

مادة تحليلية

الكود لا يطير وحده

ورقة تحليلية من مصادر مفتوحة تفصل بين استخدام الذكاء الاصطناعي، وإنتاج المخرجات، والتحقق المستقل، والقدرة التشغيلية؛ وتضع حالة شمال اليمن ضمن تاريخ إقليمي من دون تجاوز حدود الإسناد أو الدليل.

تاريخ المادة

الكاتب

22 سبتمبر 2026

صفوان المضيان

الكلمات المفتاحية: قياس أثر الذكاء الاصطناعي · الأمن السيبراني · الأنظمة متعددة الوكلاء · اليمن · إسناد التهديدات · حوكمة الذكاء الاصطناعي

ورقة تحليلية من مصادر مفتوحة حتى 22 سبتمبر 2026؛ ليست تحقيقًا جنائيًا رقميًا مستقلًا، ولا تجربة جديدة على أداء الأسلحة، ولا بحثًا محكمًا.

الملخص

تفحص هذه الورقة ادعاء أن الذكاء الاصطناعي يرفع القدرة التقنية في البيئات الأمنية الحساسة، انطلاقًا من تقرير أنثروبك عن خلية في شمال اليمن استخدمت Claude Code في عمل برمجي مرتبط بتطوير أسلحة. تقترح الورقة فصل أربع درجات كثيرًا ما تُدمج في السرد العام: استخدام مرصود، ومخرج مُنتج، وتحقيق مستقل، ونتيجة مقبولة. وتخلص القراءة إلى أن المصدر يثبت نشاطًا برمجيًا واختبارًا بدا فاشلاً وعودًا لتحليل الفشل، لكنه لا يثبت هوية حوثية مؤكدة، ولا نشر منظومة تشغيلية، ولا مقدار الإضافة السببية للذكاء الاصطناعي. [1] [2]

تضع الورقة الحالة ضمن سجل إقليمي من الهجمات السيبرانية والقدرات الصاروخية والمسيرة والآثار على الطاقة والملاحة، مع إبقاء كل واقعة في نطاق مصدرها. ثم تناقش قياس الإنتاجية، وضريبة التحقق، وحدود المراجعة متعددة الوكلاء، ورؤية مزود النموذج، وتستخلص متطلبات مؤسسية سعودية تقوم على أقل الصلاحيات، واستقلال الاختبار، وسجل القرار، وقياس الكلفة الكاملة للنتيجة المقبولة. هذه خلاصة تحليلية، لا نتائج تجربة أجراها المؤلف.

المنهج وحدود الأدلة

تستخدم الورقة مراجعة سرديّة تحليلية لمصادر مفتوحة. جُمعت مصادر أولية أو أقرب إلى الواقعة حيث أمكن: تقارير شركات عن نشاط داخل خدماتها، ووثائق أُممية وحكومية، وإعلانات جهات متضررة، وأوراق بحثية، ثم استُخدمت الصحافة الموثوقة لرَدود الأطراف والتطورات التي لا يتوفر لها سجل أولي كامل. حد الأدلة الزمني هو 22 سبتمبر 2026؛ ولا تدخل أي معلومة لاحقة في الاستنتاجات، حتى إذا أصبحت متاحة وقت النشر.

لا تدّعي هذه المراجعة الشمول المنهجي، ولم تُنفذ بروتوكول بحث قواعد بيانات أو تقييمًا كمّيًا لجودة الدراسات. وهي لا تملك سجلات أنثروبليك الخام، ولا ملفات الخلية، ولا بيانات اختبار مستقلة، ولا وصولًا ميدانيًا. لذلك تُعامل إفادات مزود الخدمة بوصفها دليلًا مباشرًا على ما رآه داخل منصبه، لا دليلًا مستقلًا على الأداء خارجها. وتُعامل بيانات الحكومات والأطراف المتحاربة بوصفها روايات منسوبة تحتاج إلى موازنة، لا حقائق محايدة تلقائيًا.

اعتمد التحليل سجلًا للادعاءات: لكل ادعاء مصدر، ونوع دليل، وحدود رؤية، وحالة إسناد. استُبعدت التفاصيل التي قد تتحول إلى إرشاد عملي لتطوير سلاح أو استهدافه، ولم تُستكمل الفراغات التقنية بالتخمين. وعند تعارض الروايات، احتفظ النص بالتعارض بدل حسمه بلغة أقوى من المصدر.

