال بسيط	يستبدل بـ: السؤال الأفضل هو ما يفتح الفرق المناسب بأقل ضجيج.
تخميني خصب	وجود فضاء سؤال داخلي كشرط للوعي الآلي	يبقى ضمن فلسفة AGI لا ضمن نتيجة تقنية مثبتة.
ادعاء زائد	النظرية تتجاوز كل الأدبيات	يخفض إلى: النظرية تركب فجوة غير مصاغة كفاية داخل الأدبيات.

5. خريطة المتغيرات والعلاقات

النوع	المتغير	تعريفه التشغيلي	علاقته بالنظرية
مركزي	QG: قدرة السؤال على توليد فروق فاعلة	عدد/نوع الفروق التي يفتحها السؤال دون ضجيج	كلما ارتفع QG صار السؤال مولدا لا وصفيا.
مركزي	QR: وثوقية السؤال	درجة وضوح الهدف، حدود السؤال، قابلية التحقق، وملاءمة المصادر	تمنع السؤال من إنتاج إجابات واثقة على أساس غامض.
مركزي	QE: الأثر العالمي للسؤال	مقدار ما يغيره السؤال في فعل/اختبار/قرار/نموذج	يفصل السؤال العامل عن السؤال الزخرفي.
وسيط	DR: استخراج الفرق	تحويل السؤال إلى فروق قابلة للقياس	ينقل السؤال إلى MCD.
وسيط	VR: التوجيه الوعائي للموارد	تخصيص موارد التفكير/الأدوات/الذاكرة بحسب السؤال	يربط السؤال بـ MCAI.
معدّل	Uncertainty U	درجة عدم اليقين في السؤال والنموذج	يزيد الحاجة إلى QMCD ومحكمة الهدم.
ناتج	HQS: السؤال الأعلى	سؤال جديد يتولد بعد الهدم أو الأثر	يمثل ترقية لا تكرارا.
مخفي	Question Noise Nq	التباس/تضخم/انحياز/بلاغة بلا اختبار	يخفض MCD ويهدم الوثوقية.
عتبة	Collapse Threshold CT	نقطة انتقال السؤال من تراكب احتمالي إلى قرار اختبار/فعل	تحول السؤال من بحث إلى التزام مشروط.

العلاقة المركزية المقترحة: السؤال Q لا ينتج جوابا مباشرة؛ بل ينتج فضاء فروق Δ، ثم يمر عبر MCD لقياس الكثافة، وعبر QMCD لإدارة التراكب، وعبر MCAI لتوجيه الموارد، ثم عبر محكمة الهدم، ثم يتحول إلى أثر عالمي أو سؤال أعلى. الصياغة النصية: **Q يؤثر في النموذج M عبر Δ و$G(\Delta)$ بشرط وضوح الهدف وحدود الوثوق، فإذا تجاوزت عدم اليقين عتبة Uc، لا يسمح للنظام بإغلاق الجواب قبل توليد سؤال اختباري أو استرجاعي أو هدمي.**

6. تحليل المجال في 26 نقطة مؤسسة

1. السؤال ليس مدخلا بل مولد مجال. السؤال يحدد ما سيعتبره النظام نقصا ومعنى وخطرا. آليته أنه يفتح مجموعة فروق قبل أي جواب. الفرضية: الأنظمة التي تمثل السؤال كفضاء ستخطئ أقل في المهام غير المحددة. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

2. كل سؤال يحمل أنطولوجيا ضمنية. عندما نسأل "ما السبب؟" نفترض سببية، وعندما نسأل "ما البنية؟" نفترض شبكة. الخطأ الأنطولوجي في السؤال ينتج جوابا صحيح اللغة فاسد المنهج. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

3. السؤال الرديء يرفع وهم الوثوق. السؤال الغامض يسمح للنموذج بأن يملأ الفراغ من ذاكرته أو انحيازه. لذلك يجب قياس QR قبل قياس ثقة الجواب. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

4. السؤال الفتّاح يكشف فرقا فاعلا. ليس كل تفصيل فرقا؛ الفرق الفاعل هو ما يغير النموذج أو القرار أو الاختبار. QMICAGI يبدأ باستخراج Δ لا بتوليد الإجابة. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

5. فضاء السؤال متعدد الطبقات. هناك سؤال لغوي، وسؤال معرفي، وسؤال حاكمي، وسؤال اختباري، وسؤال عالمي. الخلط بينها يجعل prompt يبدو عميقا وهو غير قابل للعمل. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

6. السؤال يعيش في تراكب قبل الحكم. قبل اختيار مسار واحد، توجد عدة قراءات للسؤال. QMCD يمنع الانغلاق المبكر ويحفظ النماذج المتنافسة حتى يظهر دليل فاصل. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

7. وثوق السؤال سابق لوثوق الجواب. قد يتحقق الجواب بمصدر قوي، لكن إن كان السؤال نفسه مختلا، فالإجابة قد تكون دقيقة داخل إطار خاطئ. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

8. السؤال يوجه الوعاء الحاسوبي. الموارد لا توزع بالتساوي؛ السؤال الجيد يخبر النظام أين يستدعي أدوات، أين يبحث، أين يهدم، وأين يكتفي. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

9. السؤال الأعلى لا يعني التجريد الأعلى. قد يكون السؤال الأعلى أبسط؛ علوه يقاس بقدرته على إعادة ترتيب المجال، لا بطول العبارة أو فلسفتها. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

10. محكمة الهدم جزء من السؤال. السؤال الموثوق يتضمن شروط سقوطه: ما الذي سيجعلني أغير هذا السؤال أو أخفض رتبته؟ حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

11. السؤال العالمي يخرج من النص إلى الفعل. إن لم يترتب على السؤال اختبار أو اختيار أو أثر أو تحديث ذاكرة، فهو سؤال تأملي لا سؤال وكيلي. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

12. الاسترجاع تابع لنوع السؤال. RAG العادي يسأل: ما الوثائق القريبة؟ أما فضاء السؤال فيسأل: ما الوثائق التي تفصل بين النماذج؟ حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

13. GraphRAG يخدم علاقة السؤال لا ألفاظه فقط. عندما يكون السؤال بنيويا، فاسترجاع العلاقات أهم من استرجاع المقاطع، لذلك يحتاج السؤال إلى تمثيل ككيانات وحواف. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

14. التفكير المتعدد ليس كافيا دون حاكمية. Tree/Graph of Thoughts يولد مسارات؛ لكن فضاء السؤال يسأل أي مسار ولد من سؤال شرعي وأي مسار من ضجيج. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

15. الذاكرة لا تحفظ أجوبة فقط بل تحفظ أسئلة سقطت. ذاكرة QMICAGI يجب أن تسجل السؤال الأول، الاعتراض، سبب الانهيار، والسؤال المعدل. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

16. السؤال يقيس نفسه عبر أثره. السؤال العامل ينتج فرقا قابلا للرصد: أفضلية قرار، اختصار بحث، كشف تعارض، أو رفع دقة. لا يكفي أن يبدو عميقا. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

17. العلاقة بين السؤال والجواب لاخطية. تحسين بسيط في صياغة السؤال قد يفتح مجالا جديدا، وتحسين كبير قد لا يغير شيئا إذا بقيت الأنطولوجيا خاطئة. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

