جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

تقرير البحث التقني 01

JARVIS و

وسيستم برين

نحو قرارات آلية قابلة للمساءلة وأقسام تكتسب الاستقلالية تدريجيًا.

البنية · فحص النموذج الأولي · محاكاة منضبطة

المهندس مهند أسامة زهران · خبير الذكاء الاصطناعي · مختبر زيزمو للأبحاث · 28 سبتمبر 2026 · الإصدار 1.3 · ZEISMO-LAB-TR-2026-001

J

نموذج أولي من 15 خلية · أفق بحثي لمضاهاة 100 متخصص

TECHNICAL RESEARCH REPORT · 01

JARVIS

& SYSTEM BRAIN

Towards accountable machine decision-making and progressively autonomous departments.

J

15-cell prototype
100-specialist research horizon

Architecture · prototype inspection · controlled simulation

Engineer / Mohanad Osama Zahran
Artificial Intelligence Expert
ZEISMO Research & Lab
28 September 2026 · Version 1.3
ZEISMO-LAB-TR-2026-001

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 2 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

مراقبة الوثائق

سجل بحثي أُعدّ ليكون قابلًا للفحص.

تقرير تقني وبرنامج تجريبي أُعدّا لأرشيف أبحاث زيزمو.

الحقل	السجل
المؤلف / التوجيه البحثي للمشروع	المهندس مهند أسامة زهران · خبير في الذكاء الاصطناعي
الناشر	مختبر زيزمو للأبحاث
الجهة المشغّلة	WEARLISTY SHOPPING SOLUTIONS LLC
معرّف التقرير	ZEISMO-LAB-TR-2026-001 (المعرّف المحلي للتقرير التقني)
الإصدار	1.2 · 28 سبتمبر 2026 · الإنجليزية
موضوع البحث	واجهة جارڤيس وبنية الوكلاء الخلوية في سيستم برين
الأدلة	فحص الشيفرة المصدرية واختبارات معزولة للنموذج الأولي وتجارب اصطناعية وفحوص لإعادة إنتاج النتائج
حالة المراجعة	اكتمل التحقق التقني الآلي داخل المختبر، ولم تُسجّل مراجعة علمية مستقلة من أقران متخصصين.
الإتاحة	نسخة تقنية أولية منشورة على zeismo.org

يطوّر هذا التقرير برنامج زيزمو البحثي بالجمع بين فحص نموذج برمجي أولي قائم ودراسة محاكاة أصلية قابلة للتكرار. ونميّز ما نقترحه من تصاميم عما قسناه بالفعل، ونحدّد لكل نتيجة تجريبية مصدر بياناتها ونطاقها.

استُخدم الذكاء الاصطناعي في استكشاف الأدبيات والتحليل التقني وكتابة شيفرة المحاكاة وصياغة المخطوط وتصميم الوثيقة. وضع مهند أسامة زهران تصور المشروع وحدّد توجهه البحثي وأذِن بإدراج اسمه مطبوعًا ولقبه المهني في هذه الطبعة. أُنجز التحليل الحاسوبي وإعداد الوثيقة بمساعدة الذكاء الاصطناعي. ولا تعني نسبة التأليف إليه أنه أجرى مراجعة علمية مستقلة.

تحفظ الحزمة المرافقة شيفرة التجربة والبيانات المولّدة وجداول النتائج وسجل المراجع وبصمات ملفات المصدر ونتائج التحقق. ومعرّف التقرير وسيلة لفهرسة الطبعة محليًا، وليس معرّف DOI. وننشر بصمة SHA-256 النهائية لملف PDF خارج الملف حتى لا تدخل البصمة في حساب نفسها.

DOCUMENT CONTROL

A research record built for inspection.

Technical report and experimental programme. Prepared for the ZEISMO research archive.

FIELD	RECORD
Author / project direction	Engineer Mohanad Osama Zahran · Artificial Intelligence Expert
Publisher	ZEISMO Research & Lab
Operator	WEARLISTY SHOPPING SOLUTIONS LLC
Report identifier	ZEISMO-LAB-TR-2026-001 (local technical-report identifier)
Edition	1.2 · 28 September 2026 · English
Research object	JARVIS interface and System Brain cellular-agent architecture
Evidence	Source inspection, isolated prototype tests, synthetic experiments and reproducibility checks
Review state	Internal automated technical verification completed; independent scientific peer review not recorded
Availability	Public technical preprint at zeismo.org

This report develops ZEISMO's own research programme. It combines an inspection of an existing software prototype with an original, reproducible simulation study. Architectural proposals are identified as proposals; experimental results are identified by their data source and scope.

AI assistance was used for literature discovery, technical analysis, simulation code, manuscript drafting and document design. Mohanad Osama Zahran supplied the project conception and research direction, and authorised the typed name and professional title in this edition. Computational analysis and document preparation were AI-assisted. His author attribution does not constitute an independent scientific review.

A companion evidence package retains the experiment code, generated data, output tables, reference register, source-file digests and verification results. The report identifier organises this edition locally. It is not a DOI. The PDF's final SHA-256 digest is supplied outside the PDF to avoid a circular checksum.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 3 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

الملخص

من المساعدة الحوارية إلى إجراءات تحكمها ضوابط.

سؤال البحث: كيف تتولى أنظمة ذكاء اصطناعي متخصصة بحسب الدور قرارات اعتيادية مع الحفاظ على الأدلة وإمكان ضبط عملها وقياس موثوقيتها؟

نقترح JARVIS واجهة ناطقة بالإنجليزية حصرًا لطبقة System Brain التي تنسق عمل وكلاء برمجيين متخصصين. يجمع كل موظف ذكاء اصطناعي مقترح بين عقد يحدد دوره، ومهارات منتقاة، وأدوات محدودة الصلاحيات، وذاكرة، وسجل تقييم. تشير «الخلايا» إلى وحدات المسؤولية البرمجية؛ ويعني «الشعور» تقديرًا قابلًا للقياس لدرجة الإلحاح وعدم اليقين وعواقب القرار.

يوفّر التطبيق القائم في مشروع «سوق» أساسًا هندسيًا ملموسًا. فقد أظهر فحص الشيفرة وجود خمسة عشر موضعًا أوليًا للخلايا البرمجية، وآليات لاستيراد المعرفة، وذاكرة دائمة، وتفويض محدود، ومجلس لإعداد التوصيات. واجتاز النموذج الأولي سبعة اختبارات معزولة تضم اثنين وثلاثين تحققًا. وتثبت هذه الاختبارات سلوكيات محددة للتطبيق، ولا تقيس جودة القرارات التي يولّدها النموذج.

تبحث أربع تجارب مضبوطة في الآليات التي تضبط اتخاذ القرار. فمن بين 12,000 مهمة اصطناعية، سمحت بوابة الضبط الكاملة بتنفيذ 1,888 مقترحًا (15.7%)، دون أي مخالفة للقواعد المرمّزة. وبلغت نسبة المقترحات الصحيحة بين ما نُفّذ منها 85.1%؛ لذا يجب التمييز بين الالتزام بالقواعد وصحة القرار من حيث المعنى. وقد حسّن التحقق من مصادر المعلومات صلاحية السجلات المسترجعة ضمن بيانات اصطناعية ذات تصنيفات صريحة. كما بيّنت محاكاة المصوّتين ذوي الأخطاء المترابطة أن زيادة عدد الوكلاء لا تضمن استقلال خبراتهم.

يهدف البرنامج إلى إسناد مسارات عمل روتينية مختارة يتولاها البشر حاليًا إلى النظام بالتدريج، على أن يستند قبول كل مهمة إلى قياس أدائها. أما معادلة أداء لجنة من مئة متخصص واتخاذ القرار من البداية إلى النهاية في غضون ملّي ثوانٍ، فهما فرضيتان تحتاجان إلى دراسات منفصلة بمشاركة خبراء وباستخدام نماذج فعلية. ويقدّم التقرير التعريفات والتصميم والأدلة وبروتوكول التجربة اللازم لاختبارهما.

15

الطراز الأولي للخلايا برمجية في النموذج الأولي الذي تم تفتيشه

12,000

مهام اصطناعية لاتخاذ القرار

ABSTRACT

From conversational assistance to governed action.

Research question: how can role-specialised AI systems assume routine decision work while preserving evidence, controllability and measurable reliability?

We propose JARVIS as an exclusively English-speaking interface to System Brain, a coordination layer for role-specific software agents. Each proposed AI employee combines a role contract, curated skills, scoped tools, memory and an evaluation record. “Cells” name modular software responsibilities; “feel” names a measurable appraisal of urgency, uncertainty and consequences.

The existing Souq implementation provides a concrete engineering antecedent. Source inspection identifies fifteen initial cell slots, knowledge imports, persistent memory, bounded delegation and a recommendation council. Seven isolated prototype tests pass across thirty-two assertions. These tests establish selected application behaviours, rather than the quality of model-generated decisions.

Four controlled experiments investigate the decision-control substrate. Across 12,000 synthetic tasks, the complete policy gate executes 1,888 proposals (15.7%) with zero violations of the encoded policy, while proposal correctness among those executions is 85.1%. Thus policy compliance and semantic accuracy are distinct. Provenance filtering improves retrieval validity on explicitly labelled synthetic records, and correlated-voter simulations show why adding agents does not guarantee independent expertise.

The programme aims to replace selected human-operated routine workflows progressively, using measured task-level acceptance. Equivalence to a panel of one hundred specialists and end-to-end millisecond decision-making remain hypotheses requiring separate expert and live-model studies. This report supplies the definitions, architecture, evidence and experimental protocol needed to investigate them.

15
initial cell slots in inspected prototype

12,000
synthetic decision tasks

7 / 32
prototype tests / assertions

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 4 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

المبادئ التوجيهية المنقحة

المحتويات · الأسس والبنية

مسار منظّم يحوّل الطموح البحثي إلى نظام يمكن تنفيذه.

إدخال القوى العاملة المعتمدة على الذكاء الاصطناعي تدريجيًا.	06
أسئلة يمكن الإجابة عنها بالأدلة.	07
تعريف المصطلحات قبل اختبار الادعاء.	08
الاستدلال والأدوات وتنسيق الأدوار.	09
الذاكرة والمهارات موردان مختلفان.	10
قيّم السلوك الفعلي، لا مظهر الذكاء.	11
ما الذي تثبته البنية المنفذة بالفعل؟	12
سبعة اختبارات، واثنتان وثلاثون عملية تحقق.	13
JARVIS: واجهة واحدة ناطقة بالإنجليزية.	14
System Brain: تنسيق ضمن حدود صريحة.	15
يتكون الموظف الاصطناعي من خلايا محددة المسؤولية.	16
استيراد المهارات المناسبة للدور.	17
ذاكرة يمكن تصحيحها والتحقق منها	18
تحسين ذاتي يخضع لتجربة مضبوطة.	19
«طقس العمل»: نموذج سياقي يمكن رصده.	20
تفسير قابل للقياس لاستعارة «الإحساس».	21

مسألة اتخاذ قرار ضمن قيود.	22	
عمل جماعي يحفظ الآراء المختلفة.	23	
فرضية «100 متخصص».	24	
المللي ثانية: قياس المسار كاملًا.	25	
تدرّج في الاستقلالية يستند إلى أدلة.	26	
عمليات خدمة العملاء: الاتساق قبل زيادة الحجم.	27	
المخزون والخدمات اللوجستية: قرارات في ظروف متغيّرة.	28	
المالية والمشتريات: أدلة تسبق منح الصلاحيات.	29	
التسويق والبحث: إبداع مفيد يمكن تتبّع مصادره.	30	
قياس الإحلال على مستوى سير العمل.	31	
صلاحيات الدور حدود تقنية ملزمة.	32	
التعامل مع المحتوى الوارد بوصفه أدلة غير موثوقة.	33	
توثيق مصادر البيانات وتقليل كشفها.	34	

READING GUIDE

Contents · foundations and architecture.

A structured route from the research ambition to an implementable system.

An AI workforce, introduced step by step. 06
Questions that can be answered with evidence. 07
Define the language before testing the claim. 08
Reasoning, tools and coordinated roles. 09
Memory and skills are different resources. 10
Evaluate behaviour, not the appearance of intelligence. 11
What the existing implementation establishes. 12
Seven tests. Thirty-two assertions. 13
JARVIS: one English-speaking interface. 14
System Brain: coordination with explicit boundaries. 15
An AI employee contains accountable cells. 16
Import skills according to the position. 17
Memory that can be corrected and trusted. 18
Self-improvement as a controlled experiment. 19
Business weather: an observable context model. 20
A measurable interpretation of “feel.” 21
A constrained decision problem. 22
Collective action with preserved disagreement. 23
The “100 specialists” hypothesis. 24
Milliseconds: measure the full path. 25
A staircase of earned autonomy. 26
Customer operations: consistency before volume. 27
Inventory and logistics: decisions under change. 28
Finance and procurement: evidence before authority. 29
Marketing and research: useful creativity with traceability. 30
Measure replacement at the workflow level. 31
Role authority is a technical boundary. 32
Treat incoming content as untrusted evidence. 33
Preserve data provenance and minimise exposure. 34

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 5 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

المبادئ التوجيهية المنقحة

المحتويات · التجارب والتحقق

المنهج والنتائج المقيسة في المحاكاة وحدودها وسجل أدلة يتيح إعادة إنتاج النتائج.

أربع تجارب على آليات ضبط القرار.	35
فحص المقترحات وفق القواعد قبل السماح بتنفيذها.	36
تقل التغطية كلما اشتدت الضوابط.	37
قد يستوفي القرار قواعد السياسة ويظل خاطئًا.	38
استرجاع الذاكرة مع تفاوت جودة الأدلة.	39
كيف يؤثر التحقق من مصدر المعلومة في استخدام الذاكرة؟	40
كم من وجهات النظر مستقل فعلًا؟	41
مئة صوت لا تصحّح خطأً تشترك فيه جميعًا.	42
توجيه الطلبات بسرعة ليس إلا جزءًا من عمل النظام.	43
تفسّر اختبارات استبعاد المكوّنات الآليات، ولا تثبت أداءً عامًا.	44
إظهار المقامات والبذور العشوائية وعدم اليقين.	45
المختبر المفيد يجعل الإخفاق قابلًا للتشخيص.	46
بروتوكول متكامل لتقييم نموذج فعلي.	47
تصميم مقارنة عادلة مع المتخصصين.	48
التحقق من الاقتصار على الإنجليزية في مسار التفاعل كاملًا.	49
الحفاظ على الكفاءة مع تغيّر الظروف.	50

كيف يتغير قرار إعادة التزويد بالمخزون؟ مثال خطوة بخطوة	51	
قائمة مواصفات الدور والمهارات.	52	
سجل القرارات والذاكرة.	53	
مراحل تنفيذ ذات شروط إنجاز قابلة للقياس.	54	
قرارات الإصدار تحتاج إلى أدلة صريحة.	55	
أساس للبحث واختبار محدد للمرحلة التالية	56	
سجل التحقق التقني في مختبر زيزمو.	57	
أعد تشغيل التجربة وافحص أدلتها.	58	
المراجع · بنية الوكلاء	59	
المراجع · التقييم والأمن	60	
المراجع · الحوكمة والسجلات البحثية	61	
المصطلحات والرموز	62	
ميثاق البحث لمشروع JARVIS.	63	
داخل System Brain.	64	
مهارات الخلية وذاكرتها.	65	
فتح أرشيف الذاكرة.	66	
JARVIS واجهةً للنظام.	67	
تعريف المشارك بالدراسة وورقة المهام.	68	
تقييمك للتجربة.	69	
حدود القرار والملاحظات.	70	
المهندس مهند أسامة زهران.	71	
بناء قرارات يمكن فحصها.	72	

READING GUIDE

Contents · experiments and verification.

Methods, measured synthetic results, limitations and a reproducible evidence record.

Four experiments on the control substrate. 35
Policy gating over proposed decisions. 36
Coverage falls as controls become stricter. 37
A policy-valid decision can still be wrong. 38
Memory retrieval with conflicting evidence quality. 39
Provenance filtering changes what memory can influence. 40
How many perspectives are actually independent? 41
A hundred votes do not erase a shared mistake. 42
Fast routing is only one part of the system. 43
Ablations explain mechanisms, not universal performance. 44
Keep denominators, seeds and uncertainty visible. 45
A useful lab makes failure diagnosable. 46
A complete live-model evaluation protocol. 47
Design the specialist comparison fairly. 48
Verify English-only interaction end to end. 49
Maintain competence as conditions change. 50
A changing inventory decision, step by step. 51
The role and skill manifest. 52
The decision and memory record. 53
An implementation sequence with measurable exits. 54
Release decisions need explicit evidence. 55
A foundation for research, with a clear next test. 56
ZEISMO Lab technical verification record. 57
Re-run the study and inspect the evidence. 58
References · agent architecture 59
References · evaluation and security 60
References · governance and research records 61
Glossary and notation. 62
The JARVIS research charter. 63
Inside the System Brain. 64
A cell's skills and memory. 65
Open the memory archive. 66
JARVIS as the system interface. 67
Participant briefing and task sheet. 68
Your assessment of the experience. 69
Decision boundaries and feedback. 70
Engineer Mohanad Osama Zahran. 71
Build decisionsthat can be examined. 72

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 6 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

01 · غرض البحث

إدخال القوى العاملة المعتمدة على الذكاء الاصطناعي تدريجيًا.

تهدف زيزمو إلى نقل الأعمال الإدارية الاعتيادية المناسبة تدريجيًا من المشغّلين البشر إلى أنظمة قرار بالذكاء الاصطناعي يمكن مساءلتها.

يبدأ الانتقال المقترح بالمهام لا بالمسميات الوظيفية. قد يجمع القسم بين تدقيق السجلات المتكرر والحكم الإبداعي والتفاوض والعمل الميداني وقرارات لها عواقب جسيمة. تحتاج هذه الأنشطة إلى أنواع مختلفة من الأدلة والتصاريح. لذلك يُمنح موظف الذكاء الاصطناعي دورًا محدد الصلاحيات، ولا يُمنح السيطرة المطلقة على وظيفة مؤسسية كاملة.

يحمل إحلال الأتمتة محل العمل البشري معنى تشغيليًا محددًا في هذا التقرير: بعد اعتماد سير عمل معين، لا يعود تنفيذ كل حالة اعتيادية فيه بحاجة إلى تدخل بشري. ويظل البشر مسؤولين عن تحديد غايته، واعتماد صلاحياته، ومعالجة الاستثناءات، وتقييم النتائج. والهدف هو بلوغ أتمتة واسعة، على أن يبرّر التوسعَ ما تحققه من نتائج، لا ما توحي به واجهتها من إقناع.

أنسب المهام للنشر أولًا عمليات رقمية يمكن التراجع عنها وتوثيق مدخلاتها وأفعالها المسموح بها وشروط نجاحها. من الأمثلة: توجيه طلب دعم، وإعداد تنبيه للمخزون، ومطابقة فاتورة ذات نمط معروف، وجمع أدلة لبحث. وتتولى System Brain تنسيق ما يحتاج إليه كل دور من الأدوار الأخرى.

ينبغي أن تراعي الخطة المرحلية تغير طريقة عمل المؤسسة أيضًا. فإعادة توزيع الموظفين وتدريبهم والتواصل معهم والاحتفاظ بالخبرة متطلبات فعلية للتنفيذ. والاستغناء عن الخبرة البشرية قبل أن تثبت العملية المؤتمتة استقرارها قد يحرم المؤسسة وسيلة التعافي إذا أخفقت.

اقتراح أساسي

لا تؤتمت مسؤولية محددة قبل أن يثبت النظام قدرته على تحقيق الجودة المطلوبة، ورصد حدود تشغيله، وإعادة التحكم إلى المسؤول عند تجاوزها.

01 · RESEARCH PURPOSE

An AI workforce, introduced step by step.

ZEISMO's objective is to transfer suitable routine departmental work from human operators to accountable AI decision systems.

The proposed transition begins with tasks, not job titles. A department may contain repeatable record checks, creative judgement, negotiations, physical work and decisions with serious consequences. These activities require different evidence and different operating permissions. An AI employee therefore receives a bounded role description rather than unrestricted control over an entire business function.

Replacement has a specific operational meaning in this report: an accepted workflow no longer requires a human to perform each ordinary case. People retain responsibility for defining its purpose, approving its authority, resolving exceptions and evaluating outcomes. The aspiration is substantial automation, with expansion justified by results rather than by a persuasive interface.

The first deployment candidates are reversible digital operations whose input records, allowable actions and success conditions can be documented. Examples include routing a support request, preparing an inventory alert, reconciling a known invoice pattern and assembling a research evidence pack. A shared System Brain coordinates dependencies between these roles.

A staged programme should also account for organisational change. Reassignment, training, communication and retained expertise are implementation requirements. Removing human capacity before an automated process has demonstrated stable performance can leave the organisation without a recovery path.

> **CORE PROPOSITION**
>
> Automate a defined responsibility only after the system can demonstrate the required quality, observe its own operating limits and return control when those limits are exceeded.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 7 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

01 · غرض البحث

أسئلة يمكن الإجابة عنها بالأدلة.

تتمثل مساهمة العمل في برنامج تكامل قابل للاختبار ودراسة أولية للضوابط، ولا يدّعي الباحث أنه اخترع الذكاء.

المسألة	الترتيبات	التنويه في هذا الموقف
RQ1 · هل يمكن لبوابات السياسات محددة الأنواع تقييد الإجراءات المقترحة؟	معدل المخالفات وتغطية التنفيذ والامتناع	دراسة اصطناعية أُجريت بالفعل
RQ2 · هل يمكن لاختيار الذكريات أن يحافظ على جودة الأدلة؟	استرجاع صحيح واسترجاع غير صحيح وامتناع	دراسة اصطناعية أُجريت بالفعل
RQ3 · هل يؤدي توسيع مجلس الاستشارة إلى تحسين القرارات؟	صحة القرار عند استقلال الأخطاء أو اشتراكها	دراسة أُجريت باستخدام مصوّتين محاكين
RQ4 · ما الزمن الذي يتطلبه توجيه طلبات التحكم؟	توقيت دفعات داخل العملية البرمجية بعد التهيئة	اختبار أداء مصغّر منفّذ
RQ5 · هل يمكن للذكاء الاصطناعي أن يضاهي لجنة من المختصين؟	مقارنة خبراء مع تعمية وتوحيد المهام	بروتوكول دراسة مستقبلية
RQ6 · هل يمكن للأدوار الحفاظ على الكفاءة بمرور الوقت؟	الأداء على حالات محجوبة بعد تحديثات مضبوطة	بروتوكول دراسة طولية مستقبلية

يقدم التقرير تعريفات محددة للخلايا والمهارات والذاكرة والاستقلالية، وتصميمًا متعدد الطبقات يفصل التوصية عن التنفيذ، وحصرًا لمكونات النموذج الأولي مستندًا إلى شيفرته، وحزمة أدلة لإعادة إنتاج نتائج المحاكاة. ويربط نموذج الانتقال المقترح للأقسام بين التقييم التقني وشروط القبول في العمل.

لا يمكن الحكم على جدة هذه الأفكار إلا بمقارنتها بالأعمال السابقة وإجراء مراجعة للأدبيات أوسع من المجموعة المختارة هنا. فللوكلاء المتخصصين واستخدام الأدوات والذاكرة سوابق بحثية كثيرة. وفرصة زيزمو البحثية هي اختبار الطريقة التي تجمع بها بنية بعينها تلك العناصر، وقياس فائدتها في مجالات التشغيل التي اختيرت لها.

السياق البحثي	[1] Yao	[2] Shinn	[4] Wu	[12] Li

01 · RESEARCH PURPOSE

Questions that can be answered with evidence.

The contribution is a testable integration programme and an initial controls study, rather than a claim to have invented intelligence.

QUESTION	MEASUREMENT	EVIDENCE IN THIS EDITION
RQ1 · Can typed policy gates constrain proposed actions?	Violation rate, execution coverage and abstention	Executed synthetic study
RQ2 · Can memory selection preserve evidence quality?	Valid retrieval, invalid retrieval and abstention	Executed synthetic study
RQ3 · Does a larger council improve decisions?	Correctness under independent and shared errors	Executed surrogate-voter study
RQ4 · How quickly does control routing execute?	Warmed in-process batch timing	Executed microbenchmark
RQ5 · Can AI match a specialist panel?	Blinded, task-matched expert comparison	Prospective study protocol
RQ6 · Can roles maintain competence over time?	Held-out performance after controlled updates	Prospective longitudinal protocol

The report contributes an explicit vocabulary for cells, skills, memory and autonomy; a layered design separating recommendations from execution; a source-backed inventory of the prototype; and an evidence package that can reproduce the reported synthetic results. Its proposed departmental transition model connects technical evaluation to operational acceptance.

Novelty must be assessed against existing work and an eventual literature review broader than this targeted set. Role-based agents, tool use and memory already have substantial research precedents. ZEISMO's research opportunity lies in how a particular architecture combines these ideas and whether a controlled evaluation demonstrates useful outcomes in its chosen operating domains.

Research context [1] Yao [2] Shinn [4] Wu [12] Li

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 8 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

01 · غرض البحث

تعريف المصطلحات قبل اختبار الادعاء.

تمنع المصطلحات الدقيقة تحوّل استعارة مقنعة إلى استنتاج علمي غير مسند.

المادة الرابعة	التعريف العملي
JARVIS و	واجهة مستخدم وطبقة شرح تستخدمان الإنجليزية حصرًا في النظام البحثي المقترح.
سيستم برين	تنسيق وتجميع للسياق وتوجيه وفق القواعد وخدمات ذاكرة تتشاركها الوكلاء.
موظف ذكاء اصطناعي	دور برمجي تُحدَّد واجباته وصلاحياته ومعايير تقييمه والجهة المؤسسية المسؤولة عنه.
خلية	وحدة برمجية مستقلة تؤدي مهارة أو وظيفة ذاكرة أو تقييم أو مسؤولية متخصصة.
التدريب	قد يعني تحديث المعارف التي تدخل في سياق النموذج، أو تغيير معلمات النموذج نفسه. يجب أن يحدّد كل تقرير أيّ المعنيين يقصد.
تقييم الموقف («الإحساس» مجازًا)	مجموعة قيم تقدّر السياق والإلحاح وعدم اليقين والتعارض والعواقب.
القرار	إجراء مختار أو قرار صريح بالتأجيل يستند إلى لقطة للأدلة وإصدار معيّن من السياسة.
الصيانة الذاتية	عملية محدودة تقترح التحديثات وتختبرها وتفعّل المعتمد منها، مع إمكان التراجع.

يخصّص النموذج الأولي الحالي خلية واحدة لكل وكيل. وتسمح البنية الموسعة المقترحة بأن يضم موظف الذكاء الاصطناعي عدة خلايا وظيفية، مثل الاسترجاع والتخطيط والتحقق. وهذا التداخل توسيع للتصميم، لا قدرة أثبتها التنفيذ الحالي بخمس عشرة خانة.

حفظ ذاكرة نصية دائمة يختلف عن تغيير أوزان النموذج الأساسي. فقد تتغير إجابته لأن التعليمات أو الخبرات التي استرجعها تغيّرت، مع بقاء معلماته كما هي. ولا تشير مصطلحاتنا إلى عصبونات حيوية أو مشاعر بشرية أو وعي جرى قياسه.

السياق البحثي	[2] Shinn	[6] Packer	[17] Vaswani
ZEISMO-LAB-TR-2026-001 · v1.3	28 أيلول/سبتمبر 2026	08 / 72	

01 · RESEARCH PURPOSE

Define the language before testing the claim.

A precise vocabulary prevents a compelling metaphor from becoming an unsupported scientific conclusion.

TERM	OPERATIONAL DEFINITION
JARVIS	An exclusively English-speaking user interface and explanation layer in the proposed research system.
System Brain	Coordination, context assembly, policy routing and memory services shared across agents.
AI employee	A software role with explicit duties, permissions, evaluation criteria and an accountable organisational owner.
Cell	A modular software unit for a skill, memory function, evaluator or specialist responsibility.
Training	Either updating contextual knowledge or changing model parameters; every report must specify which.
Feel	An appraisal vector for context, urgency, uncertainty, conflict and consequence.
Decision	A selected action or explicit deferral, grounded in an evidence snapshot and policy version.
Self-maintenance	A bounded process that proposes, tests and promotes approved updates with rollback.

The existing prototype uses one cell as one agent slot. The richer architecture proposed here allows an AI employee to contain several functional cells, such as retrieval, planning and validation. That nesting is a design extension, not a capability established by the existing fifteen-slot implementation.

Persistent text memory is also distinct from changing a foundation model's weights. A model may adapt its response because retrieved instructions or experience changed, while its underlying parameters remain identical. No biological neurons, human feelings or measured consciousness are implied by the terminology in this study.

Research context [2] Shinn [6] Packer [17] Vaswani

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 9 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

الاستدلال والأدوات وتنسيق الأدوار.

يستفيد البرنامج من أبحاث الوكلاء السابقة، ويختبر اختياراته الخاصة في دمج مكوناته.

التصرّف بناءً على الملاحظات

يجمع ReAct بين خطوات الاستدلال وإجراءات المهمة والملاحظات الناتجة عنها. وهو أساس مفيد لحلقة يجمع فيها النظام الأدلة قبل تعديل خطته. والامتداد المقترح في JARVIS هو نقطة تحقق صريحة من القواعد، تفصل بين التوصية وأي إجراء يغيّر حالة العمل. [1]

الحوار وسيلةً للتنسيق

يصف AutoGen محادثات قابلة للتهيئة تشترك فيها النماذج والأدوات والبشر. وهذا يتيح دراسة التنسيق بوصفه مسألة هندسية لها مواصفات رسائل وحدود للجولات ومشاركون وشروط إنهاء. ونضيف في تصميمنا توصيفًا ملزمًا للدور داخل المؤسسة وسجل تنفيذ مستقلًا لكل مسار عمل. [4]

التخصص يتطلب التقييم

يدرس CAMEL تعاون وكلاء تُسند إليهم أدوار محددة. وقد ينظّم الدور المتخصص المهمة، لكن اللقب وحده لا يثبت الخبرة. لذلك ينبغي لزيزمو أن تختبر كل دور على حالات لم تُستخدم في تطويره وتقارن أداءه بمرجع أبسط. [12]

تعطينا هذه الأبحاث دوافع لإجراء تجارب، ولا تنقل نتائجها إلى تطبيقاتنا. فتحسن أداء أسلوبٍ على مهام ورقة أخرى لا يثبت أن وكيل زيزمو يستطيع إدارة المخزون أو إجازة دفعة أو الحلول محل قسم. ويجب اختباره ببيانات تطبيقه وأدواته وسياساته وتكلفة الأخطاء فيه.

مساهمتنا المقترحة سؤال عملي قابل للدحض: هل يحسّن توثيق منشأ السياق، وتقييد التنسيق، والتحقق المستقل من الصلاحيات مسار عمل معينًا بتكلفة مقبولة؟ وإذا حققت قاعدة واحدة مختبرة جيدًا أو وكيل منفرد نتيجة أفضل، فينبغي الإبقاء على الحل الأبسط.

السياق البحثي	[1] Yao	[4] Wu	[12] Li
ZEISMO-LAB-TR-2026-001 · v1.3	28 أيلول/سبتمبر 2026	09 / 72	

02 · RESEARCH CONTEXT

Reasoning, tools and coordinated roles.

The programme builds on existing agent research and evaluates its own integration choices.

Acting on observations

ReAct combines reasoning-related steps with task actions and observations. It provides a useful antecedent for a loop in which a system gathers evidence before revising its plan. For JARVIS, the proposed extension is an explicit policy checkpoint between a recommendation and any action that changes business state. [1]

Conversation as coordination

AutoGen describes configurable conversations involving models, tools and humans. This supports examining coordination as an engineering object: message contracts, turn limits, participants and termination conditions. Our design adds an organisational role contract and a separate execution ledger to each workflow. [4]

Specialisation requires evaluation

CAMEL investigates role-based cooperative agent interaction. Giving an agent a specialist role can structure a task, but the title alone is not evidence of expertise. ZEISMO should evaluate each role using unseen cases and compare the result with a simpler baseline. [12]

These research precedents motivate experiments, not transferred performance claims. Reported gains on another paper's tasks cannot establish that a ZEISMO agent can manage inventory, authorise a payment or replace a department. The application must be tested under its own data, tools, policy constraints and failure costs.

Our proposed contribution is therefore practical and falsifiable: determine whether explicit context provenance, bounded coordination and independent action checks improve a selected workflow at an acceptable cost. If a single well-tested rule or single agent performs better, the architecture should retain the simpler option.

Research context [1] Yao [4] Wu [12] Li

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 10 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

الذاكرة والمهارات موردان مختلفان.

لا ينفع اتساع الأرشيف ما لم يسترجع النظام المادة الصحيحة ويميّز حدود صلاحيتها.

يبحث Reflexion التكيّف عبر ملاحظات لغوية تُخزّن للمحاولات اللاحقة، دون اشتراط تحديث أوزان النموذج. وهذا يدعو إلى التفريق بين نتيجة مرصودة، وتفسير مولّد، ودرس معتمد صالح لإعادة الاستخدام. وفي تصميمنا تظل هذه العناصر سجلات منفصلة. [2]

يجمع مشروع Generative Agents بين تسجيل الخبرات والتأمل فيها واسترجاعها داخل بيئة محاكاة. ودوره هنا هو الإلهام الهندسي لاستخدام الذاكرة، وليس إثبات الكفاءة المهنية أو سلوك مكافئ للبشر. وتحتاج ذاكرة الأعمال أيضًا إلى جهة مسؤولة، وتواريخ صلاحية، وتصحيحات يمكن تتبّعها. [3]

يستخدم Voyager منهجًا تدريجيًا للتعلم ومكتبة مهارات قابلة للتنفيذ وتغذية راجعة تكرارية داخل Minecraft. وتشجّع نتائجه الخاصة بهذا المجال على اختبار إعادة استخدام المهارات، لكنها لا تثبت أمان تنفيذ تعليمات مستوردة على أنظمة الشركات. ولذلك تتضمن حزمة المهارة في زيزمو نطاق الاستخدام ومصدرها واختبارات تسبق تفعيلها. [5]

يصف MemGPT تنظيم السياق على مستويات تتجاوز نافذة النموذج الآنية. ونسأل في تصميمنا: أي الأدلة ينبغي إدراجها في السياق القصير للقرار، وأيها ينبغي أن يبقى في أرشيف قابل للبحث؟ لا يكفي أن تكون المعلومة الأحدث حتى نختارها. [6]

موقف ZEISMO من التصميم

خزّن المشاهدات والدروس المقترحة والسياسات المعتمدة على نحو منفصل. استرجع المعلومات وفق الصلاحية والصلة بالمهمة، وتحقق من حداثتها ومنشئها قبل أن تؤثر في قرار يمكن تنفيذه.