اختيرت الحالات التاريخية لوظيفتها التحليلية: تمييز التخريب من التجسس، والادعاء من الإدانة، والأثر المادي من الانطباع، والدعم من القيادة المباشرة. لا تمثل القائمة حصرًا لجميع الحوادث في المنطقة. كما قورنت الصياغة النهائية بحدود المصادر حتى لا تُستخدم واقعة قديمة لتعويض نقص الدليل في الحالة الجديدة.

1. الواقعة الأصلية قبل أن تبدأ الاستعارات

وصفت أنثروبليك في تقريرها الصادر في سبتمبر 2026 خلية في شمال اليمن استخدمت Claude Code في أجزاء من عمل برمجي مرتبط بعدة برامج لتطوير أسلحة. قالت الشركة إن المستخدمين وزعوا الكتابة والبحث والمراجعة على عدة مثيلات، وأجروا اختبار إطلاق بدا فاشلاً، ثم عادوا إلى الخدمة لتحليل ما حدث. كما قالت إنها حظرت الحسابات وإن لديها دليلًا على أن المستخدمين أنشؤوا أداة محاكاة تعمل محليًا.^[1]

الحد الأكثر أهمية في التقرير هو أن الشركة لا تملك دليلًا على نجاح الفاعلين في نشر منظومة تشغيلية. هذا النفي لا يجعل النشاط غير مهم؛ فهو وصل، وفق المصدر، إلى اختبار مادي ومحاولة تعلم من الفشل. لكنه يمنع تحويل الرصد إلى ادعاء عن سلاح جاهز أو دقة محسنة. لا نعرف من المواد المفتوحة إن كان المخرج البرمجي صحيحًا، ولا ما الذي كان سيفعله الفاعلون من دون الأداة، ولا إن كانت الأداة المحلية مفيدة بعد إغلاق الحسابات.^[1]

ولم تسم أنثروبليك المستخدمين حوثيين. قالت إن الخلية في شمال اليمن، بينما قدمت أسوشيتد برس الصلة بالحوثيين بوصفها استدلالًا من الجغرافيا والسياسي، ونقلت ردًا من حزام الأسد، عضو المكتب السياسي للجماعة. رد الطرف المعني مادة ضرورية لفهم النزاع على الرواية، لكنه لا يتحقق مستقلًا من البيانات التقنية ولا ينفذها. النتيجة المنهجية: الموقع مثبت في تقرير الشركة؛ الهوية التنظيمية غير مؤكدة؛ والنجاح التشغيلي غير مثبت.^[2]

2. التاريخ الإقليمي لا يبدأ في 2026

نشأت الحركة الحوثية داخل صراعات يمنية محلية قبل سيطرتها على صنعاء في 2014، وبدأ تدخل التحالف بقيادة السعودية في 2015 دعماً للحكومة اليمنية المعترف بها دولياً. لا يجوز اختزال هذا التاريخ في علاقة تقنية جديدة، كما لا يجوز حذف العامل الخارجي منه. تقارير فرق خبراء الأمم المتحدة عرضت أدلة وتقييمات عن دعم مادي وتقني وتدريب خارجي أسهم في تطور قدرات الحوثيين، مع استمرار إيران في نفي تسليحهم. [12] [16]

الدعم، ونقل المعرفة، والتنسيق، والقيادة المباشرة ليست أسماء لشيء واحد. قد يثبت مصدر وجود مكونات أو تشابه تقني، ولا يثبت مكان اتخاذ القرار في عملية بعينها. وقد يثبت تدريباً أو مساعدة تقنية، ولا يثبت أن الجهة الداعمة تدير كل استخدام لاحق. هذا الفصل ضروري لأن التحليل السببي يفسد حين يتحول «إيران والحوثيون» إلى متغير واحد بلا فترة أو مهمة أو نوع علاقة.

أما سجل الإطلاقات والهجمات على السعودية والبحر الأحمر فيثبت أن النقاش يتعلق بقدرة مادية قائمة قبل الحالة الجديدة. وثقت الأمم المتحدة أعداداً كبيرة من الصواريخ والمسيّرات وهجمات على منشآت مدنية واقتصادية، ثم وسعت عمليات البحر الأحمر أثر النزاع إلى الملاحة الدولية. لكنها لا تثبت أن خلية أنثروبليك هي امتداد تنظيمي مباشر لأي وحدة سابقة. التاريخ يجعل الفرضية مهمة؛ ولا يمنح الواقعة الجديدة أدلة لم ينشرها مصدرها. [9] [10] [11]