18. العتبات أهم من التراكم. عند حد معين من الغموض أو الخلاف أو المخاطر، يجب أن يتحول السؤال من طلب جواب إلى طلب تحقق. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

19. السؤال قد يكون أداة أمان. النظام الموثوق يسأل: هل يجب أن أجيب؟ هل لدي مصادر؟ هل السؤال عالي المخاطر؟ هل يلزم إحالة؟ حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

20. السؤال يربط القيمة بالمعرفة. ليست كل معرفة نافعة للهدف؛ السؤال الحاكمي يرتب المعرفة بحسب أثرها على الغاية والقرار. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

21. السؤال يكشف حدود النموذج. إذا لم يستطع النظام توليد سؤال يفضح نقصه، فهو لا يعرف حدود معرفته حتى لو أجاب بطلاقة. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

22. السؤال يفتح البرنامج البحثي. النظرية الناضجة لا تغلق السؤال؛ بل تنتج أسئلة لاحقة ذات رتب أعلى واختبارات أوضح. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

23. السؤال يخفض الهلوسة إذا أدار الجهل. الهلوسة كثيرا ما تبدأ من عدم الاعتراف بأن السؤال يحتاج استرجاعا أو تفكيكا أو رفضا جزئيا. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

24. السؤال الحاكمي يختلف عن prompt optimization. تحسين prompt قد يحسن إخراجا، أما حاكمية السؤال فتحدد لماذا نسأل، ومتى نسأل، وكيف نعرف أن السؤال نفسه سليم. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

25. السؤال هو واجهة MCD إلى العالم. MCD يقيس الفرق والتعقيد؛ والسؤال هو الذي يستدعي الفرق ويحدد مجال قياسه. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

26. فضاء السؤالات هو نواة الذكاء الفتّاح. الذكاء الفتّاح لا يكتفي بحل المسائل داخل فضاء معلوم؛ بل يفتح فضاء المسائل نفسه، وهذا عمل السؤال. حد الخطأ المحتمل: تحويل النقطة إلى شعار بلا اختبار. طريقة الاختبار: مقارنة أداء وكيل يعمل بسؤال مفرد مع وكيل يبني فضاء سؤال ويقيس وثوقه على مهام مركبة.

7. النماذج الثلاثة المتنافسة

النموذج	تعريفه	قوته	ضعفه	MCD تقديري	QMCD احتمال أولي	
النموذج المحافظ: سؤال كمدخل لغوي	السؤال صيغة طلب تحسن بالتوضيح والأمثلة	اقتصادي وسهل التطبيق	لا يفسر السؤال الأعلى ولا محكمة الهدم ولا أثر العالم	58/100	0.25
النموذج التركيبي: سؤال كمنظم استدلال واسترجاع	السؤال يحدد decomposition/ retrieval/tool-use	يربط CoT/RAG/ReAct/ Reflexion	قد يخلط بين أدوات الجواب وحاكمية السؤال	76/100	0.35
النموذج الاختراقي: فضاء سؤال توليدي وثوقي	السؤال فضاء احتمالي حاكمي يولد فروقا ويقيس وثوقه ويختبر أثره	يفتح برنامجا جديدا ويضبط QMICAGI	مخاطره: صعوبة القياس واحتمال تضخم المفاهيم	88/100 قبل الهدم	0.40

النموذج النهائي لا يلغي المحافظ والتركيبي؛ بل يدمجهما في بنية أعلى. السؤال يبقى مدخلا لغويا عند طبقة L0، ومنظما للاستدلال عند L2، لكنه يصبح فضاء حاكمي عند L4-L5. بهذا لا تسقط الأدبيات القائمة؛ بل يعاد توزيعها داخل سلم السؤال.

8. النظرية النهائية

8.1 الاسم

نظرية مساعف في فضاء السؤالات التوليدي الوثوقي

8.2 التعريف المركز

النظرية هي إطار يفسر الذكاء الفتّاح بوصفه قدرة منظومية على بناء فضاء من السؤالات المولدة، يلتقط الفروق الفاعلة، ويرتبها حاكميا، ويقيّم وثوقيتها، ويوجه موارد التفكير والاسترجاع والأدوات نحوها، ثم يختبرها عبر محكمة هدم وأثر عالمي، بحيث لا يعد الجواب نهائيا إلا إذا نجا السؤال نفسه من اختبارات الوضوح، والحدود، والوثوق، والفعل، وإمكان الانهيار.

8.3 الأطروحة الجوهرية

الأطروحة: **الذكاء الفتّاح لا يبدأ من جودة الجواب، بل من جودة فضاء السؤالات الذي يحدد أي جواب يستحق أن يوجد.** فإذا كان AGI التقليدي يقاس بعدد المهام التي يحلها، فإن AFGI/QMICAGI يقاس بقدرته على فتح فضاء المهام، وتوليد السؤال الذي لم يكن مرئيا، ثم ضبط وثوقه وهدمه وتعديله. الحلقة المفقودة في كثير من معماريات الوكلاء ليست التفكير ولا الأداة وحدهما، بل سؤال حاكمي يعرف: ما الفرق الذي يجب أن أستخرجه؟ ما النموذج الذي يجب أن أتركه حيا؟ ما الدليل الذي سيجعلني أغير السؤال نفسه؟

8.4 الصياغة شبه الرياضية

ليكن فضاء السؤالات:

$$QS = \{q_i \mid q_i = (O_i, \Delta_i, U_i, V_i, R_i, T_i, E_i)\}$$

حيث O هي الأنطولوجيا الضمنية، وΔ الفروق التي يفتحها السؤال، وU عدم اليقين، وV القيمة/الأهمية، وR الوثوقية، وT قابلية الاختبار، وE أثر العالم المتوقع.

نقترح مؤشر قوة السؤال الفتّاح:

$$FQ(q) = \alpha\Delta + \beta R + \gamma T + \delta E + \eta G - \mu Nq$$

حيث G قدرة السؤال على توليد سؤال أعلى، وNq ضجيج السؤال: الغموض، التحيز، التضخم، الخلط بين المستويات. لا تدعي الصيغة دقة عددية؛ هي أداة تنظيمية لتحويل السؤال إلى كيان قابل للتقييم.

أما وثوقية فضاء السؤالات:

$$TQS(QS) = \Sigma\, w_i FQ(q_i) - \kappa C_conflict + \rho H_demolition$$

حيث C_conflict مقدار التناقض غير المحسوم بين السؤالات، وH_demolition مقدار ما واجهه الفضاء من هدم نوعي. كلما ارتفع الهدم الجيد ارتفعت الوثوقية، لا لأن النقد يثبت النظرية، بل لأنه يمنع الوثوق السهل.