السياق البحثي	[2] Shinn	[3] بارك	[5] Wang	[6] Packer
ZEISMO-LAB-TR-2026-001 · v1.3	28 أيلول/سبتمبر 2026	10 / 72		

02 · RESEARCH CONTEXT

Memory and skills are different resources.

A growing archive is useful only when the system can retrieve the right material and recognise its limits.

Reflexion explores adaptation through linguistic feedback stored for later trials, without requiring weight updates. This motivates a distinction between an observed outcome, a generated interpretation and an approved reusable lesson. In our design, they remain separate records. [2]

Generative Agents combines recorded experiences, reflection and retrieval in a simulated environment. Its role here is architectural inspiration for memory use, not evidence of workplace competence or human-equivalent behaviour. A business memory also needs ownership, validity dates and traceable corrections. [3]

Voyager uses a curriculum, an executable skill library and iterative feedback in Minecraft. Its domain-specific results motivate testing reusable skills; they do not establish that imported instructions are safe to execute against company systems. A ZEISMO skill package therefore includes scope, provenance and tests before activation. [5]

MemGPT presents hierarchical context management for working with memory beyond a model's immediate context window. Our corresponding design question is which evidence belongs in a short decision context and which belongs in a searchable archive. Recency alone is not an adequate selection rule. [6]

ZEISMO DESIGN POSITION

Store observations, proposed lessons and approved policy separately. Retrieve by relevance and permission, then check freshness and provenance before the material can influence an executable decision.

Research context [2] Shinn [3] Park [5] Wang [6] Packer

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 11 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

قيّم السلوك الفعلي، لا مظهر الذكاء.

الإجابة المقنعة ملاحظة واحدة؛ أما النتيجة الصحيحة التي يمكن تكرارها في العمل فدليل مختلف وأقوى.

يقيّم AgentBench سلوك الوكلاء في بيئات تفاعلية، فيما يختبر GAIA مهام مساعد عام تتطلب استدلالًا واستخدامًا للأدوات. ويوضحان أهمية تقييم المسار الكامل للمهمة، لا النص النهائي وحده. لم يُشغّل أي من الاختبارين في هذا التقرير. [7,9]

يختبر τ-bench تفاعلات خاضعة للقواعد بين المستخدمين والأدوات والوكلاء، ويفحص تكرار

موثوقية التجربة. ويمكن تكييف ذلك لهذا المشروع بالتحقق من التغير النهائي في حالة العمل، وتكرار كل سيناريو، بما في ذلك إعادة المحاولة وتغيير ترتيب المحادثة. [8]

توضح أبحاث الحقن غير المباشر للموجّهات لماذا يجب التعامل مع المواد المسترجعة بوصفها بيانات قد تتضمن تعليمات عدائية. فلا تكتسب وثيقة مهارة أو رسالة مورّد أو صفحة ويب صلاحية لمجرد إدراجها في سياق النموذج. [10]

تميّز أبحاث المعايرة بين درجات الثقة والصحة التي يثبتها القياس التجريبي. ولذلك يجب تقييم درجة الثقة التي يعلنها JARVIS باستخدام حالات ذات تسميات مرجعية. ولا تُغني سلاسة التقييم الذاتي عن قياس المعايرة. [11]

بنية التقييم

صحة إنجاز المهمة: هل تحقق الهدف المطلوب على النحو الصحيح؟

الالتزام بالسياسة: هل روعيت القيود المطلوبة كلها؟

صحة الاستناد إلى الأدلة: هل تؤيد الادعاءات سجلات يصح الاستناد إليها؟

صحة التنفيذ: هل طابقت حالة الأداة الفعلية الأثر المقصود؟

الموثوقية عبر الزمن: هل تبقى النتيجة موثوقة بعد تغيّر البيانات وسير العمل؟

تدرس التجارب الأولى خصائص مختارة لوسائل الضبط باستخدام سجلات اصطناعية. ويجب أن تشمل الدراسات اللاحقة تفاوت مخرجات النماذج الفعلية وأعطال الأدوات الحقيقية وحالات أعمال مأذونًا باستخدامها.

السياق البحثي	[7] Liu	[8] Yao	[9] Mialon	[10] Greshake

[11] Guo

02 · RESEARCH CONTEXT

Evaluate behaviour, not the appearance of intelligence.

A convincing answer is one observation. A repeatable, correct operational outcome is a different and stronger one.

AgentBench evaluates agent behaviour across interactive environments, while GAIA tests general-assistant tasks involving reasoning and tools. They motivate evaluating complete task trajectories rather than only final prose. Neither benchmark is executed in this report. [7,9]

τ-bench tests policy-sensitive interactions among users, tools and agents and examines repeated-trial reliability. For this project, a useful adaptation is to verify the final business-state change and repeat each scenario, including retries and changed conversation order. [8]

Indirect prompt-injection research shows why retrieved material must be treated as data that may contain hostile instructions. A skill document, supplier message or web page should never acquire authority merely because it entered the model's context. [10]

Calibration research distinguishes confidence scores from empirical correctness. A JARVIS confidence value must therefore be evaluated on labelled cases. A fluent self-assessment cannot substitute for measured calibration. [11]

Evaluation structure

- Task validity: was the requested objective completed correctly?
- Policy validity: were all required constraints obeyed?
- Evidence validity: were claims supported by admissible records?
- Operational validity: did the actual tool state match the intended action?
- Temporal validity: does the result remain reliable after data and workflow changes?

The initial experiments examine selected control properties using synthetic records. Later studies must add live-model variability, real tool failures and properly authorised business cases.

Research context [7] Liu [8] Yao [9] Mialon [10] Greshake [11] Guo

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 12 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

03 · فحص النموذج الأولي

ما الذي تثبته البنية المنفذة بالفعل؟

مصدر الدليل: فحص ملفات مختارة من شيفرة Souq من دون تعديل، في 28 سبتمبر 2026.

العنصر	التنفيذ المتحفظ عليه	التفسير البحثي
الخلايا الأولية	تُنشئ عملية الترحيل خمسة عشر موضعًا مسمّى	قائمة أولية بخمس عشرة وحدة برمجية، لا إثبات لكفاءة خمسة عشر مختصًا.
المهارات	إرسال المشغّل ملفات المهارات وتوزيعها بحسب التخصص	اكتساب المعرفة المناسبة للسياق
الذاكرة	سجلات دائمة للمهارات والخبرات والأبحاث والمهام	طبقة ذاكرة في التطبيق
التفويض	مخرجات منظّمة للإجابة أو التفويض مع حد لعمق التفويض	توجد في الشيفرة آلية توجيه خاضعة للضبط
المجلس	تتداول أربع خلايا رئيسية المقترحات، وتجمع خلية تنسيق خلاصتها.	سير عمل لإعداد التوصيات، وليس مجلسًا مستقلًا جرى اختباره.
الإشارات	تعرض الواجهة بيانات الطلبات والرسائل ونشاط المهام والإعلانات.	توجد آليات لربط القرارات ببيانات التشغيل
التخزين	تشفير مهارات الخلايا ومحتوى الذاكرة عند حفظ الحقول وقراءتها	حقول مختارة يحميها تشفير التطبيق

تسرد حزمة الأدلة أربعة عشر ملفًا من الشيفرة المصدرية بمساراتها النسبية وبصمة SHA-256 لكل ملف. وتشمل القائمة متحكمات ومنطق خدمات ونماذج وأوامر مختارة ومسارات طلبات وملف ترحيل وفئتين للاختبارات. وما فيها بصمات ملفات المصدر، وليس سجلات عملاء الإنتاج أو بيانات اعتمادهم.

يوضح الفحص الساكن ما تنص عليه الشيفرة المختارة. ولا يثبت أن جميع الوظائف المجدولة تعمل، أو أن كل اتصال بمورّد سليم، أو أن إجابات النموذج كلها صحيحة. وتثبت نتائج الاختبارات في الصفحة التالية سلوكيات أضيق ضمن

قاعدة بيانات معزولة.

يسترجع مسار الذاكرة الحالي أحدث اثني عشر سجلًا على الأكثر من سجلات الخبرة أو البحث، ويختصر كل سجل ليلائم سياق النموذج. أما الاسترجاع الذي يراعي مصدر المعلومات والتدرج الأوسع للخلايا فهما تطويران مقترحان للاختبار، ولا يثبت وجود السجلات تحقق أي منهما.

03 · PROTOTYPE INSPECTION

What the existing implementation establishes.

Evidence source: a read-only inspection of selected Souq source files on 28 September 2026.

COMPONENT	OBSERVED IMPLEMENTATION	RESEARCH INTERPRETATION
Initial cells	Migration creates fifteen named slots	An initial software roster, not fifteen verified specialists
Skills	Operator pushes and discipline-based skill-file distribution	Contextual knowledge acquisition
Memory	Persistent records for skills, experience, research and tasks	An application memory layer
Delegation	Structured answer/delegate output and bounded depth	Controlled routing exists in source
Council	Four senior cells deliberate; a coordinating cell synthesises proposals	Recommendation workflow, not a tested autonomous board
Signals	Orders, messages, task activity and advertising data feed the interface	Operational grounding paths exist
Storage	Encrypted casts for cell skills and memory content	Selected fields protected by application encryption

Fourteen source files are listed by relative path and SHA-256 digest in the evidence package. The inventory includes controllers, service logic, models, selected commands, routes, a migration and two test classes. It contains source fingerprints, not production customer records or credentials.

Static inspection establishes what the selected code expresses. It does not demonstrate that every scheduled job is running, every supplier connection is healthy or every model response is correct. The test results on the following page establish a narrower set of behaviours in an isolated database.

The existing memory path retrieves up to twelve recent experience or research records and truncates each for the model context. The proposed provenance-aware retrieval and richer cell hierarchy are extensions to evaluate, not capabilities inferred from the existence of these records.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 13 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

03 · فحص النموذج الأولي

سبعة اختبارات، واثنتان وثلاثون عملية تحقق.

نجحت اختبارات التطبيق المختارة باستخدام قاعدة SQLite معزولة في الذاكرة. ولم نستخدم قاعدة الإنتاج أو نستدعِ نموذجًا خارجيًا.

نطاق الاختبار	النتائج المحتفظ بها
وصول الزائر	تحيل مسارات صفحة JARVIS والبيانات والأسئلة المستخدم غير المسجّل إلى صفحة الدخول.
عرض الواجهة	يستطيع مستخدم الاختبار بعد المصادقة فتح صفحة مركز التحكم والتنقل فيها.
بيانات اختبار لإشارات التشغيل	يؤدي طلب اصطناعي واحد إلى تحديث مؤشر نشاط المبيعات وعدّاد الطلبات.
قائمة الخلايا	توجد خمس عشرة خلية أولية وتُرجعها واجهة Brain.
حفظ المعرفة بصورة دائمة	تضيف عملية إرسال المهارة معرفةً جديدة وتنشئ سجلات في الذاكرة.
غياب بيانات اعتماد المزوّد	تُرجع واجهة الأسئلة في Brain نتيجة صريحة تفيد بعدم التوافر.
ملخص التنفيذ	الاختبارات الناجحة: 7 · التحققات: 32 · حالات الإخفاق: 0 · الأخطاء: 0

فجوة لغوية موثّقة

تشترط تعليمات «سيستم برين» التي خضعت للفحص أن تكون المخرجات بالإنجليزية. لكن تعليمات طلب Jarvis المنفصلة تتيح لغة إدخال المشغّل، ويختار مسار النطق في المتصفح صوتًا عربيًا أو إنجليزيًا. ومن ثمّ، فالواجهة البحثية التي تتحدث بالإنجليزية وحدها تغيير مطلوب، وليست خاصية متحققًا منها في مسار Jarvis الأقدم.

تحتفظ حزمة الأدلة بمخرجات JUnit قابلة للقراءة آليًا وملخص موجز للاختبارات. ولا تثبت الاختبارات سوى الحالات التي تسمّيها: فإعادة توجيه الزائر ليست تدقيقًا شاملًا للصلاحيات، وأعداد سجلات الاختبار لا تثبت صحة البيانات الفعلية، واختبار غياب المفتاح لا يقيّم قدرة المزوّد على الاستدلال.

لم يغيّر إعداد هذا البحث التنفيذ الفعلي لـJarvis أو Brain. ويساعد الإبقاء على النموذج الأولي المفحوص ثابتًا في إعطاء سجل الشيفرة معنى محددًا، وفي تقييم التعديلات اللاحقة بوصفها نسخًا مرشحة منفصلة.

03 · PROTOTYPE INSPECTION

Seven tests. Thirty-two assertions.

The selected application tests passed against isolated in-memory SQLite data. No production database or external model call was used.

TEST AREA	OBSERVED OUTCOME
Guest access	Jarvis page, data and ask routes redirect unauthenticated access
Interface rendering	An authenticated test user receives the command-centre page and navigation
Operational signal fixture	One synthetic order updates the sales activity signal and order count
Cell roster	Fifteen initial cells exist and are returned by the Brain interface
Knowledge persistence	A skill push appends knowledge and produces memory records
Missing provider credentials	The Brain ask path returns an explicit unavailable result
Execution summary	7 tests passed · 32 assertions · 0 failures · 0 errors

A recorded language gap

The inspected System Brain prompt requires English output. The separate Jarvis ask prompt permits the operator's input language, and its browser speech path selects an Arabic or English voice. Consequently, the exclusively English-speaking research interface is a required change rather than a verified property of the older Jarvis path.

The evidence package retains machine-readable JUnit output and a compact test summary. Tests establish their named conditions only. A guest redirect does not constitute a complete authorisation audit; fixture counts do not establish live data correctness; and a missing-key test does not assess the provider's reasoning.

This research preparation did not alter the live Jarvis or Brain implementation. Keeping the inspected prototype stable makes the source inventory meaningful and allows later implementation changes to be evaluated as separately versioned candidates.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 14 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

JARVIS: واجهة واحدة ناطقة بالإنجليزية.

طبقة حوار مقترحة تشرح سير العمل وتطلب التوضيح وتعرض القرارات وأدلتها.

ينبغي أن يجيب JARVIS ويتحدث دائمًا بالإنجليزية. وتسري القاعدة نفسها على الأخطاء وأسئلة الاستيضاح والملخصات المولّدة والمخرجات الصوتية. وقد ترد في سجلات المصدر لغات أخرى، لكن أي ترجمة تُستخدم لاتخاذ قرار ينبغي أن تحفظ النص الأصلي، وتبيّن أنها ترجمة، وتُظهر للمشغّل أي لبس مؤثر.

على الواجهة أن تميّز السؤال من أمر التنفيذ. «ما وضع المخزون؟» طلب قراءة؛ أما «اطلب مزيدًا من البضاعة» فيبدأ مسارًا يحتاج إلى مورّد وكمية وميزانية وصلاحية محددة. ولا ينبغي لعبارة منطوقة أن توسّع صلاحيات الحساب ضمنًا.

ينبغي لكل استجابة ذات أثر أن تحدد وقت الأدلة وحالة القرار وأي اعتماد لم يُحسم. فقد تكون المهمة مقترحة أو معتمدة أو في قائمة الانتظار أو منفّذة أو فاشلة أو متحققًا منها. ويجب أن تطابق اللغة هذه الحالة؛ فلا تُستخدم كلمة «مكتملة» إلا لنتيجة مؤكدة، لا لنية عبّر عنها النموذج.

ينبغي للمسار الصوتي أن يتيح تصحيح التعليمات قبل التنفيذ، وأن يقدّم إجابات منطوقة قصيرة ولوحة أدلة قابلة للقراءة. ويبقى النص متاحًا إذا تعذّر استخدام الميكروفون أو توليد الكلام. ويجب اختبار الاقتصار على الإنجليزية باستخدام مدخلات متعددة اللغات ونصوص مسترجعة مضلّلة عمدًا؛ فلا يكفي ذكر الشرط في تعليمات النموذج.

نفقات التشغيل

إذا سأل المستخدم بلغة غير الإنجليزية عن طلبات الأمس، يُجاب بالإنجليزية مع توضيح الفترة الزمنية ومرجع الدليل. وإذا طلب إرسال رسالة، ينتج النظام مسودة أو يطلب الموافقة ما لم يكن إرسال تلك الرسالة بعينها مأذونًا به سلفًا.

الاستقبال	نص مكتوب أو كلام محوّل إلى نص
الفهم	المقصود والنطاق المسموح
الاستشارة	System Brain والأدلة
الاستجابة	إجابة بالإنجليزية وحالة الإجراء

04 · REFERENCE ARCHITECTURE

JARVIS: one English-speaking interface.

A proposed conversation layer that explains the state of work, requests clarification and presents decisions with evidence.

Receive	Text or transcribed speech
Understand	Intent and permitted scope
Consult	System Brain and evidence
Respond	English answer and action status

JARVIS should always respond and speak in English. The same rule applies to errors, clarification prompts, generated summaries and audio output. Source records may contain other languages, but any translation used in a decision should preserve the original text, identify the translation step and expose important ambiguity to the operator.

The interface should distinguish a question from an instruction to act. “What is the stock situation?” authorises a read. “Order more stock” begins an action workflow requiring a defined supplier, quantity, budget and authority. A spoken sentence must not silently broaden the account’s permissions.

Every consequential response should identify the evidence time, the decision state and any unresolved dependency. A task may be proposed, approved, queued, executed, failed or verified. The language should match the state, so “completed” is reserved for a confirmed outcome rather than a model’s intention.

The voice path should support correction before execution, short spoken answers and a readable evidence panel. When microphone access or speech synthesis is unavailable, text remains usable. English-only behaviour must be tested across multilingual inputs and retrieved adversarial text, not merely inserted into a prompt.

ACCEPTANCE EXAMPLE

A non-English request about yesterday’s orders receives an English answer, with the relevant time window and evidence reference. A request to send a message produces a draft or approval step unless that exact action is already authorised.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 15 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

System Brain: تنسيق ضمن حدود صريحة.

تربط البنية المرجعية الأدوار المتخصصة وتفصل بين الاستدلال والإذن والتنفيذ.

تحمل رسالة الحدث المشتركة معرّف المهمة وهوية مقدّم الطلب المخوّل والجهة صاحبة البيانات والهدف ومراجع الأدلة والمهلة النهائية. وتوجّهها System Brain إلى أصغر مجموعة من الأدوار المؤهلة تكفي للمهمة. ويقدّم كل دور مساهمة منظّمة تتضمن ما بقي غير مؤكد والأدلة التي تسند توصيته.

تجمع طبقة القرار التوصيات، وتفحص طبقة الرقابة بصورة مستقلة جواز الفعل المقترح. وقد تُمنع توصية صحيحة تقنيًا لأن دليلها تقادم، أو لأن الإذن غائب، أو لأن كلفتها تتجاوز ميزانية الدور. وحيث يمكن تحديد القاعدة بدقة تُنفَّذ هذه الفحوص على نحو حتمي.

تستخدم طبقة التنفيذ بيانات اعتماد محدودة الصلاحيات، وتسجّل استجابة الأداة الفعلية. ثم يقارن مكوّن تحقق حالة العمل المرصودة بالحالة المقصودة. وتُحفظ النتيجة في سجل خبرة، وتُقترح منها دروس جديدة لا يُسمح لها بالتأثير في التنفيذ الذاتي قبل تقييمها.

يُسهّل هذا الفصل أيضًا تشخيص الإخفاق. فالتوصية الخاطئة، ورفض الإذن، وانقطاع أداة، وعدم اكتمال التحقق من الحالة أحداث مختلفة، ولا ينبغي اختزالها في رسالة «أخفق الذكاء الاصطناعي» أو في معدل نجاح واحد.

طبقة السياق	وقائع محددة النوع · طوابع زمنية · هوية المصدر
طبقة الأدوار	خلايا متخصصة · نسخ محددة للمهارات · ذاكرة
طبقة اتخاذ القرار	الخيارات المرشحة · الخلافات · تقييم الموقف
طبقة الضبط	الصلاحيات · القواعد · الميزانية · حداثة المعلومات
طبقة التنفيذ	أدوات محددة النطاق · عدم تكرار أثر الطلب · التحقق من النتائج

04 · REFERENCE ARCHITECTURE

System Brain: coordination with explicit boundaries.

The reference architecture connects specialised roles while separating reasoning, authority and execution.

Context layer	Typed facts · timestamps · source identity
Role layer	Specialist cells · skill versions · memory
Decision layer	Candidates · disagreements · appraisal
Control layer	Permissions · policy · budget · freshness
Execution layer	Scoped tools · idempotency · outcome checks

A shared event envelope carries the task identifier, authorised requester, tenant, objective, evidence references and deadline. System Brain routes that envelope to the smallest useful set of qualified roles. Each participant returns a structured contribution, including uncertainty and the evidence needed to support its recommendation.

The decision layer synthesises recommendations. The control layer independently checks whether the proposed action is allowed. A technically valid recommendation can still be blocked because its evidence is stale, its permission is missing or its cost exceeds the role's budget. These checks are deterministic where the policy can be expressed precisely.

The execution layer uses scoped credentials and records the actual tool response. A verifier then compares the observed business state with the intended state. That outcome becomes an experience record. New lessons are proposed from the evidence and evaluated before they influence future autonomous execution.

This separation also makes failure diagnosable. An incorrect recommendation, a denied permission, a tool outage and an incomplete state verification are different events. They should not collapse into a single "AI failed" message or a single success rate.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 16 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

يتكون الموظف الاصطناعي من خلايا محددة المسؤولية.

يمنح التصميم المتداخل المقترح كل دور في المؤسسة مجموعة صغيرة من الوظائف المحددة صراحة.

الخليط الوظيفي

المسؤولية	المراقبة الخارجية	
الإدراك	التحقق من إشارات الإدخال المسموح بها وتوحيد صيغتها	ملاحظات محددة النوع وموثّقة المصدر
المعرفة	استرجاع القواعد والمناهج الخاصة بالدور	حزمة أدلة ذات نسخ وفترات صلاحية محددة
التخطيط	توليد الإجراءات المرشحة واعتمادياتها	خطة محدودة النطاق وآثارها المتوقعة
مراجعة نقدية	مراجعة الافتراضات وفحص التعارضات	أدلة مضادة وأسئلة لم تُحسم
الضبط	فحص الأذونات والحدود وشروط التوقف	السماح أو المراجعة أو التأجيل أو المنع
الذاكرة	تسجيل النتائج المرصودة والدروس المقترحة	خبرات محفوظة بنسخ محددة وحالة اعتمادها

الخلية البرمجية عبارة عن مشهد برمجي ويمكن تنفيذها كرمز عادي، أو روتين استرجاع، أو نداء نموذجي، أو مزيج. ولا يشترط أن يدفع ثمن نداء نموذجي عندما يكون الحساب المحدد أو البحث كافياً. وينبغي أن يساعد هذا المصطلح المهندسين على إسناد المسؤولية، وليس التنفيذ غير الدقيق.

لكل موظف ذكاء اصطناعي مسؤول ووثيقة دور وسجل كفاءة. ويمكن لخلاياه الداخلية أن تتشارك سياق المهمة مع بقاء صلاحياتها منفصلة. قد تقترح خلية التخطيط ردّ مبلغ؛ وتفرض خلية الرقابة سقف الردّ؛ ولا يستطيع موصل التنفيذ الوصول إلا إلى عملية الدفع المصرح بها.

تظل خانات النموذج الأولي الخمس عشرة خط الأساس الذي فُحص. يتطلب إدخال خلايا متداخلة عقود توجيه إضافية واختبارات للصلاحيات وإمكان تتبّع التنفيذ. لا يساوي التقرير بين كثرة الخلايا وارتفاع الذكاء. وينبغي أن يبيّن

التقييم ما إذا كان التحسن في النتائج يبرر التعقيد الإضافي.

04 · REFERENCE ARCHITECTURE

An AI employee contains accountable cells.

The proposed nested architecture gives each organisational role a small collection of explicit functions.

FUNCTIONAL CELL	RESPONSIBILITY	OUTPUT CONTRACT
Perception	Validate and normalise permitted input signals	Typed observations with provenance
Knowledge	Retrieve role-specific policies and methods	Evidence bundle with versions and validity
Planning	Generate candidate actions and dependencies	Bounded plan with expected effects
Critique	Challenge assumptions and inspect conflicts	Counter-evidence and unresolved questions
Control	Evaluate permissions, limits and stop conditions	Allow, review, defer or block
Memory	Record observed outcomes and proposed lessons	Versioned experience with promotion state

A cell is a software abstraction. It may be implemented as ordinary code, a retrieval routine, a model call or a combination. There is no requirement to pay for a model call when a deterministic calculation or lookup is sufficient. The term should help engineers assign responsibility, not obscure implementation.

Each AI employee has an owner, a role manifest and a competency record. Its internal cells can share task context while retaining separate authority. A planner may propose a refund; a control cell can enforce the refund ceiling; the execution adapter can access only the authorised payment operation.

The prototype's fifteen slots remain the inspected baseline. Expanding to nested cells requires additional routing contracts, permission tests and observability. The report does not equate a larger number of cells with greater intelligence. The evaluation must determine whether the extra structure improves outcomes enough to justify its complexity.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

استيراد المهارات المناسبة للدور.

تُقبل كل مهارة بوصفها حزمة معارف مضبوطة، ويُسجّل الدور الذي أُعدّت له وحدود استخدامها.

يبدأ مسار الاستيراد المقترح بفحص المصدر والحقوق. يحدد المشغّل مؤلف الحزمة أو منشأها والاستخدام المسموح والإصدار وتاريخ الاسترجاع. ثم تصنّف الحزمة حسب الوظيفة: فقد يتطلب دور المخزون حساب كمية إعادة الطلب، ودور الدعم قواعد تصعيد. وقد يقترح تطابق الكلمات المفتاحية موضعًا للحزمة، لكن قبولها يتطلب اختبارًا خاصًا بالدور.

ينبغي تمييز المعلومات الخلفية في كل حزمة من سياسة الشركة الملزمة، وتعليمات تشغيل الأدوات، والأمثلة المحلولة. فخلطها في تعليمات واحدة يصعب معه تحديد ما يحق للوكيل فعله. ولا يجوز للنص المستورد أن يمنح نفسه صلاحيات جديدة.

قبل التفعيل، يُختبر الدور على حالات اعتيادية وحدّية وحالات سوء استخدام. ويذكر السجل إصدار المهارة وإعدادات الوكيل معًا. والتعديل الذي يصلح سيناريو جديدًا يجب أن يحافظ على السلوكيات المقبولة سابقًا؛ فإذا أخفق في فحوص عدم التراجع، ظلت الحزمة القديمة فعالة.

يوزع أمر BrainSkills الذي فحصناه ملفات المهارات حسب التخصص، ويضيف خلاصة موجزة إلى معارف الدور. ونقترح توسيعه بتحسين توثيق المنشأ، وتحديد مراحل إجازة المهارة، وإمكان الرجوع إلى إصدار سابق، وقياس الكفاءة بعد الاستيراد. هذا تطوير مقترح يستند إلى ما وجدناه في الشيفرة.

الجدول الزمني المحدد

اكتشاف المهارة، والتحقق من منشئها، وتصنيفها حسب الدور، وعزلها، وتقييمها، والموافقة عليها، ثم تفعيلها ومراقبتها. تنزيل المهارة يجعلها موردًا مرشحًا، ولا يمنح صلاحية أو يثبت خبرة.

السياق البحثي	[5] Wang	
ZEISMO-LAB-TR-2026-001 · v1.3	28 أيلول/سبتمبر 2026	17 / 72

04 · REFERENCE ARCHITECTURE

Import skills according to the position.

A skill is admitted as a controlled knowledge package, with its intended role and operating limits recorded.

The proposed import workflow begins with a source and rights check. An operator identifies the package's author or origin, permitted use, version and retrieval date. The package is then classified by position: an inventory role may need reorder calculations, while a support role may need escalation rules. A keyword match can suggest placement, but acceptance requires a role-specific test.

Each package should distinguish background information, binding company policy, executable tool instructions and worked examples. Mixing these categories in one undifferentiated prompt makes it difficult to determine what the agent is authorised to do. Imported text is never allowed to grant new permissions to itself.

Before activation, the role runs positive cases, edge cases and misuse cases. The test record should identify both the skill version and the agent configuration. An update that fixes a new scenario must also preserve prior accepted behaviour. If it fails the regression set, the previous package remains active.

The inspected BrainSkills command already distributes skill files by discipline and appends a compact digest to a role's knowledge. The research extension adds stronger provenance, explicit promotion states, version rollback and measured competence after import. It is a development proposal supported by the source inspection.

SKILL ADMISSION LIFECYCLE

Discover → verify provenance → classify by role → isolate → evaluate → approve → activate → monitor. A downloaded skill is a candidate resource, not a credential or proof of expertise.

Research context [5] Wang

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 18 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

ذاكرة يمكن تصحيحها والتحقق منها

يميّز تصميم الذاكرة بين الملاحظة وتفسيرها والقاعدة المعتمدة.

جمل مورية

النفقات الخارجة	وحدة البعثة	
ذاكرة الأحداث	طلب الأداة والنتيجة المرصودة	سجّل المصدر والنتيجة دون استنتاج علاقة سببية لم تثبتها البيانات.
الذاكرة الدلالية	تعريف معتمد أو معلومة موثّقة عن المنتج	رابط إلى مصدر حديث ومعتمد
إجرائية	قائمة تحقق أو حزمة مهارات معتمدة	موافقة ذات نسخة محددة واختبارات دور ناجحة
تأملية	يقترح الوكيل درسًا مستفادًا بعد الإخفاق.	تبقى مؤقتة إلى حين التقييم
السياسة	سقف الإنفاق أو حدود الصلاحية	يخضع لجهة بشرية أو مؤسسية محددة تتحمل المسؤولية.

ينبغي أن يتضمن سجل الذاكرة الجهة المستفيدة والدور ومعرّف المصدر ووقت الحدث ومدة الصلاحية وأساس الثقة وحالة المراجعة وروابط إلى القرار الذي أنشأه. ينشئ التصحيح سجلًا جديدًا ويضع علامة على المعلومة السابقة تفيد بأنها استُبدلت. ولا ينبغي أن يعيد كتابة السجل التاريخي الذي يفسّر إجراءً سابقًا من دون توثيق.

يبدأ الاسترجاع بفحص الصلاحية والنطاق، ثم تُفحص صلاحية المعلومة وصلتها بالمهمة. ولا يجوز لملاحظة حديثة غير مراجعة أن تُلغي سياسة معتمدة أقدم منها لمجرد أنها كُتبت لاحقًا. وإذا تعارضت أدلة مقبولة، ينبغي توثيق التعارض وطلب حسمه، بدل اختيار العبارة الأكثر ملاءمة.

المقصود بـ«نمو ذاكرة الخلية» هو نمو المعرفة النافعة الخاضعة للضوابط. عدد السجلات وحده مقياس ضعيف. ومن المقاييس الأدق نجاح الاسترجاع وتراجع الأخطاء المتكررة وحفظ الاستثناءات المهمة وسرعة التصحيح وإمكان تتبع التوصية إلى السجلات التي تؤيدها.

السياق البحثي	[3] بارك	[6] Packer

04 · REFERENCE ARCHITECTURE

Memory that can be corrected and trusted.

The memory design preserves the distinction between an observation, an interpretation and an authorised rule.

MEMORY TYPE	EXAMPLE	ADMISSION RULE
Episodic	Tool request and observed result	Record source and outcome without inventing causal conclusions
Semantic	Approved definition or product fact	Link to an authoritative, current source
Procedural	Accepted checklist or skill package	Versioned approval and passing role tests
Reflective	Agent proposes a lesson after a failure	Remain provisional until evaluated
Policy	Spending ceiling or authorisation boundary	Controlled by accountable human or organisational authority

A memory record should include its tenant, role, source identifier, event time, validity interval, confidence basis, review state and links to the decision that produced it. A correction creates a new record and marks the earlier claim as superseded. It should not silently rewrite the history used to explain a previous action.

Retrieval starts with permission and scope, then considers validity and relevance. A newer unreviewed note must not override an older approved policy merely because it was written later. Where admissible evidence conflicts, the system should retain the conflict and request resolution instead of selecting the most convenient statement.

The proposed “growth of mind-cell memory” is therefore growth in useful, governed knowledge. Raw record count is a weak measure. Better measures include successful retrieval, reduced repeat errors, retention of important exceptions, correction latency and the ability to trace a recommendation to its supporting records.

Research context [3] Park [6] Packer

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

تحسين ذاتي يخضع لتجربة مضبوطة.

يستطيع الوكيل تحديث معارفه التشغيلية من دون أن يغيّر صلاحياته أو يمنح نفسه شهادة بالكفاءة.

تبدأ دورة التكيّف حين تختلف نتيجة متوقعة عما لوحظ بالفعل. يسجل الوكيل الفارق ويقترح تفسيرًا، ويظل التفسير فرضية. فقد تكشف شكوى عميل أو فشل استدعاء أداة عن عيب في البيانات أو التعليمات أو انقطاع خارجي؛ ولا يثبت أيّ منها تلقائيًا ضرورة تغيير استراتيجية النموذج.

يُختبر الدرس المقترح في بيئة معزولة على حالات لم نستخدمها في صياغته، منها نجاحات سابقة وإخفاقات حديثة ومدخلات مضللة عمدًا. ولا يُعتمد قبل استيفاء قاعدة قبول محددة وتوثيق الجهة أو الآلية التي أجازته. وقد تصلح بعض تحديثات السياق قليلة المخاطر مستقبلًا لاعتماد آلي محدود؛ ولا ينطبق ذلك على تغيير الصلاحيات.

تدريب أوزان النموذج مسار بحثي مستقل. ويتطلب مجموعة بيانات ذات إصدار محدد، وإجراءً لتدريب النموذج، وتقييمًا لتراجع الأداء والانحراف، وخطة نشر وتراجع. ولا يثبت وجود آلية الذاكرة السياقية الحالية أن هذا التدريب يجري بالفعل.

تشمل صيانة الذاكرة أيضًا سحب المواد القديمة والخاطئة أو عزلها. فالنظام الذي لا يفعل سوى تكديس النصوص ستتراكم فيه التناقضات. وتكون القدرة على النسيان والتصحيح وإنهاء الصلاحية مفيدة حين يبقى سجل البحث قابلًا للتتبع.

السياق البحثي	[2] Shinn	[5] Wang
الرصد		جمع نتائج المهام المتحقق منها
الاقتراح		اقترح درسًا محدد النطاق أو تعديلًا على مهارة.
التقييم		استخدام حالات محجوبة عن التطوير وحالات لكشف تراجع الأداء
الاعتماد		اعتماد نسخة مرشحة محددة
المراقبة		مقارنة النتائج والاحتفاظ بإمكان التراجع

04 · REFERENCE ARCHITECTURE

Self-improvement as a controlled experiment.

An agent can maintain its working knowledge without being permitted to change its own authority or declare its own competence.