3. من محو الأقراص إلى سرقة المعرفة

تعرضت أرامكو في أغسطس 2012 لهجوم «شمعون» الذي عطل عشرات الآلاف من الأجهزة، ثم ظهرت موجات محدثة استهدفت جهات سعودية في 2016. فرّقت المصادر بين ما أعلنته التحقيقات السعودية، وما نسبته تقييمات استخباراتية أو شركات أمنية إلى فاعلين مرتبطين بإيران. الثابت هو الأثر التخريبي الكبير؛ أما درجة الإسناد فيجب أن تُذكر مع مصدرها بدل تقديمها كحكم قضائي شامل. [18]

في 2017 قيّمت مانديانت أن APT33 تعمل لصالح الحكومة الإيرانية، ووثقت اهتماماً بقطاعات الطيران والطاقة والبتروكيماويات، وبين الأهداف مؤسسات مرتبطة بالسعودية. وفي 2022 وصفت جهات أمريكية MuddyWater بأنها عنصر تابع لوزارة الاستخبارات الإيرانية. تختلف هاتان الحالتان في المصدر والوقائع ودرجة الثقة، ولا تصبحان حادثة واحدة لأن خلفيتهما السياسية متقاربة. [19] [20]

وفي 18 أغسطس 2026 أعلنت وزارة العدل الأمريكية لائحة اتهام بديلة موسعة ضد سبعة عشر عضواً في معهد مينا، بينهم تسعة سبق اتهامهم عام 2018. تتهمهم اللائحة بحملة واسعة لسرقة بيانات أكاديمية وملكية فكرية لصالح جهات إيرانية. تظل الوقائع ادعاءات قضائية إلى أن يُبت فيها؛ إلا أن القضية توضح أن المعرفة البحثية نفسها هدف استراتيجي. [7]

هذا السجل يفسر لماذا يصبح مزود النموذج جزءاً من مشهد استخبارات التهديدات. كانت الأسئلة التقليدية: من دخل الشبكة، وماذا سرق، وماذا عطل؟ يضاف اليوم سؤال: ما القدرة التي حاول مستخدم بناءها بمساعدة المنصة؟ ولا يجوز أن يقود التشابه في الأسئلة إلى دمج الجهات أو الحالات بلا دليل. الإسناد هو ربط واقعة محددة بأثر محدد ومصدر محدد، لا اختيار الاسم الأكثر حضوراً في الذاكرة السياسية.

4. حين تخرج المخاطر من الشاشة

في 14 سبتمبر 2019 أعلنت أرامكو أن الهجمات على بقيق وخريص أدت إلى تعليق إنتاج قدره 5.7 ملايين برميل يوميًا. وقوع الهجوم وأثره التشغيلي ثابتان في إعلان الجهة المتضررة، لكن إعلان المسؤولية، ومنشأ المعدات، ومكان الإطلاق، والقرار السياسي أسئلة مختلفة تحتاج إلى أدلة مختلفة.^[21]

وفي 17 يناير 2022 أدى هجوم على أبوظبي تبناه الحوثيون إلى مقتل ثلاثة مدنيين. ثم اتسع الأثر الاقتصادي مع هجمات البحر الأحمر: قدّر صندوق النقد الدولي أن حجم التجارة العابرة لقناة السويس انخفض قرابة النصف في أول شهرين من 2024 مقارنة بالفترة نفسها من 2023، مع تحويل مسارات شحن وإطالة زمن التسليم. الرقم يصف فترة محددة ولا يصلح مؤشرًا آنيًا لعام 2026، لكنه يبين كيف ينتقل أثر أمني من موقع الهجوم إلى التأمين والمخزون وسلسلة الإمداد.^[22] [23]

وفي سبتمبر 2026 وسع برنامج الأغذية العالمي استجابته في اليمن مع تجدد القتال والنزوح. لا تربط مصادر البرنامج أو رويترز هذا التصعيد بحالة كلود، ولا يجوز فعل ذلك تحريريًا. إدراج السياق الإنساني يذكرنا بأن قياس القدرة لا ينتهي عند أداء قطعة تقنية؛ فالنتيجة الاجتماعية قد تظهر في الغذاء والتنقل واستمرارية الخدمات.^[4] [5]

من زاوية تحليل النظم، القدرة ليست قطعة بل سلسلة اعتماد: معلومة، وبرمجيات، وعتاد، وتكامل، واختبار، وتشغيل، ولوجستيات. لا يحتاج التحليل العام إلى وصف تقنيات الاستهداف أو دمج المستشعرات. يكفي أن ندرك أن ضعف أي حلقة قد يحد النتيجة، وأن تحسن العمل البرمجي لا يثبت تحسن السلسلة كلها.

