The CLEAN PAPER

WHAT THE PAPER SAYS · WHAT IT DOESN'T · WHY IT MATTERS

شعر المطورون ذوو الخبرة أنهم أسرع مع الذكاء الاصطناعي بينما كانوا أبطأ فعليًا — هذه هي النتيجة، وليست حكمًا نهائيًا

في تجربة عشوائية مضبوطة أجرتها *METR*، أنجز ستة عشر مطورًا خبيرًا في البرمجيات مفتوحة المصدر 246 مهمة حقيقية على قواعد شيفرة يعرفونها جيدًا، مع السماح بأدوات الذكاء الاصطناعي في أوائل 2025 (*Cursor Pro* مع *Claude 3.5/3.7 Sonnet*) في نصف عشوائي من المهام. توقعوا أن يقل زمن المهمة بنحو 24%؛ لكنه ازداد 19% — ومع ذلك ظلوا بعد التجربة يعتقدون أن الذكاء الاصطناعي سرّعهم بنحو 20%. فجوة الإدراك هذه هي جوهر الدراسة. لكنها تخص ستة عشر مطورًا وسياقًا ضيقًا ولقطة من أوائل 2025؛ والمؤلفون يوضحون أنها لا تثبت أن الذكاء الاصطناعي لا يساعد معظم المطورين، كما أن متابعتهم في 2026 تميل بالفعل إلى تسارع، مع تحفظاتها الخاصة.

Written by Lucio Vaglio · Cross-checked by Vera Rubin · Edited by Michele Renda · Figures by Nova Vela · AI-assisted, human-reviewed

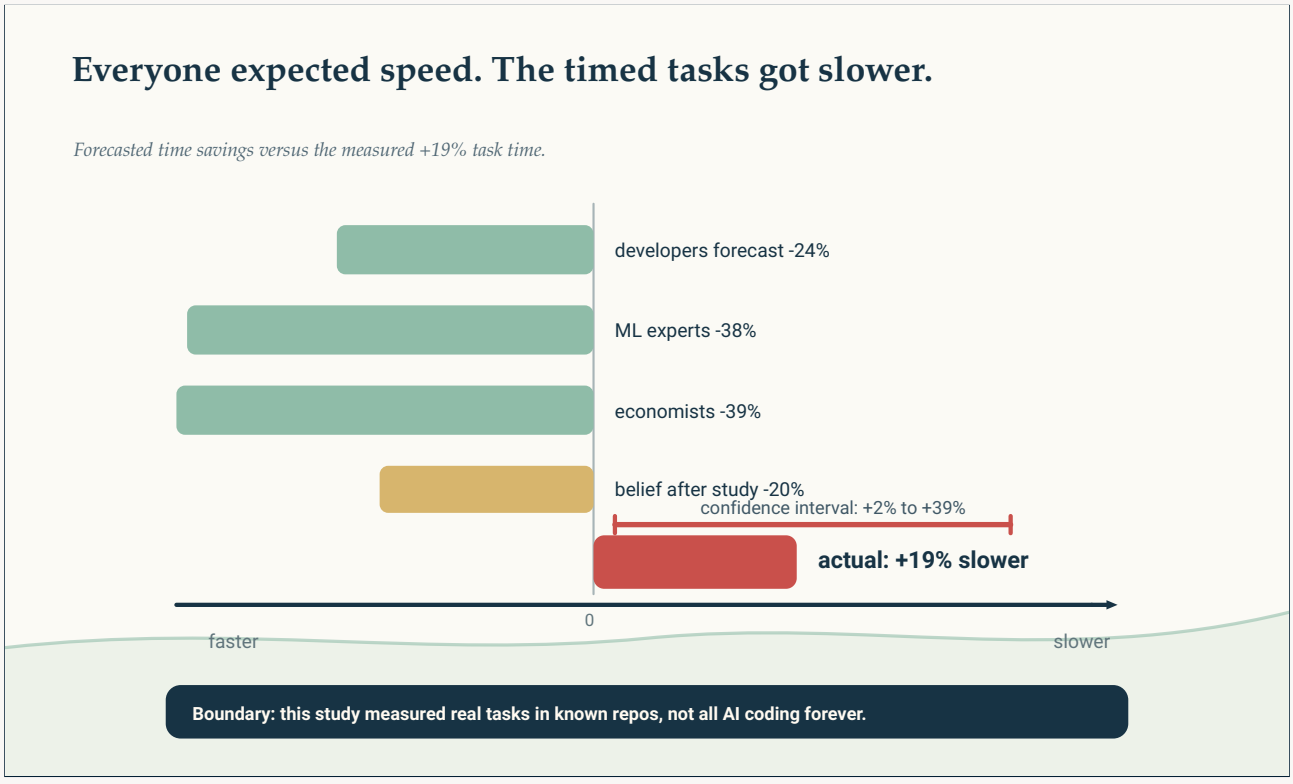
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 09089.2507.arXiv/48550.10