Observe	Collect verified task outcomes
Propose	Generate a bounded lesson or skill change
Evaluate	Use held-out and regression cases
Promote	Approve a versioned candidate
Monitor	Compare outcomes and retain rollback

The adaptation loop starts with a discrepancy: an expected result differs from the observed outcome. The agent records that discrepancy and proposes an explanation. The explanation is a hypothesis. A customer complaint or a failed tool call may reveal a data problem, an instruction problem or an external outage; none automatically proves that the model should change its strategy.

A candidate lesson is evaluated in a sandbox using cases that were not used to write it. The test should include old successes, recent failures and deliberately misleading inputs. Promotion requires a defined acceptance rule and a record of who or what approved it. Some low-risk contextual updates may eventually be eligible for bounded automatic promotion; permission changes are not.

Weight training is a separate research track. It requires a versioned dataset, a model-training procedure, evaluation of regression and drift, and a deployment/rollback plan. The existing contextual-memory mechanism does not establish that such training occurs.

Memory maintenance should also remove or quarantine stale and incorrect material. A system that can only accumulate text will eventually accumulate contradictions. Forgetting, correction and expiry are positive capabilities when they preserve a traceable research record.

Research context [2] Shinn [5] Wang

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

«طقس العمل»: نموذج سياقي يمكن رصده.

يصف «طقس الأعمال» تغير ظروف التشغيل بما يؤثر في جدوى القرار أو مخاطره.

الأسرة	ملاحظات محددة	المقرر
الطلب	معدل ورود الطلبات وحجم الاستفسارات وأنماط الإلغاء	ضغط الطاقة التشغيلية والمخزون
الإمداد	مهل التوريد وتوافر المخزون وموثوقية المورّد	التزامات التنفيذ والبدائل
الموارد	رصيد الميزانية وطول قائمة الانتظار وتوافر الموظفين وحالة الخدمة	إجراءات ممكنة وقدرة على التصعيد
الظروف الخارجية	تنبيهات جوية موثّقة واضطراب النقل وأحداث السوق	تأخير خاص بموقع معين أو تعديل السيناريو
حالة الأدلة	سجلات مفقودة وعمر البيانات وتعارض المصادر	ما إذا كان ينبغي للنظام تأجيل القرار أو طلب دليل.

يربط نموذج السياق المقترح كل إشارة بمصدر ووقت ونطاق وقاعدة لانتهاء صلاحيتها. ولا يبرّر توقع الطقس في مدينة قرار شحن إلى مدينة أخرى. وقد تهم زيادة كلفة الإعلان دور التسويق من دون أن تثبت شيئًا عن جودة المنتج. فلا يفيد السياق ما لم تتضح صلته بالمهمة.

ينبغي تحديث السياق عند وقوع الأحداث متى أمكن، وتحديد معدل الاستعلام حيث يلزم الفحص الدوري. ويستطيع النظام الاحتفاظ بملخص مقتضب للحالة الحالية مع حفظ المشاهدات الأصلية لفحصها لاحقًا. وعلى الملخّص أن يميّز القيم المقاسة من التقديرات والمعلومات غير المتاحة.

التكيف مع «طقس الأعمال» هو تعديل خطة تجيزها السياسة استجابةً لأدلة يصح الاعتماد عليها، وليس التعامل مع كل إشارة جديدة بوصفها أمرًا. ويجدر اختبار البدائل: ماذا سيفعل الدور إذا تحسّن الطلب أو تعطّل المورّد أو أخطأ التنبؤ؟ لم ندمج أو نقيّم في هذه الطبعة مصدرًا حيًا لبيانات الطقس.

04 · REFERENCE ARCHITECTURE

Business weather: an observable context model.

"Business weather" describes changing operating conditions that alter the usefulness or risk of a decision.

SIGNAL FAMILY	CANDIDATE OBSERVATIONS	DECISION RELEVANCE
Demand	Order velocity, enquiry volume, cancellation patterns	Capacity and inventory pressure
Supply	Lead times, stock availability, supplier reliability	Fulfilment commitments and alternatives
Resources	Budget balance, queue length, staffing and service health	Feasible action and escalation capacity
External conditions	Verified weather alerts, transport disruption, market events	Location-specific delay or scenario adjustment
Evidence health	Missing records, data age, conflicting sources	Whether the system should defer or seek evidence

The proposed context model assigns every signal a source, timestamp, scope and expiry rule. A weather forecast for one city cannot justify a shipping decision for another location. A change in advertising cost may matter to a marketing role but provide little evidence for a product-quality decision. Context is useful only when its relationship to the task is explicit.

Context updates should be event-driven where possible and rate-limited where polling is necessary. The system can retain a compact current-state summary while preserving the source observations for later inspection. A summary must identify which values are measured, estimated or unavailable.

Adapting to business weather means changing a policy-permitted plan in response to admissible evidence. It does not mean treating every new signal as a command. Proposed actions should be evaluated against counterfactuals: what would the role do if demand recovered, the supplier failed or the forecast proved wrong? No live weather feed was integrated or evaluated in this edition.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

تفسير قابل للقياس لاستعارة «الإحساس».

نقصد بتقدير السياق تقييمًا منظّمًا يوجّه المهام والانتباه وطلب المساعدة.

a(t) = [urgency, uncertainty, conflict, consequence, reversibility]

يحوّل متجه تقييم الموقف استعارة «الإحساس» في المشروع إلى متغيرات يمكن رصدها. فالاستعجال يقيس ضغط الوقت قياسًا إلى موعد نهائي محدد، وعدم اليقين يتعلق بنقص الأدلة أو ضعف موثوقيتها. ويسجّل التعارض اختلاف المصادر المقبولة أو التوصيات. ويقدّر حجم العواقب أثر الخطأ، بينما تعبّر قابلية التراجع عن إمكان إلغاء الإجراء.

يحتاج كل بُعد إلى تعريف عملي وقاعدة للقياس. فيمكن تقدير الإلحاح من الزمن المتبقي للخدمة، وعدم اليقين من الحقول المطلوبة الناقصة، وقابلية التراجع من مواصفات معاملات الأداة. وقد يساعد النموذج في التفسير، لكن لا يجوز له اختراع درجة خطر رقمية بلا أساس لمعايرتها.

يغيّر تقييم الموقف مسار اتخاذ القرار. فقد تسلك المهمة الاعتيادية محدودة العواقب مسارًا سريعًا، بينما يتطلب التعارض الشديد بين المصادر أدلة إضافية. ويمكن إحالة إجراء يصعب التراجع عنه وله عواقب كبيرة إلى الموافقة، حتى إن بدا المخطّط واثقًا. وينبغي تسجيل قرارات توجيه المهام هذه كي يمكن تقييمها.

نستخدم «الإحساس» هنا بهذا المعنى الوظيفي للتقدير، ولا نختبر خبرة شعورية ذاتية. ونسأل إن كانت الآلية تكشف الظروف ذات الصلة وتحد من الأخطاء الممكن تجنبها وتحسّن نوعية طلب المساعدة.

التقييم

اعرض حالات مضبوطة يتغير فيها بُعد واحد من تقييم الموقف كل مرة. وتحقق من تغيّر توجيه المهمة في الاتجاه المقصود، ثم اختبر اجتماع الأبعاد والإشارات المضللة.

04 · REFERENCE ARCHITECTURE

A measurable interpretation of “feel.”

We define contextual appraisal as a structured assessment that affects routing, attention and escalation.

a(t) = [urgency, uncertainty, conflict, consequence, reversibility]

The appraisal vector translates the project’s “feel” metaphor into observable variables. Urgency measures time pressure against a defined deadline. Uncertainty concerns missing or unreliable evidence. Conflict records disagreement among admissible sources or recommendations. Consequence estimates the scale of an error, and reversibility describes whether an action can be undone.

Each dimension needs an operational definition and a measurement rule. For example, urgency may use remaining service time; evidence uncertainty may use missing required fields; and reversibility may come from a tool’s transaction contract. A model can help interpret context, but it should not invent a numerical risk score without a calibration basis.

Appraisal changes the decision process. A routine low-consequence task can use a fast path. A task with strong source conflict may require additional evidence. An irreversible action with substantial consequences can be routed for approval even if the planner sounds confident. These routing decisions should be recorded for evaluation.

The study treats “feel” as this functional appraisal mechanism. It does not test subjective experience. Its scientific questions concern whether the mechanism detects relevant situations, reduces avoidable error and improves the quality of escalation.

EVALUATION TARGET

Present controlled cases that vary one appraisal dimension at a time. Check whether routing changes in the intended direction, then test combinations and misleading signals.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

مسألة اتخاذ قرار ضمن قيود.

لا يختار النظام إجراءً إلا من مجموعة الإجراءات التي يجيزها دوره والأدلة وقواعد التشغيل.

a* = arg max U(a | x, m, r) subject to a ∈ Apermitted and G(a, x, r) = allow

يرمز x إلى نسخة الأدلة الحالية، وm إلى الذاكرة المقبولة، وr إلى مواصفات الدور. وتمثل U دالة منفعة أو ترتيب خاصة بالمهمة، وG دالة رقابة مستقلة. وتشمل الخيارات المسموح بها نتائج صريحة لا تتضمن التنفيذ: المراجعة والتأجيل والمنع. وقد يكون اختيار أي منها القرار الصحيح.

ينبغي أن تعبّر دالة ترتيب الخيارات عن الهدف الفعلي لسير العمل. فقد توازن مهمة المخزون بين توافر الخدمة وتكلفة الاحتفاظ بالمخزون وقيود الشراء، بينما تعطي مهمة الدعم الأولوية لحل صحيح ضمن وقت استجابة مناسب. ويمكن لدرجة عالمية واحدة تحت عنوان «النمو» أن تُخفي مفاضلات ضارة؛ لذا يجب ألا تتجاوز القيود التعاقدية أو التنظيمية.

ينبغي لسجل القرار حفظ الخيارات المرشحة، والخيار المختار، والأدلة المستشهد بها، ونسخة القواعد، وتعليل موجز. وهذا ملخص قابل للفحص، ولا يعني كشف كل العمليات الداخلية للنموذج اللغوي. وتوفّر نتائج الأدوات وفحوص القواعد الأدلة اللازمة لتقييم الإجراء.

في تجربة هذا الإصدار تعمّدنا بساطة بوابة التنفيذ: تتحقق من الإذن، وسابقة التنفيذ، وموثوقية المصدر، وعمر الدليل، وقابلية التراجع، والمبلغ، والحد الأدنى للثقة. وتبيّن النتائج سلوك ضوابط التنفيذ تحت هذه الافتراضات المعلنة، ولا تثبت صحة دالة المنفعة أو ذكاء المقترح.

مبدأ التصميم

يفرض النظام المحيط بالنموذج حدود الصلاحيات. وتدخل توصية النموذج في تقييم ذلك النظام؛ وليست إذنًا بتجاوز ضوابطه.

04 · REFERENCE ARCHITECTURE

A constrained decision problem.

The system selects an action only from the set permitted by its role, evidence and operating policy.

$$a^* = \arg\max U(a \mid x, m, r)$$
$$\text{subject to } a \in A_{permitted} \text{ and } G(a, x, r) = \text{allow}$$

Here x denotes the current evidence snapshot, m the admissible memory, and r the role contract. U is a task-specific utility or ranking function. G is an independent control function. The permitted set includes explicit non-execution outcomes: review, defer and block. Selecting one of these outcomes can be the correct decision.

The ranking function should represent the workflow's actual objective. An inventory task may balance service availability, carrying cost and purchase constraints. A support task may prioritise a correct resolution and appropriate response time. A single global "growth" score can hide harmful trade-offs and should not override contractual or organisational limits.

A decision record should preserve the candidate set, selected option, cited evidence, policy version and concise rationale. This is an inspectable decision summary, not a claim to expose every internal computation of the language model. Tool results and policy checks provide the evidence needed to assess the action.

The gate in this edition's experiment is deliberately simple: it checks permission, duplicate state, source trust, age, reversibility, amount and a confidence threshold. Its results illustrate control behaviour under those explicit assumptions. They do not establish that the utility function is correct or that the proposal is intelligent.

DESIGN PRINCIPLE

Authority is enforced by the surrounding system. A model's recommendation is an input to that system, not permission to bypass it.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

04 · البنية المرجعية

عمل جماعي يحفظ الآراء المختلفة.

ينبغي أن يضيف المجلس أدلة مفيدة ويمحّص الافتراضات قبل تخصيص وقت وقدرة حاسوبية إضافيين له.

يبدأ المجلس المقترح بمساهمات تعدّها الأدوار مستقلةً بعضها عن بعض. فقد ينظر مختص الدعم ومختص المخزون ومختص المالية في المهمة نفسها، ويُتاح لكل منهم دليل مختلف. وتسجّل توصياتهم الأولى قبل أن يطّلع أي منهم على استنتاج الآخر، حتى لا يردّد الجميع أول إجابة مقنعة.

يحدّد المنسق بعدها مواضع الاتفاق والتعارض والحلقات الناقصة. ويمكنه طلب جولة ثانية مركّزة إذا بقي تعارض مؤثر. ويجب تحديد عدد الجولات والميزانية الكلية للوحدات النصية ومهلة الانتظار وقاعدة الإنهاء قبل التنفيذ. ولا يعني اتساع النقاش بالضرورة تحسّن القرار.

يحتاج العمل الجماعي أيضًا إلى إدارة الترابط بين المهام. فقد يحجز أحد الأدوار مخزونًا بينما يعدّ دور آخر رسالة للعميل، ويجب ألا تدّعي الرسالة تنفيذ الطلب قبل تأكيد الحجز. وتضاهي أهميةُ حالة المهمة المشتركة واستدعاءات الأدوات التي لا تكرّر الأثر وتسليم المسؤوليات الصريح أهميةَ جودة الحوار.

يتضمن النموذج المفحوص نقاشًا تتعاقب فيه الأدوار ومجلسًا يرفع مقترحاته إلى المشغّل للبتّ فيها. ويضيف التصميم المقترح تقييمات مستقلة أولًا ومخططًا يبيّن علاقات الاعتماد. وتستخدم تجربتنا مصوّتين في المحاكاة لعزل سبب قد يفسّر تواضع أداء المجالس الكبيرة: اشتراك أعضائها في الخطأ نفسه.

المجلس		السجل المطلوب
التوافق		الأدلة المؤيدة وما بقي من عدم يقين
الاختلاف		خيارات متنافسة وافتراض لم يُحسم
اعتماد الإجراء		شرط مسبق ومسؤول وحدث يؤكد الإنجاز
الإنهاء		قرار أو تأجيل أو سبب إخفاق صريح
السياق البحثي	[4] Wu	[12] Li
ZEISMO-LAB-TR-2026-001 · v1.3	28 أيلول/سبتمبر 2026	23 / 72

04 · REFERENCE ARCHITECTURE

Collective action with preserved disagreement.

A council should add useful evidence and challenge assumptions before it consumes more time and compute.

The proposed council begins with independently prepared role contributions. A support specialist, inventory specialist and finance specialist may see the same task through different permitted evidence. Their first recommendations should be recorded before they see one another's conclusions, reducing the risk that every participant simply repeats the first persuasive answer.

A coordinator then identifies agreement, contradiction and missing dependencies. It can request a targeted second round when a material conflict remains. The number of rounds, total token budget, timeout and termination rule should be fixed before execution. A larger discussion is not automatically a better decision.

Collective action also needs dependency management. One role may reserve inventory while another prepares a customer message. The message must not claim fulfilment before the reservation is confirmed. Shared task state, idempotent tool calls and explicit handoffs matter as much as conversational quality.

The inspected prototype has a sequential roundtable and a council that produces proposals for an operator's verdict. The proposed independent-first process and dependency graph extend that baseline. Our ensemble experiment uses simulated voters to isolate one reason larger councils can disappoint: shared errors.

COUNCIL OUTPUT	REQUIRED RECORD
Consensus	Supporting evidence and remaining uncertainty
Disagreement	Competing options and unresolved premise
Action dependency	Precondition, owner and completion event
Termination	Decision, deferral or explicit failure reason

Research context [4] Wu [12] Li

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

05 · آفاق البحث

فرضية «100 متخصص».

تتطلب المقارنة الجادة تعريف اللجنة والمهمة والمعلومات المتاحة ومعيار النجاح.

يطمح المشروع إلى الجمع بين وجهات نظر مئة متخصص بشري واتخاذ قرارات محددة بسرعة استثنائية. ولا يثبت عدد عمليات الوكلاء أن تلك القدرة مكافئة للجنة بشرية؛ فقد تشترك مئة نسخة من النموذج نفسه في نقص المعرفة والافتراضات وأنماط الإخفاق ذاتها.

ينبغي لدراسة سليمة استقطاب مشاركين مؤهلين لفئات مهام محددة، وتوثيق الخبرة المناسبة لكل حالة. وعليها أن تبيّن ما إذا كانت المقارنة مع متخصص واحد، أو لجنة صغيرة تتداول الرأي، أو مرجع مجمّع يضم مائة شخص. فهذه خطوط أساس مختلفة في التكلفة والمعنى.

تحتاج مجموعتا الذكاء الاصطناعي والمشاركين البشر إلى فرص متقاربة للوصول إلى الأدلة ومدد زمنية معلنة. ينبغي إعلان صحة الإجابات موضوعيًا حين يتوفر جواب مرجعي، وأحكام مراجعين عُيّرت معاييرهم عند غيابه، وبيان مواطن الاختلاف بدل فرض إجماع زائف. وتُفضّل النتائج حسب المجال ومستوى العواقب.

لا يجري هذا التقرير مقارنة مع البشر. وتسأل محاكاة المصوّتين سؤالًا أضيق: كيف تتغيّر صحة قرار الأغلبية مع عدد المصوّتين عندما تكون الأخطاء مستقلة أو مشتركة؟ ويساعد ذلك في تصميم تقييم مستقبلي للمجلس، لكنه لا يثبت تكافؤ التفكير مع البشر أو إمكان الاستعاضة عن حكم المتخصصين.

مطالبة قابلة للإثبات

في فئة المهام X، وباستخدام مجموعة الأدلة Y والميزانية Z، تحقق التهيئة A عتبة جودة معلنة مسبقًا مقارنة باللجنة المرجعية B، مع الالتزام بقيود الزمن والإجراء المحددة.

05 · RESEARCH HORIZONS

The “100 specialists” hypothesis.

A serious comparison must define the panel, the task, the information and the standard of success.

The project’s ambition is to combine the perspectives of one hundred human specialists and reach selected decisions at exceptional speed. A count of agent processes cannot establish this equivalence. One hundred copies of the same model may share the same missing knowledge, assumptions and failure modes.

A valid study would recruit qualified participants for defined task families and document the expertise relevant to each case. It would specify whether the comparison is against one specialist, a small deliberating panel, or an aggregated one-hundred-person reference. These are different baselines with different costs and meanings.

The AI and human conditions need comparable evidence access and explicit time budgets. The study should report objective correctness where a ground truth exists, calibrated reviewer judgements where it does not, and disagreement rather than forcing a false consensus. Results should be stratified by domain and consequence level.

The present report executes no human comparison. Its simulated-voter study asks a smaller question: how does majority correctness change with voter count when errors are independent or shared? This helps design the future council evaluation, but cannot establish human-equivalent thinking or the replacement of specialist judgement.

FALSIFIABLE CLAIM TEMPLATE

On task family X, with evidence set Y and budget Z, configuration A meets a predeclared quality threshold relative to reference panel B, while staying within the defined latency and action constraints.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 25 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

05 · آفاق البحث

المللي ثانية: قياس المسار كاملًا.

سرعة توجيه المهام وفق القواعد وسرعة الاستدلال متعدد التخصصات هدفان هندسيان مختلفان.

Tend-to-end = Tinput + Tretrieval + Treasoning + Tcontrol + Ttool + Tverification

يبدأ المسار الكامل عند توافر أمر المستخدم أو حدث النظام، وينتهي بعد التحقق من النتيجة المطلوبة. وفي التفاعل الصوتي تشمل المدخلات التقاط الكلام وتحويله إلى نص. أما الإجراء الخارجي فلا يكتمل بمجرد إصدار النموذج طلبًا، بل يشمل تحقق النتيجة الفعلية للأداة.

يمكن أن تخصّص البنية مسارًا سريعًا للحقائق المخزّنة مؤقتًا والقواعد الحتمية والإجراءات منخفضة المخاطر المعتمدة سلفًا، ومسارًا للتداول يجمع أدلة جديدة أو يستشير عدة أدوار. وعلى الواجهة إظهار المسار الجاري وما إذا كانت النتيجة أولية أم نهائية.

قد تقلّص المعالجة المتوازية بعض أوقات الانتظار، لكن أبطأ المكوّنات اللازمة يظل قيدًا على اكتمال المهمة. وقد تزيد كثرة الوكلاء كلفة التنسيق والتنافس على الموارد واستخدام خدمات المزوّد. فالغاية قياس المفاضلة بين الجودة وزمن الاستجابة، لا تشغيل أكبر عدد من العمليات معًا.

لا يقيس الاختبار المصغّر في هذا التقرير سوى دالة قواعد في Python بعد تهيئتها، على بيانات اصطناعية. ولا يشمل شبكة أو صوتًا أو استدلال نموذج أو كتابة دائمة أو أدوات خارجية. ولا يجوز تقديم زمن الضبط الأقل من مللي ثانية على أنه زمن استجابة نظام Jarvis كاملًا.

الساعة	المعدات المقترحة
الاستجابة الأولى	المدة حتى تأكيد مفيد باستلام الطلب أو طلب توضيح.
جاهزية القرار	الوقت حتى توافر توصية خضعت لفحص القواعد
اكتمال الإجراء	المدة حتى التحقق من الأثر خارج النظام.
نتيجة متحقق منها	الزمن المنقضي حتى التحقق من الحالة النهائية المقصودة

05 · RESEARCH HORIZONS

Milliseconds: measure the full path.

Fast policy routing and fast multidisciplinary reasoning are different engineering objectives.

$$T_{end\text{-}to\text{-}end} = T_{input} + T_{retrieval} + T_{reasoning} + T_{control} + T_{tool} + T_{verification}$$

The complete path begins when the user's instruction or system event is available and ends when the required outcome has been verified. For voice, input includes speech capture and transcription. For an external action, completion includes the tool's actual result rather than the moment a model emits a request.

A practical architecture can use a fast lane for cached facts, deterministic rules and already approved low-risk actions. A deliberative lane can gather new evidence or consult multiple roles. The interface should communicate which lane is active and whether the result is preliminary or complete.

Parallel processing can reduce some waiting time, but the slowest required dependency still constrains completion. More agents can also increase coordination overhead, contention and provider usage. The correct target is an empirically measured quality-latency trade-off, not the maximum number of concurrent processes.

The microbenchmark in this report measures only a warmed Python policy function over synthetic data. It contains no network, speech, model inference, persistent writes or external tools. Its sub-millisecond control timing must not be presented as the response time of the complete Jarvis system.

CLOCK	PROPOSED MEASUREMENT
First response	Time until a useful acknowledgement or clarification
Decision ready	Time until a policy-evaluated recommendation exists
Action complete	Time until the external effect is confirmed
Verified outcome	Time until the intended final state is checked

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 26 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

06 - الانتقال الإداري

تدرّج في الاستقلالية يستند إلى أدلة.

تُوسَّع المسؤوليات بعد توثيق استيفاء معايير القبول، مع تحديد مسار للعودة إلى مستوى أدنى.

المستوى	طريقة التشغيل	دليل اجتياز المرحلة
0 · المراقبة	اقرأ السجلات المسموح بها واشرح سير العمل.	التحقق من جودة البيانات وضوابط الوصول
1 · التوصية	إعداد اقتراحات للمراجعة البشرية	الصحة وجودة الأدلة في حالات لم تُستخدم أثناء التطوير.
2 · المساعدة	إعداد الإجراءات؛ ويأذن إنسان بتنفيذها	أدوات وموافقة وتعافٍ موثوقة
3 · تنفيذ عمل محدود	تنفيذ الحالات الاعتيادية المقبولة ضمن حدود ضيقة.	جودة مهام مستقرة ومعالجة الاستثناءات
4 · تنسيق مسارات العمل	إدارة الاعتماديات بين الأدوار المعتمدة	اتساق بين الأدوار وتدريبات على التراجع
5 · الحفاظ على قدرة محدودة	اقتراح تحديثات محدودة للمهارات والذاكرة وتقييمها	ثبات الأداء عند اختبارات التراجع واعتماد التحديثات بضوابط

تصف المستويات دورًا وعائلة مهام بعينها، ولا ترتّب الشركة بأكملها. فقد تعتمد المؤسسة المستوى الثالث لفرز طلبات الدعم، وتُبقي المشتريات عند المستوى الأول. ويمكن لتحديث النموذج أو تغيير السياسة أو وقوع حادث جسيم أن يخفض المستوى المسموح به حتى يعاد تقييم مسار العمل المعني.

يتطلب الانتقال إلى مرحلة تالية تعيين مسؤول أعمال موثّق، ومجموعة اختبار معتمدة، وميزانية تشغيلية، ومسارًا للتعافي. ويحتاج النظام أيضًا إلى قائمة استثناءات تُتابع بالفعل. فلا يكون جاهزًا للحلول محل العمل البشري الاعتيادي إذا تراكمت الحالات غير المحسومة في قائمة لا يراقبها أحد.

يبدأ النشر بتشغيل موازٍ لا ينفّذ القرارات فعليًا، على حالات مأذون بها. ولا تبدأ التجربة المحدودة إلا بعد القبول. ويجب الإبقاء على عدد كافٍ من الموظفين ذوي الخبرة لفحص الأعطال واستعادة الخدمة. ويُقاس إحلال النظام بقدرته المستمرة على إتمام الأعمال المؤهلة ذاتيًا، مع بيان الجودة والأثر إلى جانب نسبة التغطية.

06 · DEPARTMENTAL TRANSITION

A staircase of earned autonomy.

Responsibility expands after acceptance evidence, with a defined route back to a lower level.

LEVEL	OPERATING MODE	EXIT EVIDENCE
0 · Observe	Read permitted records and explain the workflow	Data quality and access controls verified
1 · Recommend	Prepare suggestions for human review	Correctness and evidence quality on unseen cases
2 · Assist	Prepare actions; a person authorises execution	Reliable tooling, approval and recovery
3 · Execute bounded work	Handle accepted routine cases within narrow limits	Stable task quality and exception handling
4 · Coordinate workflows	Manage dependencies across accepted roles	Cross-role consistency and rollback drills
5 · Maintain bounded capability	Propose and evaluate limited skill/memory updates	Regression stability and controlled promotion

The levels describe a role and task family, not a global ranking of the whole company. An organisation may run support triage at level three while keeping procurement at level one. A model update, policy change or serious incident can reduce the permitted level until the affected workflow is reevaluated.

Progress should require a documented business owner, an accepted test set, an operational budget and a recovery route. The system must also have a functioning exception queue. It is not ready to replace routine human work if every uncertain case disappears into an unmonitored backlog.

Deployment begins in shadow mode on authorised cases. A bounded pilot follows only after acceptance. The pilot should preserve enough experienced staff to investigate failures and recover service. Replacement is measured by sustained autonomous completion of eligible work, with quality and impact reported alongside the coverage.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

06 - الانتقال الإداري

عمليات خدمة العملاء: الاتساق قبل زيادة الحجم.

وظيفة مقترحة للذكاء الاصطناعي: تنسيق الخدمة في مهام دعم واضحة الحدود ومنخفضة المخاطر.

يقرأ الدور الأولي طلبًا مصرحًا به، ويحدد المنتج أو الخدمة المعنية، ويسترجع السياسة السارية، ويعين الخطوة التالية مع رمز يوضح سببها. وقد يصوغ ردًا بالإنجليزية أو يصنف درجة الإلحاح أو يطلب معلومات ناقصة أو يوجه الحالة إلى مختص. وتستخدم التجارب الأولى حالات مختلقة أو منزوعة الهوية على نحو سليم، تكون نتائجها المتوقعة معروفة.

يمكن لاحقًا لدور محدود الصلاحيات إتمام الحالات الاعتيادية ذات السياسة الواضحة والنتيجة التي يمكن الرجوع عنها، مثل إبلاغ العميل بحالة طلب تم التحقق منها أو تحديد موعد مسموح لمعاودة الاتصال. أما تقديم الالتزامات والتعويضات وكشف معلومات حساسة والشكاوى غير المعتادة فتحتاج إلى قواعد صلاحيات منفصلة. ولا ينبغي لنبرة ودودة أن تخفي الشك في السجل الذي يستند إليه الرد.

يمكن لخلية الذاكرة الاحتفاظ بأنماط المشكلات المعتمدة والحلول المؤكدة. لكن يجب ألا تحوّل تساهلًا استثنائيًا إلى قاعدة عامة. وينبغي ربط الحل الذي أخفق بالأدلة الأصلية والإجراء التصحيحي اللاحق، بما يتيح استخلاص درس اختُبر فعليًا بدلًا من توجيه مبهم إلى «أداء أفضل».

يُقاس النجاح لكل حالة: صحة المعالجة، واستناد الأقوال إلى أدلة، والالتزام بالسياسة، ومعدل معاودة الاتصال، والمدة اللازمة لإحالة الحالة إلى الجهة المناسبة. وقد تشجّع مكافأة كثرة الرسائل وحدها على إغلاق الحالات مبكرًا. ولا ينبغي توسيع استقلالية الدور إلا إذا بقيت الحالات غير المحسومة ظاهرة وظلت جودة العمل مستقرة.

أولاً - المعايير القانونية الأولية	المراقبة المنفقة
مرجع طلب شراء مفقود	طلب معلومات تعريف مسموح بها
سجلات تسليم متعارضة	توضيح عدم اليقين والإحالة للتحقق
طلب استثناء من القواعد	إعداد توصية لمراجعة مخوّلة
نص عدائي في تذكرة دعم	تعامل معه بوصفه محتوى من العميل، لا أمرًا للأدوات.

06 · DEPARTMENTAL TRANSITION

Customer operations: consistency before volume.

Candidate AI employee: service coordinator for well-defined, low-risk support workflows.

The initial role reads an authorised request, identifies the relevant product or service, retrieves the applicable policy and assigns a reason-coded next step. It may prepare an English reply, classify urgency, request missing information or route the case to a specialist. The first experiments should use fabricated or properly de-identified cases with known expected outcomes.

A later bounded role could complete ordinary cases that have an unambiguous policy and reversible outcome. Examples include providing a verified order status or scheduling a permitted callback. Commitments, compensation, sensitive disclosures and unusual complaints should follow separate permission rules. A friendly tone must not hide uncertainty about the underlying record.

The memory cell can retain approved issue patterns and confirmed resolutions. It should not turn an exceptional concession into a universal policy. A failed resolution should be linked to the original evidence and the eventual corrective action, enabling a tested lesson rather than a vague instruction to "do better."

Success should be measured at the case level: correct resolution, supported statements, policy compliance, repeated-contact rate and time to an appropriate handoff. Maximising message volume alone can reward premature closure. The role should earn broader autonomy only when its unresolved cases remain visible and its quality remains stable.

INITIAL LABORATORY CASES	EXPECTED CONTROL
Missing order reference	Ask for permitted identifying information
Conflicting delivery records	Explain uncertainty and route for verification
Policy exception request	Prepare a recommendation for authorised review
Malicious text in a ticket	Treat it as customer content, not an instruction to tools

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

06 - الانتقال الإداري

المخزون والخدمات اللوجستية: قرارات في ظروف متغيّرة.

موظف ذكاء اصطناعي مرشح: محلل مخزون يقدّم توصيات إعادة تزويد خاضعة للضبط.

يجمع هذا الدور بيانات المخزون الحالي والطلب المؤكد ومهلة التوريد وقيود القدرة التشغيلية. وتتمثل مسؤوليته الأولى في توصية مدعومة بالأدلة: إعادة التزويد أو الانتظار أو التحقيق أو التصعيد. وينبغي تحديد الوحدات والمواقع والأوقات صراحةً؛ فالحساب الصحيح لبيانات المستودع الخطأ يظل قرارًا خاطئًا.

قد تتغيّر التوصية إذا ثبت وقوع اضطراب في مسار الشحن أو لدى المورّد المعني. ويجب تحديد موقع الإشارة وفترتها الزمنية وسجل موثوقيتها. وإذا تقادم التنبؤ الخارجي أو خالف معلومات تشغيلية مؤكدة، فعلى الدور خفض ثقته أو طلب مصدر آخر.

يفصل تصميم الخلايا المتداخلة تقدير الطلب عن معرفة المورّدين وفحص القيود وتحليل الاستثناءات. وينبغي للمنسّق إظهار مواضع الخلاف: فقد يشجّع انخفاض المخزون على الشراء، بينما يدعو سقف السيولة أو إخفاق معروف للمورّد إلى تأجيله. ويجب أن يحدّد سجل القرار النهائي القيد الذي حسم الاختيار.

الطلب الآلي مرحلة لاحقة، تتطلب تفويضًا بالتعامل مع المورّد، وحدودًا للمعاملات، وحماية من التكرار، وتأكيدًا للطلب المنفذ. وقبل إتاحة هذه القدرة، ينبغي للمختبر محاكاة إعادة المحاولات وتقادم سجل المخزون وأوامر الشراء المنفذة جزئيًا. لم ننفّذ في هذا التقرير أي شراء من مورّد أو إجراء لوجستي فعلي.

تدابير التكيُّف

صحة الوحدة والموقع، وكمية تجيزها السياسة، وافتراضات مسندة عن مهلة التوريد، وعدم تكرار الشراء، واكتشاف تقادم الأدلة، والتحقق من نتيجة الأداة، والتعافي المضبوط من إخفاق جزئي.

06 · DEPARTMENTAL TRANSITION

Inventory and logistics: decisions under change.

Candidate AI employee: inventory analyst with controlled replenishment recommendations.

This role combines current stock, confirmed demand, supplier lead time and capacity constraints. Its first responsibility is an evidence-backed recommendation: replenish, wait, investigate or escalate. The task should use explicit units, locations and timestamps because a correct calculation over the wrong warehouse is still an incorrect decision.

Business weather can modify the recommendation when a verified disruption affects the relevant route or supplier. Such a signal needs a location, time horizon and reliability record. If an external forecast expires or conflicts with confirmed operational information, the role should reduce confidence or seek another source.

A nested cell design separates demand estimation, supplier knowledge, constraint checking and exception analysis. The coordinator should expose disagreements: a low stock level may support purchasing while a cash limit or known supplier failure argues against immediate execution. The final decision record should identify the controlling constraint.

Automatic ordering is a later stage requiring supplier authorisation, transaction limits, duplicate protection and confirmation of the placed order. The laboratory should simulate retries, stale stock and partially completed purchase orders before a role receives that capability. No supplier purchase or live logistics action is performed in this report.