5. تعريف الإضافة قبل قياسها

قولنا إن الذكاء الاصطناعي «زاد القدرة» لا يكتمل قبل تحديد النتيجة. قد نعني تقليل زمن مهمة، أو خفض ساعات الخبير، أو تحسين الجودة، أو زيادة عدد المحاولات، أو توسيع قائمة المشروعات الممكنة. هذه المقاييس ليست مترادفة. يمكن أن تزيد كمية المخرجات وتخفض موثوقيتها، أو تختصر الكتابة بينما تنتقل الكلفة إلى المراجعة، أو ينجح جزء تقني بينما يبقى المشروع مقيدًا بالاختبار والاعتماد.

في لغة الاستدلال السببي نريد الفرق بين النتيجة مع الأداة والنتيجة من دونها. لكن التاريخ يعرض مسارًا واحدًا، ولا ينفذ الواقعة نفسها مرتين. رؤية مستخدم يطلب مساعدة لا تخبرنا كم كان سيحتاج من الوقت من دونها. ورؤية اختبار فاشل لا تكشف إن كانت الأداة قريبته من النجاح أو أبعدته. لذلك لا توفر حالة شمال اليمن معاملاً رقميًا لرفع القدرة.

يمكن مع ذلك صياغة فرضية أضيق: قد تقلل النماذج احتكاك العمل المعرفي، فتجعل البحث والكتابة والمقارنة وإنتاج المسودات أسرع أو أرخص. إذا ثبت ذلك فالأثر الاستراتيجي قد يكون زيادة عدد المحاولات، لا ضمان نجاح كل محاولة. وهذه فرضية قابلة للاختبار، وليست نتيجة نستخرجها من تقرير واحد.

في 9 سبتمبر 2026 وصفت مجموعة استخبارات التهديدات في غوغل انتقال بعض الخصوم من الطلبات البسيطة إلى سير عمل توكيلي وأتمتة مترابطة. يدعم التقرير وجود اتجاه أوسع نحو تنظيم العمل بالوكلاء، لكنه لا يتحقق من الحالة اليمنية ولا يثبت أن هذه البنية أفضل من العمل البشري أو من أداة واحدة.^[3]

6. عندما تخالف ساعة الإيقاف شعور المستخدم

قدمت METR مثالاً مفيداً على المسافة بين الإحساس والقياس. في تجربة عشوائية منشورة في يوليو 2025، أنجز ستة عشر مطوراً متمرساً 246 مهمة في مستودعات يعرفونها. في سياق الدراسة استغرقوا وقتاً أطول بنسبة 19% حين أُتيحت لهم أدوات الذكاء الاصطناعي، مع أنهم اعتقدوا بعد التجربة أنها سرعتهم. لا تثبت النتيجة أن الذكاء الاصطناعي يبطل جميع المبرمجين؛ فهي تخص مستخدمين ومهام وأدوات وفترة بعينها. [26]

في فبراير 2026 نشرت METR تحديثاً قالت فيه إن اتساع تبني الوكلاء غير من يقبل المشاركة وأي مهام يقدمها للدراسة، وأضعف صلاحية التصميم السابق لتقدير مقدار الأثر. رجحت أن أدوات أواخر 2025 قد تكون سرعت بعض العمل، لكنها وصفت البيانات بأنها ضعيفة لتحديد الحجم بسبب تحيز الاختيار ومشكلات قياس الوقت المتوازي. الباحث الجاد لا يمنح نتيجة قديمة إقامة دائمة، ولا يحول تحديثاً غير حاسم إلى نصر للرواية المقابلة. [27]

وفي مذكرة لاحقة فرقت المؤسسة بين أثر الأداة على المهام القديمة، وأثرها على المهام الجديدة التي يختارها الناس بعد توفرها، والأثر على القيمة الكلية. قد تغير التقنية قائمة ما نحاول تنفيذه، لا سرعة القائمة السابقة وحدها. هذا التفريق شديد الصلة بحالة أمنية: الأداة قد تجعل محاولات لم تكن تستحق الكلفة سابقاً قابلة للتجربة، من دون أن تحسن جودة كل نتيجة. [28]

الدروس هنا منهجية. استطلاع الرأي يقيس الاعتقاد؛ وسجل الوقت يقيس مدة محددة؛ واختبار الجودة يقيس خصائص مخرج؛ والنتيجة التشغيلية تقيس قيمة في بيئة حقيقية. لا ينبغي لمقياس واحد أن يتنكر في صورة المقاييس كلها.