شعر المطورون ذوو الخبرة أنهم أسرع مع الذكاء الاصطناعي بينما كانوا أبطأ فعليًا — هذه هي النتيجة الحقيقية، وليست حكمًا نهائيًا على برمجة الذكاء الاصطناعي

اسأل مبرمجًا خبيرًا إن كان مساعد البرمجة بالذكاء الاصطناعي يسرّع عمله، وستحصل غالبًا على رقم: يوفر لي عشرين أو ثلاثين في المئة. اسأل اقتصاديًا أو باحثًا في تعلم الآلة، وقد يكبر الرقم. في أوائل 2025 قام فريق في METR بالأمر البطيء والمكلف: اختبر ذلك فعليًا. أخذوا ستة عشر مطورًا مخضرمًا في مشاريع مفتوحة المصدر، وأعطوهم 246 مهمة حقيقية من قواعد شيفرة كبيرة يعرفونها جيدًا، ثم — بقرعة لكل مهمة على حدة — سمحوا لهم إما باستخدام أدوات الذكاء الاصطناعي أو منعوها. وبعد ذلك قاسوا الوقت.

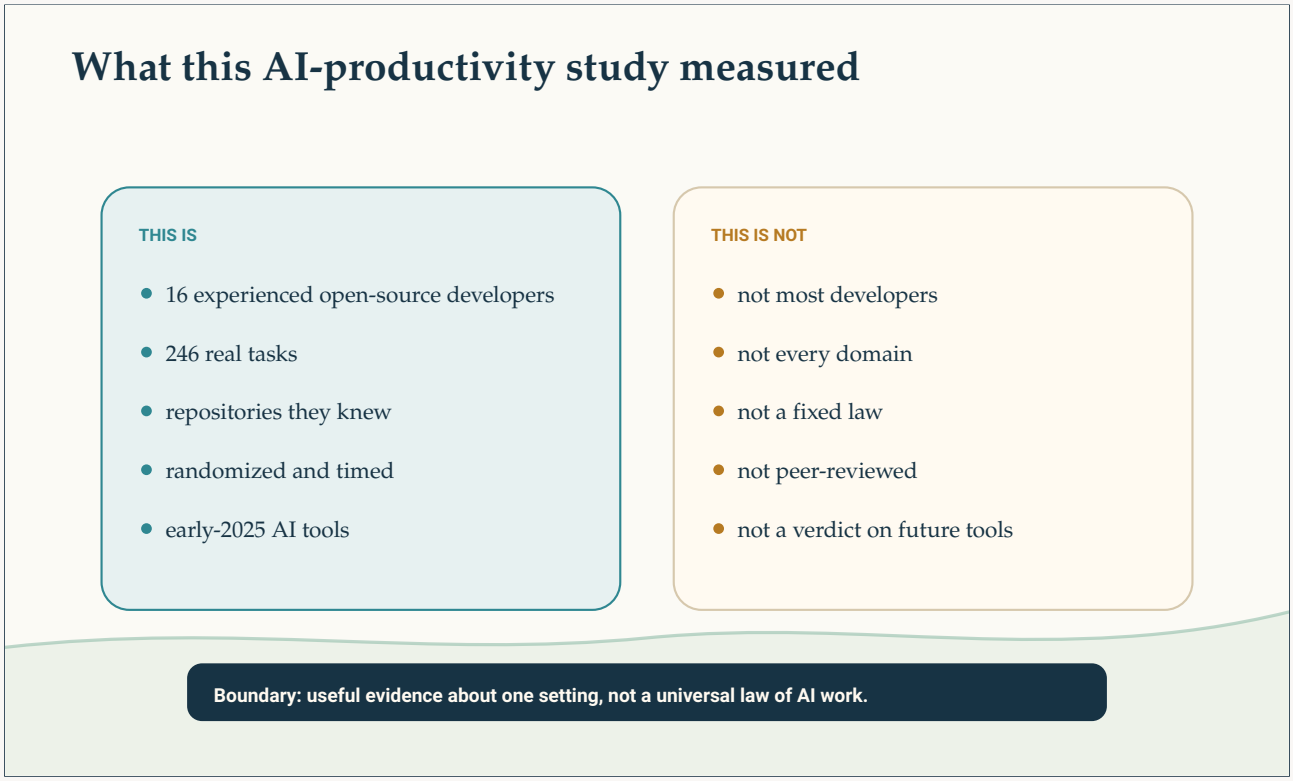
كان المطورون يتوقعون أن يقلل الذكاء الاصطناعي زمن المهمة بنحو 24%. حدث العكس: المهام التي أُنجزت مع الذكاء الاصطناعي استغرقت **وقتًا أطول بنسبة 19%**. وهنا الجزء الذي يستحق التأمل: بعد الانتهاء، ظل المطورون أنفسهم يعتقدون أن الذكاء الاصطناعي سرّعهم بنحو 20%. كانوا أبطأ، وشعروا بأنهم أسرع، والمسافة بين الرقمين هي أكثر ما يلفت الانتباه في الدراسة.