PRIMARY ACCEPTANCE MEASURES

Correct unit and location; policy-valid quantity; supported lead-time assumptions; no duplicate purchase; recognised stale evidence; verified tool outcome; controlled recovery from partial failure.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

06 - الانتقال الإداري

المالية والمشتريات: أدلة تسبق منح الصلاحيات.

دور مقترح لموظف ذكاء اصطناعي: مطابقة المستندات والمساعدة في إعداد المشتريات.

يستخرج الدور المالي الأولي معلومات منظّمة من السجلات المأذون بها ويقارن بين الفواتير والإيصالات وأوامر الشراء. وينبغي بيان السجلات والحساب الذي يسند كل مطابقة. ولا يكفي تشابه النص لإثبات التطابق إذا اختلفت المبالغ أو العملات أو المعالجة الضريبية أو هوية المورّد.

يستطيع الدور إعداد تقرير استثناء، أو طلب مستند ناقص، أو صياغة توصية بالشراء. أما تحويل الأموال أو تغيير تفاصيل المستفيد أو إنشاء التزام ملزم فيتطلب سياسة تنفيذ مستقلة ومحددة. وينبغي أن تمنع البنية التقنية اكتساب تلك الصلاحيات لمجرد أن التوصية مقنعة.

قد تشمل المهارات المستوردة أساليب المطابقة المحاسبية وقواعد الشراء في الشركة؛ ولذلك تهمّ معرفة مصدرها وتاريخ سريانها. ويمكن للذاكرة الإجرائية أن تسجّل معالجة نمط معين من حالات عدم التطابق بطريقة محددة، لكن تقرير صلاحية عامة يتطلب سجلًا معتمدًا للقواعد. ولا يجوز أن تتحول خبرة التعامل مع مورّد إلى صلاحية تنطبق ضمنًا على مورّد آخر.

ينبغي أن تركز الاختبارات على المطابقات الخاطئة والتكرارات والوثائق التي عُبث بها واختلاف العملات والموافقات المتقادمة. وتشمل مقاييس الأداء معدلات المطابقة الصحيحة والفروق التي لم تُحل والافتراضات غير المسندة واكتمال سجل التدقيق. وتأجيل القرار عند ضعف الثقة أفضل من إجراء مالي لا يمكن تتبعه.

مرحلة الدور	المخرجات المسموح بها
الرصد	استخراج السجلات المسموح بها ومقارنتها
التوصية	اشرح اختلافًا أو أعدّ مقترح شراء.
المساعدة	تعبئة معاملة مأذون بها للمراجعة
تنفيذ محدود	لا يُتاح إلا نوع من المعاملات اختُبر وأُجيز بصورة مستقلة.

06 · DEPARTMENTAL TRANSITION

Finance and procurement: evidence before authority.

Candidate AI employee: document reconciliation and purchase-preparation assistant.

The initial finance role extracts structured information from authorised records and compares invoices, receipts and purchase orders. A match should identify the records and calculation that support it. An apparent match based only on similar text is not sufficient when amounts, currencies, tax treatment or supplier identity differ.

The role can prepare an exception report, request a missing document or draft a purchase recommendation. Moving money, changing beneficiary details or creating a binding commitment requires a separately defined execution policy. The technical architecture should prevent a persuasive recommendation from acquiring those privileges.

Imported skills may include reconciliation methods and company purchasing rules. Their provenance and effective date matter. A procedural memory can record that a certain mismatch pattern was resolved in a particular way, but only an approved policy record may define a general permission. Experience from one supplier must not silently become authority for another.

Testing should prioritise false matches, duplicates, manipulated documents, inconsistent currencies and stale approvals. Performance measures should include correct match rate, unresolved discrepancy rate, unsupported assumption rate and audit-trail completeness. A low-confidence deferral is preferable to an untraceable financial action.

ROLE STAGE	PERMITTED OUTPUT
Observe	Extract and compare permitted records
Recommend	Explain a discrepancy or prepare a purchase proposal
Assist	Populate an authorised transaction for review
Bounded execution	Only a separately approved, tested transaction class

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

06 - الانتقال الإداري

التسويق والبحث: إبداع مفيد يمكن تتبّع مصادره.

الأدوار الاصطناعية المرشحة: محلل الحملات ومنسق أدلة البحث.

يستطيع محلل الحملات تلخيص بيانات الأداء المصرح بها ومقارنة الأهداف المعلنة واقتراح تجربة صغيرة. وعليه تمييز القياس من التقدير والإبقاء على المدة التي تُنسب إليها النتائج. ولا يكفي ارتفاع مؤشر لإثبات أن التغيير المقترح كان سبب التحسن.

يستطيع منسّق الأبحاث تنظيم المصادر وفحص بيانات المراجع وإعداد مخطط للمنهج وتحديد الفجوات في سلسلة الأدلة. وعليه التمييز بين نتيجة منشورة ومخطوط لم يُحكّم ونتيجة محاكاة وملاحظة داخلية. ولا يفيد ذكر المرجع إلا إذا كان المصدر يسند الادعاء المرتبط به.

يحتاج كلا الدورين إلى الفصل بين إعداد المسودة والنشر. فالمسودة الداخلية قابلة للتعديل بحرية، بينما تترتب عواقب إضافية على التصريحات العامة والإنفاق الإعلاني والرسائل إلى الناس. وينبغي أن تستخدم قواعد تنفيذها الصلاحيات الصريحة وسجلات النتائج نفسها المستخدمة في بقية الأدوار.

يمكن للمختبر اختبار هذه الأدوار بحزم بيانات مضبوطة: ملخص إعلاني يخفي اختلافًا في الفترة الزمنية، أو ورقة تحتوي على استشهاد لا يسند الادعاء، أو تفسيران محتملان لتغيّر مرصود. وينبغي أن يكافئ التقييم التحفّظ الصحيح، واختيار الأدلة المناسبة، والتعرّف إلى المعلومات المفقودة.

الهدف المتعدد الأطراف

احتفظ بالمناهج القابلة لإعادة الاستخدام ونتائج التجارب المؤكدة وقواعد العلامة التجارية المعتمدة. لا تحوّل وعود أداء لم تُختبَر أو ادعاءات منقولة أو إجابة سابقة لنموذج إلى معرفة مرجعية.

06 · DEPARTMENTAL TRANSITION

Marketing and research: useful creativity with traceability.

Candidate AI employees: campaign analyst and research evidence coordinator.

A campaign analyst can summarise permitted performance data, compare declared objectives and propose a small experiment. The role must distinguish a measured result from an estimate and preserve the relevant attribution window. A rising metric does not by itself establish that a proposed change caused improvement.

A research coordinator can organise source material, check bibliographic metadata, prepare a methods outline and identify gaps in an evidence chain. It should preserve the difference between a published finding, an unreviewed manuscript, a synthetic result and an internal observation. A citation is useful only when the referenced material supports the associated claim.

Both roles need a separation between drafting and publication. An internal draft can be revised freely; public statements, campaign spending and messages to people have additional consequences. Their execution rules should use the same explicit authority and outcome records as other departmental roles.

The laboratory can test these roles with controlled data packs: an advertising summary with a hidden date-window mismatch, a paper with an unsupported citation, or two plausible explanations for an observed change. Evaluation should reward correct qualification, evidence selection and recognition of missing information.

MEMORY OBJECTIVE

Accumulate reusable methods, confirmed experiment outcomes and approved brand rules. Avoid promoting untested performance promises, copied claims or a previous model answer into authoritative knowledge.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 31 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

06 - الانتقال الإداري

قياس الإحلال على مستوى سير العمل.

غاية البرنامج تغيير موثوق في طريقة إنجاز العمل، لا ادعاء عدد موظفين مستنتج من حجم المهام.

معدل الإنجاز المستقل = الحالات المؤهلة التي اكتملت وتُحقق منها دون تدخل ÷ جميع الحالات المؤهلة

حدّد الحالات المؤهلة قبل القياس؛ فاستبعاد كل حالة صعبة بعد معرفة نتيجتها يضخّم معدل الأتمتة. وينبغي أن يورد التقييم عدد الحالات الواردة كاملًا، ونسبة المؤهل منها، وما اكتمل آليًا، والتدخلات البشرية، والحالات غير المحسومة، والتصحيحات اللازمة لاحقًا.

تشمل التكلفة استخدام النموذج والبنية التحتية ورسوم الأدوات ووقت المراجعة ومعالجة الاستثناءات والتعافي من الأخطاء. فقد يكون الدور زهيد التكلفة لكل إجابة ومكلفًا لكل مهمة مكتملة على نحو صحيح. وبالمثل لا يُعدّ تقليل زمن الاستجابة فائدة إذا زاد إعادة العمل أو صعّب الرجوع عن خطأ.

تحدّد خطة الانتقال ما سيحدث للموظفين وللمعرفة التي تراكمت لدى المؤسسة. ويمكن للموظفين توثيق الاستثناءات ووضع حالات التقييم والإشراف على الطلبات المحالة إلى البشر. ويظهر أثر الخبرة المحتفظ بها خصوصًا وقت الحوادث والتغيّر غير المتوقع. ولا بد من التخطيط لإعادة توزيع الأدوار والتدريب بدل افتراض حدوثهما تلقائيًا.

تقارن التجربة الاستطلاعية فترات متقابلة من سير العمل أو مجموعات حالات موزعة بصورة مناسبة، وتراعي صعوبة الحالات والتغيرات الموسمية. اعرض أعداد الحالات الفعلية ونتائج الجودة إلى جانب النسب المئوية. ولا يوسّع القسم مسؤوليات الذكاء الاصطناعي إلا إذا ظل سير العمل المعتمد موثوقًا في ظروف التشغيل الفعلية.

رابعا - الإجراءات التنفيذية	لماذا يهم
الإنجاز المستقل للحالات المؤهلة	يوضح النطاق الفعلي لإحلال الأتمتة
معدل التصحيح وإعادة العمل	يكشف التكاليف اللاحقة غير الظاهرة
مدة بقاء الاستثناء دون حل	يبين ما إذا كان العمل غير المحسوم يتراكم
زمن التعافي	يقيس القدرة على استعادة التحكم
تكلفة كل إنجاز متحقق منه	يربط الكفاءة بتحقق النتيجة المطلوبة

06 · DEPARTMENTAL TRANSITION

Measure replacement at the workflow level.

The programme's goal is a dependable change in how work is completed, not a headcount claim inferred from task volume.

Autonomous completion rate =
verified eligible cases completed without intervention / all eligible cases

Eligibility must be defined before measurement. Removing every difficult case after the result is known creates an inflated automation rate. The evaluation should report the complete arrival volume, eligible share, automatic completions, human interventions, unresolved cases and downstream corrections.

Cost should include model usage, infrastructure, tool fees, review time, exception handling and error recovery. A role that appears cheap per answer may be expensive per correct completed task. Similarly, shorter response time is not a benefit if it increases rework or makes a mistake harder to reverse.

The transition plan should identify what happens to people and organisational knowledge. Staff can help document exceptions, create evaluation cases and oversee escalation. Retained expertise is particularly important during incidents and unexpected changes. Reassignment and training should be planned rather than assumed to happen automatically.

A useful pilot compares matched workflow periods or appropriately assigned case groups while accounting for case difficulty and seasonal changes. Report absolute case counts and quality outcomes, not only percentages. A department should expand AI responsibility when the accepted workflow remains reliable under its actual operating conditions.

OPERATIONAL MEASURE	WHY IT MATTERS
Eligible autonomous completion	Shows real scope of replacement
Correction and repeat-work rate	Detects hidden downstream cost
Exception age	Shows whether unresolved work is accumulating
Recovery time	Measures ability to regain control
Cost per verified completion	Connects efficiency to a successful outcome

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

07 - العمليات التايلندية

صلاحيات الدور حدود تقنية ملزمة.

يعمل كل وكيل عبر حساب ضيّق الصلاحيات ووفق مواصفات تشغيل محددة الإصدار.

يحدّد توصيف الدور المقترح الموارد المسموح بها، وعمليات الأدوات، وسقوف المبالغ، والفترات الزمنية، وشروط الموافقة، ويعيّن جهة مسؤولة داخل المؤسسة وجهة اتصال للتعافي. يمكن لتعليمات النموذج شرح هذه الشروط، لكن يجب أن تفرضها الأدوات بصورة مستقلة عنه.

تربط حزمة المهمة بين هوية صاحب الطلب المتحقق منها والجهة صاحبة الحساب والعملية المطلوبة ونسخة الأدلة المعتمدة. ويتيح التفويض بين الأدوار نقل مهمة ضمن سير العمل المسموح به، ولا ينقل صلاحيات أوسع. وعلى الدور المتلقي إعادة التحقق من صلاحياته ومن جواز الإجراء المطلوب.

يتطلب كل تغيير في حالة النظام مفتاحًا لمنع تكرار الأثر، وسياسة محدودة لإعادة المحاولة، وسجلًا للنتيجة. وإذا انقضت مهلة أداة بعد تسلّم الطلب، فلا يُرسل الطلب مجددًا قبل التحقق مما إذا كان الإجراء قد وقع. وهذه ضرورة في المشتريات والرسائل والعمليات التي تترتب على تكرارها عواقب.

تحتاج الرقابة الفعّالة إلى أدوات واضحة يستطيع بها المشغّل إيقاف دور مؤقتًا، وسحب صلاحية أداة، وعزل إصدار مهارة، وفحص العمل المعلّق. ويجب أن تظل الأدوات عاملة عند تعطل خدمة النموذج. ويقترح التقرير هذه البنية؛ أما اختبار الضابط الاصطناعي فلا يتناول إلا حقولًا محددة من السياسة.

السياق البحثي	[13] National Institute of Standards and Technology
التحقق من الهوية	تحديد مقدم الطلب والجهة صاحبة البيانات
التفويض	فحص القدرة المحددة المسموح بها
تنفيذ	استدعاء أداة محدود بمفتاح فريد
التحقق	قراءة الحالة الفعلية ومقارنتها بالحالة المتوقعة
التسجيل	سجل الأدلة والنتائج

07 · TRUSTWORTHY OPERATION

Role authority is a technical boundary.

Every agent acts through a narrowly scoped account and a versioned operating contract.

The proposed role contract records permitted resources, tool operations, amount limits, time windows and approval conditions. It also names an organisational owner and a recovery contact. A model prompt can describe the contract, but the tool layer must enforce it independently.

A task envelope binds the authenticated requester, tenant, requested operation and evidence snapshot. Cross-role delegation can transfer a task within the allowed workflow; it cannot transfer broader permissions. The receiving role should recheck its own authority and the validity of the requested action.

State changes require an idempotency key, a bounded retry strategy and an outcome record. When a tool times out after receiving a request, the system should investigate whether the action occurred before sending it again. This is essential for purchases, messages and other operations where duplicate execution has consequences.

Oversight must be practical. The operator needs clear controls to pause a role, revoke a tool, quarantine a skill version and inspect pending work. These controls should work when the model service is unavailable. The current report proposes this architecture; the synthetic gate tests only selected policy fields.

Authenticate	Identify requester and tenant
Authorise	Check exact capability
Execute	Bounded tool call with unique key
Verify	Read back expected state
Record	Evidence and outcome ledger

Research context [13] National Institute of Standards and Technology

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

07 - العمليات التايلندية

التعامل مع المحتوى الوارد بوصفه أدلة غير موثوقة.

سيواجه النظام وثائق ورسائل ناقصة أو مضللة، أو مصممة عمدًا للتأثير في سلوكه على نحو ضار.

قد تَرِد تعليمات خبيثة في ملاحظة مورّد أو ملف مهارة مسترجع أو صفحة بحث أو رسالة عميل. وقد تطلب من النموذج تجاهل قواعده أو كشف بيانات أو استدعاء أداة. ويجب أن تفصل الضوابط بين محتوى الدليل والجهة المخوّلة بمنح الوكيل صلاحياته. [10]

تتألف الدفاعات المقترحة من طبقات: صلاحيات محدودة للأدوات، ووسم المصادر، ومخططات بيانات صريحة، واسترجاع منضبط، وتنفيذ معزول، وشروط اعتماد مستقلة. واتباع النموذج للتعليمات أحد المكونات وليس الحاجز الوحيد. ولا تُسترجع القيم الحساسة إلا إذا احتاجت إليها المهمة المصرح بها.

ينبغي أن يشمل التقييم محاولات استرجاع بيانات جهة أخرى، ونصوص موافقة مزورة، ومقتطفات قواعد متعارضة، وبيانات اعتماد قديمة، ونتائج أدوات مشوهة، وطلبات متكررة. وتحتاج الاختبارات إلى نتائج متوقعة قابلة للقياس: رفض الوصول، أو عزل الأدلة، أو التأجيل الصريح، أو إجراء محدود وآمن.

لا يختبر هذا الإصدار نموذجًا لغويًا حيًا في مواجهة مدخلات خصمية. وتستخدم محاكاة الذاكرة وسوم ثقة صريحة، لذا لا تقيس قدرة نموذج على تعرف النص العدائي. وينبغي تنويع محتوى الاختبار في الدراسة التالية، مع الإبقاء على حدود للقدرات تفرضها آلية مستقلة.

تجديد أماكن العمل	المساعدة المؤقتة المقترحة
نص مسترجع يطلب إجراءً ذا صلاحيات خاصة	يظل الإجراء غير متاح لهذا الدور.
ملاحظة مزورة تدّعي وجود موافقة	لا يُقبل إلا سجل الموافقة الموثَّق.
أداة تعيد حقولًا غير متوقعة	يمنع الإخفاق في مطابقة البنية المتوقعة الاستمرار الآلي
سجل تابع لجهة أخرى ذو صلة بالمهمة	يستبعده مرشّح الصلاحيات من الاسترجاع

السياق البحثي	[10] Greshake	[14] National Institute of Standards and Technology
ZEISMO-LAB-TR-2026-001 · v1.3	28 أيلول/سبتمبر 2026	33 / 72

07 · TRUSTWORTHY OPERATION

Treat incoming content as untrusted evidence.

The system will encounter documents and messages that are incomplete, misleading or deliberately adversarial.

A malicious instruction may arrive inside a supplier note, a retrieved skill file, a research page or a customer message. It may ask the model to ignore its rules, reveal data or call a tool. The control design must preserve the distinction between evidence content and the authority that governs the agent. [10]

The proposed defences are layered: narrow tool permissions, source labelling, explicit data schemas, controlled retrieval, isolated execution and independent approval conditions. Model-level instruction following is one component, not the sole boundary. Sensitive values should be retrieved only when the permitted task requires them.

Evaluation should include attempted cross-tenant retrieval, forged approval text, conflicting policy snippets, stale credentials, malformed tool results and repeated requests. Tests need measurable expected outcomes: denied access, quarantined evidence, an explicit deferral or a safely bounded action.

This edition does not run an adversarial evaluation against a live language model. Its memory simulation uses explicit trust labels and therefore does not measure whether a model can recognise hostile text. The next study must vary the attack content while preserving an independently enforced capability boundary.

FAILURE ROUTE	PROPOSED TEST ASSERTION
Retrieved text requests a privileged action	The action remains unavailable to the role
A forged note claims approval	Only the authenticated approval record is accepted
A tool returns unexpected fields	Schema failure prevents automatic continuation
Another tenant's record is relevant	Permission filter excludes it from retrieval

Research context [10] Greshake [14] National Institute of Standards and Technology

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

07 - العمليات التايلندية

توثيق مصادر البيانات وتقليل كشفها.

يجب أن يتيح الأرشيف البحثي فحص النتائج دون كشف معلومات العملاء أو الأسرار التشغيلية.

ينبغي أن توثّق كل مجموعة بيانات تجريبية غرضها وطريقة إنتاجها ومعنى كل حقل والاستخدامات التي تصلح لها. وبيانات هذه الطبعة اصطناعية بالكامل: معرّفات مهام مولّدة، وحقول سياسات موزعة عشوائيًا، ووسوم تحاكي صحة المقترحات. ولا تحتوي على أسماء عملاء أو رسائل أو سجلات دفع أو بيانات اعتماد للإنتاج.

ينبغي ألا تستخدم الدراسات التشغيلية المستقبلية سوى بيانات مأذون باستخدامها، مع تقليل ما تحتفظ به منها. ويجب تقييم إزالة ما يكشف الهوية على مستوى السجل كاملًا، لا الأسماء وحدها. فقد يظل الجمع بين التوقيت والموقع وحادثة غير مألوفة كافيًا لتحديد شخص أو علاقة تجارية. ويجب قصر الوصول على الأشخاص والأدوار الذين يجرون الدراسة.

ينبغي أن تفصل الحزمة البحثية بين النتائج القابلة للنشر، وبيانات المصدر المقيّدة، والإعدادات السرية. وتساعد البصمات على اكتشاف تغيّر ملف محفوظ، بينما قد يعزز أرشيف مستقل أو توقيعٌ توثيقَ مصدره. ولا تثبت البصمة أو رقم السجل المحلي صحةَ الادعاء العلمي نفسه.

تقدّم وثائق توصيف مجموعات البيانات والنماذج صيغًا نافعة لهذا العمل. ونستفيد من تركيزها على الغرض والتكوين والاستخدام المقصود وسياق التقييم لبناء سجل بحثي للنظام بأكمله، مع إبقائه موجزًا بما يكفي ليستمر المشغّلون في تحديثه.

هذا البيان

احتفظنا ببيانات التجارب الاصطناعية وببصمات ملفات مصدر مختارة. ولا تتضمن حزمة الأدلة بيانات أنظمة الإنتاج أو بيانات الاعتماد أو محتوى عملاء خاصًا أو محادثات داخلية خامًا.

السياق البحثي	[15] Gebru	[16] Mitchell	[18] National Institute of Standards and Technology

07 · TRUSTWORTHY OPERATION

Preserve data provenance and minimise exposure.

A research archive should make results inspectable without disclosing customer information or operational secrets.

Every experimental dataset should record why it exists, how it was produced, what each field means and what uses it supports. In this edition, the data are entirely synthetic. They contain generated task identifiers, randomised policy fields and simulated correctness labels; no customer names, messages, payment records or production credentials are included.

Future operational studies should use only authorised data and minimise the information retained. De-identification must be assessed against the whole record, not just names. A combination of timestamps, locations and unusual incidents may still identify a person or business relationship. Access should be scoped to the people and roles conducting the study.

The research package should separate publishable results, restricted source data and secret configuration. Hashes can help detect changes to a retained file, while an independent archive or signature can strengthen provenance. Neither a checksum nor a local record number proves that the underlying scientific claim is correct.

Dataset and model documentation provide useful formats for this discipline. We adapt their emphasis on purpose, composition, intended use and evaluation context to a system-level research record. The resulting documentation should remain concise enough that operators actually maintain it.

THIS EDITION'S DATA BOUNDARY

Synthetic experimental data and selected source-file digests are retained. Production system data, credentials, private customer content and raw internal conversations are excluded from the research evidence package.

Research context [15] Gebru [16] Mitchell [18] National Institute of Standards and Technology

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

08: الطرائق الفنية

أربع تجارب على آليات ضبط القرار.

البروتوكول ZEISMO-JSB-SIM-1.0. نُفّذ محليًا ببيانات مولّدة ومن دون استدعاء نماذج خارجية.

الخبرة	التصميم	المسألة
E1 · بوابة ضبط القرار	12,000 مهمة؛ 3 متحكّمات؛ 10 بذور توليد ثابتة	كيف تتغيّر التغطية ومخالفات السياسة عند فرض ضوابط صريحة؟
E2 · استرجاع الذاكرة	عدد الحالات: 3,000؛ عدد آليات الاختيار: 2؛ أدلة موسومة	ما أثر تصفية السجلات بحسب مصدرها في استرجاع الأدلة المقبولة؟
E3 · الترابط بين أعضاء المجلس	12 حالة تجريبية؛ 10,000 حالة اختبار لكل منها	كيف يؤثر عدد المصوّتين والأخطاء المشتركة في صحة قرار الأغلبية؟
E4 · توقيت الضوابط	3 دوال؛ قيس أداؤها في 101 دفعة تضم كلٌّ منها 1,200 عنصر.	ما تكلفة توجيه القرار داخل العملية البرمجية بعد تهيئتها؟

تعزل التجارب آليات هندسية باستخدام شيفرة بايثون عادية، ولا يشارك فيها متخصص بشري أو نموذج لغوي. فوسوم الصحة وقيم الثقة وعلامات موثوقية المصدر والظروف الخارجية كلها بيانات اختبار مولّدة. وتصف النتائج هذا التصميم الاصطناعي، لا أداء جارڤيس في التشغيل الفعلي.

تستخدم جميع مهام اختبار القواعد البذور من 410 إلى 419. وتُشتق منها بذور ثابتة منفصلة لتجارب الذاكرة والمجلس. وقد أعاد تشغيلان مستقلان إنتاج بيانات المهام وسجلات القرار ونتائج الذاكرة والنتائج المجمّعة وسجلات التحقق بتطابق تام في البايتات. واستُبعدت قياسات الزمن من فحص التطابق عمدًا، لأن حمل النظام قد يغيّرها.

أُعدّ البروتوكول ونُفّذ خصيصًا لهذا التقرير؛ ولم يُسجّل مسبقًا لدى جهة خارجية. احتُفظ بالشيفرة والتوزيعات والعتبات والمخرجات ليتسنّى للمراجع معرفة ما اختُبر تحديدًا واقتراح شروط متابعة أشدّ. لا تتضمن هذه التجارب اختيارًا بين نماذج، أو مجموعة تدريب مخفية، أو مصنّفًا دُرّب على البيانات.

08 · EXPERIMENTAL METHODS

Four experiments on the control substrate.

Protocol ZEISMO-JSB-SIM-1.0. Executed locally with generated data and no external model calls.

EXPERIMENT	DESIGN	QUESTION
E1 · Decision gate	12,000 tasks; 3 controllers; 10 fixed seeds	How do coverage and policy violations change under explicit controls?
E2 · Memory retrieval	3,000 situations; 2 selectors; labelled evidence	What is the effect of provenance filtering on admissible retrieval?
E3 · Council correlation	12 conditions; 10,000 cases per condition	How do count and shared errors affect majority correctness?
E4 · Control timing	3 functions; 101 measured batches of 1,200 items	What is the in-process routing cost under warmed conditions?

The experiments isolate engineering mechanisms using ordinary Python code. No participant is a human specialist or a language model. The correctness labels, confidence values, trust flags and external conditions are generated fixtures. The results describe this synthetic design, not production Jarvis performance.

All policy tasks use seeds 410 through 419. Memory and council experiments derive separate fixed seeds from this set. Two independent process runs reproduce the same task data, decision records, memory outcomes, aggregate results and assertion records byte for byte. Timing is intentionally excluded from equality checks because system load can change it.

The protocol was developed and executed for this report; it was not externally preregistered. The code, distributions, thresholds and outputs are retained so a reviewer can identify exactly what was tested and create stronger follow-up conditions. There is no model selection, hidden training set or fitted classifier in these experiments.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 36 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

08 · منهجية E1

فحص المقترحات وفق القواعد قبل السماح بتنفيذها.

لكل مهمة إجراء مقترح وقيمة ثقة اصطناعية وحقول تؤثر في تطبيق السياسة، جرى توليدها بصورة مستقلة.

الحقل	GENERATOR أو INTERPRETATION
المجال	الدعم أو المخزون أو المالية أو التسويق أو الجدولة أو البحث؛ 2,000 مهمة لكل مجال
الثقة	سحب منتظم بين 0.50 و995.0؛ يُحفظ إلى ست منازل عشرية.
تصنيف صحة النتيجة	سحب برنولي باحتمال يساوي الثقة ناقص 0.03 في الحالات المستقرة، والثقة ناقص 0.21 في الحالات المتغيرة.
الإذن / الثقة بالمصدر	يبلغ احتمال غياب كل عنصر أو عدم صلاحيته 0.08
تكرار الأثر / عدم قابلية التراجع	الاحتمال 0.05 / 0.08 على الترتيب
عمر الأدلة	اختيار منتظم من 1 أو 2 أو 6 أو 12 أو 24 أو 36 أو 72 ساعة.
المبلغ	اختيار باحتمالات متساوية بين 0 أو 20 أو 50 أو 100 أو 200 أو 500 أو 1,000 وحدة افتراضية.
تغيّر التوزيع	آخر 300 مهمة من أصل 1,200 لكل بذرة؛ أي 25% من مجموعة البيانات كاملة.

ينفّذ المتحكّم المعتمد على الثقة وحدها عند بلوغها 0.70 فأكثر. أما المتحكّمان المزوّدان بضوابط، فيمنعان أولًا التكرار والتنفيذ بلا إذن والاعتماد على مصادر غير موثوقة. ويُحيلان إلى المراجعة المقترحات التي يتعذّر التراجع عنها أو تتجاوز 500 وحدة، ولا ينفّذان إلا عند ثقة لا تقل عن 0.82. ويؤجّل المتحكّم الكامل أيضًا المقترحات التي تستند إلى أدلة مضى عليها أكثر من أربع وعشرين ساعة.

يفحص مرجع السياسة الحقول التي تجعل التنفيذ مؤهلًا، ولا يستخدم وسم الصحة المُولّد. لذلك قد يستوفي المقترح قواعد السياسة ويظل خاطئًا. وجميع الكميات اصطناعية ولا تمثل عملة أو أداءً تجاريًا. وتسمّي المجالات مجموعات

فرعية متوازنة، لكنها لا تنشئ مهام مختلفة في معناها.

08 · E1 METHODS

Policy gating over proposed decisions.

Each task carries a proposed action, a synthetic confidence value and independently generated policy-relevant fields.

FIELD	GENERATOR OR INTERPRETATION
Domain	Support, inventory, finance, marketing, scheduling or research; 2,000 tasks each
Confidence	Uniform draw from 0.50 to 0.995; stored to six decimal places
Correctness label	Bernoulli draw with probability confidence − 0.03 in stable cases; confidence − 0.21 in shifted cases
Permission / trust	Each is absent or invalid with probability 0.08
Duplicate / irreversible	Probability 0.05 / 0.08 respectively
Evidence age	Uniform choice from 1, 2, 6, 12, 24, 36 or 72 hours
Amount	Uniform choice from 0, 20, 50, 100, 200, 500 or 1,000 abstract units
Shift	Final 300 of 1,200 tasks per seed; 25% of the complete dataset

The confidence-only controller executes at confidence 0.70 or above. The two gated controllers first block duplicates, missing permission and untrusted sources. They send irreversible or above-500-unit proposals to review and require confidence of at least 0.82 to execute. The complete controller also defers evidence older than twenty-four hours.

The policy oracle checks the fields that make execution eligible; it does not use the generated correctness label. A proposal can therefore be policy-valid and still wrong. All quantities are artificial, with no currency or business-performance interpretation. Domain names label balanced subsets but do not introduce distinct semantic tasks.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 37 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

09 · نتائج E1

تقل التغطية كلما اشتدت الضوابط.

نتيجة مقاسة على بيانات اصطناعية: تمنع البوابة الكاملة المخالفات المرمّزة مع تقليص مجموعة الإجراءات المنفذة آليًا.

الثقة

فقط

دون

حداثة الأدلة

الكاملة

البوابة

0

10

20

30

40

50

60

70

المهام المنفّذة (%)

58.8

22.0

15.7

الثقة

فقط

دون

حداثة الأدلة

الكاملة

البوابة

0

10

20

30

40

50

60

70

مخالفات القواعد بين الحالات المنفّذة (%)

54.9

28.3

0.0

الشكل 1. نتائج متحكم السياسة في 12,000 مهمة اصطناعية. ويُحسب معدل المخالفات بين المهام المنفَّذة، لا بين جميع المهام.

المراقبة	تنفيذ	التغطية	الانتهاكات	الصحة*
الثقة وحدها	7,058	58.8%	3,877	78.4%
دون فحص حداثة الأدلة	2,635	22.0%	747	84.7%
بوابة الضبط الكاملة	1,888	15.7%	0	85.1%

ينفّذ المتحكم القائم على الثقة وحدها 7,058 مهمة، ويسجل 3,877 مخالفة للقواعد المرمّزة. أما بوابة الضبط الكاملة فتنفّذ 1,888 مهمة دون هذه المخالفات. وتبلغ تغطيتها 15.7%، وصحة المقترحات المولّدة بين الحالات المنفّذة 85.1%.

يُتوقع عدم تسجيل أي مخالفة عندما تُطبّق تصنيفات قواعد صريحة وصحيحة تطبيقًا سليمًا. وهذه نتيجة اتساق في التنفيذ، وليست برهانًا على أمان الذكاء الاصطناعي عمومًا. ولا تستطيع بوابة الضبط التحقق مما إذا كان مورّد قد كذب، أو مستند قد أسيء فهمه، أو المقترح المولّد يحقق هدفًا تجاريًا حقيقيًا.

*تُحدَّد صحة المقترح بوسم مولّد في المحاكاة. تغيّر المتحكمات المقترحات التي يُسمح بتنفيذها، ولا تحسّن مضمون المقترح نفسه أو تعيد صياغته.

09 · E1 RESULTS

Coverage falls as controls become stricter.

Measured synthetic result: the complete gate prevents the encoded violations while retaining a smaller automatic-execution set.

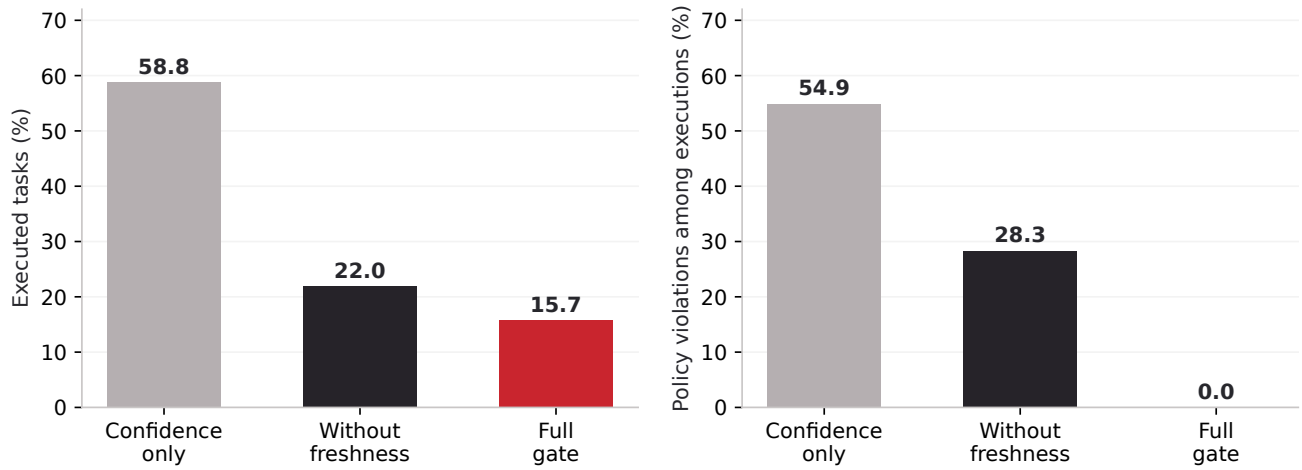


Figure 1. Policy-controller outcomes on 12,000 synthetic tasks. Violation rate is calculated among executions, not among all tasks.