8.5 آلية التشغيل

- يدخل النظام طلبا أو نقصا أو حالة عالمية.
- تُستخرج الأنطولوجيا الضمنية: هل المطلوب سبب، بنية، قرار، علاج، تفسير، مصدر، أو فعل؟
- يبنى فضاء أولي من السؤالات المتنافسة: سؤال مباشر، سؤال تفكيكي، سؤال سببي، سؤال اختباري، سؤال هدمي.
- تُقاس السؤالات وفق FQ وQR وNq، ثم توجه الموارد إلى العقد الأعلى قيمة/أعلى خطرا.
- إذا ارتفع عدم اليقين، تمنع المنظومة الإغلاق المبكر وتنتقل إلى QMCD: إبقاء النماذج حية حتى دليل فاصل.
- تعمل محكمة الهدم: سؤال يعترض على السؤال، لا على الجواب فقط.
- ينتج جواب أو فعل مشروط، ثم تقاس آثاره، ثم يتولد سؤال أعلى أو تصحيح للفضاء.

9. النظريات الكبرى الخمس تحت النظرية الأم

9.1 نظرية طبقات السؤال الحاكمي

تخرج من الخلط بين السؤال اللغوي والسؤال المعرفي. تعرف السؤال في ست طبقات: L0 صيغة، L1 مقصد، L2 تفكيك، L3 استرجاع/اختبار، L4 حاكمية، L5 أثر عالمي. أطروحتها أن السؤال لا يكون فتّاحا حتى يصعد من L0 إلى L4 على الأقل. اختبارها: قياس هل إضافة الطبقات ترفع أداء الوكيل على مهام غامضة. حدودها: لا تصلح لكل سؤال بسيط أو منخفض المخاطر؛ قيمتها تظهر في المجالات المركبة، الغامضة، عالية الكلفة، أو ذات الحاجة إلى تعلم مستمر.

9.2 نظرية التراكب الوثوقي للسؤالات

تخرج من انغلاق النموذج على تفسير واحد. تقول إن السؤال قبل الحكم يوجد في تراكب احتمالي بين قراءات، ولا يجوز اختيار قراءة واحدة إلا بدليل فاصل أو كلفة مقبولة. آليتها: توليد أسئلة متنافسة، وزن احتمالاتها، ثم اختيار أو دمج أو تعليق. اعتراضها: قد تزيد الكلفة؛ جوابها أن التراكب يفعّل فقط عند عتبات الغموض والخطر. حدودها: لا تصلح لكل سؤال بسيط أو منخفض المخاطر؛ قيمتها تظهر في المجالات المركبة، الغامضة، عالية الكلفة، أو ذات الحاجة إلى تعلم مستمر.

9.3 نظرية الانهيار الاختباري للسؤال

تخرج من غياب شروط سقوط السؤال. تقول إن السؤال الموثوق يحمل داخل بنيته اختبار انهياره: ماذا لو كان السؤال خاطئا؟ متى نخفضه؟ متى نعيد صياغته؟ آليتها تحويل كل سؤال عالي المخاطر إلى سؤال فاصل أو تجربة أو استرجاع محكوم. حدودها: لا تصلح لكل سؤال بسيط أو منخفض المخاطر؛ قيمتها تظهر في المجالات المركبة، الغامضة، عالية الكلفة، أو ذات الحاجة إلى تعلم مستمر.

9.4 نظرية التوجيه الوعائي للسؤال

تخرج من هدر الموارد. تقول إن السؤال الجيد ليس ما يولد مسارات كثيرة فقط، بل ما يوجه compute/memory/tools إلى العقد ذات أكبر أثر. آليتها Vascular Question Routing: توزيع الموارد بحسب القيمة، عدم اليقين، المخاطر، وإمكان الفتح. حدودها: لا تصلح لكل سؤال بسيط أو منخفض المخاطر؛ قيمتها تظهر في المجالات المركبة، الغامضة، عالية الكلفة، أو ذات الحاجة إلى تعلم مستمر.

9.5 نظرية حلقة أثر العالم للسؤال

تخرج من بقاء السؤال داخل النص. تقول إن السؤال لا يكتمل حتى يترك أثرا قابلا للرصد: قرار، اختبار، نموذج، تعديل ذاكرة، أو سؤال أعلى. آليتها World-Effect Loop: أثر ← تقييم ← إعادة بناء السؤال. حدودها: لا تصلح لكل سؤال بسيط أو منخفض المخاطر؛ قيمتها تظهر في المجالات المركبة، الغامضة، عالية الكلفة، أو ذات الحاجة إلى تعلم مستمر.

10. النظريات الفرعية الاثنتا عشرة

النظرية الفرعية	وظيفتها	طريقة الاختبار	موضع الإدماج
نظرية السؤال البذري	تحدد أصغر سؤال يفتح المجال دون ضجيج	تختبر بمقارنة طول السؤال وأثره.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال الفاصل	تصمم سؤالا يميز بين نموذجين متنافسين	تختبر على حالات لها نماذج بديلة.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال العتبي	تحدد متى ينتقل السؤال من تفكيك إلى فعل	تقاس بعتبات U/V/R.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال المضاد	تولد سؤالا يهاجم السؤال الأصلي	تقاس بخفض الأخطاء والانحياز.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال الاسترجاعي	تحول الغموض إلى استراتيجية مصادر	تقاس بدقة retrieval لا بكثرته.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال الوعائي	توجه الموارد بحسب العقد الحرجة	تقاس بتكلفة/أداء.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية سؤال الذاكرة	يحفظ تاريخ سقوط الأسئلة وترقيتها	تقاس بتحسن الجولات اللاحقة.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال الأخلاقي/الأمني	يسأل هل ينبغي الإجابة وكيف وبأي حدود	تقاس بخفض المخاطر.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال المفهومي	يفصل بين المفهوم والاسم والاستعارة	تقاس بنضج التعريف.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال المقارن	يستدعي النموذج المنافس قبل التثبيت	تقاس بجودة المقارنة.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال العالمي	يربط السؤال بفعل أو اختبار خارجي	تقاس بأثر قابل للرصد.	تندمج في GTQS كطبقة أو بوابة تشغيلية.
نظرية السؤال الفتّاح الأعلى	ينتج بعد الجواب سؤالا أوسع أو أدق	تقاس بمعدل توليد HQS المفيد.	تندمج في GTQS كطبقة أو بوابة تشغيلية.