7. ضريبة التحقق والواقع الذي لا يقرأ العرض

لنفترض مثالاً مدنياً: مشروع يحتاج ساعتين لإنتاج مسودة برمجية وثمانية ساعات لمراجعتها واختبارها ودمجها. حتى إذا اختفت الساعتان تماماً، لا يختفي زمن المشروع. هذا مثال توضيحي لا نتيجة تجربة، لكنه يكشف كيف قد نحتفل بتسارع الجزء المرئي بينما يبقى القيد الحقيقي في الاختبار والموافقة والتشغيل.

في الأعمال الحساسة يوجد سؤالان: هل بُني النظام وفق المواصفات؟ وهل المواصفات تمثل المشكلة الواقعية؟ قد تتجاز البرمجيات اختباراً صحيح التنفيذ لكنه مبني على افتراضات ناقصة. وقد تنتج المحاكاة مخرجات متسقة داخل عالمها، بينما يهدم الواقع افتراضاً لم يدخل إليها. لذلك لا تكفي عبارة «نجحت المحاكاة» من دون معرفة نطاقها وما استبعدته.

كتب ريتشارد فاينمان في ملحق تقرير تشالنجر أن الواقع يجب أن يتقدم على العلاقات العامة، لأن الطبيعة لا يمكن خداعها. قيمة الفكرة في السؤال الذي تفرضه: أين يوجد الاختبار القادر على مخالفة روايتنا؟ إذا كانت المنظومة تنتج الحل، وتفسر صحته، ثم تمنحه شهادة النجاح، فقد أنشأنا دائرة متماسكة لغوياً، لا تحققاً مستقلاً. [29]

ولهذا يقترح سلم الدليل أربع درجات: استخدام مرصود، ثم مخرج مُنتج، ثم تحقق مستقل، ثم نتيجة مقبولة. الانتقال بين الدرجات يحتاج دليلاً جديداً. حالة شمال اليمن تصل في المصدر إلى استخدام ومخرجات واختبار بدا فاشلاً؛ ولا تصل إلى قبول تشغيلي مثبت.

8. تعدد الوكلاء لا يضمن تعدد مصادر الخطأ

يمكن أن يفيد تقسيم العمل بين وكيل يكتب وآخر يبحث وثالث يراجع. لكن الاستقلال لا يأتي من عدد النوافذ. إذا اشتركت الوكلاء في البيانات نفسها، أو معيار نجاح واحد، أو افتراض خاطئ، فإن اتفاقها لا يصحح الافتراض. وإذا رأى المراجع الإجابة قبل أن يبني اختبارها فقد يصبح دوره تحسين الصياغة، لا تحدي النتيجة.

حللت نسخة أولية منشورة على arXiv بعنوان «لماذا تفشل أنظمة النماذج اللغوية متعددة الوكلاء؟» ألقاً وستمنه واثنين وأربعين سجلاً عبر سبعة أطر، وحددت أربعة عشر نمطاً للفشل ضمن تصميم النظام، وسوء التنسيق بين الوكلاء، والتحقق من المهمة وإنائها. طوّر الخبراء التصنيف من عينة أولية، ثم وُسع الوسم إلى المجموعة الكاملة بأداة آلية جرى التحقق منها مقابل وسم بشري؛ ولم يوسم الخبراء كل السجلات يدوياً. لا تتناول الدراسة الخلية البيئية، لكنها تفند القفزة من «عدة وكلاء» إلى «مراجعة موثوقة».^[25]

تحتاج المراجعة الحقيقية إلى اختلاف في الأدلة والمهام. يبدأ المراجع من مواصفات واضحة، ويستطيع الوصول إلى مصدر أولي أو اختبار لا ينتجه الكاتب، ويملك صلاحية إعلان أن الدليل غير كاف. ويمكن قياس قيمته بعدد العيوب التي اكتشفها قبل النشر، لا بعدد التعليقات التي أضافها إلى المحادثة.

في هندسة الاعتمادية، لا يكفي أن تسأل هل أكملت المنظومة المهمة. يجب أن تسأل هل اكتشفت الحالة التي ينبغي عندها أن تتوقف، وهل حفظت أثر القرار، وهل نقلت عدم اليقين بوضوح. الاستقلال هنا خاصية تصميم واختبار، لا اسم دور في مخطط العمل.