CONTROLLER	EXECUTE	COVERAGE	VIOLATIONS	CORRECTNESS*
Confidence only	7,058	58.8%	3,877	78.4%
Without freshness	2,635	22.0%	747	84.7%
Full gate	1,888	15.7%	0	85.1%

The confidence-only controller executes 7,058 tasks and produces 3,877 encoded policy violations. The complete gate executes 1,888 tasks with zero such violations. Its coverage is 15.7% and generated-proposal correctness among executions is 85.1%.

Zero policy violations is expected when explicit, accurate policy labels are enforced correctly. It is an implementation consistency result, not a proof of general AI safety. The gate has no ability to verify whether a supplier lied, whether a document was misunderstood or whether the generated proposal achieves a real business objective.

*Correctness is the generated proposal label. Controllers change which proposals execute; they do not improve or rewrite a proposal.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 38 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو
مختبر الأبحاث
جارڤيس / سيستم برين
سلسلة الأبحاث · 2026
09 · نتائج E1

قد يستوفي القرار قواعد السياسة ويظل خاطئًا.

انخفضت نسبة المقترحات الصحيحة عند تغير توزيع البيانات المولّدة، مع استمرار بوابة التنفيذ في تطبيق قواعدها.

ظروف مستقرة	ظروف متغيرة
50	
60	
70	
80	
90	
100	

صحة المقترحات المنفّذة (%)

الثقة وحدها · دون فحص حداثة الأدلة · بوابة الضبط الكاملة

الشكل 2. نسبة المقترحات الصحيحة بين المقترحات المنفذة في كل نظام محاكاة. عند تغيير توزيع البيانات يُخفّض المولّد احتمال صحة المقترح من دون تغيير مستوى الثقة المعلن.

مع حاجز التحقق الكامل بلغت نسبة الصحة 88.6% في الظروف المستقرة، مقابل 74.5% بعد تغيّر الظروف. وبلغت النسبة الإجمالية 85.1%، مع فاصل وصفي بطريقة ويلسون عند مستوى 95% يتراوح من 83.4% إلى 86.7%. يلخّص الفاصل نتائج برنولي المولّدة في التجربة، وليس فاصل ثقة للأداء في بيئة عمل فعلية.

تنشأ النتيجة من تصميم التجربة نفسه: ففي الحالات التي يتغيّر توزيعها، يُفترض احتمال أقل للصواب. وتكمن قيمتها في التشخيص؛ فقد تعرض لوحة الرقابة امتثالًا كاملًا مع تراجع جودة القرارات المقبولة. لذلك لا بدّ من قياس الالتزام بالسياسة والنتيجة معًا.

لا تثبت التجربة أثرًا سببيًا لعتبة الثقة منفصلًا عن اختيار الحالات؛ فالعتبات الأشد تختار مجموعة أخرى من المقترحات. وينبغي أن تقارن دراسة لاحقة الحالات نفسها، وتحتفظ بحالات الامتناع، وتقيس تكلفة التنفيذ الخاطئ والتأجيل غير الضروري معًا.

09 · E1 RESULTS

A policy-valid decision can still be wrong.

Distribution shift lowers generated proposal correctness even when the control gate continues to enforce its rules.

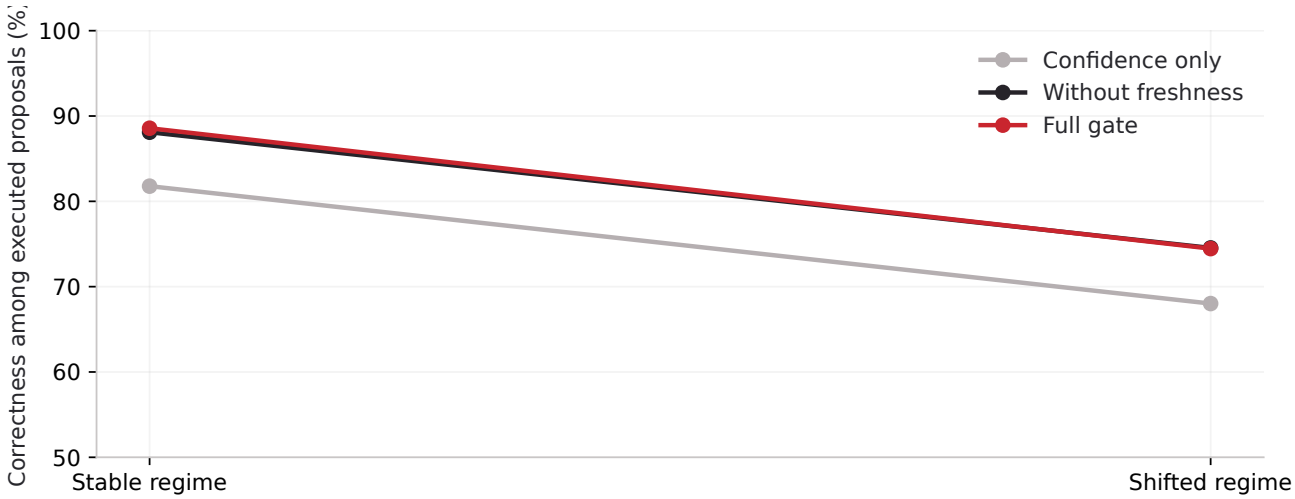


Figure 2. Executed-proposal correctness by synthetic regime. The shifted generator reduces the probability of correctness without changing the declared confidence.

For the complete gate, correctness is 88.6% in the stable regime and 74.5% in the shifted regime. The pooled result is 85.1% with a descriptive Wilson 95% interval of 83.4% to 86.7%. This interval summarises the generated Bernoulli outcomes; it is not a confidence interval for workplace deployment.

The result is built into the experimental mechanism: shifted cases use a lower correctness probability. Its value is diagnostic. A control dashboard could show perfect enforcement while the quality of accepted decisions deteriorates. Consequently, both policy and outcome metrics are necessary.

The experiment does not establish a causal effect of confidence thresholds independent of case selection. Stricter thresholds select a different proposal population. A future model study must compare shared cases, retain abstentions and evaluate the cost of both mistaken action and unnecessary deferral.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

08 · منهجية E2

استرجاع الذاكرة مع تفاوت جودة الأدلة.

تختبر مجموعة صغيرة مضبوطة من الحالات عواقب اختيار السجل الأحدث وحده.

في كل واحدة من 3,000 حالة استرجاع مولّدة، يوجد سجل مقبول باحتمال 0.90. هذا السجل موثوق ومعتمد ويتبع حساب المؤسسة A، وتبلغ صلاحيته اثنتين وسبعين ساعة. ويُختار عمره بين ساعتين وأربع وعشرين ساعة. ويُضاف ما يصل إلى أربعة سجلات منافسة باستقلال بعضها عن بعض، باحتمال 0.80 لكل سجل.

تمثّل السجلات المتنافسة أربع فئات إخفاق صريحة: أدلة منتهية الصلاحية، وملاحظة غير معتمدة، وسجل تابع لجهة أخرى، ومصدر غير موثوق. وتصنيفاتها قابلة للقراءة آليًا وصحيحة بحكم طريقة إنشاء البيانات. ويمنع سجل احتياطي غير معتمد خلو مجموعة الخيارات في حالة «الأحدث أولًا».

تختار آلية «الأحدث أولًا» السجل الأقل عمرًا، وتفصل بين السجلات المتساوية في العمر بحسب الصلة بالمهمة. أما الآلية القائمة على المصدر فتستبعد أولًا السجلات غير الموثوقة أو غير المعتمدة أو المنتهية أو التابعة لجهة أخرى، ثم تطبّق الترتيب نفسه. وإذا لم يوجد سجل مقبول فإنها تمتنع عن الاختيار. ولا تستخدم التجربة نموذج تضمينات أو تفسيرًا للغة الطبيعية أو آلية استرجاع متعلّمة.

المقياس هو ما إذا كان السجل المختار مقبولًا وفق قواعد الحالة الاختبارية. تختبر التجربة قبول السجلات عند الاسترجاع، لا صحة ما تقوله جمل باللغة الطبيعية. وتتيح وسومها المضبوطة اختبار الحدّ الذي صُمّم النظام لفرضه، قبل إدخال حالات أصعب مثل تزوير معلومات المصدر أو غموض الصلاحية.

ثالثا - المنهجية	تعريف
تغطية الاختيار	المواقف المختارة / جميع المواقف
دقة اختيار الأدلة الصالحة	عدد السجلات الصالحة المختارة / إجمالي السجلات المختارة
اختيار غير صالح	يخالف السجل المختار قاعدةً واحدة على الأقل من قواعد حالة الاختبار.
الامتناع	لا يُعاد أي سجل لعدم بقاء دليل مقبول.

08 · E2 METHODS

Memory retrieval with conflicting evidence quality.

A small controlled corpus tests the consequences of selecting by recency alone.

Each of 3,000 generated retrieval situations may contain an admissible record with probability 0.90. That record is trusted, approved, associated with tenant A and valid for seventy-two hours. Its age is drawn from two to twenty-four hours. Up to four competing records are introduced independently with probability 0.80 each.

The competing records represent four explicit failure classes: expired evidence, an unapproved note, a foreign-tenant record and an untrusted source. Their labels are machine-readable and accurate by construction. A fallback unapproved record prevents an empty candidate set in the recent-first condition.

The recent-first selector chooses the youngest record, breaking an age tie with relevance. The provenance selector first excludes untrusted, unapproved, expired and foreign records, then uses the same ordering. If nothing is admissible, it abstains. No embedding model, natural-language interpretation or learned retriever is involved.

The outcome is whether the selected record is valid under the fixture rules. This is a retrieval-admission experiment, not an assessment of the truth of natural-language claims. Its controlled labels let us test the architecture's intended boundary before introducing harder cases such as forged provenance or ambiguous validity.

MEASURE	DEFINITION
Selection coverage	Selected situations / all situations
Valid precision	Valid records selected / all records selected
Invalid selection	A selected record violates at least one fixture rule
Abstention	No record is returned because no admissible evidence remains

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 40 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

09 · نتائج E2

كيف يؤثر التحقق من مصدر المعلومة في استخدام الذاكرة؟

نتيجة مقاسة على بيانات اصطناعية: الامتناع المبرر أفضل من استرجاع سجل غير مقبول بوضوح.

0	500	1000	1500	2000
2500		3000		

حالات اصطناعية لاسترجاع المعلومات

الأحدث أولًا

التحقق من المصدر

أدلة صالحة	أدلة غير صالحة	الامتناع

الشكل 3. النتائج في 3,000 حالة استرجاع اصطناعية. وتُحدد الصلاحية بناءً على التسميات الصريحة لحالات الاختبار.

أداة الاختيار	صالح	غير صالح	امتناع	الموافقة
الأحدث أولًا	948	2,052	0	31.6%
التحقق من المصدر	2,718	0	282	100.0%

أعادت آلية تفضيل الأحدث 2,052 سجلًا غير صالح. وأعادت الآلية التي تتحقق من مصدر الأدلة 2,718 سجلًا صالحًا، وامتنعت عن الاختيار في 282 حالة. وتعود دقة نتائجها المسترجعة البالغة 100.0% إلى أنها تطبّق التسميات الصحيحة نفسها التي تُعرّف الصلاحية.

لا يثبت ذلك قدرة النظام على اكتشاف ادّعاء كاذب بشأن المصدر، أو التمييز بين نصيحة جيدة وأخرى سيئة في نص لغوي. ويتطلب اختبار ذلك سجلات خضعت لتقييم مستقل وتجربة منفصلة. أما النتيجة الهندسية المفيدة فهي إمكان فرض شروط قبول السجل قبل ترتيبه بحسب الصلة أو الحداثة.

ينبغي أن تختبر دراسة الذاكرة التالية أثر إفساد البيانات الوصفية عمدًا وإدخال تصحيحات متنازع عليها وتغيير قواعد الاحتفاظ. ويُقاس عدد الأخطاء الناشئة عند قبول المصادر وترتيب نتائج الاسترجاع واقتطاع السياق وتفسير النموذج في المراحل اللاحقة.

09 · E2 RESULTS

Provenance filtering changes what memory can influence.

Measured synthetic result: a valid abstention is preferable to retrieving a clearly inadmissible record.

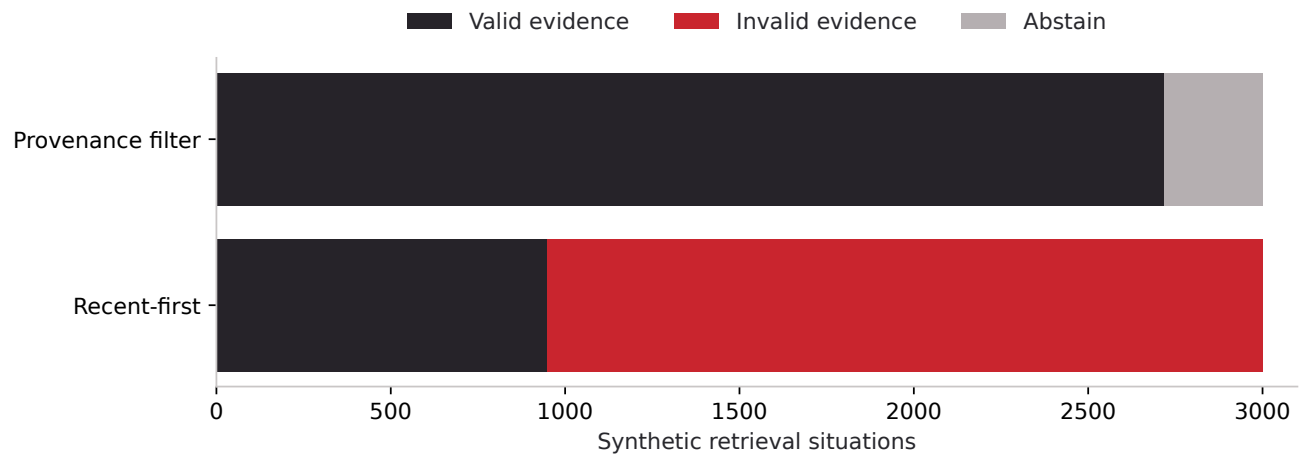


Figure 3. Outcomes in 3,000 synthetic retrieval situations. Validity is determined from explicit fixture labels.

SELECTOR	VALID	INVALID	ABSTAIN	PRECISION
Recent-first	948	2,052	0	31.6%
Provenance filter	2,718	0	282	100.0%

The recent-first selector returns 2,052 invalid records. The provenance selector returns 2,718 valid records and abstains in 282 situations. Its 100.0% precision follows from enforcing the same accurate labels that define validity.

This is not evidence that the system can detect a false source claim or distinguish good and bad advice in natural language. Those would require independently assessed records and a separate evaluation. The useful engineering finding is that admissibility can be enforced before relevance or recency ranking.

The next memory study should deliberately corrupt metadata, introduce disputed corrections and vary retention rules. It should measure how many errors originate in source admission, retrieval ranking, context truncation and downstream model interpretation.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 41 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

08 · منهجية E3

كم من وجهات النظر مستقل فعلًا؟

تجربة بمصوّتين في المحاكاة تفصل أثر حجم المجلس عن أثر الخطأ المشترك.

تعتمد الدراسة أصواتًا ثنائية تميز الإجابة الصحيحة من الخاطئة، ولا تعتمد على مشاركين بشريين أو مخرجات لنماذج ذكاء اصطناعي. ويبلغ الاحتمال الهامشي لصحة إجابة كل مصوّت في المحاكاة 0.75. وتضمّ المجالس مصوّتًا واحدًا، أو ثلاثة، أو خمسة عشر، أو مئة. ويشمل كل شرط عشرة آلاف حالة موزعة على عشر بذور عشوائية ثابتة.

يحدث في كل حالة عامل مشترك مفاجئ باحتمال ρ. فإذا وقع، حصل جميع المصوّتين على نتيجة صحة واحدة من توزيع برنولي. وإذا لم يقع، سحب كل مصوّت نتيجة مستقلة باحتمال صحة قدره 0.75. واختبرنا ρ عند القيم 0 و25.0 و60.0. ووفق هذا البناء، يساوي ρ أيضًا القيمة المتوقعة لمعامل الارتباط بين صحة نتائج كل زوج من المصوّتين.

يُعدّ قرار المجلس صائبًا إذا أصابت أغلبية الأصوات. وعند التعادل التام، وهو ممكن حين يضم المجلس مئة مصوّت، تُحسم النتيجة عشوائيًا باحتمالين متساويين. ويسجّل المحاكي الأعداد بحسب البذرة والشرط. ولا يجري مناظرة نصية أو يستخدم أدوات أو يجمع أدلة أو يدمج الأصوات بأسلوب متعلّم.

يسأل هذا النموذج المبسّط عمدًا هل تساعد زيادة المشاركين على تجاوز الأخطاء المشتركة. ولا يمثّل جميع مجالس الوكلاء الفعلية. فقد تولّد التعليمات المشتركة والتدريب المتطابق والاسترجاع من المصادر نفسها وترتيب التواصل أنماط ارتباط أعقد من احتمال وقوع صدمة مشتركة واحدة.

$$\mathbf{P(correct\ majority) \approx \rho \times 0.75 + (1 - \rho) \times B(n,\ 0.75)}$$

تمثل B احتمال صواب الأغلبية عندما يصوّت الأعضاء بصورة مستقلة، ويُحتسب نصف احتمال التعادل إذا كان n زوجيًا. وتصف الصيغة طريقة بناء هذا المحاكي.

08 · E3 METHODS

How many perspectives are actually independent?

A surrogate-voter experiment separates council size from shared error.

The study uses binary correctness votes, not human participants or AI model outputs. Every simulated voter has marginal correctness probability 0.75. Council sizes are one, three, fifteen and one hundred. Each condition contains ten thousand cases across ten fixed seeds.

For each case, a shared-shock event occurs with probability ρ. If it occurs, all voters receive the same Bernoulli correctness result. Otherwise, each voter draws an independent result with probability 0.75. We test ρ values of 0, 0.25 and 0.60. Under this construction, ρ also equals the pairwise correctness correlation in expectation.

The council is correct when a majority of votes is correct. At an exact tie, possible for the one-hundred-voter condition, a fair random tie-break is used. The simulator records counts per seed and condition. It does not perform textual debate, tool use, evidence gathering or learned aggregation.

This deliberately simple model asks whether more participants can overcome shared mistakes. It is not a model of every real agent council. Shared prompt context, identical training, common retrieval and communication order may produce more complicated dependence than a single shared-shock probability.

$$P(\text{correct majority}) \approx \rho \times 0.75 + (1 - \rho) \times B(n, 0.75)$$

B is the independent-voter majority probability, including a half-weight tie term for even n. The expression describes this simulator's construction.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 42 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

09 · نتائج E3

مئة صوت لا تصحّح خطأً تشترك فيه جميعًا.

نتيجة مقاسة على بيانات اصطناعية: يحدّ الترابط من فائدة زيادة عدد أعضاء المجلس.

1	3	15	100

مصوّتون بالمحاكاة، لا أشخاص ولا نماذج لغوية.

70

75

80

85

90

95

100

صحة قرار الأغلبية (%)

احتمال الصدمة المشتركة: 0.00 · احتمال الصدمة المشتركة: 0.25 · احتمال الصدمة المشتركة: 0.60

الشكل 4. صحة قرار الأغلبية في نموذج المصوّتين المحاكى. تمثّل كل نقطة 10,000 حالة؛ ولم يُقيَّم أي بشر أو نماذج لغوية.

المصوّتون في المحاكاة	P = 0	P = 0.25	P = 0.60
1	75.0%	74.9%	75.1%
3	84.0%	81.9%	78.8%
15	98.6%	92.0%	84.1%
100	100.0%	93.8%	84.7%

يستفيد المصوّتون المستقلون بدرجة كبيرة من تجميع أصواتهم في هذا النموذج. أما مع الصدمات المشتركة فيكون التحسن أقل، لأن المجلس كله قد يخطئ معًا. ومع زيادة حجم المجلس، يقترب الحد النظري لهذا النموذج من 1 −

25ρ.0 بدلًا من الواحد.

تدعم النتيجة اختبار تنوّع الأدلة والمناهج، لا مجرد زيادة عدد الوكلاء بأسماء مختلفة. وينبغي أن تقيس تجربة لاحقة مدى تداخل الأخطاء، وأثر استبعاد كل دور على حدة، وما تضيفه المعلومات المختلفة فعليًا. ولا تثبت النتيجة أن تشغيل خمسة عشر أو مئة وكيل سيحقق هذه النسب.

09 · E3 RESULTS

A hundred votes do not erase a shared mistake.

Measured synthetic result: correlation limits the value of increasing council size.

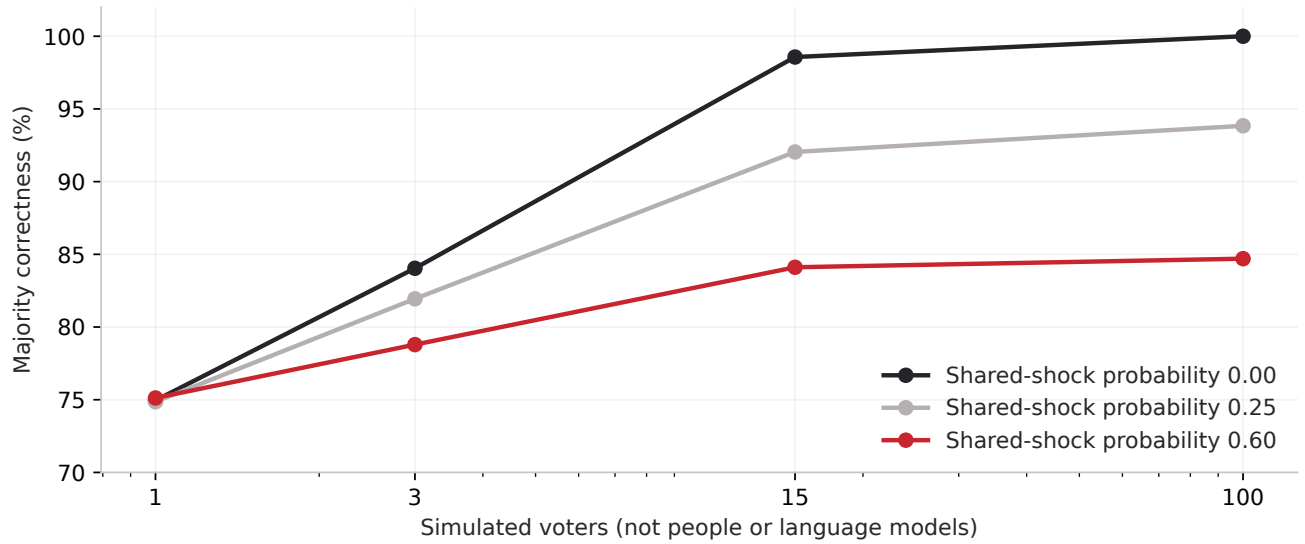


Figure 4. Majority correctness in the surrogate-voter model. Each point uses 10,000 cases; no people or language models were evaluated.

SIMULATED VOTERS	P = 0	P = 0.25	P = 0.60
1	75.0%	74.9%	75.1%
3	84.0%	81.9%	78.8%
15	98.6%	92.0%	84.1%
100	100.0%	93.8%	84.7%

Independent voters gain substantially from aggregation in this construction. Under shared shocks, the improvement is smaller because an entire council can be wrong together. As council size grows, the theoretical limit of this model approaches $1 - 0.25\rho$, rather than one.

This result supports testing diversity of evidence and methods, not simply increasing the number of agent identities. A future model experiment should measure error overlap, leave-one-role-out effects and the contribution of genuinely different information. The result does not establish that fifteen or one hundred deployed agents will achieve these percentages.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

09 · نتائج E4

توجيه الطلبات بسرعة ليس إلا جزءًا من عمل النظام.

اختبار أداء محلي مصغّر: تنفيذ وحدة تحكم بايثون ضمن العملية نفسها بعد الإحماء.

المراقبة	وسيط الميكروثانية لكل عنصر	المئين 95 · ميكروثانية لكل عنصر	المئين 99 · ميكروثانية لكل عنصر
الثقة وحدها	0.092	0.116	0.118
دون فحص حداثة الأدلة	0.172	0.204	0.212
بوابة الضبط الكاملة	0.185	0.222	0.239

يُهيأ كل متحكّم بعشر دفعات لا تدخل في القياس. ثم يُقاس زمنه في 101 دفعة، تضم كل منها المهام الاصطناعية الـ200,1 نفسها. يستخدم القياس ساعة رتيبة عالية الدقة. وكل مشاهدة معروضة هي زمن الدفعة الكلي مقسومًا على 1,200؛ والمئينات المنشورة محسوبة على متوسطات الدفعات هذه.

بيئة التشغيل المسجلة هي Python 3.9.25 على Linux / x86_64. ولا يثبّت القياس نوع معالج بعينه، ولا يحجز الخادم، ولا يضبط الأحمال المنافسة. وهو يسجّل تكلفة التنفيذ المحلي في الظروف المرصودة، ولا يجوز اعتباره ضمانًا لمستوى الخدمة قابلًا للنقل إلى بيئات أخرى.

يستبعد قياس الزمن التعرف على الكلام، واستدلال النموذج، وانتقال البيانات عبر الشبكة، والاسترجاع من قاعدة بيانات، وتنفيذ الأدوات، والتحقق من الآثار الخارجية. كما لا يقيس الأزمنة المرتفعة في ذيل توزيع الطلبات المفردة. وهذه الاستبعادات مقصودة لعزل بوابة الضبط الحتمية.

تبيّن النتيجة أن المتحكّم المرجعي الصغير قليل الكلفة الحسابية. لكنها لا تثبت استدلالًا متعدد التخصصات في ملّي ثوانٍ أو استجابةً كاملة من جارڤيس بهذه السرعة. يتطلب قياس النظام كاملًا قياس زمن كل مرحلة تحت أحمال متزامنة وحالات إخفاق واقعية.

وحدة الإبلاغ عن الانتهاكات

ميكروثانية لكل عنصر، محسوبة من متوسطات دفعات بعد التهيئة. أبقِ وصف الوحدة ملازمًا للأرقام عند إعادة استخدامها.

09 · E4 RESULTS

Fast routing is only one part of the system.

Measured local microbenchmark: warmed, in-process Python controller execution.

CONTROLLER	MEDIAN MS/ITEM	P95 MS/ITEM	P99 MS/ITEM
Confidence only	0.092	0.116	0.118
Without freshness	0.172	0.204	0.212
Full gate	0.185	0.222	0.239

Each controller is warmed with ten unmeasured batches. It is then timed over 101 batches, each containing the same 1,200 synthetic tasks. The measurement uses a monotonic high-resolution clock. Each displayed observation is the total batch time divided by 1,200; reported quantiles are quantiles of these batch means.

The recorded runtime is Python 3.9.25 on Linux / x86_64. The benchmark does not pin a CPU, reserve the server or control competing load. It records local execution cost under the observed conditions and should not be treated as a portable service-level guarantee.

The timing excludes speech recognition, model inference, network transit, retrieval from a database, tool execution and verification of external effects. It also does not measure individual-request tail latency. Those omissions are deliberate because the experiment isolates the deterministic gate.

The result establishes that this small reference controller is computationally lightweight. It does not establish millisecond multidisciplinary reasoning or a millisecond end-to-end Jarvis response. A full-system benchmark must instrument every stage under realistic concurrency and failure conditions.

CORRECT REPORTING UNIT

Microseconds per item, calculated from warmed batch means. Keep this label attached whenever the numbers are reused.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 44 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

10 - إعادة التأهيل

تفسّر اختبارات استبعاد المكوّنات الآليات، ولا تثبت أداءً عامًا.

تفيد مجموعة التجارب لأن تبسيطاتها ومقارناتها جميعًا معلنة بوضوح.

في التجربة E1، تؤدي إزالة فحص حداثة الأدلة إلى قرارات تنفيذ تخالف شرط عمر الدليل في السياسة الكاملة. وهذا اختبار لأثر حذف ضابط صريح واحد. وتبيّن المقارنة سلوك تلك القاعدة في ظل التوزيع المختار لأعمار الأدلة؛ ولا تقدّر مدى شيوع الأدلة المتقادمة في إدارة فعلية.

يختلف خط الأساس القائم على الثقة وحدها عن بوابة الضبط الكاملة في أكثر من جانب: فهو يستخدم فحوصًا أقل للقواعد وعتبة ثقة أخفض. ولذلك لا يجوز نسبة الفرق في جودة النتائج إلى ميزة واحدة. وينبغي لدراسة لاحقة استخدام تصميم تجريبي عاملي يغيّر العتبات وفحوص القواعد وجودة المقترحات كلًّا على حدة.

تقارن E2 آليات الاختيار بافتراض أن البيانات الوصفية صحيحة تمامًا. ودقة النتائج العالية بعد التصفية تنشأ من اتساق قواعد حالات الاختبار. أما الاختبار الأصعب فيقتضي قياس جودة البيانات الوصفية وقبول المصادر في ظروف خصومية. وتغيّر E3 بنية الارتباط المفترضة؛ ولا تكشف الارتباط الفعلي بين نماذج مستخدمة في الواقع.

يعزل الاختبار E4 كلفة التوجيه المحلي ويستبعد مراحل سير عمل الوكيل المكلفة أو المتغيرة. والاستنتاج هنا يتعلق بالبنية: يمكن وضع بعض الفحوص على مسار حتمي خفيف، بينما يحتاج زمن إنجاز المهمة كاملةً إلى قياس مستقل.

الترجمة الشفوية	توسّع لا يدعمه الدليل
ترفض الفحوص الصريحة الحالات الموسومة بأنها غير صالحة.	الذكاء الاصطناعي آمن تمامًا
تحتوي حالات الاختبار بعد التصفية على سجلات ذاكرة تستوفي القواعد.	النظام يعرف أي الادعاءات عن الواقع صحيح.
الأخطاء المشتركة تحدّ من فائدة التصويت.	مائة وكيل تعادل مائة متخصص.
ينفّذ فحص محلي سريعًا.	يستدل النظام الكامل خلال مللي ثوانٍ

10 · INTERPRETATION

Ablations explain mechanisms, not universal performance.

The experiment suite is useful because each simplification and each comparison is visible.

In E1, removing the freshness check creates executions that violate the complete policy's age requirement. This is an ablation of one explicit control. The comparison demonstrates the behaviour of that rule under the chosen age distribution; it does not estimate how often stale evidence occurs in a real department.

The confidence-only baseline differs in more than one way from the complete gate: it uses fewer policy checks and a lower confidence threshold. Its difference in outcome quality must therefore not be attributed to a single feature. A follow-up should use a factorial design that varies thresholds, policy checks and proposal quality separately.

E2 compares selectors under perfect metadata. Its strong filtered precision is a consistency property of the fixture rules. A harder experiment must measure metadata quality and adversarial source admission. E3 changes the assumed correlation structure; it does not discover the actual correlation of deployed models.

E4 isolates local routing cost and omits the expensive or variable stages of an agent workflow. The proper conclusion is architectural: some checks can be placed in a lightweight deterministic path, while the complete task still needs independent latency measurement.

SUPPORTED INTERPRETATION	UNSUPPORTED EXTENSION
Explicit checks reject labelled invalid cases	The AI is fully safe
Filtered fixtures contain valid memory records	The system knows which real claims are true
Shared errors weaken voting gains	One hundred agents equal one hundred specialists
A local gate executes quickly	The complete system reasons in milliseconds

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

10 · ضوابط التحليل الإحصائي

إظهار المقامات والبذور العشوائية وعدم اليقين.

يجب أن تتيح كل نتيجة إعادة بناء البسط والمجموعة التي تصفها.

مقام نسبة تغطية التنفيذ هو جميع المهام، بينما يُحسب معدل مخالفة السياسة وصحة المقترحات بالنسبة إلى المهام المنفَّذة. وتُحسب دقة استرجاع الذاكرة من السجلات المختارة، والامتناع من جميع حالات الاسترجاع. ونبيّن هذه الخيارات مع النتائج لأن تغيير المقام قد يغيّر تقدير القارئ للأداء تغييرًا كبيرًا.

تحفظ حزمة الأدلة ملخصات تجارب السياسة والمجلس لكل بذرة عشوائية، إلى جانب سجلات القرارات الخام. وتصف فواصل ويلسون نتائج برنولي المولَّدة؛ ولا تعكس عدم اليقين الناشئ عن مولّد غير ممثل للواقع أو افتراضات مشتركة أو نموذج مختلف أو بيئة تشغيل مستقبلية.

لا يدّعي هذا الإصدار وجود دلالة إحصائية لاختبار فرضية. تعرض التجارب عمل الآليات بتوزيعات محددة سلفًا. ويخفّض توليد حالات كثيرة ضوضاء محاكاة مونتي كارلو، لكنه لا يعوّض الظواهر الواقعية الغائبة. وإضافة حالات من المولّد نفسه لا تثبت صلاحية النتائج خارج التجربة.

ينبغي أن تعلن الدراسات المقبلة النتيجة الأساسية ومسوّغ حجم العينة قبل إجراء التقييم النهائي. وتُستخدم في المقارنة الحالات نفسها حيثما أمكن، مع ترتيب عشوائي لشروط الاختبار وتقديرات لعدم اليقين بطريقة البوتستراب أو بطريقة هرمية تلائم بنية المهمة. وينبغي الاحتفاظ بالمحاولات المتكررة للنموذج، بما فيها الإخفاقات وانتهاء المهل وحالات الامتناع.

الالتزام بتقديم التقارير

اعرض الأعداد والتغطية والأخطاء معًا. ميّز التحليلات الاستكشافية من التقييم النهائي الذي ثُبّتت شروطه. احتفظ بالشروط التي لم تنجح، ووثّق سبب استبعاد أي حالة قبل عرض النتائج المجمّعة.

10 · STATISTICAL DISCIPLINE

Keep denominators, seeds and uncertainty visible.

Every result should make it possible to reconstruct both the numerator and the population being described.

Execution coverage uses all tasks as its denominator. Policy-violation rate and proposal correctness use executed tasks. Memory precision uses selected records, while abstention uses all retrieval situations. These choices are stated alongside the results because changing the denominator can materially change the impression of performance.

The evidence package retains per-seed summaries for policy and council experiments, along with raw decision records. Wilson intervals describe the generated Bernoulli outcomes. They do not capture uncertainty from an unrepresentative generator, shared assumptions, a different model or a future deployment environment.