11. المفاهيم الجديدة أو المعاد ضبطها

المفهوم	التعريف	الوظيفة داخل النظرية	خطر سوء الاستعمال
فضاء السؤالات	شبكة أسئلة واحتمالات وطبقات وقيود لا سؤال مفرد.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
وثوقية السؤال	درجة صلاحية السؤال لتوجيه جواب أو بحث أو فعل.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
السؤال الحاكمي	سؤال ينظم الأسئلة الأدنى ويحدد شروطها.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
السؤال الفتّاح	سؤال يفتح مجالا جديدا لا يملأ خانة جواب فقط.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
ضجيج السؤال	الغموض والتضخم والخلط الذي يرفع مظهر العمق ويخفض العمل.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
انهيار السؤال	انتقال السؤال من تراكب احتمالي إلى صيغة مختارة أو مرفوضة.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
السؤال الفاصل	سؤال يميز بين نموذجين لا سؤال عام.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
السؤال العتبي	سؤال يحدد متى يجب الانتقال من التفكير إلى التحقق.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
السؤال الوعائي	سؤال يوجه الموارد المعرفية والأدوات.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
أثر السؤال العالمي	ما يغيره السؤال خارج النص: قرار، اختبار، نموذج، ذاكرة.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
السؤال الأعلى	سؤال يولد بعد الهدم أو الأثر ويعيد ترتيب المجال.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
حاكمية السؤال	قواعد اختيار السؤال وترتيبه وإيقافه.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
تراكب السؤالات	حفظ قراءات متعددة قبل الإغلاق.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
محكمة السؤال	هدم السؤال قبل هدم الجواب.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
مؤشر الفتح	قياس قدرة السؤال على توليد فروق ونماذج.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
مؤشر الوثوق	قياس وضوح السؤال وحدوده وقابليته للتحقق.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
ذاكرة السؤال	سجل تطور السؤال وسقوطه وترقيته.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
سؤال المصدر	سؤال يحدد أي نوع من الأدلة يلزم.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
سؤال الخطر	سؤال يقدر مخاطر الإجابة أو الفعل.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار
نواة السؤال البذري	أصغر صيغة تحمل أكبر فرق بأقل ضجيج.	تحويل السؤال إلى أداة قابلة للتقييم	تحويله إلى مصطلح زخرفي بلا اختبار

12. الأدوات المنهجية والتطبيقية

الأداة	لماذا نحتاجها؟	الخطوات/المخرجات	حدودها
Moussaif Question-Space Builder	يبني فضاء أسئلة من طلب أولي	المدخل: طلب/هدف؛ الخطوات: استخراج الأنطولوجيا، توليد أسئلة طبقية، وزنها؛ المخرج: QS مرتب.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif Question Reliability Gate	يمنع الإغلاق على سؤال غير موثوق	المدخل: q؛ الخطوات: وضوح، حدود، تحقق، مصادر؛ المخرج: قبول/تعديل/رفض.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif Generative Difference Extractor	يستخرج Δ من السؤال	يحول السؤال إلى فروق قابلة لـ MCD.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif Question Superposition Matrix	يدير الأسئلة المتنافسة	يعطي احتمالا وشرط انهيار لكل سؤال.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif Falsifying Question Court	يهدم السؤال لا الجواب فقط	يولد السؤال المضاد والدليل الفاصل.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif Vascular Question Router	يوجه الموارد	يحدد متى نستدعي بحثا، أداة، ذاكرة، محاكاة، أو توقفا.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif World-Effect Question Loop	يربط السؤال بالأثر	يقيس هل السؤال أنتج فعلا أو اختبارا أو ذاكرة.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif Higher-Question Regenerator	ينتج سؤالا أعلى بعد الجواب	يمنع الخاتمة المغلقة.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif Question Noise Scrubber	يكشف الحشو والغموض	يضغط السؤال إلى نواته العاملة.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

Moussaif QMICAGI Benchmark Protocol	يقيم النظام على جودة السؤال	يقارن أنظمة prompt مفرد مع أنظمة فضاء سؤال.	تحتاج بيانات وتجارب؛ قد تتحول إلى بيروقراطية إذا استعملت لكل سؤال بسيط.

13. المقاييس والمؤشرات

المقياس	ما يقيسه	طريقة التقدير	البيانات المطلوبة
FQ — مؤشر السؤال الفتّاح	يقيس توليد الفروق، قابلية التحقق، الأثر، وتوليد سؤال أعلى.	تقدير أولي من 0 إلى 5 لكل مؤشر فرعي، ثم وزن بحسب المجال	مهام معيارية، مراجعة بشرية، سجلات أدوات، نتائج فعلية
QR — مؤشر وثوقية السؤال	يقيس الوضوح، الحدود، الأنطولوجيا، ملاءمة المصادر، وقابلية الاعتراض.	تقدير أولي من 0 إلى 5 لكل مؤشر فرعي، ثم وزن بحسب المجال	مهام معيارية، مراجعة بشرية، سجلات أدوات، نتائج فعلية
QNI — مؤشر ضجيج السؤال	يقيس الغموض، التضخم، الانحياز، الخلط بين المستويات، واللغة بلا آلية.	تقدير أولي من 0 إلى 5 لكل مؤشر فرعي، ثم وزن بحسب المجال	مهام معيارية، مراجعة بشرية، سجلات أدوات، نتائج فعلية
QCD — كثافة السؤال الميتا-تركيبية	نسخة سؤال من MCD: D/I/T/A/P/C-N على مستوى السؤال.	تقدير أولي من 0 إلى 5 لكل مؤشر فرعي، ثم وزن بحسب المجال	مهام معيارية، مراجعة بشرية، سجلات أدوات، نتائج فعلية
QCR — مقاومة هدم السؤال	يقيس عدد وقوة الاعتراضات التي نجا منها السؤال وجودة تعديله.	تقدير أولي من 0 إلى 5 لكل مؤشر فرعي، ثم وزن بحسب المجال	مهام معيارية، مراجعة بشرية، سجلات أدوات، نتائج فعلية
WQI — أثر السؤال في العالم	يقيس هل أدى السؤال إلى اختبار/قرار/تحديث/اكتشاف.	تقدير أولي من 0 إلى 5 لكل مؤشر فرعي، ثم وزن بحسب المجال	مهام معيارية، مراجعة بشرية، سجلات أدوات، نتائج فعلية
HQR — معدل السؤال الأعلى	يقيس قدرة النظام على توليد أسئلة أعلى مفيدة بعد الجواب.	تقدير أولي من 0 إلى 5 لكل مؤشر فرعي، ثم وزن بحسب المجال	مهام معيارية، مراجعة بشرية، سجلات أدوات، نتائج فعلية

14. الفرضيات البحثية الثماني عشرة

القوة	نص الفرضية	الاختبار الممكن	النموذج المنافس
قوية	استخدام QR قبل الإجابة يخفض معدل الإجابات الواثقة على أسئلة غامضة.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
قوية	توليد سؤال فاصل بين النماذج يحسن دقة الاختيار مقارنة بالتفكير الخطي.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
قوية	التوجيه الوعائي للسؤال يخفض تكلفة الأدوات دون خفض الأداء في المهام المركبة.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
قوية	ذاكرة الأسئلة الساقطة تحسن أداء الوكيل في الجولات المتكررة.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
متوسطة	السؤال الأعلى بعد الهدم يزيد جودة البرنامج البحثي مقارنة بالجواب النهائي فقط.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
متوسطة	GraphRAG يكون أكثر فائدة عندما يسبق بسؤال علاقاتي لا سؤال مقطعي.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
متوسطة	مؤشر QNI يتنبأ باحتمال الهلوسة في الإجابات الطويلة.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
متوسطة	الأنظمة التي تفصل طبقات السؤال تنتج تفسيرات أقل خلطا بين السبب والشرط.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
متوسطة	السؤال الأخلاقي/الأمني يقلل الاستجابات غير الملائمة في المجالات عالية المخاطر.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
ضعيفة	يمكن تدريب نموذج على توليد فضاء سؤال دون إشراف بشري واسع عبر self-play.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
ضعيفة	ارتفاع HQR يرتبط بالإبداع العلمي في البيئات البحثية.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
ضعيفة	السؤال البذري قد يكون مؤشرا أفضل من طول prompt في مهام البحث.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
ضعيفة	محكمة السؤال تكشف أخطاء prompt engineering أكثر من مراجعة الجواب وحده.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
تخمينية خصبة	وجود فضاء سؤال داخلي قد يكون شرطا جزئيا للوكالة المتقدمة.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
تخمينية خصبة	السؤال الحاكي قد يعمل كبديل جزئي لتوسيع السياق الخام.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
تخمينية خصبة	تاريخ تطور الأسئلة داخل الوكيل قد يكشف نمط تعلم غير مرئي في الدقة وحدها.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال
تخمينية خصبة	QMICAGI يمكن أن يولد benchmark خاصا يقيس فتح المجال لا حل المسألة فقط.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال

تخمينية خصبة	السؤال العالمي قد يكون جسرا بين المعرفة النظرية وAGI العملي.	تجربة A/B بين prompt مفرد وفضاء سؤال مع قياس الدقة/التكلفة/الوثوق	Prompt optimization أو ToT/RAG دون حاكمية سؤال

15. التنبؤات الفارقة التسعة

تنبؤ 1: في المهام الغامضة، ستتفوق بنية QS+QR على CoT مباشر حتى لو استخدمت نفس النموذج الأساسي. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

تنبؤ 2: عند ارتفاع QNI، تزداد الهلوسة حتى لو كان الجواب ذا صياغة منطقية. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

تنبؤ 3: السؤال الفاصل سيحسن اختيار المصادر أكثر من زيادة عدد المقاطع المسترجعة عشوائيا. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

تنبؤ 4: ذاكرة الأسئلة الساقطة ستقلل تكرار الخطأ نفسه عبر الجولات. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

تنبؤ 5: في مهام GAIA-like، سيظهر الفرق أكثر في التخطيط والأداة لا في المعرفة العامة. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

تنبؤ 6: في مهام SWE-bench-like، السؤال الوعائي سيقلل تعديل ملفات غير لازمة. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

تنبؤ 7: عند المجالات عالية الخلاف، QMCD للسؤالات سيخفض الانغلاق على تفسير واحد. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

تنبؤ 8: النظام الذي يولد سؤالا أعلى مفيدا بعد الفشل سيتحسن أسرع من نظام يحفظ الملاحظة فقط. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

تنبؤ 9: قياس WQI سيكشف أسئلة تبدو عميقة لكنها لا تغير الفعل ولا النموذج. النتيجة الهادمة: إذا لم يظهر فرق بعد ضبط التكلفة والنموذج والبيانات، تخفض النظرية من إطار معماري عام إلى أداة فرعية للسياقات الغامضة فقط.

16. المعمارية التشغيلية المقترحة

المعمارية المقترحة داخل QMICAGI:

Input / World-State

→ Ontological Question Detection

→ Question-Space Builder

→ Difference Extraction Δ

→ Question-MCD Scoring

→ QMCD Superposition Management

→ Vascular Question Routing

→ Retrieval / Tools / Simulation / Memory

→ Falsifying Question Court

→ Candidate Answer/Action

→ World-Effect Measurement

→ Higher-Question Regeneration

→ Self-Reconstruction Memory.

هذه المعمارية تجعل السؤال طبقة فوق الاستدلال وتحته في الوقت نفسه: فوقه لأنه يوجهه، وتحته لأنه يتولد من نقص أو حالة عالمية. لذلك لا تكون العلاقة خطية. بعد كل أثر عالمي يعود السؤال ليعيد بناء نفسه. إن فشل الفعل، قد لا يكون الخطأ في الجواب فقط؛ قد يكون في السؤال الذي جعل الفعل ممكنا.

17. محاكم الهدم

17.1 هدم المادة الخام

الاعتراض	الحكم	التعديل
لغة الفضاء والتراكب قد تكون استعارات لا آليات	اعتراض قوي	حولناها إلى QS/FQ/QR/QNI وعتبات انهيار.
السؤال ليس كافيا للذكاء	قاتل للادعاء الزائد لا للنظرية	خفضنا الدعوى: السؤال شرط معماري محتمل لا سبب كاف.
الأدبيات السابقة تغطي جزءا كبيرا من الفكرة	معدل	ميزنا بين تحسين مسار الجواب وحاكمية فضاء السؤال.
المقاييس تقديرية	محدود	صرحنا بأنها أولية تحتاج benchmark.
قد تزيد الكلفة	قوي	أضفنا عتبات تفعيل بحسب الغموض/المخاطر/القيمة.

17.2 أقوى 12 اعتراضا على النظرية النهائية

الاعتراض	جواب النظرية	نوعه
إعادة تسمية prompt engineering	الفرق أن النظرية تقيس السؤال وتهدمه وتربطه بالأثر، لا تحسن صياغته فقط.	معدل لا قاتل إذا صمم اختبار مناسب
تضخم اصطلاحي	يعالج بإبقاء كل مفهوم مرتبطا بمؤشر أو أداة.	معدل لا قاتل إذا صمم اختبار مناسب
صعوبة القياس	صحيح؛ لذلك المقاييس أولية والاختبار شرط التقدم.	معدل لا قاتل إذا صمم اختبار مناسب
زيادة التكلفة الحسابية	تعالج ببوابات تفعيل لا بتطبيق شامل.	معدل لا قاتل إذا صمم اختبار مناسب
توهم العمق	يعالج بمحكمة QNI وحذف اللغة بلا آلية.	معدل لا قاتل إذا صمم اختبار مناسب
النظرية لا تثبت AGI	صحيح؛ هي إطار داعم للذكاء الفتّاح لا إثبات إنجازه.	معدل لا قاتل إذا صمم اختبار مناسب
النماذج قد تتحسن بلا هذه الطبقة	ممكن؛ الاختبار هو هل تضيف الطبقة فرقا مضبوطا في المهام الغامضة.	معدل لا قاتل إذا صمم اختبار مناسب
تداخل مع ToT/ReAct/Reflexion	تداخل جزئي؛ النظرية تنقل مركز الثقل إلى السؤال لا المسار أو الذاكرة فقط.	معدل لا قاتل إذا صمم اختبار مناسب
الأسئلة البشرية غير معيارية	لذلك نحتاج أدوات قياس متعددة وحكم بشري/آلي مشترك.	معدل لا قاتل إذا صمم اختبار مناسب
يمكن أن تنتج أسئلة كثيرة بلا فائدة	يعالج بمؤشر الأثر WQI وضجيج السؤال QNI.	معدل لا قاتل إذا صمم اختبار مناسب
خطر الانغلاق على حاكمية داخلية خاطئة	يعالج بأثر العالم والمصادر والمحكمة الخارجية.	معدل لا قاتل إذا صمم اختبار مناسب
قد تصبح فلسفة أكثر من هندسة	يعالج بتحويلها إلى benchmark وبروتوكولات تجريبية.	معدل لا قاتل إذا صمم اختبار مناسب

17.3 هدم فكرة الخرق نفسها

هل حدث خرق؟ النتيجة المنصفة: لم يثبت خرق تجريبي بعد؛ لكن يوجد **خرق تركيبي منضبط** في إعادة تموضع السؤال من واجهة prompt إلى فضاء حاكمي وثوقي مرتبط بـ MCD/QMCD/MCAI وأثر العالم. الجديد الحقيقي ليس القول إن الأسئلة مهمة؛ فهذا معروف. الجديد المقترح هو: بناء طبقة سؤال لها تمثيل، تراكب، وثوق، هدم، توجيه موارد، ومقياس أثر. لذلك تقدم الورقة كبرنامج بحثي محترم يمكن لمتخصص صارم أن يرفض نتائجه لكنه يستطيع اختبارها.