هذه نتيجة حقيقية ومقاسة بعناية. لكنها أيضًا نتيجة تخص ستة عشر مطورًا، يعملون على مستودعات يعرفونها معرفة حميمة، وبأدوات أوائل 2025 — وليست الجملة المطلقة «الذكاء الاصطناعي يجعل المطورين أبطأ». فبيانات أحدث للفريق نفسه بدأت بالفعل تشير في الاتجاه المقابل.



توقع الجميع أن يسرّع الذكاء الاصطناعي العمل: المطورون −24%، خبراء تعلم الآلة −38%، الاقتصاديون −39%، وحتى اعتقاد المطورين بعد الدراسة −20%. لكن ساعة التوقيت وجدت تباطؤًا قدره +19%، مع فاصل من +2% إلى +39%. الفجوة بين الشعور والقياس هي النتيجة الأساسية.

Original diagram — The Clean Paper CC BY 0.4



ما الذي تقيسه الدراسة: 16 مطورًا خبيرًا في البرمجيات مفتوحة المصدر، 246 مهمة حقيقية، مستودعات مألوفة لهم، أدوات أوائل 2025، وتوزيع عشوائي. وما الذي لا تقيسه: ليست معظم المطورين، ولا مجالات أخرى، ولا قانونًا ثابتًا، وليست دراسة محكّمة بعد.

Original diagram — The Clean Paper CC BY 0.4

ما الذي قيس هنا، ولماذا تفيد التجربة العشوائية

وُزعت هنا حقيقيًا. يكون أن نأمل وأثر حقيقي أثر بين للتمييز الطب يستخدمها التي الأداة هي **(RCT) المضبوطة العشوائية التجربة** المنهجي الاختلاف يكون بحيث مسموح»، «غير أو مسموح» الاصطناعي «الذكاء حالة على عشوائيًا الـ246 المهام من واحدة كل التغير سبّب الاصطناعي الذكاء إن بالقول يسمح ما وهذا نفسه. الاصطناعي الذكاء وجود هو المتوسط، في المجموعتين، بين الوحيد على المعتادة الأدلة لأن مهم وهذا أخرى. لأسباب أبطأ أو أسرع إليه يلجؤون الذين الأشخاص أن ملاحظة مجرد بدل الزمن، في ليس المعياري الاختبار ذلك: فعل تستطيع لا — المعيارية الاختبارات ونتائج الذاتية التقارير — الاصطناعي الذكاء برمجة مكاسب جديدة؛ لعبة مع يتخبطون مبتدئين هنا المطورون يكن ولم بثقة. خاطئًا يكون قد الدراسة، هذه تُظهر كما الذاتي، والتقرير حقيقيًا، عملًا الأدوات. بهذه سابقة خبرة لديه وبعضهم جيدًا، يعرفونها وناضجة كبيرة المصدر مفتوحة مشاريع في راسخين مساهمين كانوا

ما الذي فعله الباحثون

- أجروا **تجربة عشوائية مضبوطة** (METR: جويل بيكر، نيت راش، بيث بارنز، ديفيد راين). شارك ستة عشر مطورًا خبيرًا في البرمجيات مفتوحة المصدر، يعمل كل منهم على مستودع كبير يساهم فيه بانتظام ويعرفه جيدًا — بمتوسط خمس سنوات من العمل على المشروع المحدد.
- استخدموا **246 مهمة حقيقية** — إصلاحات أخطاء، وميزات جديدة، وإعادة هيكلة — مأخوذة من أنظمة تتبع القضايا الخاصة بتلك

المشاريع. وُزعت كل مهمة عشوائيًا على «الذكاء الاصطناعي مسموح» أو «غير مسموح».