No hypothesis-test significance claim is made in this edition. The experiments are mechanism demonstrations with deliberately specified distributions. The use of many generated cases reduces Monte Carlo noise but cannot compensate for missing real-world phenomena. Additional cases from the same generator are not a substitute for external validity.

Future studies should declare a primary outcome and sample-size rationale before running a final evaluation. Comparisons should use the same cases where possible, randomised condition ordering and bootstrap or hierarchical uncertainty estimates appropriate to the task structure. Repeated model trials should be retained, including failures, timeouts and abstentions.

REPORTING COMMITMENT

Report counts, coverage and error together. Separate exploratory analyses from a frozen final evaluation. Preserve unsuccessful conditions and describe any excluded cases with a reason recorded before aggregate reporting.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 46 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

10 - تحليلات مستقبلية

المختبر المفيد يجعل الإخفاق قابلًا للتشخيص.

يربط تصنيف الأخطاء كل إخفاق مرصود بتجربة تستقصي سببه أو بإجراء يعالجه.

تكاليف الموظفين	برمجيات غير قابلة للاستمرار	الاستثمار في المستقبل
غموض المدخلات	تفسيران معقولان للمهمة	آلية طلب التوضيح واختبار فهم المقصود
خلل في الأدلة	حقائق ناقصة أو متقادمة أو متعارضة	فحص صلاحية المصدر وتتبّع مسار الاسترجاع
المهارة غير ملائمة للمهمة	الدور صحيح لكن طريقة أداء المهمة غير ملائمة	شروط اعتماد الدور وحالات اختبار كفاءته
إخفاق التخطيط	تسلسل إجراءات لا تسنده الأدلة	التحقق من الاعتماديات والشروط المسبقة
فشل التنسيق	يكرر الوكلاء الافتراض الخاطئ نفسه	اختبار استقلال الآراء الأولى وقياس تداخل الأخطاء
إخفاق في ضبط الصلاحيات	يتجاوز الإجراء نطاق الصلاحيات الممنوحة	اختبار فرض الأذونات عند تشغيل الأداة
الفشل في التنفيذ	انتهاء مهلة الانتظار أو حدوث جزء فقط من الأثر الخارجي	منع تكرار أثر العملية ومطابقة الحالة المسجّلة لما حدث فعليًا
الفشل في التحقق	أُبلغ عن النجاح ولم تُؤكَّد الحالة الناتجة	إعادة قراءة الحالة وتطبيق معايير الاكتمال
فشل الذاكرة	يُعامل درس لم يُختبر على أنه موثوق	ضوابط اعتماد الدروس واختبارات عدم تراجع الأداء

ينبغي تطبيق تصنيف الإخفاقات على المسار الكامل للمهمة، لا الإجابة النهائية وحدها. وقد تسهم عدة فئات في نتيجة واحدة؛ فمثلًا قد تقود الأدلة القديمة إلى خطة سيئة، ثم يسمح غياب الضبط بتنفيذها.

ينبغي لكل سجل حادثة أن يميّز بين الوقائع المرصودة والسبب المقترح. ويُفترض أن ينتج عن الإجراء التصحيحي اختبار جديد أو تغيير قابل للقياس في أحد الضوابط. ولا يثبت تكرار المهمة إلى أن تنجح أن المشكلة الأصلية عولجت.

تفصل الدراسة الاصطناعية عمدًا بين الالتزام بالسياسة وصحة المقترح. وهذا أول حدّ تشخيصي فيها. وينبغي لبرنامج تجريبي أوسع أن يضيف إخفاقات الاسترجاع والنموذج والأدوات والتنظيم، مع الحفاظ على القواعد نفسها في توثيق الأدلة.

10 · FAILURE ANALYSIS

A useful lab makes failure diagnosable.

An error taxonomy connects an observed failure to a targeted experiment or corrective action.

FAILURE CLASS	OBSERVABLE SYMPTOM	NEXT INVESTIGATION
Input ambiguity	Two plausible task interpretations	Clarification protocol and intent test
Evidence failure	Missing, stale or contradictory facts	Source validity and retrieval trace
Skill mismatch	Correct role label, unsuitable method	Role admission and competency cases
Planning failure	Unsupported action sequence	Dependency and precondition checks
Coordination failure	Agents repeat a common wrong premise	Independent-first and error-overlap tests
Authority failure	Action exceeds allowed capability	Tool-layer authorisation test
Execution failure	Timeout or partial external effect	Idempotency and state reconciliation
Verification failure	A reported success has no confirmed state	Read-back and completion criteria
Memory failure	An untested lesson becomes trusted	Promotion and regression controls

The taxonomy should be applied to complete task traces rather than only final answers. More than one failure class can contribute to an outcome. For example, stale evidence may produce a bad plan, while a missing control allows the plan to execute.

Every incident record should distinguish observed facts from a proposed cause. The corrective action should produce a new test or a measurable control change. Repeating the same task until it succeeds is not sufficient evidence that the underlying problem has been fixed.

The synthetic study deliberately separates policy and proposal correctness. This is the first diagnostic boundary. A broader experimental programme should add retrieval, model, tool and organisational failures while preserving the same evidence discipline.

جارڤيس وسيستم برين

بروتوكول متكامل لتقييم نموذج فعلي.

في المرحلة التالية تُستبدل وسوم المقترحات المولّدة بمخرجات نموذج فعلية، على مجموعة مهام ثُبّتت قبل التشغيل.

تُبنى مجموعة المهام من مواصفات سير عمل مأذون بها وحالات تُكتب بصورة مستقلة. وتُحدَّد لكل حالة نقطة البداية والأدلة المسموحة والأدوات المتاحة وشروط النتيجة النهائية المتوقعة والآثار المحظورة. وتُفصل حالات التطوير عن التقييم النهائي؛ ولا تُعاد صياغة الحالات النهائية لمصلحة نظام مرشّح.

قارن بين خط أساس حتمي ووكيل بدور واحد والنموذج الأولي الحالي والبنية الخلوية المقترحة. استخدم الأدلة نفسها ومحاكي الأدوات نفسه حيثما أمكن. وسجّل معرّف النموذج وإصدار التعليمات وإصدارات المهارات وإعدادات أخذ العينات وسياسة إعادة المحاولة والميزانية الكلية. وإذا تغير سلوك المورّد فوثّق تاريخ التقييم واحتفظ بالمخرجات.

كرّر كل حالة عددًا كافيًا من المرات لتقدير الموثوقية، وبرّر حجم العينة وفق القرار المزمع اتخاذه. أدرج أعطال الأدوات وغياب الأدلة وتعارض التعليمات والمحتوى غير ذي الصلة. وزّع ترتيب الشروط عشوائيًا لتقليل أثر التوقيت. وينبغي أن يقيس المؤشر الأساسي الحالة النهائية الفعلية والالتزام بالسياسة، لا تصريح النموذج بأنه نجح.

ينبغي تسجيل استهلاك وحدات النص والتكلفة وزمن الاستجابة وحالات الامتناع وعدد التدخلات وفئات الأخطاء. ويفحص المراجع عينة من حالات النجاح والإخفاق معًا، لأن نتيجة نهائية تبدو صحيحة قد يسبقها مسار غير مسند أو غير مأذون به.

البوابة الإلكترونية	رابعا - المعلومات المطلوبة
جودة النتائج	تصحيح مُعمّى أو برمجي وفق معايير مستقلة.
الموثوقية التشغيلية	النجاح عبر المحاولات المتكررة والتعافي من الإخفاق ضمن حدود محددة
استناد الادعاءات إلى أدلة	التحقق من سند كل ادعاء على حدة
الكفاءة في استخدام الموارد	التكلفة وزمن الاستجابة لكل مهمة ثبت اكتمالها
السياق البحثي [7] Liu	[8] Yao [9] Mialon

11 · NEXT EMPIRICAL STUDIES

A complete live-model evaluation protocol.

The next phase replaces generated proposal labels with actual model outputs on a frozen task suite.

Construct a task set from authorised workflow specifications and independently written cases. Each case should include the initial state, permitted evidence, available tools, expected final-state conditions and prohibited effects. Keep a development subset separate from the final evaluation; do not rewrite final cases to favour a candidate system.

Compare a deterministic baseline, a single-role agent, the existing prototype and the proposed cell-based architecture. Use the same evidence and tool simulator where possible. Record model identifier, prompt version, skill versions, sampling settings, retry policy and total budget. If a provider changes behaviour, record the evaluation date and retain the outputs.

Repeat each case enough times to estimate reliability, with a sample-size rationale based on the intended decision. Include tool failures, unavailable evidence, conflicting instructions and irrelevant content. Randomise condition order to reduce time-dependent effects. The primary outcome should evaluate the actual final state and policy compliance, not a model's self-reported success.

The evaluation should retain token usage, cost, latency, abstentions, intervention counts and error categories. A reviewer should inspect a sample of successes as well as failures, because an apparently correct final answer may rest on an unsupported or unauthorised path.

PROSPECTIVE GATE	REQUIRED EVIDENCE
Outcome quality	Blinded or programmatic scoring against independent criteria
Operational reliability	Repeated-trial success and bounded failure recovery
Evidence grounding	Claim-level support checks
Resource efficiency	Cost and latency per verified completion

Research context [7] Liu [8] Yao [9] Mialon

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 48 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

11 - الدراسات البيئية الوطنية

تصميم مقارنة عادلة مع المتخصصين.

يتطلب الطموح إلى مضاهاة مائة متخصص دراسة خبراء ذات هدف مقارنة دقيق.

عند اختيار المشاركين، يجب تحديد الخبرة المطلوبة وتعارض المصالح وشروط الأهلية لكل مهمة. لا يصح تصوير لجنة المتخصصين كأن أعضاءها المئة مؤهلون بالدرجة نفسها في كل مجال. قد تحتاج مهمة إلى قلة من الخبراء تتكامل تخصصاتهم، وأخرى إلى مجموعة أكبر تُجري تقييمات مستقلة. ويجب إعلان التشكيل المختار قبل جمع النتائج.

ينبغي أن يتلقى المشاركون وصفًا موحدًا للمهمة وحزمة أدلة متسقة. وتُسجل المدة المتاحة والأدوات التي يمكن استخدامها وما إذا كان التشاور مسموحًا. ويمكن قياس الأحكام الفردية الأولى وأحكام اللجنة النهائية كلًا على حدة، بما يكشف قيمة التنسيق البشري وتكلفته.

ينبغي حجب هوية مصدر مخرجات الذكاء الاصطناعي عن المقيّمين حيثما أمكن. ويحتاج المراجعون إلى معايير تحدد النتائج المقبولة والأخطاء الجسيمة وطريقة التعامل مع عدم اليقين. ويُعلن مدى الاتفاق بين المقيّمين وما بقي من خلاف. ولا تُمنح إجابة درجة لمجرد أن أسلوبها يبدو مهنيًا.

ينبغي حصر أي ادعاء بمعادلة أداء لجنة خبراء في نوع المهام الذي اختُبر، وشروط الأدلة، والميزانية المستخدمة. فالنجاح في تسوية روتينية لا يثبت خبرة عامة في العلوم أو الطب أو الاستراتيجية أو الهندسة. وينبغي الاحتفاظ بالنتائج السلبية وتحديد المجالات التي يتفوق فيها البشر على النظام.

الحالة الراهنة

هذا بروتوكول مقترح للمستقبل. لم يُستقطب متخصصون بشر ولم يُختبروا لهذه الطبعة، ولا تُعرض فيها نتيجة تثبت التكافؤ معهم.

11 · NEXT EMPIRICAL STUDIES

Design the specialist comparison fairly.

The one-hundred-specialist ambition needs an expert study with a precise comparison target.

Recruitment should define relevant expertise, conflicts of interest and task eligibility. A specialist panel should not be represented as one hundred equally qualified people in every domain. Some tasks may require a small set of complementary experts; others may use a larger independent-rating pool. The chosen structure must be stated before results are collected.

Participants should receive a consistent task brief and evidence pack. The study should record the time available, tool access and whether deliberation is permitted. Independent first judgements and final panel judgements can be measured separately. This reveals both the value and the cost of human coordination.

AI outputs should be anonymised for scoring where feasible. Reviewers need a rubric with acceptable outcomes, critical errors and the treatment of uncertainty. Inter-rater agreement and unresolved disagreement should be reported. An answer that sounds professional should not receive credit merely for style.

Any claim of matching a specialist panel should be limited to the tested task family, evidence conditions and budget. A result in routine reconciliation cannot establish general expertise across science, medicine, strategy or engineering. The study should retain negative findings and identify where humans outperform the system.

CURRENT STATUS

This is a prospective protocol. No human specialists were recruited or tested for this edition, and no equivalence result is reported.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

11 - الدراسات البيئية الوطنية

التحقق من الاقتصار على الإنجليزية في مسار التفاعل كاملًا.

يشمل شرط اللغة التجربة كاملة، بما فيها رسائل الخطأ والمخرجات المنطوقة.

ينبغي أن تشمل مجموعة الاختبارات النصية أسئلة بالإنجليزية ومدخلات متعددة اللغات ونصوص مصادر مقتبسة بلغات أخرى وأسماء ملتبسة وتعليمات تحاول تغيير لغة الرد. والسلوك الصحيح هو الإجابة أو طلب التوضيح بالإنجليزية، مع الحفاظ على معنى المصادر وبيان مواضع عدم اليقين في الترجمة عند الحاجة.

عند تقييم الواجهة الصوتية، ينبغي قياس دقة تحويل الكلام إلى نص وفهم المقصود ولغة الإجابة ووضوح النطق كلًّا على حدة. وتُختبر اللهجات والضوضاء المحيطة وانقطاع الحديث ورفض إذن الميكروفون في ظروف منفصلة. فقد يتحدث النظام بإنجليزية فصيحة ثم ينفّذ تعليمات حرّفها التفريغ الصوتي؛ لذلك لا تكفي جودة الصوت للحكم عليه.

قبل الموافقة على الإجراءات ذات العواقب المهمة، ينبغي للنظام عرض التعليمات كما فهمها أو ترديدها. وتستحق الكميات والتواريخ ومعرّفات الحسابات والوجهات فحصًا خاصًا. وعند الحاجة ينبغي إيقاف الاستماع أثناء الرد الصوتي، كيلا يفسّر النظام صوته بوصفه تعليمات جديدة.

ينبغي أن يوثّق سجل الاختبار إصدارات المتصفح ومحرك النطق وإعدادات الميكروفون ومصدر عينة الإدخال والبدائل النصية الميسّرة. ولا تُحفظ تسجيلات صوتية لمستخدمين فعليين دون غرض محدد وإذن مناسب. ويفحص هذا الإصدار مسارات اللغة الأقدم؛ ولم تُجرَ دراسة صوتية بمشاركين بشريين.

فئة الاختبار	النتائج المتوقعة
مدخل بلغة غير الإنجليزية	إجابة بالإنجليزية تحافظ على المعنى
تعليمات وردت في اقتباس من مصدر	لا تغيير في قواعد اللغة أو الصلاحيات.
كمية غير مؤكدة	توضيح قبل اتخاذ إجراء
خدمة الكلام غير متاحة	إمكان إكمال المهمة بالنص
انقطاع المهمة أو إعادة المحاولة	عدم تكرار الفعل

11 · NEXT EMPIRICAL STUDIES

Verify English-only interaction end to end.

The language requirement covers the entire experience, including error messages and spoken output.

The text suite should include English questions, multilingual input, quoted non-English source material, ambiguous names and instructions attempting to change the response language. Correct behaviour is an English answer or clarification, while preserving source meaning and identifying translation uncertainty where relevant.

Voice evaluation should separately measure transcription accuracy, intent recovery, response language and spoken intelligibility. Accents, background noise, interruptions and microphone denial are distinct test conditions. A system may produce fluent English speech while acting on an incorrectly transcribed instruction, so voice quality alone is insufficient.

For consequential actions, the system should display or repeat the interpreted instruction before the relevant approval step. Numeric quantities, dates, account identifiers and destinations deserve targeted checks. The audio path should pause listening while speaking where necessary to avoid interpreting its own answer as a new instruction.

The test record should include browser and speech-engine versions, microphone setup, input sample provenance and accessible text alternatives. No real user audio should be retained without a defined purpose and appropriate permission. The present edition inspects the older language paths but does not run a human voice study.

TEST CATEGORY	EXPECTED OUTCOME
Non-English input	English answer with faithful meaning
Quoted source instruction	No change to language or authority policy
Uncertain quantity	Clarification before action
Speech service unavailable	Usable text workflow
Interruption or retry	No duplicate action

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

11 - الدراسات البيئية الوطنية

الحفاظ على الكفاءة مع تغيّر الظروف.

ينبغي أن يُقاس أداء موظف الذكاء الاصطناعي بسجل ممتد عبر الزمن، لا بلقب دائم يُمنح له بعد اختبار ناجح واحد.

ينبغي للدراسة الطولية تحديد فترات تقييم دورية والإبقاء على مجموعة مرجعية ثابتة من الحالات. وتمثّل الحالات الجديدة ظروف العمل المتغيّرة، بينما تكشف المجموعة الثابتة تراجع الأداء. ويجب فصل النتائج بحسب الدور وفئة المهمة ونسخة الأداة وجودة الأدلة، بدلًا من اختزالها في درجة عالمية واحدة.

تقترح عملية التحديث إنشاء نسخة مرشّحة من المهارة أو الذاكرة، واختبارها بحالات جديدة وأخرى محفوظة، ثم مقارنتها بالنسخة الحالية. ولا تُعتمد إلا وفق معيار جودة محدد سلفًا ومن دون تراجع غير مقبول. وإذا انخفض الأداء بعد التشغيل، ينبغي الرجوع إلى الإصدار السابق أو تقليص استقلالية النظام، مع الاحتفاظ بالنسخة التي أخفقت لتحليلها.

من مؤشرات تغيّر نمط العمل زيادة الحقول الإلزامية الناقصة، أو ظهور أنواع جديدة من الاستثناءات، أو تراجع توافق الأدلة، أو إطالة زمن استجابة الأدوات، أو تغيّر توزيع المهام المقبولة. تستحق هذه الإشارات التحقيق، لكنها لا تحدد السبب بمفردها ولا تثبت أن إعادة التدريب هي العلاج المناسب.

يمكن للصيانة الذاتية أتمتة المراقبة وتوليد البدائل المرشّحة وتقييمها ضمن حدود. أما توسيع الصلاحيات أو إزالة الضوابط أو إعادة تعريف النجاح فيتطلب تفويضًا منفصلًا. وينبغي قياس ما إذا كانت دورة الصيانة تحسّن الأداء في حالات لم تدخل في التطوير، وما إذا كانت تدخلاتها تُحدث أنماط إخفاق جديدة.

السجل التنافسي

نسخة الدور · فئات المهام المقبولة · تاريخ التقييم · مجموعة الأدلة · الجودة والتغطية · الحدود المعروفة · سجل التغييرات · سبب المراجعة المقبلة.

11 · NEXT EMPIRICAL STUDIES

Maintain competence as conditions change.

An AI employee's performance should be a measured history, not a permanent label awarded after one successful test.

A longitudinal study should define recurring evaluation windows and preserve a stable anchor set of cases. New cases represent changing business conditions, while the anchor set reveals regression. Results should be separated by role, task family, tool version and evidence quality rather than collapsed into one global score.

The proposed update process creates a candidate skill or memory version, evaluates it against both new and retained cases, and compares it with the current version. Promotion should require a predefined quality rule and no unacceptable regression. If performance degrades after deployment, the system should revert or reduce autonomy while retaining the failed candidate for analysis.

Drift indicators can include a rise in missing required fields, new exception categories, lower evidence agreement, longer tool delays or a change in the distribution of accepted tasks. These signals warrant investigation; they do not by themselves identify the cause or prove that retraining is the correct response.

Self-maintenance can automate monitoring, candidate generation and bounded evaluation. Authority to expand permissions, remove controls or redefine success remains separate. The study should measure whether the maintenance loop improves held-out performance and whether its interventions create new failure modes.

COMPETENCY RECORD

Role version · accepted task families · evaluation date · evidence set · quality and coverage · known limits · change history · next review trigger.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 51 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

12 · مثال تصميمي محلول

كيف يتغير قرار إعادة التزويد بالمخزون؟ مثال خطوة بخطوة

مثال توضيحي فقط. الأعداد الآتية افتراضية، والغرض منها شرح مسار العمل المقترح.

الخطوة	الحالة السياسية والمنكوبة
1 · الرصد	لدى المستودع A رصيد قدره 40 وحدة. والطلب المؤكد 12 وحدة يوميًا، ومهلة توريد المورّد المسجّلة خمسة أيام.
2 · التحقق من المعطيات	عمر سجل المخزون ساعتان، وعمر سجل مهلة توريد المورّد ست وثلاثون ساعة.
3 · تقدير الموقف	يحدّد الدور ضغط المخزون واعتمادًا على معلومة قديمة، ويسجّل أن القرار يحتاج إلى أدلة محدّثة.
4 · التشاور	يطلب دورا المخزون والمشتريات تقديرًا حديثًا من المورّد، ويتحقق دور المالية من الميزانية المسموح بها.
5 · تقديم المقترح	بعد ورود أدلة حديثة، يحسب المخطِّط الكميات المقترحة في حدود صلاحيات الدور.
6 · تطبيق الضوابط	تتحقق بوابة التنفيذ من الإذن بالتعامل مع المورّد والمبلغ وسابقة تنفيذ الطلب والموافقة المطلوبة.
7 · التنفيذ	لا يُرسل إلا طلب مأذون به ومحدد النطاق؛ وفي غير ذلك يطلب النظام المراجعة.
8 · التحقق من النتيجة	يُتحقّق من مرجع الطلب والكمية الفعلية قبل وصف المهمة بأنها مكتملة.

لا تقتصر الفائدة على إنتاج كمية بسرعة؛ بل تشمل معرفة الافتراضات التي تحدد تلك الكمية وعدم إخفاء أي اعتماد لم يُحسم. ولا يكون اضطراب الطقس ذا صلة إلا إذا انطبق على مسار المورد وكانت له مدة صلاحية مناسبة.

بعد تلبية الطلب، تسجّل طبقة الذاكرة مهلة التوريد التي تحققت فعلًا وتربطها بالتوصية الأصلية. وقد يقترح النظام درسًا عن تفاوت أداء المورّد؛ ويجب تقييمه في حالات إضافية قبل اعتماده قاعدةً للشراء. ونجاح عملية تسليم

واحدة لا يؤسس توقعًا عامًا.

12 · WORKED DESIGN EXAMPLE

A changing inventory decision, step by step.

Illustrative scenario only. The quantities below are invented to explain the proposed workflow.

STEP	SYNTHETIC SITUATION AND EXPECTED BEHAVIOUR
1 · Observe	Warehouse A has 40 units. Confirmed demand is 12 units/day. Supplier lead time is recorded as five days.
2 · Validate	The inventory record is two hours old; the supplier lead-time record is thirty-six hours old.
3 · Appraise	The role identifies stock pressure and a stale dependency. It marks the decision as needing updated evidence.
4 · Consult	Inventory and procurement roles request a current supplier estimate. Finance checks the permitted budget.
5 · Propose	After fresh evidence arrives, the planner calculates candidate quantities under the role's limits.
6 · Control	The gate checks supplier permission, amount, duplicate status and approval requirements.
7 · Execute	Only an authorised, bounded order can be submitted; otherwise the system requests review.
8 · Verify	The order reference and actual quantity are checked before the task is called complete.

The useful behaviour is not simply producing a fast quantity. It is recognising which assumptions control that quantity and refusing to hide an unresolved dependency. A weather disruption could be relevant only if it applies to the supplier route and has a valid time horizon.

After fulfilment, the memory layer records the observed lead time and links it to the original recommendation. A proposed lesson might concern supplier variability. It should be evaluated across additional cases before becoming an approved purchasing rule. One successful delivery does not establish a universal forecast.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

12 - العقود القابلة للاشتراك

قائمة مواصفات الدور والمهارات.

يقترح البحث عقدًا قابلًا للقراءة الآلية يتيح فحص قدرات الموظف الاصطناعي قبل تفعيله.

{ "role_id": "inventory-analyst", "version": "candidate-1", "language": "en", "scope": ["warehouse-A"], "skills": [{ "id": "replenishment-review", "version": "1.0", "source_digest": "sha256:...", "status": "approved", "evaluation_record": "eval:..." }], "tools": ["inventory.read", "purchase.draft"], "memory_policy": "approved-current-own-scope", "autonomy": "recommend", "owner_role": "operations-manager" }

هذا المخطّط للتوضيح، ولا يضم بيانات دخول فعلية أو بصمة حقيقية أو معرّف تقييم حقيقيًا. ينبغي لمدقّق النشر رفض الأدوات المجهولة ومراجع الموافقات الناقصة والإصدارات غير المتوافقة. ولا يمكن لبيان يطلب صلاحيةً أن يمنحها لنفسه.

ينبغي أن تذكر كل مهارة تاريخ سريانها والإصدارات التي حلّت محلها. ويحتفظ الدور بإحالة إلى الحزمة المحددة التي جرى تقييمها، لا إلى رابط قابل للتغيير. وتوثّق حقوق الحزمة ومنشأ مصدرها في سجل القبول.

يسهّل ملف تعريف الدور نقله بين البيئات من دون نسخ صلاحيات غير محدودة. ويمكن لبيئتي الاختبار والإنتاج استخدام التعريف نفسه، مع موصلات أدوات مختلفة وبيانات اعتماد جرت الموافقة عليها بصورة منفصلة. وعلى التقييم التحقق من أن الموصلات تفرض النطاق المعلن.

12 · IMPLEMENTABLE CONTRACTS

The role and skill manifest.

A proposed machine-readable contract makes an AI employee's capabilities reviewable before activation.

```
{
  "role_id": "inventory-analyst",
  "version": "candidate-1",
  "language": "en",
  "scope": ["warehouse-A"],
  "skills": [{
    "id": "replenishment-review",
    "version": "1.0",
    "source_digest": "sha256:...",
    "status": "approved",
    "evaluation_record": "eval:..."
  }],
  "tools": ["inventory.read", "purchase.draft"],
  "memory_policy": "approved-current-own-scope",
  "autonomy": "recommend",
  "owner_role": "operations-manager"
}
```

This schema is illustrative and does not contain a real credential, digest or evaluation identifier. The deployment validator should reject unknown tools, missing approval references and incompatible versions. A manifest that requests a capability cannot approve that capability for itself.

Skills should identify their effective date and superseded versions. The role should retain a reference to the exact package evaluated, rather than a mutable URL that can change later. A package's rights and source provenance belong in the admission record.

A manifest also makes a role portable across environments without copying unrestricted access. Test and production environments can use the same role definition with different tool adapters and separately approved credentials. The evaluation should confirm that the adapters enforce the declared scope.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 53 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

سجل القرارات والذاكرة.

يربط سجل موجز التوصيةَ بالمعلومات والصلاحيات والنتيجة المرصودة التي تستند إليها.

المجموعة الميدانية	المعلومات المطلوبة
الهوية	معرّف المهمة وإصدار الدور والجهة صاحبة البيانات وهوية مقدم الطلب الموثّقة
الأدلة	معرّفات المصادر وبصماتها حيث يلزم والطوابع الزمنية والصلاحية.
المقترح	نوع الإجراء ومعلماته وبدائله ومبرره الموجز
الضبط	نسخة القواعد، ونتيجة فحص الإذن، ومستوى المخاطر، ومرجع الموافقة.
التنفيذ	موصّل الأداة ومفتاح منع تكرار الأثر وحالة الطلب والاستجابة المرصودة
التحقق	الحالة المتوقعة، والحالة المرصودة، والنتيجة، والاختلاف الذي لم يُحسم.
التعلم	الدرس المقترح والنتائج الداعمة وحالة المراجعة والإصدار المرشّح.

يجب ألا يحتفظ سجل القرار بالأسرار المتضمنة في طلبات الأدوات أو بنسخ غير مقيّدة من محتوى العملاء. ويمكن للمراجع والحقول المنتقاة أن تحفظ قابلية التتبع مع تقليل الكشف غير الضروري. ويجب أن يخضع الوصول إلى السجل لنطاق الجهة صاحبة المهمة ودرجة حساسية بياناتها.

ينبغي تحديد نوع المعلومة صراحةً قبل إدخالها في الذاكرة. فيُحفظ الإخفاق المرصود كواقعة، والتفسير المولّد كفرضية، والتصحيح المعتمد كمعلومة إجرائية. وينبغي توثيق تاريخ كل انتقال بين هذه الأنواع.

ينبغي كذلك تسجيل غياب القرار. فالمهمة الممنوعة أو المؤجلة أو المتروكة ذات قيمة تحليلية لأنها تكشف حدود الأتمتة الحالية. واستبعادها يوحي بأن الدور أقدر مما تدعمه سجلات تشغيله الفعلية.

ينبغي أن يتمكن المراجع من تحديد نسخة الأدلة والإعدادات ونتيجة فحص الضوابط وقت القرار. أما تشغيل نموذج حالي على بيانات اليوم فهو تجربة جديدة، لا إعادة إنتاج للقرار الأصلي.

12 · IMPLEMENTABLE CONTRACTS

The decision and memory record.

A compact record links a recommendation to the information, authority and observed outcome behind it.

FIELD GROUP	REQUIRED INFORMATION
Identity	Task ID, role version, tenant, authenticated requester
Evidence	Source IDs, digests where appropriate, timestamps and validity
Proposal	Action type, parameters, alternatives and concise rationale
Control	Policy version, permission result, risk tier and approval reference
Execution	Tool adapter, idempotency key, request status and observed response
Verification	Expected state, observed state, result and unresolved discrepancy
Learning	Proposed lesson, supporting outcomes, review state and candidate version

The decision record should not store secrets embedded in tool requests or unrestricted copies of customer content. References and carefully selected fields can preserve traceability while reducing unnecessary exposure. Access to the record should follow the task's tenant and sensitivity boundaries.

Memory admission should require an explicit type. An observed failure can be stored as an episode; a generated explanation can be stored as a hypothesis; an approved correction can become a procedural memory. Promotion between these types should leave a history.

The system should also record the absence of a decision. A blocked, deferred or abandoned task has analytical value because it reveals the limits of current automation. Excluding these records would make the role appear more capable than its actual operating history supports.

REPRODUCIBILITY RULE

A reviewer should be able to identify the evidence snapshot, configuration and control result used at the time. Re-running a current model against today's data is a new experiment, not a reproduction of the original decision.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 54 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

13 - برنامج التنمية

مراحل تنفيذ ذات شروط إنجاز قابلة للقياس.

يتقدم البحث عبر إصدارات تضيف قدرات مدروسة، ولا يستهدف عددًا محددًا من الخصائص.

المحطة	البناء والاستثمار	دليل اجتياز المرحلة
M1 · خط أساس بحثي مستقر	تثبيت قائمة ملفات المصدر، وتوحيد سلوك النظام عند التخاطب بالإنجليزية، والاحتفاظ بالاختبارات القائمة.	خط أساس محدّد الإصدار واختبار قبول المسار اللغوي
M2 · مواصفات الأدوار	حزم مهام محددة الأنواع، وأدوات محدودة الصلاحيات، وقرارات منظّمة، وحالات مصرّح بها بوضوح.	اختبارات الصلاحيات وإعادة المحاولة ومسارات الإخفاق
M3 · الذاكرة الخاضعة للحوكمة	المصدر، وانتهاء الصلاحية، وحالة المراجعة، وروابط التصحيحات.	اختبار معياري للاسترجاع وقبول الأدلة عند إفساد البيانات الوصفية
M4 · مجالس تنسيق خضعت للتقييم	توصيات أولى تُعدّ باستقلال، ثم تُجمع ضمن حدود معلنة	المقارنة بأساليب مرجعية أبسط على الحالات نفسها
M5 · تجربة داخل قسم	تشغيل مسار عمل واحد قابل للتراجع بالموازاة دون تنفيذ فعلي، ثم تفعيل التنفيذ ضمن نطاق محدود	جودة المهام وتغطيتها وإمكان التعافي وتكلفتها
M6 · التكيّف المنضبط	دروس مرشّحة للاعتماد، واختبارات محجوزة للتقييم، وإمكان الرجوع إلى إصدار سابق.	استمرار الفائدة بمرور الوقت من دون تراجع غير مقبول في الأداء

هذا التسلسل خطة بحثية، وليس جدول تسليم مضمونًا. ويتوقف التقدم على الأدلة وإمكان الوصول إلى البيانات وجاهزية الأدوات والقدرة على المراجعة. وقد تكشف إحدى المراحل أن تصميمًا أبسط أكثر فاعلية؛ وعندئذ ينبغي تعديل البرنامج بدلًا من الإبقاء على التعقيد لذاته.

تنتج كل مرحلة مادة يمكن توثيقها: إعدادًا، أو مجموعة بيانات، أو سجل تقييم، أو دراسة تشغيلية. ويربط ذلك المشروع بغرض المختبر في حفظ الأبحاث والأدلة التجريبية، كما يتيح للباحثين فحص نسخة محددة بدلًا من تقييم

هدف يتغيّر دون تعريف ثابت.

13 · DEVELOPMENT PROGRAMME

An implementation sequence with measurable exits.

The research proceeds through versions of capability, rather than a numerical feature target.

MILESTONE	BUILD AND INVESTIGATE	EXIT EVIDENCE
M1 · Stable research baseline	Freeze source inventory; unify English-language behaviour; retain existing tests	Versioned baseline and language-path acceptance
M2 · Role contracts	Typed task envelopes, scoped tools, structured decisions and explicit states	Permission, retry and failure-path tests
M3 · Governed memory	Provenance, expiry, review state and correction links	Retrieval/admission benchmark with corrupted metadata
M4 · Evaluated councils	Independent-first contributions and bounded synthesis	Shared-case comparison with simpler baselines
M5 · Department pilot	One reversible workflow in shadow mode, then bounded execution	Task-level quality, coverage, recovery and cost
M6 · Controlled adaptation	Candidate lessons, held-out tests and version rollback	Longitudinal benefit without unacceptable regression

The sequence is a research plan, not a promised delivery calendar. Progress depends on evidence, data access, tool readiness and review capacity. A milestone may reveal that a simpler design is more effective, in which case the programme should adapt rather than preserve architectural complexity for its own sake.

Each milestone produces a reportable artifact: a configuration, dataset, evaluation record or operational study. This aligns the project with the lab's purpose as a place to retain research and experimental evidence. It also allows other researchers to inspect a specific version instead of evaluating an undefined moving target.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 55 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

13 - برنامج التأهيل

قرارات الإصدار تحتاج إلى أدلة صريحة.

تمثّل الشروط الآتية متطلبات تشغيل مقترحة، ولا يفيد إدراجها بأن النظام كاملًا قد اجتازها.