9. المخرجات تبقى، والمنصة ترى نافذتها

قالت أنثروبك إن المستخدمين أنشؤوا أداة محاكاة محلية قبل حظر حساباتهم. لا يثبت وجود الأداة أنها صحيحة أو نافعة. لكنه يعني أن إنهاء الوصول لا يسترجع المخرجات تلقائياً. المؤسسة التي تنهي اشتراكاً خارجياً قد تبقي ملفات أو كوداً أو قرارات أنتجتها بمساعدته. لذلك ينبغي تقييم التدخل عبر ثلاثة أسئلة: ما الوصول الذي توقف؟ ما المخرجات التي نعرف أنها خرجت؟ وما الدليل على استخدامها اللاحق؟^[1]

في المقابل، تمنح سجلات مزود النموذج رؤية لا تملكها أدوات أمنية أخرى. قد يرى المزود أجزاء من تسلسل حل المشكلة، وتغير الطلبات، واستعمال الأدوات. هذه الرؤية تبرر التعامل مع تقارير المزود بوصفها مصدراً مهماً، لكنها لا تمنحه علماً بما حدث بعد تنزيل الملف أو في بيئة لا تمر بخدمته.

ويذكر تقرير أنثروبك أن الحالات المختارة ليست ممثلة للاستخدام المعتاد، بل اختيرت لأنها لافتة أو جديدة. لا يمكن إذن استخدام عدد الحالات المنشورة لتقدير معدل الانتشار. زيادة الإفصاحات قد تعكس زيادة النشاط، أو تحسن الكشف، أو تغير سياسة النشر، أو مزيجاً منها. المقام المفقود يمنع تحويل العدد إلى نسبة.^[1]

الصحافة لا تحل هذه المشكلة بمجرد التكاثر. خمسون مادة تعيد الإفصاح نفسه قد توسع الشرح وتضيف مقابلات، لكنها لا تصبح خمسين شاهداً مستقلاً على الادعاء الأصلي. استقلال الدليل يُفحص لكل ادعاء، لا لكل شعار مؤسسة نشر.

10. حوكمة سعودية تقيس القيمة الكاملة

تملك السعودية سببًا مباشرًا للاهتمام بإساءة استخدام الذكاء الاصطناعي، وسببًا مباشرًا أيضًا لاستخدامه في تحسين الإنتاج والصمود. لذلك لا ينبغي أن تصبح الحوكمة مرادفًا للمنع. المطلوب هو تصميم استخدام يستطيع إثبات قيمته، واحتواء خطئه، والحفاظ على الخدمة حين تفشل الأداة أو تتغير.

أطلقت الهيئة الوطنية للأمن السيبراني في 5 يوليو 2026 مشاوره عامة على إرشادات للأمن السيبراني للذكاء الاصطناعي، تشمل الحوكمة والدفاع والصمود والأطراف الخارجية، وتمتد إلى الأنظمة التوليدية والتوكيلية. المصدر يثبت المشاورة وموضوعاتها، ولا يثبت أن نصًا نهائيًا ملزمًا صدر في ذلك التاريخ.^[8]

المبدأ التنفيذي الأول هو أقل الصلاحيات: يحصل الوكيل على الحد اللازم للمهمة، ويُفصل بين القراءة والتعديل، وبين الاختبار والإنتاج، وبين التوصية والتنفيذ. لا تمنح البلاغة ولا تسمية «الوكيل المستقل» حق الوصول. وكل انتقال إلى إجراء مرتفع الأثر يحتاج شرطًا صريحًا ومراجعة بشرية موثقة أو ضابطًا تقنيًا مناسبًا للمخاطر.^[30]

المبدأ الثاني هو سجل النتيجة. عند مراجعة قرار ساعد فيه الذكاء الاصطناعي، ينبغي معرفة نسخة الأداة، ومصادر المعلومات ذات الصلة، والتعديلات التي أجريت، والاختبار الذي طُبّق، ومن أجاز الاستخدام. لا يعني ذلك حفظ كل كلام الموظفين؛ بل الاحتفاظ بالقدر المهني اللازم للتدقيق، مع ضبط الوصول والمدة.

المبدأ الثالث هو قياس الكلفة لكل نتيجة مقبولة وفق اختبار مستقل. تدخل في الكلفة مرحلة إعداد المهمة، وتشغيل الأداة، والمراجعة، وإعادة العمل، ومعالجة العيوب اللاحقة. وتدخل في النتيجة جودة محددة، لا مجرد اكتمال ملف أو إغلاق تذكرة. بهذا القياس نستطيع التمييز بين زيادة النشاط وزيادة القيمة.