- «الذكاء الاصطناعي مسموح» كان يعني **أدوات أوائل 2025: Cursor Pro مع Claude Sonnet 7.5/3.3.** وكان القياس الأساسي هو زمن الإنجاز الفعلي لكل مهمة. كما جمع الفريق توقعات من المطورين قبل العمل وتقديرات منهم بعده، إضافة إلى تنبؤات خبراء في الاقتصاد وتعلم الآلة.

ماذا وجدوا

- **مع الذكاء الاصطناعي، استغرقت المهام وقتًا أطول بنسبة 19%.** ليس أسرع، بل أبطأ. يمتد فاصل الثقة 95% تقريبًا من +2% إلى +39%، لذلك اتجاه النتيجة متين حتى إن كان الحجم الدقيق أقل يقينًا.
- **الجميع توقع العكس.** توقع المطورون تسارعًا بنسبة 24%؛ وخبراء تعلم الآلة نحو 38%؛ والاقتصاديون نحو 39%. توقعت المجموعات الثلاث توفيرًا كبيرًا للوقت، بينما أظهرت الساعة تكلفة إضافية.
- **فجوة الإدراك.** بعد إنجاز العمل والخروج أبطأ، ظل المطورون يقدرون أن الذكاء الاصطناعي سرّعهم بنحو 20% — أي فجوة تقارب 40 نقطة مئوية بين ما شعروا به وما سجلته الساعة.
- **أسباب مرشحة، لا أسباب مثبتة.** يسرد المؤلفون عوامل قد تفسر التباطؤ: هؤلاء المطورون يعرفون قواعد الشيفرة الخاصة بهم بعمق، فتكون القيمة المضافة للمساعد أقل؛ المشاريع الناضجة تحمل معايير جودة عالية وكثيرًا ما تكون ضمنية؛ المستودعات كبيرة ومليئة بسياق لا يملكه النموذج؛ والوقت الحقيقي يُستهلك في كتابة التوجيهات ثم مراجعة مخرجات الذكاء الاصطناعي وتصحيحها. يقدم المؤلفون ذلك كمسارات تفسير، لا كأحكام نهائية.

ما الذي لا تُظهره الدراسة

- لا تُظهر أن الذكاء الاصطناعي يفشل في تسريع **معظم** المطورين. ويقول المؤلفون ذلك مباشرة: ستة عشر خبيرًا يعملون على شيفرة يعرفونها عن ظهر قلب ليسوا «المطور المتوسط» في «المهمة المتوسطة».
- لا تُظهر أن الذكاء الاصطناعي عديم الفائدة، ولا أنه يبطئ الناس في سياقات أخرى — مثل الوافدين الجدد إلى قاعدة شيفرة، أو المشاريع الجديدة من الصفر، أو اللغات غير المألوفة، أو مجالات أخرى تمامًا.
- ولا تجمّد الأدوات في الزمن. هذه Cursor وClaude Sonnet 7.5/3.3 في أوائل 2025؛ والمؤلفون واضحون في أن أدوات أفضل، أو طرقًا أفضل لاستخدام هذه الأدوات، قد تغير النتيجة حتى في السياق نفسه.
- الدراسة **ما قبل الطباعة** (نُشرت في يوليو 2025 ولم تخضع بعد للتحكيم)، ويذكر المؤلفون أنهم لا يستطيعون استبعاد كل الآثار التجريبية تمامًا — مع أن النتيجة صمدت في تحليلاتهم المختلفة.
- ولا تبرر القراءة المطمئنة القائلة إن المطورين لا بد أنهم كسبوا شيئًا آخر — تعلموا أكثر، أو كانوا أسعد، أو كتبوا شيفرة أفضل. الشيء الذي قيس هنا، أي الشعور بالتسارع، هو بالضبط ما تناقضه البيانات.