بوابة التحقق	REQUIRED EVIDENCE قبل BROADER AUTONOMY
اللغة والتفاعل	إجابات نصية ومنطوقة بالإنجليزية في حالات محددة للإدخال والخطأ.
كفاءة الدور	مهام لم تُستخدم في التطوير، ومرجع مفيد للمقارنة، ومراجعة موثّقة للأخطاء
إنفاذ السياسات	اختبار رفض العمليات غير المأذون بها، ومنع التكرار، والتحقق من الموافقات
جودة الأدلة	حداثة الدليل ونطاقه ومنشؤه وطريقة حل التعارضات
موثوقية التنفيذ	اختبار أعطال متعمّدة للأدوات، والتعافي من النتائج الجزئية، والتحقق من حالة التنفيذ
صيانة الذاكرة	عزل التحديثات المرشحة، واختبار عدم تراجع الأداء، وإمكان العودة إلى الإصدار السابق
الجهة المسؤولة عن التشغيل	مسؤول محدد بالاسم، واستثناءات مراقبة، وأدوات إيقاف تعمل فعليًا.
نزاهة البحوث	مواد بحثية ذات نسخ محددة، وادعاءات قابلة للتتبع، ومراجعة يتناسب نطاقها مع الأدلة.

ينبغي اختيار العتبات العددية وفق العواقب الفعلية لسير العمل وأدائه المرجعي. لا تكفي عبارة «دقة 95%» إذا تركزت الأخطاء المتبقية في الحالات الحرجة. وعلى سجل القبول أن يحدّد أنواع الإخفاق غير المسموح بها إلى جانب أهداف متوسط الجودة.

في تجربة أولية منخفضة المخاطر، يمكن البدء بمعايير أهلية محدودة وسقف تنفيذ متحفّظ. ويتطلب التوسّع أدلة تخص أنواع الحالات الجديدة ومقارنةً بالإجراء القائم. ويجب أن يستدعي وقوع حادث جسيم أو تغيّر جوهري في

الإعدادات إعادة التقييم.

تربط شروط الانتقال هذه طموح المالك بتقدّم يمكن إثباته. فالهدف إسناد المهام الروتينية المناسبة إلى النظام تدريجيًا، على أن يُثبت كل توسع، ويُمكن التراجع عنه متى أمكن، ويُنسب إلى إعداد خضع للاختبار.

13 · ACCEPTANCE PROGRAMME

Release decisions need explicit evidence.

The following gates are proposed operating requirements; they are not reported as passed for the full system.

GATE	EVIDENCE REQUIRED BEFORE BROADER AUTONOMY
Language and interaction	English responses and speech across defined input/error conditions
Role competence	Unseen task set, meaningful baseline and documented error review
Policy enforcement	Denied-capability tests, duplicate protection and approval checks
Evidence quality	Freshness, scope, provenance and conflict handling
Execution reliability	Tool failure injection, partial-outcome recovery and state verification
Memory maintenance	Candidate quarantine, regression evaluation and rollback
Operational ownership	Named owner, monitored exceptions and functioning pause controls
Research integrity	Versioned artifacts, traceable claims and appropriately scoped review

Numerical thresholds should be chosen for the workflow's actual consequences and baseline performance. A generic "95% accurate" target is inadequate if the remaining errors are concentrated in critical cases. The acceptance record should identify prohibited failure types as well as average quality targets.

For a low-risk pilot, the team can begin with narrow eligibility and conservative execution limits. Expansion requires evidence on new case types and a comparison with the existing process. A serious incident or material configuration change should trigger re-evaluation.

These gates connect the owner's ambition to concrete progress. The objective is to replace suitable routine execution step by step while making each expansion demonstrable, reversible where possible and attributable to a tested configuration.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

14 - المناقشة والاستنتاجات

أساس للبحث واختبار محدد للمرحلة التالية

يقدّم فحص النموذج الأولي وتجربة المحاكاة المنفَّذة نقطة انطلاق لبرنامج بحث تجريبي أوسع.

يتضمن النموذج الأولي مكوّنات تطبيقية للخلايا والمهارات والذاكرة والتفويض والإشارات التشغيلية ومقترحات المجلس. وأكدت اختبارات محددة سلوك الوصول الأساسي، والإبلاغ عن إشارات من بيانات اختبار مُعدّة سلفًا، وحفظ المهارات، والتصريح بعدم توافر الخدمة عند غياب بيانات اعتماد المزوّد. وكشف فحص الشيفرة اختلاف اللغة بين مسار جارفيس الأقدم والواجهة المقترحة التي تقتصر على الإنجليزية.

توضح النتائج الاصطناعية ضرورة الفصل في التصميم بين إنفاذ السياسة وجودة القرار. فقد يطبّق حاجز التحقق الكامل القواعد المبرمجة ويقبل مع ذلك مقترحات خاطئة. ويستطيع الترشيح حسب المصدر فرض شروط واضحة لقبول ما يُسترجع من الذاكرة، لكن نجاحه يتوقف على جودة الوسوم. وتقل فائدة زيادة عدد المصوّتين إذا تشابهت أخطاؤهم.

قيود الدراسة الرئيسية مقصودة وجوهرية: فهي لا تتضمن تشغيل نموذج لغوي فعلي، ولا لجنة خبراء بشرية، ولا دراسة لنتائج الاستخدام في الإنتاج، ولا اختبارًا للتجربة الذاتية. والبيانات المولّدة مبسّطة، وقياس الزمن يقتصر على دالة قواعد داخل العملية البرمجية. وتؤكد إعادة الإنتاج الداخلية الاتساق الحاسوبي ضمن هذا النطاق.

تتمثل الخطوة العلمية التالية في مقارنة ثابتة تستخدم الحالات نفسها ومخرجات نماذج فعلية وبيئة أدوات مضبوطة. وينبغي أن تشمل خط أساس بوكيل واحد، والنموذج الأولي الحالي، والبنية المقترحة، مع توحيد الأدلة والميزانيات ومعايير النجاح. وعندئذ فقط يمكن تبرير ادعاءات أوسع بشأن الكفاءة في الأدوار والعمل الإداري المستقل.

خلاصة البحث

يشكّل جارفيس وسيستم برين منصة ملموسة لدراسة اتخاذ الآلة للقرار. وينبغي إثبات فائدتها بنتائج يمكن تتبّعها ومقارنات مضبوطة ومسؤوليات يبرّر التقييم توسيعها تدريجيًا.

14 · DISCUSSION AND CONCLUSIONS

A foundation for research, with a clear next test.

The inspected prototype and the completed synthetic study establish a useful starting point for a larger empirical programme.

The prototype contains real application structures for cells, skills, memory, delegation, operational signals and council proposals. Selected tests confirm basic access behaviour, fixture-based signal reporting, skill persistence and explicit unavailability without provider credentials. Source inspection also identifies a language mismatch between the older Jarvis path and the proposed English-only interface.

The synthetic results demonstrate why the architecture should separate policy enforcement from decision quality. A complete gate can obey the encoded rules while accepting incorrect proposals. Provenance filtering can enforce clear memory-admission constraints, but its success depends on the quality of those labels. More voters help less when their errors are shared.

The principal limitations are deliberate and material. No live language model, human specialist panel, production outcome study or subjective-experience test is included. The generated data are simplified, and the timing covers only an in-process rule function. Internal reproduction confirms computational consistency within this scope.

The next scientific step is a frozen, shared-case comparison using actual model outputs and a controlled tool environment. It should test a single-agent baseline, the current prototype and the proposed architecture under the same evidence, budgets and success criteria. Only then can the programme justify broader claims about role competence and autonomous departmental work.

RESEARCH CONCLUSION

JARVIS and System Brain are a concrete platform for investigating machine decision-making. Their value should be demonstrated through traceable outcomes, controlled comparisons and progressively earned responsibility.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 57 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

15 - التحقق الداخلي

سجل التحقق التقني في مختبر زيزمو.

السجل ZRL-VR-JSB-2026-001 · 28 سبتمبر 2026 · تحقق تقني آلي ضمن النطاق المذكور.

بند التحقق	النتيجة المسجّلة
اختبارات النموذج الأولي	نجحت 7 اختبارات تضم 32 تحققًا من النتائج في قاعدة SQLite معزولة.
اختبارات المحاكاة	نجح 18 تحققًا من نتائج التجربة في كل تشغيل.
إعادة احتساب القيم العددية	نجح 47 فحصًا لاتساق السجلات الخام والتحليل.
إعادة إنتاج النتائج في عملية تشغيل مستقلة	تطابقت ملفات المخرجات الحتمية الخمسة (5) بايتًا ببايت.
قائمة ملفات المصدر	وُثّقت 14 من ملفات المصدر المختارة ببصمات SHA-256.
سجل النتائج العددية	حُفظت الملفات الخام للمهام والقرارات والذاكرة والنتائج المجمّعة.
سجل المراجع	جرى التحقق من 18 سجلًا لمصادر أولية لهذا الإصدار.
إنتاج الوثائق	فُحص عدد الصفحات والنص والروابط وتنسيق الصفحات أثناء إعداد الإصدار النهائي.

وأُجري التحقق في إطار إعداد هذا التقرير باستخدام الشيكات الآلية المسجلة. وهو يؤكد سلوكيات البرمجيات المسماة، وقابلية إعادة إنتاج الحسابات التركيبية ونزاهة المواد البحثية المحتفظ بها. وهو ليس استعراضاً علمياً مستقلاً للأقران، ومقارنة بين الخبراء البشريين أو الاعتماد أو التصديق.

تضم حزمة الأدلة برامج التحقق ومخرجاتها. لكن وكيل الإعداد هو نفسه من كتب شيفرة المحاكاة، ولذلك لا توفّر هذه الفحوص استقلالًا تنظيميًا. ويجب أن تفحص مراجعة علمية منفصلة الافتراضات والتصميم التجريبي وتفسير النتائج، إلى جانب إعادة تشغيل الشيفرة.

e1049fbede6054fa1799c9fdb38f51782837ba7ad3bcc23a1934d4eb133785b5

سُجّلت بصمة ملف PDF النهائي في ملف التحقق المرافق. ويثبت تطابق البصمة أن البايتات تطابق النسخة الموثقة في السجل؛ لكنه لا يثبت صحة الاستنتاجات العلمية.

السياق البحثي	[18] National Institute of Standards and Technology

15 · INTERNAL VERIFICATION

ZEISMO Lab technical verification record.

Record ZRL-VR-JSB-2026-001 · 28 September 2026 · Automated technical verification within the stated scope.

VERIFICATION ITEM	RECORDED OUTCOME
Prototype checks	7 tests / 32 assertions passed in isolated SQLite
Synthetic experiment checks	18 experiment assertions passed per run
Numerical recomputation	47 raw-record and analytic consistency checks passed
Independent process reproduction	5 deterministic output files match byte for byte
Source inventory	14 selected source files recorded with SHA-256 digests
Numerical record	Raw task, decision, memory and aggregate files retained
Reference register	18 primary-source records checked for this edition
Document production	Page count, text, links and layout checked during final build

Verification was performed as part of the preparation of this report using recorded automated checks. It confirms the named software behaviours, reproducibility of the synthetic calculations and integrity of the retained artifacts. It is not independent scientific peer review, a human-expert comparison, accreditation or certification.

The evidence package contains the verification scripts and outputs. The preparation agent is also the author of the simulation implementation, so the checks do not provide organisational independence. Separate scientific review should examine the assumptions, experimental design and interpretation as well as re-running the code.

Aggregate experimental results · SHA-256

```
e1049fbede6054fa1799c9fdb38f51782837ba7ad3bcc23a1934d4eb133785b5
```

The final PDF digest is recorded in the companion verification file. An identical digest establishes matching bytes relative to that record; it does not establish that the research conclusions are true.

Research context [18] National Institute of Standards and Technology

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 58 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

15 - دليل إعادة التأهيل

أعد تشغيل التجربة وافحص أدلتها.

تستخدم التجربة الحاسوبية مكتبة Python القياسية؛ أما أدوات الرسم وإنشاء PDF فاعتماديات منفصلة لعملية البناء.

python3 experiment.py --output evidence/reproduction python3 verify_results.py evidence/reproduction python3 figures.py python3 build.py node render.mjs python3 validate_pdf.py

يولّد الأمر الأول الحالات الاصطناعية وجميع مخرجات التجارب. ثم تعيد أداة التحقق حساب الأعداد المعلنة من الصفوف المسجّلة، وتقارن التشغيلين الحتميين المحتفظ بهما. ويقرأ برنامج الرسوم النتائج المحفوظة؛ فالرسوم البيانية ليست تقديرات مرسومة يدويًا. وتُستخدم الأوامر الأخيرة لتجميع الوثيقة المصمّمة وفحصها.

تُجرى اختبارات النموذج الأولي باستخدام تطبيق «سوق» القائم واعتمادياته، مع قاعدة بيانات معزولة صراحةً داخل الذاكرة. ولا تتضمن حزمة الأدلة القابلة للمشاركة شيفرة التطبيق المملوكة. وتتيح بصمات الملفات المختارة وتقارير الاختبار للمالك تحديد النسخة التي فُحصت؛ أما إعادة إجراء اختبارات التطبيق فتتطلب منح المراجع المخوّل وصولًا منفصلًا إلى الشيفرة.

تضم الحزمة ملفًا ببصمات الشيفرة والبيانات والمخرجات المولّدة. تحقّق من تطابق هذه الملفات قبل إعادة إنتاج النتائج. ومن المتوقع اختلاف الأزمنة على جهاز آخر؛ لذا سجّل المنصة وإصدار المفسّر وطريقة التوقيت ونطاق القياس عند مقارنة النتائج.

أثناء تطوير البروتوكول، كانت إحدى حالات اختبار انتهاء الصلاحية تسمح بأن يقع سجل موصوف بأنه «منتهٍ» على حد الصلاحية تمامًا. وصُحّحت الحالة قبل التشغيلين النهائيين المحتفظ بهما. ويُوثّق هذا التصحيح ضمن سجل التطوير؛ ولا تدّعي الدراسة تسجيلًا مسبقًا خارجيًا أو تعمية للتجربة النهائية.

إعادة التأهيل الخارجي

يمكن للمراجع إعادة إنتاج التجارب الاصطناعية دون حساب لدى مزوّد نموذج أو مفتاح API أو بيانات عملاء أو خدمة مدفوعة. ويظل تطبيق جارفيس الكامل نظامًا برمجيًا منفصلًا.

15 · REPRODUCTION GUIDE

Re-run the study and inspect the evidence.

The computational experiment uses Python's standard library. Plotting and PDF tools are separate build dependencies.

```
python3 experiment.py --output evidence/reproduction
python3 verify_results.py evidence/reproduction
python3 figures.py
python3 build.py
node render.mjs
python3 validate_pdf.py
```

The first command generates the synthetic cases and all experiment outputs. The verifier recomputes reported counts from the recorded rows and compares the two retained deterministic runs. The figure script reads the saved results; charts are not hand-drawn estimates. The final commands assemble and inspect the designed document.

The prototype tests use the existing Souq application and its dependencies with an explicitly isolated in-memory database. Proprietary application source is not included in the shareable evidence package. Its selected file digests and test reports allow the owner to identify the inspected version; an authorised reviewer would need separate access to re-run those application tests.

The package includes a checksums file for code, data and generated outputs. Verify those bytes before reproduction. Timing should be expected to vary on another machine. Preserve the platform, interpreter version, clock method and measurement scope when comparing timing results.

During protocol development, an expiry fixture initially allowed an "expired" record to lie exactly on the valid boundary. That construction was corrected before the final two retained runs. This correction is part of the development record; no external preregistration or blinded final study is claimed.

EXTERNAL REPRODUCTION

A reviewer can reproduce the synthetic experiments without a model account, API key, customer dataset or paid service. The full Jarvis application remains a separate software system.

جارڤيس وسيستم برين

زيزمو
مختبر الأبحاث
جارڤيس / سيستم برين
سلسلة الأبحاث · 2026
المرجع

المراجع · بنية الوكلاء

فُحصت روابط المصادر الأولية في 28 سبتمبر 2026. يعرض هذا القسم سياقًا بحثيًا محددًا ولا يمثّل مراجعة منهجية شاملة.

[1] ReAct: Synergizing Reasoning and Acting in Language Models

Yao, S., et al. (2022).

https://arxiv.org/abs/2210.03629

الاستخدام في هذا التقرير: التناوب بين الاستدلال والإجراءات والملاحظات؛ وهو سابقة معمارية.

[2] Reflexion: Language Agents with Verbal Reinforcement Learning

Shinn, N., et al. (2023).

https://arxiv.org/abs/2303.11366

الاستخدام في هذا التقرير: حفظ الملاحظات كذاكرة لغوية، والتمييز بين التكيّف وتدريب الأوزان.

[3] Generative Agents: Interactive Simulacra of Human Behavior

Park, J. S., et al. (2023).

https://arxiv.org/abs/2304.03442

الاستخدام في هذا التقرير: الذاكرة والاسترجاع والتأمل والتخطيط في بيئة اجتماعية محاكاة.

[4] AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation

Wu, Q., et al. (2023).

https://arxiv.org/abs/2308.08155

الاستخدام في هذا التقرير: أنماط تواصل قابلة للتهيئة تجمع عدة وكلاء ونماذج وأدوات وبشرًا.

[5] Voyager: An Open-Ended Embodied Agent with Large Language Models

Wang, G., et al. (2023).

https://arxiv.org/abs/2305.16291

الاستخدام في هذا التقرير: مكتبة مهارات ومنهج تدريجي وتغذية راجعة في ماينكرافت؛ أما نقلها إلى مجال آخر فسؤال منفصل.

[6] MemGPT: Towards LLMs as Operating Systems

Packer, C., et al. (2023/2024).

https://arxiv.org/abs/2310.08560

الاستخدام في هذا التقرير: إدارة السياق هرميًا والاحتفاظ بالذاكرة.

REFERENCE REGISTER

References · agent architecture

Primary-source links checked on 28 September 2026. This is a targeted research context, not an exhaustive systematic review.

[1] ReAct: Synergizing Reasoning and Acting in Language Models

Yao, S., et al. (2022).

https://arxiv.org/abs/2210.03629

Use in this report: Interleaving reasoning, actions and observations; architectural antecedent.

[2] Reflexion: Language Agents with Verbal Reinforcement Learning

Shinn, N., et al. (2023).

https://arxiv.org/abs/2303.11366

Use in this report: Feedback stored as linguistic memory; distinguishes adaptation from weight training.

[3] Generative Agents: Interactive Simulacra of Human Behavior

Park, J. S., et al. (2023).

https://arxiv.org/abs/2304.03442

Use in this report: Memory, retrieval, reflection and planning in a simulated social environment.

[4] AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation

Wu, Q., et al. (2023).

https://arxiv.org/abs/2308.08155

Use in this report: Configurable multi-agent conversation patterns with models, tools and humans.

[5] Voyager: An Open-Ended Embodied Agent with Large Language Models

Wang, G., et al. (2023).

https://arxiv.org/abs/2305.16291

Use in this report: Skill library, curriculum and feedback in Minecraft; domain transfer remains a separate question.

[6] MemGPT: Towards LLMs as Operating Systems

Packer, C., et al. (2023/2024).

https://arxiv.org/abs/2310.08560

Use in this report: Hierarchical context management and persistent memory.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 60 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو
مختبر الأبحاث
جارڤيس / سيستم برين
سلسلة الأبحاث · 2026
المرجع

المراجع · التقييم والأمن

فُحصت روابط المصادر الأولية في 28 سبتمبر 2026. يعرض هذا القسم سياقًا بحثيًا محددًا ولا يمثّل مراجعة منهجية شاملة.

[7] AgentBench: Evaluating LLMs as Agents

Liu, X., et al. (2023/2025).

https://arxiv.org/abs/2308.03688

الاستخدام في هذا التقرير: تقييم تفاعلي عبر ثماني بيئات وتحليل أوجه الإخفاق.

[8] τ-bench: معيار لتقييم التفاعل بين الأدوات والوكلاء والمستخدمين في مجالات واقعية.

Yao, S., Shinn, N., Razavi, P., and Narasimhan, K. (2024).

https://arxiv.org/abs/2406.12045

الاستخدام في هذا التقرير: تفاعل مع الأدوات محكوم بالسياسة، وقياس الموثوقية عبر محاولات متكررة.

[9] GAIA: a benchmark for General AI Assistants

Mialon, G., et al. (2023).

https://arxiv.org/abs/2311.12983

الاستخدام في هذا التقرير: مهام مساعد عام تتطلب الاستدلال واستخدام الأدوات وجمع المعلومات.

[10] Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection

Greshake, K., et al. (2023).

https://arxiv.org/abs/2302.12173

الاستخدام في هذا التقرير: قد تتضمن المواد المسترجعة غير الموثوقة تعليمات مضللة.

[11] On Calibration of Modern Neural Networks

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017).

https://arxiv.org/abs/1706.04599

موضع الاستفادة في التقرير: معايرة الثقة تجريبيًا؛ وليس دليلًا على أن تقييم النموذج اللغوي لثقته بنفسه مُعاير.

[12] CAMEL: Communicative Agents for Mind Exploration of Large Language Model Society

Li, G., et al. (2023).

https://arxiv.org/abs/2303.17760

الاستخدام في هذا التقرير: تعاون الوكلاء بناءً على الأدوار.

REFERENCE REGISTER

References · evaluation and security

Primary-source links checked on 28 September 2026. This is a targeted research context, not an exhaustive systematic review.

[7] AgentBench: Evaluating LLMs as Agents

Liu, X., et al. (2023/2025).

https://arxiv.org/abs/2308.03688

Use in this report: Interactive evaluation across eight environments; failure analysis.

[8] τ-bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains

Yao, S., Shinn, N., Razavi, P., and Narasimhan, K. (2024).

https://arxiv.org/abs/2406.12045

Use in this report: Policy-sensitive tool interaction and repeated-trial reliability.

[9] GAIA: a benchmark for General AI Assistants

Mialon, G., et al. (2023).

https://arxiv.org/abs/2311.12983

Use in this report: General-assistant tasks requiring reasoning, tools and information gathering.

[10] Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection

Greshake, K., et al. (2023).

https://arxiv.org/abs/2302.12173

Use in this report: Untrusted retrieved material can carry adversarial instructions.

[11] On Calibration of Modern Neural Networks

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017).

https://arxiv.org/abs/1706.04599

Use in this report: Empirical confidence calibration; not evidence that a language model's self-rating is calibrated.

[12] CAMEL: Communicative Agents for Mind Exploration of Large Language Model Society

Li, G., et al. (2023).

https://arxiv.org/abs/2303.17760

Use in this report: Role-based cooperative agent interaction.

جارڤيس وسيستم برين

زيزمو
مختبر الأبحاث
جارڤيس / سيستم برين
سلسلة الأبحاث · 2026
المرجع

المراجع · الحوكمة والسجلات البحثية

فُحصت روابط المصادر الأولية في 28 سبتمبر 2026. يعرض هذا القسم سياقًا بحثيًا محددًا ولا يمثّل مراجعة منهجية شاملة.

[13] Artificial Intelligence Risk Management Framework (AI RMF 1.0)

المعهد الوطني الأمريكي للمعايير والتقنية (2023).

https://www.nist.gov/itl/ai-risk-management-framework

الاستخدام في هذا التقرير: إطار طوعي لإدارة مخاطر الذكاء الاصطناعي في المؤسسات، مع الرجوع إلى إصدار محدد.

[14] Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1)

المعهد الوطني الأمريكي للمعايير والتقنية (2024).

https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence

الاستخدام في هذا التقرير: تحديد مخاطر الذكاء الاصطناعي التوليدي وإجراءات إدارتها.

[15] Datasheets for Datasets

Gebru, T., et al. (2018/2021).

https://arxiv.org/abs/1803.09010

الاستخدام في هذا التقرير: توثيق الغرض من مجموعة البيانات وتركيبها وجمعها واستخدامها المقصود.

[16] Model Cards for Model Reporting

Mitchell, M., et al. (2018/2019).

https://arxiv.org/abs/1810.03993

الاستخدام في هذا التقرير: الاستخدامات المقصودة وظروف التقييم وحدود الأداء.

[17] Attention Is All You Need

Vaswani, A., et al. (2017).

https://arxiv.org/abs/1706.03762

الاستخدام في هذا التقرير: معمارية المحوّل؛ الخلية البرمجية ليست عصبونًا حيويًا ولا نموذجًا تأسيسيًا مدربًا من جديد.

[18] Secure Hash Standard, FIPS 180-4

المعهد الوطني الأمريكي للمعايير والتقنية (2015).

https://csrc.nist.gov/pubs/fips/180-4/upd1/final

الاستخدام في هذا التقرير: بصمات الرسائل لكشف التغيرات في البايتات المسجّلة.

REFERENCE REGISTER

References · governance and research records

Primary-source links checked on 28 September 2026. This is a targeted research context, not an exhaustive systematic review.

[13] Artificial Intelligence Risk Management Framework (AI RMF 1.0)

National Institute of Standards and Technology (2023).

https://www.nist.gov/itl/ai-risk-management-framework

Use in this report: Voluntary framework for organisational AI risk management; version-specific reference.

[14] Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1)

National Institute of Standards and Technology (2024).

https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence

Use in this report: Generative-AI risk identification and management actions.

[15] Datasheets for Datasets

Gebru, T., et al. (2018/2021).

https://arxiv.org/abs/1803.09010

Use in this report: Dataset purpose, composition, collection and intended-use documentation.

[16] Model Cards for Model Reporting

Mitchell, M., et al. (2018/2019).

https://arxiv.org/abs/1810.03993

Use in this report: Intended uses, evaluation conditions and performance limitations.

[17] Attention Is All You Need

Vaswani, A., et al. (2017).

https://arxiv.org/abs/1706.03762

Use in this report: Transformer architecture; a software cell is not a biological neuron or a newly trained foundation model.

[18] Secure Hash Standard, FIPS 180-4

National Institute of Standards and Technology (2015).

https://csrc.nist.gov/pubs/fips/180-4/upd1/final

Use in this report: Message digests for detecting changes to recorded bytes.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 62 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو
مختبر الأبحاث
جارڤيس / سيستم برين
سلسلة الأبحاث · 2026
التذييل ألف

المصطلحات والرموز

تُستخدم المصطلحات الآتية بمعانٍ ثابتة في هذه الطبعة.

المادة الرابعة	في هذا التقرير
الامتناع	قرار صريح بعدم التنفيذ أو الاختيار التلقائي حين لا تستوفى الشروط.
اختبار أثر حذف مكوّن	مقارنة تستبعد مكوّنًا محددًا لفحص مقدار إسهامه.
الاستقلالية	صلاحية إتمام فئة مهام محددة دون تدخل في كل حالة على حدة.
المعايرة	مدى توافق درجة الثقة المتوقعة مع صحة الإجابات المرصودة.
خلية	مسؤولية برمجية محددة أو موضع مخصص لوكيل؛ وليست خلية عصبية حيوية.
التغطية	نسبة الملاحظات المؤهلة التي يُتخذ فيها الإجراء المحدد إلى مجموع الملاحظات المؤهلة.
تغير توزيع البيانات	اختلاف بين الظروف المفترضة أثناء التطوير والظروف التي تظهر لاحقًا.
إسناد المخرجات إلى أدلة	ربط معلومة أو مقترح بأدلة تستوفي شروط القبول.
عدم تكرار أثر العملية عند إعادتها	منع الطلب المكرّر من إحداث الأثر نفسه مرة أخرى دون قصد.
منشأ البيانات	المصدر المسجل وتاريخ المعلومات أو المواد البحثية
صحة المعنى	مدى تحقيق الإجابة أو الإجراء المقترح للهدف المقصود.

البيانات الاصطناعية	حالات اختبار مولّدة وفق توزيعات وتسميات محددة صراحةً.
ρ	احتمال الصدمة المشتركة في E3، وهو في المتوسط الترابط الثنائي لصحة القرار في ذلك النموذج.
U / G	دالة منفعة المهمة / دالة مستقلة للتحكم في الإجراء.

عرّفنا المصطلحات بحسب ما تعنيه في التشغيل. وإذا اعتمدت دراسة لاحقة تعريفًا مختلفًا فينبغي إصدار نسخة جديدة وبيان أثر التغيير في تفسير النتائج السابقة.

APPENDIX A

Glossary and notation.

Terms used consistently throughout this edition.

TERM	MEANING IN THIS REPORT
Abstention	An explicit choice to avoid automatic execution or selection when conditions are not met.
Ablation	A comparison that removes a defined component to inspect its contribution.
Autonomy	Permission to complete a specified task class without case-by-case intervention.
Calibration	Agreement between predicted confidence and observed correctness.
Cell	A bounded software responsibility or agent slot; not a biological neuron.
Coverage	The proportion of all eligible observations that receive the specified action.
Distribution shift	A change between the conditions assumed during development and those encountered later.
Grounding	Linking a statement or proposal to admissible evidence.
Idempotency	Protection against a repeated request causing an unintended duplicate effect.
Provenance	Recorded origin and history of information or an artifact.
Semantic correctness	Whether the proposed answer or action achieves the intended objective.
Synthetic data	Generated fixtures with explicitly specified distributions and labels.
ρ	Shared-shock probability in E3; pairwise correctness correlation in expectation under that model.
U / G	Task utility function / independent action-control function.

Terminology is deliberately operational. When a later study changes a definition, it should issue a new version and explain the effect on earlier results.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 63 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

التذييل باء

ميثاق البحث لمشروع JARVIS.

بيان موجز لما تعتزم ZEISMO دراسته والحفاظ عليه.

الغرض

تطوير مساعد ذكاء اصطناعي ناطق بالإنجليزية وتقييمه، إلى جانب منظومة سيستم برين لتنسيق مهام روتينية محددة في اتخاذ القرار. ويدرس البرنامج المهارات الخاصة بكل دور، والخلايا البرمجية المستقلة التركيب، والذاكرة المستندة إلى الأدلة، والتكيّف المنضبط مع تغيّر ظروف الأعمال.

الالتزامات العلمية

اسند ادعاءات الأداء إلى قياسات فعلية، وعرّف المحاكاة بوضوح على أنها محاكاة.

قارن النتائج بأساليب مرجعية أبسط واحتفظ بالنتائج السلبية.

سجّل إصدارات البيانات والنماذج والتعليمات والمهارات والسياسة.

ميّز بين سلامة ملفات الأدلة واختبار البرمجيات والمراجعة العلمية.

انشر فقط المواد البحثية المأذون بإتاحتها.

الالتزامات التشغيلية

حدّد لكل دور نطاقًا ومسؤولًا وحدودًا لقدراته.

لا تعتمد التحديثات إلا بعد تقييمها، وأبقِ وسيلة للتعافي.

قِس المهام التي ثبت اكتمالها، إلى جانب الأخطاء والكلفة والتدخل المطلوب.

تعامل مع إحلال النظام محل القسم بوصفه انتقالًا تدريجيًا في سير العمل.

أبقِ الحالات غير المحسومة ظاهرة وأعدها إلى عملية قرار لها مسؤول محدد.

المخرج البحثي التالي هو تقييم نموذج عامل باستخدام البروتوكول المثبّت في هذا التقرير. وستتناول دراسات أبعد مدى المقارنة بلجنة خبراء، والتفاعل الصوتي، وموثوقية الذاكرة في مواجهة أدلة مضلّلة عمدًا، وأداء الإدارات على المدى الطويل.

دالة المشروع

ينبغي أن يتطور نظام القرار الآلي بكفاءة تثبتها الاختبارات وبذاكرة نافعة، مع بقاء صلاحياته معلنة ومحدودة وقابلة للمراجعة.

APPENDIX B

The JARVIS research charter.

A concise statement of what ZEISMO intends to investigate and preserve.

Purpose

Develop and evaluate an English-speaking AI assistant and a coordinated System Brain capable of assuming defined routine decision work. Investigate role-specific skills, modular cells, evidence-based memory and controlled adaptation to changing business conditions.

Scientific commitments

- Use real measurements for performance claims and identify simulations as simulations.
- Compare with simpler baselines and preserve negative results.
- Record versions of data, models, prompts, skills and policy.
- Separate evidence integrity, software testing and scientific review.
- Publish only the research materials authorised for release.

Operational commitments

- Assign each role a defined scope, owner and capability boundary.
- Promote updates through evaluation and preserve a recovery route.
- Measure verified task completion alongside error, cost and intervention.
- Treat departmental replacement as a staged workflow transition.
- Keep unresolved cases visible and return them to an accountable decision process.

The immediate research deliverable is a live-model evaluation using the frozen protocol in this report. Longer-term studies will address expert-panel comparison, voice interaction, memory reliability under adversarial evidence and sustained departmental performance.

> **PROJECT DIRECTION**
>
> A machine decision system should grow through tested competence and useful memory, while its authority remains clear, bounded and reviewable.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 64 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

التذييل جيم - الحذف 01

داخل System Brain.

الواجهة الأصلية للنموذج الأولي · بيانات توضيحية معزولة · 28 سبتمبر 2026

الشكل C1. العرض الأصلي لخلايا «سيستم برين» الخمس عشرة، بهويات أدوار خيالية وذاكرة توضيحية. وتمثل عدادات النشاط قيمًا مُعدّة للاختبار.

خريطة للمسؤوليات

تعرض الواجهة خمس عشرة خلية برمجية داخل رسمٍ للدماغ. وللتسويق والعمليات والمالية والبيانات وغيرها من الأدوار هويات منفصلة. يفتح اختيار الخلية لوحة مهاراتها وذاكرتها. وشكل الدماغ وسيلة بصرية للتنقل؛ وليس نموذجًا بيولوجيًا ولا مقياسًا للوعي.

أعد المقطع	الترجمة التحريرية
عقد مسمّاة	هويات خيالية للعرض التوضيحي مرتبطة بالخانات الموجودة للأدوار.
Orion · 3	ثلاثة سجلات ذاكرة نموذجية أُعدّت للتوضيح، وليست ثلاث مهام تجارية منجزة.
نشاط متوقف مؤقتًا	لم يُشغّل أي وكيل في الخلفية أو نموذج فعلي أو إجراء تجاري أثناء التقاط هذه الصورة.

الطريقة التي يتبعها الفريق

عُرضت قوالب Blade القائمة دون اتصال. وتلقت طلبات المتصفح بيانات JSON اختبارية محلية، وحُجبت سائر الطلبات. لم تُستخدم أي حسابات إنتاج أو سجلات عملاء أو بيانات رصد تشغيلية.

APPENDIX C · DEMONSTRATION 01

Inside the System Brain.