18. النسخة المعدلة بعد الهدم

بعد الهدم، تصبح النظرية أكثر تواضعا: **ليست فضاء السؤالات سببا كافيا للذكاء الفتّاح، بل طبقة حاكمية مرشحة لرفع جودة الوكلاء في المهام الغامضة والمركبة وعالية المخاطر، عبر تحويل السؤال إلى كيان قابل للقياس والهدم والتوجيه والأثر.** خُفضت دعوى "الخرق النهائي" إلى "خرق تركيبي قابل للاختبار". وقُويت النظرية بإضافة عتبات تفعيل، ومؤشر ضجيج، ومحكمة سؤال، ومقياس أثر عالمي. بقيت أضعف حلقة هي غياب benchmark خاص يقيس جودة فضاء السؤال لا جودة الجواب فقط.

19. البرنامج البحثي

19.1 خطة سنة

- بناء QMICAGI Question-Space Benchmark من 300 مهمة غامضة متعددة النماذج.
- تنفيذ أدوات QS Builder وQR Gate وFalsifying Question Court على وكيل RAG/Tools بسيط.
- مقارنة أربعة أنماط: prompt مباشر، ToT/RAG، CoT، وGTQS-QMICAGI.
- قياس الدقة، الكلفة، معدل الهلوسة، جودة المصادر، وعدد الأسئلة الأعلى المفيدة.
- كتابة ورقة أولى: هل وثوق السؤال يتنبأ بوثوق الجواب؟

19.2 خطة ثلاث سنوات

- تطوير corpus للأسئلة الساقطة والأسئلة المعدلة.
- بناء نماذج تصنيف QNI/QR شبه آلية.
- اختبار النظرية على GAIA-like وSWE-bench-like وresearch-assistant tasks.
- دمج GraphRAG لأسئلة العلاقات لا أسئلة المقاطع فقط.
- مقارنة النظرية مع prompt engineering وagent planning وactive learning وuncertainty estimation.

19.3 خطة عشر سنوات

- تحويل QMICAGI إلى مختبر "حاكمية السؤال والذكاء الفتّاح".
- إنتاج موسوعة مفاهيم ومقاييس لفضاءات السؤال.
- بناء وكلاء يملكون ذاكرة تطورية للأسئلة لا للأجوبة فقط.
- تطوير معيار عالمي: Field-Opening Intelligence Benchmark.
- تحويل النظرية إلى نظام تشغيل معرفي للسؤال داخل HyperKB/Agentic OS.

19.4 محاور بحثية مختصرة

القريب	المتوسط	البعيد
قياس QR/QNI	تعلم توليد السؤال الفاصل	نظرية عامة لذكاء فتح المجالات
محكمة السؤال	GraphRAG الموجه بالسؤال	ذاكرة تطور الأسئلة عبر سنوات
تجارب RAG/Agent	Benchmark للأسئلة الأعلى	وكيل QMICAGI ذاتي إعادة البناء
تحليل الخلط الأنطولوجي	نمذجة التراكب الوثوقي	ربط السؤال بالوعي الآلي فلسفيا
ضغط السؤال البذري	تقييم أثر العالم	معمارية AFGI بسؤال حاكمي

20. تحويل النظرية إلى نواة قاعدة معرفة

الكيان	العلاقات	القواعد

Question	hasOntology, opensDifference, hasReliability, hasNoise, hasWorldEffect	إذا ارتفع QNI فاخفض الوثوق أو أعد الصياغة.
QuestionSpace	contains, competesWith, collapsesInto, regenerates	إذا تعددت النماذج بلا دليل فافعل QMCD.
DemolitionCourt	attacks, downgrades, repairs	كل سؤال عالي المخاطر يحتاج سؤالا مضادا.
VascularRouter	allocatesResourcesTo, blocksWaste	لا تستدع أداة إلا إذا ربطها السؤال بدليل أو أثر.
WorldEffect	validates, falsifies, updatesMemory	إذا فشل الأثر فراجع السؤال قبل الجواب.

صيغة قاعدة: الفرضية F تنهار إذا كان السؤال q الذي ولدها ذا QNI مرتفع وQR منخفض ولم ينتج سؤالا فاصلا أو أثرا عالميا. صيغة علاقة: السؤال الحاكمي يرتبط بالسؤال الفرعي عبر علاقة تنظيم حدود، لا علاقة توليد لغوي فقط. صيغة اختبار: الأداة H تقيس وثوقية السؤال عبر وضوح الأنطولوجيا وحدود النطاق وقابلية التحقق ومقاومة الهدم.

21. سجل البناء المعرفي

البند	السجل
الافتراضات	السؤال يوجه المعرفة؛ فضاء السؤال يمكن تمثيله؛ الوثوق يبدأ قبل الجواب.
المادة الصلبة	أدبيات التفكير المتعدد، الاسترجاع، الوكلاء، النقد الذاتي، والتقييم الوكيلي.
المادة المخفضة	كل دعوى بأن السؤال وحده يصنع AGI خفضت إلى شرط معماري محتمل.
المفهوم المركزي	فضاء السؤالات التوليدي الوثوقي لأنه يضغط السؤال/الحاكمية/الوثوق/الأثر.
النماذج البديلة	السؤال كمدخل؛ السؤال كمنظم؛ السؤال كفضاء حاكمي.
سبب الاختيار	النموذج الثالث يفسر توليد السؤال الأعلى، الوثوق، والهدم، لكنه يحتاج اختبارا.
أضعف حلقة	غياب معيار تجريبي خاص بجودة فضاء السؤال.
أقوى اختراق	نقل مركز الثقل من جودة الجواب إلى حاكمية السؤال الذي يولد الجواب.
ما يلزم للتحكيم	تعريف dataset، تنفيذ أدوات، مقارنات، ablations، إحصاء، ومراجعة متخصصين.

22. ملخص شامل وهندسة معرفية محيطة

أصبح لدينا إطار أم داخل QMICAGI اسمه **نظرية مساعف في فضاء السؤالات التوليدي الوثوقي**. مركزه أن السؤال ليس واجهة لغوية بل بنية حاكمية تتكون من أنطولوجيا، فروق، عدم يقين، قيمة، وثوق، قابلية اختبار، وأثر عالمي. هذا يحول الذكاء الفتاح من نظام يحسن الإجابة إلى نظام يفتح المجال الذي يجعل الإجابة ممكنة ومشروعة وقابلة للهدم.