11. القيود وما لا تسمح به هذه الورقة

تعتمد الواقعة المركزية على إفصاح أنثروبك، ولا توجد في مجموعة المصادر مراجعة مستقلة لسجلات الحساب أو الاختبار المادي. لا يمكن التحقق من هوية الخلية التنظيمية، أو قياس جودة الكود، أو معرفة مصير المخرجات بعد الحظر. كما لا يمكن تقدير السيناريو المضاد: ما الذي كان سينجزه المستخدمون، وفي أي زمن، من دون كلود.

المصادر الإقليمية متفاوتة. تقارير الأمم المتحدة، وإعلانات الحكومات والجيش، وتصريحات الجماعات، وتقارير الشركات والصحافة، لكل منها نافذة ومصلحة ومعيّار إثبات مختلف. جمعت الورقة هذه المصادر لتحديد السياق، لا لدمجها في سجل واحد متجانس. وقد بقيت روايات متعارضة بشأن بعض الهجمات والإسنادات من دون حسم.

هذه مراجعة سردية مفتوحة المصدر، لا مراجعة منهجية ولا تحليلًا إحصائيًا تجميعيًا. اختيار الأمثلة قد يعطي وزنًا أكبر للحالات المعروفة أو القابلة للنشر. كما أن تقارير مزودي النماذج تخضع لقدرات الكشف وسياسات الإفصاح، ولا تقدم مقامًا يسمح بقياس الانتشار.

ولا تقدم الورقة تقييمًا تقنيًا لمنظومة سلاح أو وصفًا لكيفية تطويرها. جرى تعميم التفاصيل المتعلقة بالتصميم والاستهداف والتشويش لأنها غير لازمة للحجة وقد تحول التحليل إلى مادة تشغيلية. الاستنتاجات تتعلق بقياس العمل المعرفي والحوكمة وحدود الأدلة، لا بهندسة السلاح.

كما يحتمل السجل المفتوح انحياز النشر: الشركات تعرض الحالات التي اكتشفتها واختارت الإفصاح عنها، والضحايا قد لا يعلنون كل أثر، والنجاحات والفشل لا يظهران بالمعدل نفسه. لا تقدم المصادر مقامًا مشتركًا يسمح بحساب الانتشار أو المقارنة بين المزودين. لذلك لا تستنتج الورقة اتجاهًا كمياً من تزايد التقارير وحده.

12. أجندة بحث مدنية، ثم السؤال الباقي

المسار الأول هو أثر الخبرة السابقة. يمكن مقارنة مجموعات متفاوتة الخبرة في مهام برمجية مدنية آمنة، مع توزيع عشوائي للأداة وتقييم أعمى للجودة وقياس الزمن الكامل. السؤال: هل تمنح الأداة أكبر فائدة لمن يستطيع اكتشاف أخطائها، أم تقلص الفجوة فعلاً؟ يجب أن يسمح التصميم بظهور النتيجة.

المسار الثاني هو انتقال العبء إلى التحقق. تسجل مراحل العمل منفصلة: إعداد المهمة، وإنتاج المسودة، والمراجعة، والاختبار، وإعادة العمل. ثم تُقاس العيوب التي تظهر قبل القبول وبعده. قد يتبين أن الأداة سرعت الإنتاج والمراجعة معاً؛ عندها يجب تعديل فرضية «ضريبة التحقق»، لا حماية الفكرة لأنها أنيقة.

المسار الثالث هو استقلال المراجعة الآلية. يمكن مقارنة مراجع يرى إجابة الكاتب بمراجع يبدأ من المواصفات والأدلة الخام، ثم قياس العيوب المكتشفة والمتفق على تفويتها. والمهم ألا يكون عدد الوكلاء هو المتغير الوحيد؛ بل اختلاف الأدلة، وحق الاعتراض، ومعيار التوقف.

المسار الرابع هو صلاحية النتائج بالعربية وفي بيئات العمل المحلية. لا نفترض مسبقاً ضعف الأداء أو تفوقه؛ نختبر اللغة والمجال والوثائق التي تستخدمها المؤسسة فعلاً. والغاية ليست إنتاج ترتيب تسويقي للنماذج، بل معرفة أين تتغير الجودة والكلفة وخطر السهو حين تتغير اللغة والسياق.