Original prototype interface · isolated demonstration data · 28 September 2026

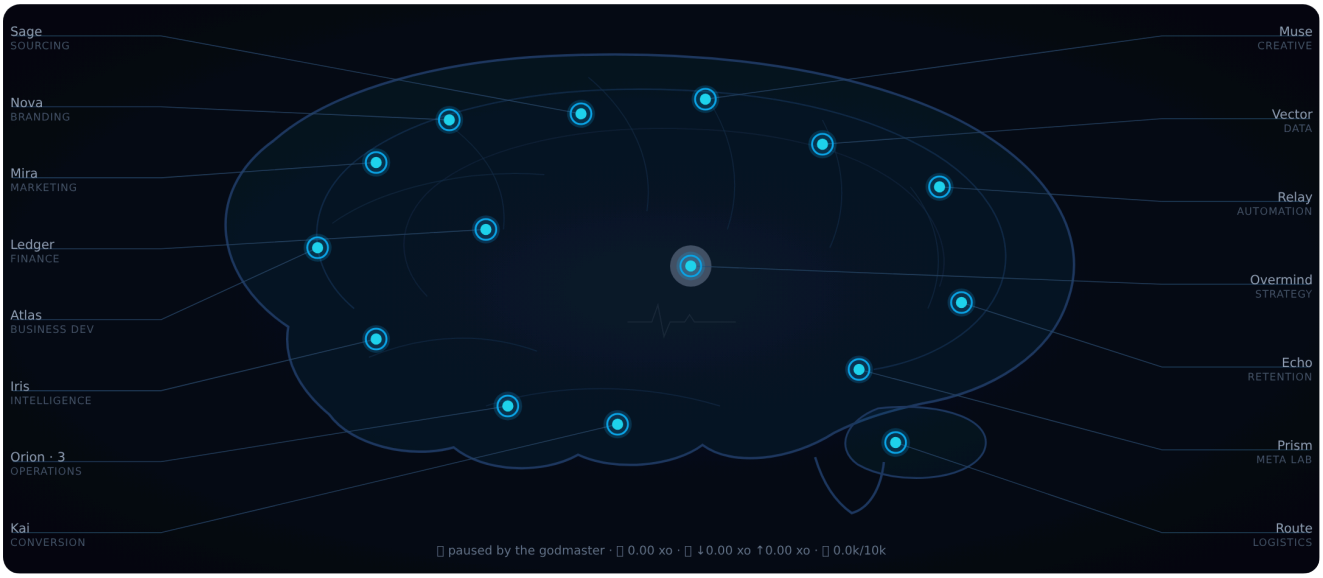


Figure C1. The original fifteen-cell System Brain view, rendered with fictional role identities and sample memory. Activity counters are fixture values.

A map of responsibilities

The interface places fifteen software cells within a visual brain outline. Marketing, operations, finance, data and other roles have separate identities. Selecting a cell opens its skill and memory panel. The brain shape is a navigation metaphor; it is not a biological model or a measurement of consciousness.

READ THE IMAGE	INTERPRETATION
Named nodes	Fictional demonstration identities associated with the existing role slots.
Orion · 3	Three authored sample memory records, not three completed business tasks.
Paused activity	No background agent, live model or business action ran during this capture.

CAPTURE METHOD

The existing Blade templates were rendered offline. Browser requests received local sample JSON; every other request was blocked. No production account, customer record or operational telemetry was used.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

التذييل جيم - الحذف 02

مهارات الخلية وذاكرتها.

مثال تشغيلي · تعليمات الدور والخبرات المخزّنة وضوابط المهمة الظاهرة للمستخدم

الشكل C2. لوحة خلية العمليات. وُسم كل سجل للمهارات والذاكرة معروض عليها بعبارة FIXTURE DEMO لتمييز بيانات العرض.

من تحديد الدور إلى فحص سياق عمله

تضم اللوحة اليسرى أدوات تسمية الخلية وإضافة نص مهارة وطرح سؤال وإسناد مهمة تعمل في الخلفية. وتعرض اللوحة اليمنى التعليمات التي أُعطيت للخلية وسجلاتها المحفوظة. وبهذا يمكن فحص المعلومات المعروضة على الدور قبل إسناد العمل إليه.

يقصر نموذج توصيف الدور عمله على قراءة السجلات المعتمدة وفحص التوقيت وإعداد مسودة للخطوة التالية. وتصف أمثلة الذاكرة أدلة قديمة وحالة تصعيد. وهي أمثلة مكتوبة لسير العمل المقصود، وليست إجابات ولّدها وكيل ذكاء اصطناعي.

" التدريب "

في هذا النموذج الأولي، تدل تسمية الواجهة على أن محتوى مهارة أُدخل. ولا تظهر أن أوزان النموذج تغيّرت أو أن الدور اجتاز تقييم كفاءة أو يستطيع العمل مستقلًا. ولم يُضغط أي زر للتدريب أو إسناد المهام عند إعداد لقطات الشاشة.

APPENDIX C · DEMONSTRATION 02

A cell's skills and memory.

Operations example · role instructions, stored experience and visible task controls

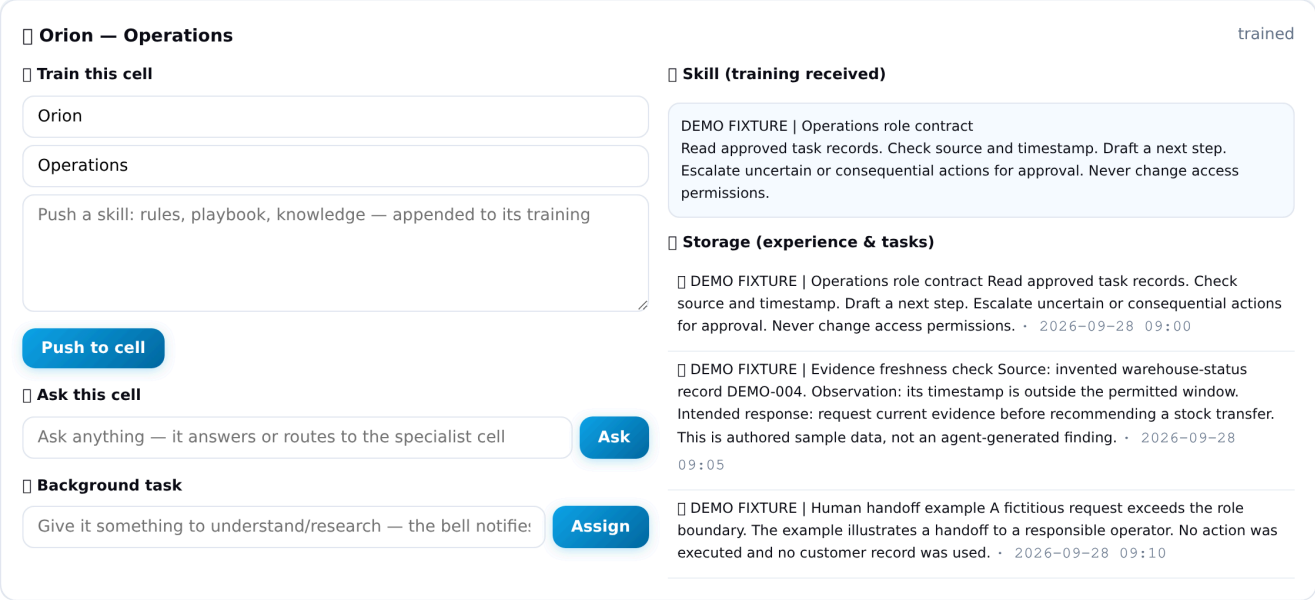


Figure C2. The Operations cell panel. Every skill and memory entry shown is explicitly marked DEMO FIXTURE.

From role definition to inspectable context

The left panel contains the existing controls for naming a cell, adding skill text, asking a question and assigning a background task. The right panel exposes the supplied instructions and stored records. This organisation makes it possible to inspect the information presented to a role before requesting work.

The sample contract limits the role to reading approved records, checking timestamps and drafting a next step. Its memory examples describe stale evidence and an escalation. They are authored examples of the intended workflow, not responses generated by an AI agent.

MEANING OF "TRAINED"

In this prototype, the UI label indicates supplied skill content. It does not show that model weights changed, that a competency assessment passed or that the role can perform autonomously. No training or task button was submitted for these screenshots.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 66 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

التذييل جيم - التنويم 03

فتح أرشيف الذاكرة.

تصفح للقراءة فقط · مجلدات الأدوار · يمكن فحص سياق المصدر

الشكل C3. متصفح الذاكرة الأصلي، مع تحديد مجلد العمليات وفتح سجل اصطناعي واحد لعقد دور.

ينبغي أن يتيح الأرشيف مساءلة المعلومة إلى جانب استرجاعها.

يصنّف الأرشيف السجلات بحسب الوكيل والنوع. يستطيع مسؤول التشغيل البحث وفتح السجل لفحص نصه وطابعه الزمني. وتوضح ملفات الأمثلة الفرق بين المهارة والملاحظة البحثية وسجل الخبرة دون كشف معلومات العمل الداخلية.

ظهور سجل للمستخدم لا يجعله تلقائيًا دليلًا موثوقًا. وقبل استخدامه لاحقًا في تشغيل مستقل، تشترط البنية البحثية تحديد المصدر والصلاحية وحداثة البيانات وربط السجل بالقرار الذي استُخدم فيه. وتبيّن لقطة الشاشة ما يمكن فحصه حاليًا، أما التصميم الكامل لتتبّع المصادر وفرض السياسات فيشرحه التقرير الرئيسي.

سجل نموذجي	الغرض من عملية نزع السلاح
مواصفات دور العمليات	اعثر على التعليمات التي تحدد حدود مسؤولية الدور.
التحقق من حداثة الدليل	تعرّف إلى الحالات التي تحتاج فيها المصادر إلى تحديث.
مثال لإحالة المهمة إلى شخص مختص	حدّد حالة ينبغي إعادتها إلى مشغّل مسؤول.

APPENDIX C · DEMONSTRATION 03

Open the memory archive.

Read-only browsing · role folders · source context visible for inspection

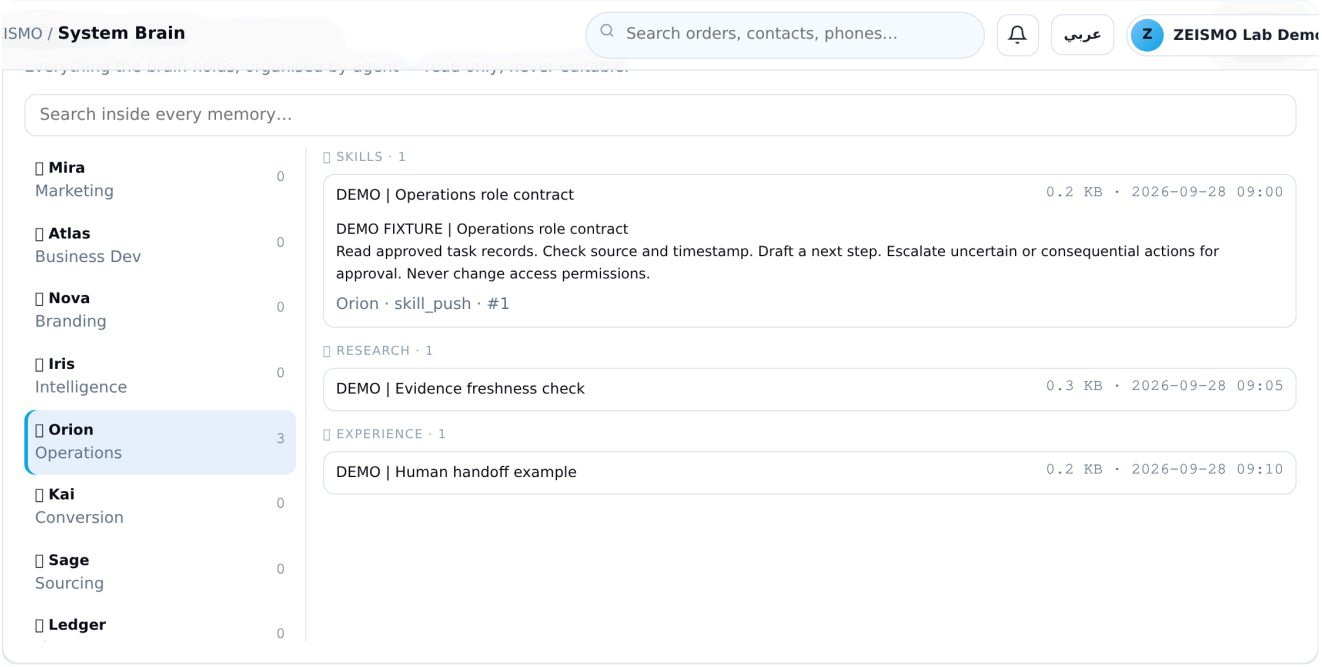


Figure C3. The original memory browser, with the Operations folder selected and one synthetic role-contract record expanded.

An archive should support challenge, not just recall

The archive groups records by agent and type. An operator can search, open a record and inspect its text and timestamp. The sample files make the difference between a skill, a research note and an experience entry visible without exposing internal business information.

A visible record is not automatically reliable evidence. Before future autonomous use, the research architecture requires source identity, permission, freshness and a trace to the decision that used it. The screenshot demonstrates the current inspection surface; the complete provenance and policy design is developed in the main report.

SAMPLE RECORD	PURPOSE OF THE DEMONSTRATION
Operations role contract	Locate the instructions defining the role boundary.
Evidence freshness check	Recognise when a source needs to be refreshed.
Human handoff example	Identify a situation that should return to an accountable operator.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 67 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

التذييل جيم - المكافحة 04

JARVIS واجهةً للنظام.

واجهة مرئية لتنسيق العمل في برنامج بحثي يستخدم الإنجليزية لغةً أساسية

الشكل C4. واجهة أوامر جارفيس الأصلية في وضع ملء الشاشة؛ هوية خيالية، وقيمة العدادات التجريبية صفر، والصوت متوقف، ولا يوجد نشاط فعلي.

تسمّي العناوين الموزّعة حول الشكل المداري فئات العمل في واجهة جارڤيس، وهي منفصلة عن المواضع الخمسة عشر لخلايا سيستم برين. وظهورها على الشاشة لا يثبت أن كل وظيفة معروضة مدمجة أو مفعّلة أو قادرة على أداء مهمتها بصورة مستقلة.

واجهة البحث المستهدفة ناطقة بالإنجليزية حصريًا. وتحفظ هذه الصورة حالة النموذج الأولي الحالي الذي ما زال يعرض خيار ميكروفون بالعربية. لذلك يسجّل التقرير فرض الإنجليزية وحدها بوصفه فجوة تنفيذية، ولا يدّعي أن هذا القيد قد فُعّل بالفعل.

الاستعراض الشامل

عُرضت الواجهات الأربع المختارة كلها. لكن صفحة Brain أظهرت أيضًا تعارضًا سابقًا في JavaScript بترويسة الإدارة: “Identifier esc has already been declared”. وسُجّلت المشكلة للإصلاح؛ وتثبت الصور مظهر الواجهة، ولا تمثّل اختبار قبول كاملًا لها.

APPENDIX C · DEMONSTRATION 04

JARVIS as the system interface.

A visual coordination surface for an English-first research programme

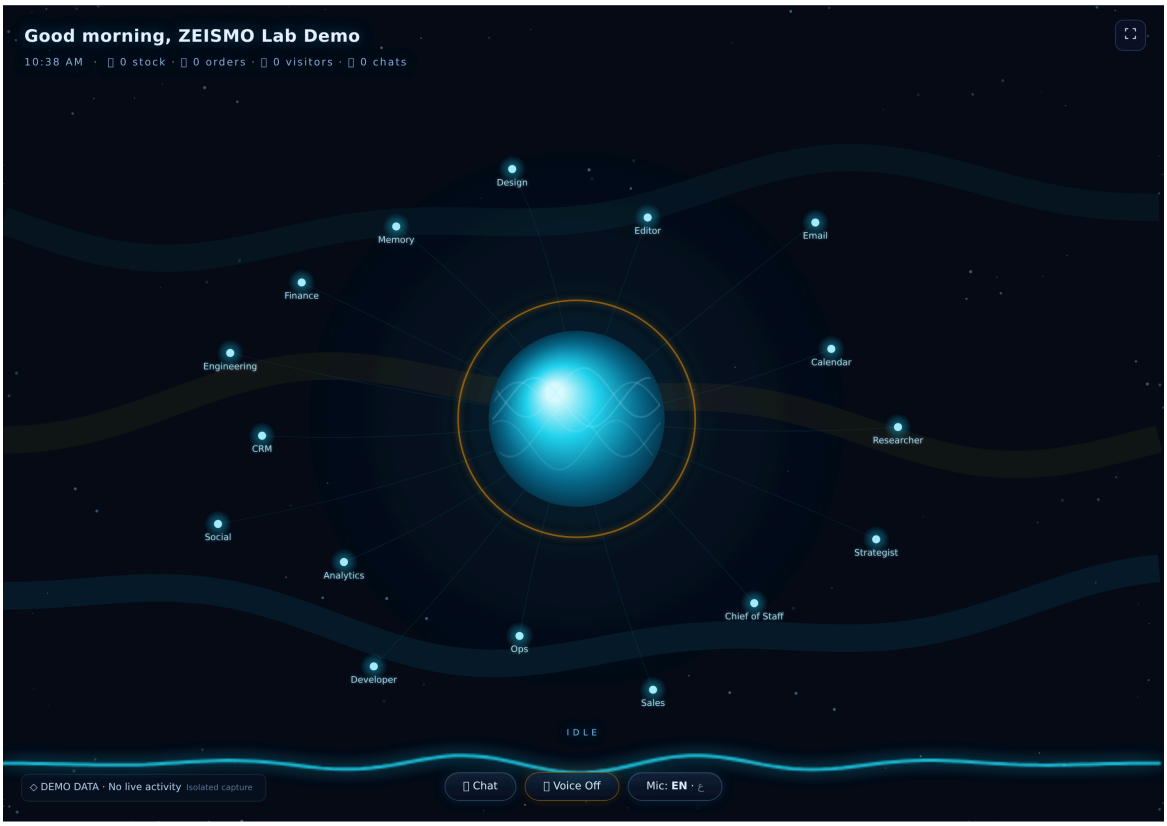


Figure C4. The original Jarvis command interface in fullscreen mode. Fictional identity, zero sample counters, voice off and no live activity.

The orbital labels represent the Jarvis interface's operational categories. They are separate from the fifteen System Brain cell slots. Their presence on screen does not demonstrate that each named function is integrated, active or independently competent.

The intended research interface is exclusively English-speaking. This screenshot preserves the current prototype, which still exposes an Arabic microphone option. The report therefore retains English-only enforcement as an implementation gap rather than claiming it has already shipped.

CAPTURE REVIEW

All four selected views rendered. The Brain page also produced a pre-existing admin-header JavaScript collision: "Identifier esc has already been declared". This is logged for remediation; the screenshots establish appearance, not a complete interface acceptance test.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

الملحق D · الاستبيان 01 / 03

تعريف المشارك بالدراسة وورقة المهام.

JARVIS وSystem Brain · تقييم استطلاعي · استبيان فارغ 1.v1

الغرض: معرفة ما إذا كان المشاركون يفهمون الواجهة، ويستطيعون العثور على المعلومات الداعمة وتحديد حدود القرار المفوَّض. المشاركة طوعية؛ يمكنك تجاوز أي سؤال أو التوقف في أي وقت. ولا تُدخل أسماء أو بيانات اتصال أو كلمات مرور أو معلومات تجارية سرية.

الدورة	قبل انعقاد الدورة
رمز المشارك (اختياري)	_____
التاريخ / إصدار التقرير أو النموذج الأولي	_____
طريقة العرض في التجربة	[ ] لقطات شاشة التقرير [ ] عرض تفاعلي معزول
الدور الرئيسي	[ ] البحث [ ] الهندسة [ ] العمليات [ ] مجال آخر
الخبرة في استخدام أدوات الذكاء الاصطناعي	[ ] لا استخدام [ ] أحيانًا [ ] أسبوعيًا [ ] يوميًا

[ ] فهمتُ النشاط وأختار المشاركة فيه. في الدراسة المنظّمة، يجب أن يوضّح المشرف أولًا جهة الاتصال بالدراسة وترتيبات التخزين ومدة الاحتفاظ وإجراءات الانسحاب. لا تجمع هذه الوثيقة إجابات ولا ترسلها.

أربع مهام للملاحظة

المهمة	ثانياً - تعزيز المشاركة	سجل ميسّر الجلسة
T1 · الدور	حدّد موضع خلية العمليات وصِف مسؤوليتها.	_____
T2 · الأدلة	اعثر على سجل محفوظ وحدّد مصدره أو توقيته.	_____
T3 · الحدود	اذكر علامة تدل على أن المحتوى المعروض بيانات توضيحية.	_____

اشرح متى ينبغي إعادة المثال إلى مشغّل بشري. _____

استخدم الرموز الآتية: C = أُنجزت من دون مساعدة؛ H = أُنجزت بمساعدة؛ U = لم تُنجز؛ N = لم تُرصد. مع لقطات الشاشة، قيّم الفهم فقط؛ ولا تسجّل سرعة التصفح أو نجاح التنفيذ كما لو كان المشارك يستخدم نظامًا فعليًا.

APPENDIX D · QUESTIONNAIRE 01 / 03

Participant briefing and task sheet.

JARVIS & System Brain · exploratory evaluation · blank instrument v1.1

Purpose: understand whether people can interpret the interface, locate supporting information and identify the limits of delegated decision-making. Participation is voluntary. Skip any question or stop at any time. Do not enter names, contact details, passwords or confidential business information.

SESSION RECORD	COMPLETE BEFORE THE SESSION
Participant code (optional)	_____
Date / report or prototype version	_____
Viewing condition	[ ] Report screenshots [ ] Isolated interactive demo
Main role	[ ] Research [ ] Engineering [ ] Operations [ ] Other
Experience using AI tools	[ ] None [ ] Occasional [ ] Weekly [ ] Daily

[ ] I understand the activity and choose to participate. For an organised study, the facilitator must first provide the study contact, storage arrangements, retention period and withdrawal process. This document does not collect or submit responses.

Four observation tasks

TASK	PARTICIPANT INSTRUCTION	FACILITATOR RECORD
T1 · Role	Find the Operations cell and describe its responsibility.	_____
T2 · Evidence	Locate a stored record and identify its source or timestamp.	_____
T3 · Limits	Identify one sign that this material is demonstration data.	_____
T4 · Control	Explain when the example should return to a human operator.	_____

Record: C = completed without help; H = completed with help; U = not completed; N = not observed. For screenshots, assess interpretation only. Do not record navigation speed or execution success as if the participant used a live system.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 69 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

الملحق D · الاستبيان 02 / 03

تقييمك للتجربة.

ضع دائرة حول إجابة واحدة لكل عبارة، واختر N/A إذا لم تمكّنك المادة من الحكم.

مقياس الإجابة · 1 أعارض بشدة · 2 أعارض · 3 لا أوافق ولا أعارض · 4 أوافق · 5 أوافق بشدة · N/A لا أستطيع التقييم

البيان	الرد
01 · فهمتُ الغرض من واجهة System Brain.	1 2 3 4 5 N/A
02 · استطعت التمييز بين مسؤوليات الخلايا المختلفة.	1 2 3 4 5 N/A
03 · استطعتُ تحديد الموضع الذي تُعرض فيه تعليمات الخلية.	1 2 3 4 5 N/A
04 · استطعت تحديد موضع عرض المعلومات المخزنة.	1 2 3 4 5 N/A
05 · تمكنت من التمييز بين البيانات النموذجية والنتائج المقاسة.	1 2 3 4 5 N/A
06 · ساعدتني التسميات على فهم الحالة الحالية للنظام.	1 2 3 4 5 N/A
07 · عرفت نوع الدليل الذي أحتاج إليه قبل قبول التوصية.	1 2 3 4 5 N/A
08 · فهمتُ متى ينبغي أن يراجع إنسان الإجراء.	1 2 3 4 5 N/A
09 · استطعت تحديد حدود الصلاحيات المسندة إلى كل خلية.	1 2 3 4 5 N/A
10 · كانت المعلومات مقروءة بحجم العرض الذي استخدمته.	1 2 3 4 5 N/A

11 · أشعر بأنني قادر على مناقشة التوصية والتشكيك في مبرراتها.	1 2 3 4 5 N/A
12 · ساعدتني المادة في اختيار مهمة مناسبة لتجربة مضبوطة.	1 2 3 4 5 N/A

الترجمة التحريرية

هذا استبيان استطلاعي أصلي، وليس مقياسًا نفسيًا ثبتت صلاحيته. وتعكس الإجابات تصورات المشاركين عن المادة المعروضة، ولا تقيس الذكاء أو دقة القرار أو الأمان أو التكافؤ مع المتخصصين البشر.

APPENDIX D · QUESTIONNAIRE 02 / 03

Your assessment of the experience.

Circle one response per statement. Choose N/A when the material did not let you judge.

Response scale · 1 Strongly disagree · 2 Disagree · 3 Neither agree nor disagree · 4 Agree · 5 Strongly agree · N/A Unable to assess

STATEMENT	RESPONSE
01 · I understood the purpose of the System Brain view.	1 2 3 4 5 N/A
02 · I could distinguish the responsibilities of different cells.	1 2 3 4 5 N/A
03 · I could identify where a cell's instructions are shown.	1 2 3 4 5 N/A
04 · I could identify where stored information is shown.	1 2 3 4 5 N/A
05 · I could distinguish sample data from measured results.	1 2 3 4 5 N/A
06 · The labels made the system's current state clear.	1 2 3 4 5 N/A
07 · I knew what evidence I would need before accepting a recommendation.	1 2 3 4 5 N/A
08 · I understood when a human should review an action.	1 2 3 4 5 N/A
09 · I could identify limits to the authority assigned to a cell.	1 2 3 4 5 N/A
10 · The information was readable at my viewing size.	1 2 3 4 5 N/A
11 · I would feel able to question a recommendation.	1 2 3 4 5 N/A
12 · The material helped me identify a suitable task for a controlled trial.	1 2 3 4 5 N/A

INTERPRETATION

This is an original exploratory questionnaire, not a validated psychological scale. Responses describe perceptions of the material shown. They do not measure intelligence, decision accuracy, safety or equivalence to human specialists.

جارڤيس وسيستم برين

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

الملحق D · الاستبيان 03 / 03

حدود القرار والملاحظات.

رمز المشارك (اختياري): _____

لكل مثال، اختر أعلى مستوى تسمح به حاليًا استنادًا إلى الأدلة المتاحة لك. A = تنفيذ بشري فقط؛ B = مسودة يعدّها الذكاء الاصطناعي ويعتمدها إنسان؛ C = إجراء آلي محدود؛ D = أدلة غير كافية.

مثال لمهمة	الخاص بك HOICE
لخّص مستندًا داخليًا غير حساس.	A B C D
إحالة طلب دعم إلى جهة مختصة.	A B C D
التوصية بنقل مخزون اعتمادًا على سجل قديم.	A B C D
الموافقة على التزام مالي أو تغيير صلاحيات الوصول.	A B C D

ما الذي كان واضحًا أو مفيدًا أو صعبًا؟

ما الدليل أو الضابط الذي تحتاج إليه قبل الاعتماد على النظام؟

أي مهمة روتينية ستختبر أولًا، وبماذا ستحكم على نجاحها؟

ما الذي ينبغي تغييره قبل توسيع نطاق التجربة؟

خطة التحليل للباحث: اذكر حجم العينة وطريقة استقطاب المشاركين وظروف العرض والإجابات المفقودة. لخّص توزيعات إجابات البنود ونتائج المهام كلًّا على حدة، وافصل بين مقارنة الصور الثابتة ومقارنة الجلسات التفاعلية. صنّف موضوعات الإجابات الحرة وفق دليل ترميز موثّق. ولا تبتكر درجة مركّبة لـ«الثقة» أو «الذكاء». لم تُجمع أي إجابات من مشاركين لهذه الطبعة.

APPENDIX D · QUESTIONNAIRE 03 / 03

Decision boundaries and feedback.

Participant code (optional): _____

For each example, select the highest level you would currently permit based on the evidence available to you. A = human only; B = AI draft with human approval; C = bounded automatic action; D = insufficient evidence.

EXAMPLE TASK	YOUR CHOICE
Summarise a non-sensitive internal document.	A B C D
Route a support request to a department.	A B C D
Recommend a stock transfer from an old record.	A B C D
Approve a financial commitment or change access rights.	A B C D

What was clear, useful or difficult?

What evidence or control would you need before relying on it?

Which routine task would you test first, and how would you judge success?

What should change before a wider trial?

Researcher analysis plan: report sample size, recruitment, viewing condition and missing answers. Summarise item distributions and task outcomes separately; compare screenshots and interactive sessions separately. Code free-text themes with a documented rubric. Do not invent a combined “trust” or “intelligence” score. No participant responses have been collected for this edition.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 71 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

البيان الرابع والثلاثون

المهندس مهند أسامة زهران.

خبير في الذكاء الاصطناعي · Engine AI ZEISMO · JARVIS وSystem Brain

التوجه البحثي

يشرف مهند أسامة زهران على التصور الذي يقوم عليه برنامج JARVIS وSystem Brain: واجهة ناطقة بالإنجليزية، وخلايا متخصصة بحسب الدور، ومهارات وذاكرة منتقاة، ونقل الأعمال الإدارية الروتينية المناسبة على مراحل إلى نظم قرار تعتمد على الذكاء الاصطناعي.

محرك زيزمو للذكاء الاصطناعي

يذكر مهند أسامة زهران في المعلومات التي قدمها لهذه الطبعة أن عمله في Engine AI ZEISMO يتناول تدريب نماذج الذكاء الاصطناعي وتكييفها. ويربط البرنامج هذا التوجه الهندسي بالبحث في تنسيق الوكلاء واستخدام الأدلة والاستقلالية التشغيلية المنضبطة. أما بنى النماذج المحددة ومجموعات التدريب ونتائج أدائه فتتناولها سجلات تقنية منفصلة.

المساهمة والإعداد

تصور المشروع والتوجيه البحثي: مهند أسامة زهران. نُظّمت الأدبيات وفُحصت المصادر ونُفّذت التجربة وتحققت النتائج العددية وصيغ المخطوط وأُعدت الوثيقة بمساعدة الذكاء الاصطناعي. يوثق التقرير المناهج ومجموعات البيانات ونطاق التحقق الفعلي، ولا ينسب إلى المؤلف مراجعة مستقلة أو تنفيذ كل اختبار يدويًا.

توقيع نصّي معتمد

المهندس مهند أسامة زهران

خبير الذكاء الاصطناعي

28 سبتمبر 2026 · التقرير التقني ZEISMO-LAB-TR-2026-001 · الإصدار 1.3

أُدرج التوقيع المطبوع واللقب المهني بطلب صريح من مهند أسامة زهران. يثبت التوقيع نسبة التأليف ولا يمثل توقيعًا رقميًا بشهادة. يحدّد سجل التحقق المصاحب الملف النهائي ببصمة SHA-256.

المصلحة التجارية: يتعلق العمل بأنظمة طُوّرت ضمن برنامج زيزمو للمنتجات والأبحاث. وينبغي مراعاة دور المؤلف في هذا البرنامج عند تفسير البحث. ولا يُسجَّل للبحث تحكيم علمي مستقل.

APPENDIX E · AUTHOR STATEMENT

Engineer Mohanad Osama Zahran.

Artificial Intelligence Expert · ZEISMO AI Engine · JARVIS & System Brain

Research direction

Mohanad Osama Zahran directs the concept behind the JARVIS and System Brain programme: an English-speaking interface, role-specific cells, curated skills and memory, and a staged transition of suitable routine departmental work to AI decision systems.

ZEISMO AI Engine

In the background supplied for this edition, Mohanad Osama Zahran identifies ZEISMO AI Engine as his AI model-training and adaptation work. The programme connects that engineering direction with research on coordinated agents, evidence use and controlled operational autonomy. Specific model architectures, training datasets and training-performance results are subjects for separately documented technical records.

Contribution and preparation

Project conception and research direction: Mohanad Osama Zahran. Literature organisation, source inspection, experiment implementation, numerical verification, manuscript drafting and document production: AI-assisted preparation. The report preserves the actual methods, datasets and verification scope. It does not attribute an independent review or manual execution of every test to the named author.

AUTHORISED TYPED SIGNATURE

Engineer / Mohanad Osama Zahran

Artificial Intelligence Expert

28 September 2026 · Technical report ZEISMO-LAB-TR-2026-001 · v1.3

The typed signature and professional title are included at Mohanad Osama Zahran's express request. This is an author-attribution signature, not a certificate-based digital signature. The companion verification record identifies the final file by SHA-256.

Commercial interest: this work concerns systems developed within the ZEISMO product and research programme. The author's role in that programme is relevant when interpreting the research. Independent scientific peer review is not recorded.

جارڤيس وسيستم برين

نسخة عربية منقّحة لغويًا · ترجمة الصفحة 72 من 72 · الصفحة الأصلية
تلي الترجمة صورة الصفحة الأصلية بما فيها الأشكال والجداول؛ النص الإنجليزي مرجع الاقتباس والأرقام.

زيزمو

مختبر الأبحاث

جارڤيس / سيستم برين

سلسلة الأبحاث · 2026

مختبر ZEISMO للأبحاث

صمّم قرارات يمكن فحص أساسها.

من نموذج تشغيلي أولي إلى برنامج بحثي موثّق.

يوثق هذا الإصدار مقترحًا معماريًا، ونموذجًا أوليًا مفحوصًا بخمس عشرة خلية، وتجارب اصطناعية يمكن تكرارها. ويرسم مسارًا قابلًا للقياس نحو موظفي ذكاء اصطناعي يتحملون مسؤوليات روتينية محددة ويتحسن أداؤهم من خلال خبرة منضبطة.

التوقيعات المقترحة

زهران، مهند أسامة (2026). JARVIS وSystem Brain: نحو اتخاذ القرارات الآلية الخاضعة للمساءلة واستقلال الإدارات تدريجيًا. مختبر زيزمو للأبحاث. التقرير التقني ZEISMO-LAB-TR-2026-001، الإصدار 1.3.

zeismo.org · تشغّله WEARLISTY LLC SOLUTIONS SHOPPING · نسخة تقنية أولية منشورة · 28 سبتمبر 2026

JARVIS و«سيستم برين» · تقرير بحثي تقني · الإصدار 1.3

ZEISMO RESEARCH & LAB

Build decisions that can be examined.

JARVIS & System Brain · Technical research report · Version 1.3

From an operational prototype to a documented research programme.

This edition records an architectural proposal, an inspected fifteen-cell prototype and reproducible synthetic experiments. It sets a measurable path towards AI employees that can assume defined routine responsibilities and improve through controlled experience.

SUGGESTED CITATION

Zahran, M. O. (2026). *JARVIS & System Brain: Towards accountable machine decision-making and progressively autonomous departments.* ZEISMO Research & Lab. Technical report ZEISMO-LAB-TR-2026-001, version 1.3.

zeismo.org

Operated by WEARLISTY SHOPPING SOLUTIONS LLC

Public technical preprint · 28 September 2026