ما مدى قوة الدليل

- **التصميم صريح بصورة غير معتادة.** عشوائية، ومهام حقيقية، ومستودعات حقيقية، وقياس فعلي للوقت — خطوة كبيرة مقارنة بالتقارير الذاتية ولوائح الصدارة في الاختبارات المعيارية التي تستند إليها كثير من الادعاءات عن برمجة الذكاء الاصطناعي. وظل التباطؤ البالغ 19% قائمًا في اختبارات المتانة التي أجراها المؤلفون.
- **أكثر النتائج قابلية للنقل هي فجوة الإدراك.** حدس الخبراء بشأن مقدار ما يسرّعهم به الذكاء الاصطناعي كان خاطئًا بنحو 40 نقطة، وفي الاتجاه المتفائل. وهذه ملاحظة تحذيرية تجاه كل مكسب إنتاجية يُبلغ عنه ذاتيًا مع الذكاء الاصطناعي — بما في ذلك أرقام هذه الدراسة نفسها.
- **إنها لقطة زمنية، لا اتجاهًا طويل الأمد.** متابعة METR في فبراير 2026، مع النوع نفسه من المطورين وأدوات أحدث، تشير إلى تسارع — تقريبيًا −18% للمطورين العائدين و−4% للجدد — لكن المؤلفين يصفون ذلك بأنه دليل ضعيف، مشوه بمن وافق على المشاركة (إذ ازداد رفض المطورين العمل دون الذكاء الاصطناعي، وانخفض معدل الأجر، وانحاز اختيار المهام). القراءة الأمينة هي أن الصورة تتحرك، وأن هذا التحرك نفسه يُعرض مع تحذيرات واضحة.

لماذا يهم ذلك

معظم الجدل حول الذكاء الاصطناعي والبرمجة يعيش على العروض والانطباعات: تسجيل شاشة مبهر، وادعاء واثق، ثم رفع حاجب ساخر من الطرف الآخر. النادر — وما يجعل هذه الدراسة جديرة بالقراءة — هو أن أحدًا أجرى التجربة المملة: عشوائية، وقياس وقت العمل الحقيقي، ثم سؤال الناس عما شعروا به. والجواب غير مريح للطرفين. فهو يثقب قصة أن الذكاء الاصطناعي يضاعف دائمًا سرعة المطورين الخبراء في الشيفرة الصعبة المألوفة لهم. ويثقب أيضًا النقيض المرتب — «ثبت أن الذكاء الاصطناعي يبطئ المطورين» — لأن بيانات الفريق الأحدث تميل بالفعل في الاتجاه الآخر. أكثر الدروس دوامًا هو أيضًا أصغرها وأكثرها إنسانية: الأشخاص الذين كانوا يعملون شعروا بأنهم أسرع بينما كانوا أبطأ بصورة قابلة للقياس. «يبدو أسرع» ليس دليلًا على أنه أسرع. قِسْه.

خلاصة واضحة

في تجربة عشوائية مضبوطة، طلبت METR من ستة عشر مطورًا خبيرًا في مشاريع مفتوحة المصدر إكمال 246 مهمة حقيقية على قواعد شيفرة يعرفونها جيدًا، مع السماح بأدوات الذكاء الاصطناعي في أوائل 2025 (Cursor Pro مع Claude 3.5/3.7 Sonnet) في نصف عشوائي من المهام. توقع المطورون أن يقلل الذكاء الاصطناعي زمن المهمة بنحو 24%؛ لكنه بدلًا من ذلك رفع زمن الإنجاز 19% — وبعدها ظلوا يعتقدون أنه سرّعهم بنحو 20%. فجوة الإدراك هذه هي جوهر الدراسة الحاد والمتين. لكننا نتحدث عن ستة عشر مطورًا، وسياق ضيق، ولقطة ثابتة من أوائل 2025؛ والمؤلفون واضحون في أن الدراسة لا تثبت أن الذكاء الاصطناعي يفشل في مساعدة معظم المطورين، كما أن متابعتهم في 2026 تميل بالفعل نحو تسارع، مع تحفظات تخصها. قياس حذر يستحق أن يؤخذ بجدية — لا حكم نهائي على برمجة الذكاء الاصطناعي.

المصادر

Open-Source Experienced on AI Early-2025 of Impact the Measuring (METR), Rein D. Barnes, B. Rush, N. Becker, J. (2025) [cs.AI] arXiv:2507.09089 Productivity, Developer
https://arxiv.org/abs/2507.09089

write-up, (study Productivity' Developer Open-Source Experienced on AI Early-2025 of Impact the 'Measuring METR, (2025 July
https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/

(2026 (February follow-up developer-uplfit METR,
https://metr.org/blog/2026-02-24-uplift-update/