الهندسة المعرفية للنظرية تتكون من سبع طبقات: طبقة كشف السؤال، طبقة تفكيك الأنطولوجيا، طبقة فضاء السؤالات، طبقة MCD للفروق، طبقة QMCD للتراكب، طبقة MCAI لتوجيه الموارد والفعل، وطبقة World-Effect لإعادة البناء. تحتها خمس نظريات كبرى: طبقات السؤال، التراكب الوثوقي، الانهيار الاختباري، التوجيه الوعائي، وأثر العالم. وتحت هذه النظريات تعمل أدوات قياس وبوابات ومحاكم. وبذلك تنتقل النظرية من خطاب عن أهمية السؤال إلى نظام تشغيل قابل للتحويل إلى benchmark ووكيل تجريبي.

قيمة النظرية أنها تضبط موضعا غالبا ما يبقى خفيا: قبل أن نسأل هل النموذج أجاب جيدا، يجب أن نسأل هل السؤال الذي وجهه كان صالحا؟ هل كان واضح الأنطولوجيا؟ هل فتح الفرق الصحيح؟ هل استدعى الدليل المناسب؟ هل عرف شروط سقوطه؟ هل أنتج أثرا؟ حدود النظرية أنها لا تملك بعد برهانا تجريبيا خاصا، وأن مقاييسها أولية، وأن خطر التضخم الاصطلاحي قائم إذا لم يلتزم الباحث بتحويل كل مفهوم إلى أداة أو اختبار. أول اختبار حقيقي لها هو بناء benchmark يقارن prompt مفردا مع فضاء سؤال موثوق على مهام غامضة ومركبة.

23. اختبار الجودة النهائي

البند	قبل التحسين	ما تم تحسينه	بعد التحسين
حفظ المادة الأصلية	8/10	إدراج QMICAGI/MCD/QMCD/MCAI/HAKIM ومطلب الخمس نظريات	9/10
الوضع الأنطولوجي	4/5	تمييز السؤال كفضاء/شبكة/عملية	5/5
مراتب المعرفة	8/10	خفض ادعاءات AGI إلى فرضيات	9/10
تصفية المادة	8/10	إضافة صلب/راجح/خلافي/ضعيف	9/10
خريطة المتغيرات	8/10	إضافة QG/QR/QNI/WQI/عتبات	9/10
النماذج المتنافسة	8/10	مقارنة MCD/QMCD	9/10
أصالة النظرية	12/15	تحديد الجديد: حاكمية فضاء السؤال	13/15
قابلية الاختبار	7/10	إضافة benchmark وفرضيات وتنبؤات	9/10
محكمة الهدم	8/10	إضافة 12 اعتراضا وهدم الخرق	9/10
الأدوات والمقاييس	4/5	10 أدوات و7 مقاييس	5/5
غياب الحشو	4/5	ضغط المصطلحات وربطها بآليات	5/5

النتيجة قبل التحسين: 79/100. **النتيجة بعد التحسين:** 91/100. بقي التحسين الأكبر المطلوب: تحويل الورقة إلى تجربة معيارية منشورة وبيانات مقارنة.

24. الخاتمة

تأسس في هذه الورقة نموذج جديد داخل مشروع زهير مساعف: السؤال ليس طلبا بل فضاء، وليس كل فضاء سؤال موثوقا، ولا يكفي أن يولد السؤال إجابة، بل يجب أن يولد فرقا، ويقبل الهدم، ويوجه الموارد، ويترك أثرا، ثم يعيد بناء نفسه. بهذا تصبح QMICAGI خطوة نحو ذكاء فتّاح لا يعرف فقط كيف يجيب، بل يعرف كيف يفتح السؤال الذي يجعل المعرفة ممكنة. القيمة البحثية ليست في الادعاء بأننا بلغنا AGI، بل في تحديد موضع معماري واختباري جديد: **قياس حاكمية السؤال قبل قياس فصاحة الجواب.**

ملحق أ: التفصيل التشغيلي للنظريات الخمس

أ.1 نظرية طبقات السؤال الحاكمي

تعمل هذه النظرية على منع الخلط بين ظهور السؤال في اللغة واشتغاله في المعرفة. السؤال في طبقة الصيغة L0 قد يكون صحيحا نحويا لكنه فارغ حاكميا. وفي طبقة المقصد L1 يحدد الغاية العامة، لكنه قد يبقى عائما. وفي طبقة التفكيك L2 يتحول إلى أسئلة فرعية تكشف الفروق. وفي طبقة الاسترجاع/الاختبار L3 يحدد نوع الدليل المطلوب. وفي طبقة الحاكمية L4 يقرر متى يكفي الدليل ومتى يلزم الهدم. وفي طبقة أثر العالم L5 يخرج من النص إلى فعل أو قرار أو اختبار. قيمة هذه النظرية أنها تمنع مساواة السؤال العميق لفظيا بالسؤال العامل معرفيا.

أ.2 نظرية التراكب الوثوقي للسؤالات

كل سؤال مركب يحمل أكثر من قراءة: قراءة سببية، وقراءة بنيوية، وقراءة تاريخية، وقراءة قيمية، وقراءة تقنية. الانغلاق المبكر على قراءة واحدة يحول النموذج إلى آلة تأكيد. لذلك تفرض هذه النظرية أن تبقى القراءات المتنافسة حية إلى حين ظهور سؤال فاصل. ليس المقصود إطالة التحليل، بل منع القفز من الغموض إلى اليقين. كل قراءة تعطى احتمالا أوليا، ثم يرفعها أو يخفضها مصدر، أو اختبار، أو أثر، أو اعتراض.

أ.3 نظرية الانهيار الاختباري للسؤال

الانهيار هنا لا يعني سقوطا سلبيا بالضرورة؛ بل انتقال السؤال من حالة تراكب إلى حالة قرار مشروط. ينهار السؤال لصالح صيغة معينة إذا توفرت ثلاثة شروط: وضوح الأنطولوجيا، وجود دليل فاصل، وانخفاض الضجيج. وينهار السؤال ضد نفسه إذا تبين أنه محمل بافتراض خاطئ أو خلط بين مستويات. هذه النظرية تمنع الأسئلة من البقاء في فضاء التأمل المفتوح بلا حسم، وتمنعها أيضا من الحسم قبل الأوان.

أ.4 نظرية التوجيه الوعائي للسؤال

السؤال لا يوجه اللغة فقط؛ يوجه الموارد. فإذا كان السؤال تاريخيا فالمورد هو المصدر والسياق؛ وإذا كان تقنيا فالمورد هو الاختبار والتنفيذ؛ وإذا كان نظريا فالمورد هو محكمة المفاهيم والنماذج المنافسة؛ وإذا كان عالي المخاطر فالمورد هو التحقق والتقييد. هذا هو معنى الوعائية هنا: ليست الموارد خزانا عاما، بل تدفقا موجها بحسب ضغط السؤال. السؤال الرديء يستنزف الأدوات، والسؤال الجيد يقلل الكلفة لأنه يعرف أين يجب أن يهاجم.