ويمكن إضافة مسار خامس للحوكمة نفسها: مقارنة فرق تستخدم الصلاحيات الدنيا وسجل قرار واختباراً مستقلاً بفرق تستخدم الأداة من دون هذه الضوابط، في مهام مدنية آمنة ومتقاربة. يقاس الأثر في الأخطاء المتأخرة، وزمن الاسترداد، وكلفة التدقيق، لا في الرضا وحده. بهذه الطريقة تصبح الحوكمة فرضية قابلة للقياس، لا قائمة بمبادئ جميلة معلقة على الجدار.

تضعنا حالة كلود أمام مسألتين: تاريخ إقليمي موثق من الهجمات وتطور الأدوات، وحالة جديدة لا تكفي أدلتها لإثبات أثر عسكري ناجح. التاريخ يجعل السؤال ملحاً، لكنه لا يملأ فراغات المصدر. القراءة الاستراتيجية تحدد ما تغير، وما لم نعرف بعد إن كان قد تغير، وما الاختبار الذي يفصل بينهما.

الخطر لا يحتاج إلى مبالغة كي يستحق الانتباه، والثقة لا تصبح حماية لأنها صيغت ببلاغة. الكود، مهما بدا مقنعاً على الشاشة، لا يطير وحده. بينه وبين النتيجة عتاد وبيانات واختبار ومسؤولية. وفي تلك المسافة تقع قيمة البحث والحوكمة معاً.

سُلّم الدليل من الاستخدام إلى الأثر

4
نتيجة مقبولة3
تحقق مستقل2
مخرج مُنتج1
استخدام مرصود

لا يجوز القفز بين الدرجات: كل انتقال يحتاج معيارًا ودليلاً منفصلين عن الجهة التي أنتجت المخرج.

المراجع

1. Detecting and countering misuse of AI: September 2026
Anthropic
2. Users in Houthi-held Yemen tried to develop advanced weapons with AI, Anthropic says
Associated Press, 11 سبتمبر 2026
3. GTIG AI Threat Tracker: From Prompting to Autonomy – The Evolution of Adversarial AI
Google Threat Intelligence Group, 09 سبتمبر 2026
4. World Food Programme scales up response in Yemen as fighting displaces thousands
World Food Programme, 17 سبتمبر 2026
5. World Food Programme ramps up response in Yemen as hunger fears escalate, official says
Reuters, 21 سبتمبر 2026
7. 17 Iranians Charged With Conducting Massive Cyber Theft Campaign On Behalf Of The Islamic Revolutionary Guard Corps And Other Iranian Entities
U.S. Department of Justice, 18 أغسطس 2026
8. NCA Launches Public Consultation on AI Cybersecurity Guidelines
National Cybersecurity Authority, 05 يوليو 2026
9. Final report of the Panel of Experts on Yemen
United Nations Security Council
10. United Nations Country Team in Yemen Annual Report 2021
United Nations in Yemen
11. Red Sea ballistic missile attacks trigger Asian interest in defences
Reuters, 21 فبراير 2024

<u>Final report of the Panel of Experts on Yemen</u>	.12
United Nations Security Council, 11	أكتوبر 2024
<u>Final report of the Panel of Experts on Yemen</u>	.16
United Nations Security Council, 17	أكتوبر 2025
<u>Shamoon 2: Return of the Distrack Wiper</u>	.18
Unit 42, 30	نوفمبر 2016
<u>APT33: Insights into Iranian Cyber Espionage</u>	.19
Mandiant, 20	سبتمبر 2017
<u>Iranian Government-Sponsored Actors Conduct Cyber Operations Against Global Government and Commercial Networks</u>	.20
CISA, 24	فبراير 2022
<u>Incidents at Abqaiq and Khurais</u>	.21
Saudi Aramco, 14	سبتمبر 2019
<u>Security Council Press Statement on Terrorist Attacks in Abu Dhabi, United Arab Emirates</u>	.22
United Nations Security Council, 21	يناير 2022
<u>Red Sea Attacks Disrupt Global Trade</u>	.23
International Monetary Fund, 07	مارس 2024
<u>Why Do Multi-Agent LLM Systems Fail?</u>	.25
arXiv, 17	مارس 2025
<u>Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity</u>	.26
METR, 10	يوليو 2025
<u>We are Changing our Developer Productivity Experiment Design</u>	.27
METR, 24	فبراير 2026
<u>Task Substitution and Uplift</u>	.28
METR, 08	مايو 2026
<u>Report of the Presidential Commission on the Space Shuttle Challenger Accident, Appendix F</u>	.29
NASA	
<u>The Protection of Information in Computer Systems</u>	.30
Proceedings of the IEEE	