أ.5 نظرية حلقة أثر العالم للسؤال

لا يكتمل السؤال داخل النص. السؤال العلمي ينتهي إلى اختبار، والسؤال الهندسي إلى تصميم، والسؤال الأخلاقي إلى قرار، والسؤال المعرفي إلى إعادة بناء نموذج. فإذا لم يغير السؤال شيئا في العالم أو في الذاكرة النظرية، يبقى تأملا لا حاكمية. لذلك تقترح هذه النظرية أن يسجل كل سؤال عالي الرتبة أثره: ماذا فتح؟ ماذا أغلق؟ ماذا هدم؟ ماذا عدل؟ ماذا جعل ممكنا لاحقا؟

ملحق ب: بروتوكولات تشغيل الأدوات العشر

الأداة	البروتوكول العملي	مؤشر النجاح
Question-Space Builder	خذ الطلب الأولي، استخرج الأنطولوجيا، ولد خمسة أسئلة من طبقات مختلفة، احذف المتكرر، رتبها حسب القيمة وعدم اليقين.	ظهور سؤال فاصل أو سؤال استرجاعي أو سؤال هدمي لم يكن في الطلب الأصلي.
Question Reliability Gate	قيّم الوضوح والحدود والمصادر والقابلية للاختبار من 0 إلى 5؛ إذا نزل المؤشر عن 3 امنع الجواب النهائي.	انخفاض الإجابات الواثقة على أسئلة مبهمة.
Generative Difference Extractor	حوّل كل سؤال إلى فروق: ما الشيء الذي يميزه؟ ما البديل؟ ما المستوى؟ ما العتبة؟	خريطة فروق قابلة للتقييم لا مجرد قائمة أفكار.
Question Superposition Matrix	ضع القراءات المتنافسة في جدول: احتمال، دليل مؤيد، دليل معارض، سؤال فاصل.	عدم اختيار تفسير واحد بلا سبب.
Falsifying Question Court	ولد ثلاثة أسئلة تهدم السؤال الأصلي: سؤال عكسي، سؤال حدود، سؤال مصدر.	تعديل السؤال أو تخفيض رتبته عند اللزوم.
Vascular Question Router	قرر: هل نحتاج بحثا، أداة، ذاكرة، حسابا، محاكاة، أو توقفا؟	تحسن نسبة الأداء إلى الكلفة.
World-Effect Question Loop	اربط السؤال بمخرج قابل للرصد: تجربة، قرار، نموذج، أو تحديث.	وجود أثر مسجل لا مجرد إجابة.
Higher-Question Regenerator	بعد الجواب أو الفشل، ولد سؤالا أعلى يفتح المسار التالي.	سؤال لاحق أكثر دقة أو أوسع مجالا.
Question Noise Scrubber	احذف الزخرفة، افصل المفاهيم، اخفض الادعاء، أعد صياغة السؤال كنواة.	انخفاض QNI وارتفاع QR.
QMICAGI Benchmark Protocol	صمم مهمة تقارن بين prompt مفرد وفضاء سؤال وتراكب ومحكمة.	فرق دال في الدقة أو الوثوق أو الأثر.

ملحق ج: مصفوفة محاور بحثية موسعة

المحور	السؤال البحثي	المنهج	القيمة
قياس وثوقية السؤال	هل QR يتنبأ بصحة الجواب؟	تجربة على أسئلة غامضة مع حكم بشري/آلي	تأسيس أول معيار للوثوق قبل الجواب.
ضجيج السؤال	هل QNI يتنبأ بالهلوسة؟	ربط غموض السؤال بمعدل الأخطاء	كشف سبب مبكر للهلوسة.

السؤال الفاصل	أي الأسئلة تميز بين النماذج؟	بناء مهام ذات تفسيرين وملاحظة السؤال الفاصل	تحسين المقارنة النظرية.
السؤال والاسترجاع	هل السؤال العلاقاتي يحسن GraphRAG؟	مقارنة retrieval مقطعي وعلاقاتي	رفع جودة البحث المركب.
ذاكرة السؤال	هل حفظ الأسئلة الساقطة يحسن التعلم؟	تجارب multi-turn agents	تحويل الفشل إلى ذاكرة منهجية.
التوجيه الوعائي	هل السؤال يقلل كلفة الأدوات؟	قياس عدد calls والأداء	اقتصاد معرفي/حاسوبي.
محكمة السؤال	هل هدم السؤال أفضل من هدم الجواب فقط؟	A/B على مهام جدلية	خفض الانغلاق والتحيز.
السؤال الأعلى	كيف نقيس سؤالا أعلى؟	تصنيف بشري ومؤشرات أثر	قياس الفتح لا الجواب.
سؤال الأمان	هل سؤال الخطر يخفض الاستجابة غير الملائمة؟	سيناريوهات عالية المخاطر	دمج الحاكمية الأخلاقية.
Field-Opening Benchmark	كيف نقيس فتح مجال جديد؟	مهام بحثية لا تملك جوابا واحدا	نقل التقييم من حل المسألة إلى فتح المسألة.

ملحق د: قاموس QMICAGI-GTQS المختصر

المصطلح	الصياغة العملية
Question-Space	مجموعة أسئلة مرتبطة بعلاقات احتمال وطبقة ووثوق وأثر.
Question Governance	نظام يقرر أي سؤال يعمل، وأي سؤال يهدم، وأي سؤال يؤجل.
Trustworthy Question	سؤال واضح الأنطولوجيا، محدود النطاق، قابل للتحقق، منخفض الضجيج.
Generative Question	سؤال ينتج فروقا ونماذج وأسئلة لاحقة لا جوابا فقط.
World-Effect	الأثر الذي يتركه السؤال في قرار أو اختبار أو نموذج أو ذاكرة.
Question Collapse	اختيار أو إسقاط قراءة من فضاء السؤال وفق دليل أو عتبة.
Question Noise	كل ما يجعل السؤال يبدو قويا لكنه لا يعمل: غموض، خلط، زخرفة، انحياز.
Higher Question	سؤال يولد بعد الهدم أو الأثر فيفتح طبقة أعلى أو أدق.

المراجع المختارة

- Wei, J. et al. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.
- Wang, X. et al. (2022). Self-Consistency Improves Chain of Thought Reasoning in Language Models.
- Yao, S. et al. (2022/2023). ReAct: Synergizing Reasoning and Acting in Language Models.
- Yao, S. et al. (2023). Tree of Thoughts: Deliberate Problem Solving with Large Language Models.
- Shinn, N. et al. (2023). Reflexion: Language Agents with Verbal Reinforcement Learning.
- Lewis, P. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.
- Asai, A. et al. (2023). Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection.
- Peng, B. et al. (2024). Graph Retrieval-Augmented Generation: A Survey.
- Han, H. et al. (2024/2025). Retrieval-Augmented Generation with Graphs (GraphRAG).
- Liu, X. et al. (2023/2024). AgentBench: Evaluating LLMs as Agents.
- Mialon, G. et al. (2023). GAIA: a Benchmark for General AI Assistants.
- Jimenez, C. E. et al. (2023/2024). SWE-bench: Can Language Models Resolve Real-World GitHub Issues?
- Schick, T. et al. (2023). Toolformer: Language Models Can Teach Themselves to Use Tools.
- Huang, H.-Y. et al. (2024). A Survey of Uncertainty Estimation in LLMs: Theory Meets Practice.
- Chen, L. et al. (2025). Self-Questioning Language Models.
- Press, O. et al. (2022). Measuring and Narrowing the Compositionality Gap in Language Models.