



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

توسعه‌دهندگان با تجربه با AI احساس می‌کردند سریع‌ترند، اما اندازه‌گیری نشان داد کندتر کار می‌کنند — یافته واقعی همین است، نه حکم نهایی درباره کدنویسی با AI

در یک کارآزمایی تصادفی کنترل‌شده از *METR*، شانزده توسعه‌دهنده با تجربه متن‌باز ۲۴۶ کار واقعی را روی کدیس‌هایی که خوب می‌شناختند انجام دادند و در نیمی تصادفی از کارها ابزارهای AI (اوایل ۲۰۲۵ *Cursor* با *Pro* *Claude* 7.5/3.3 و *Sonnet*) مجاز بود. آن‌ها انتظار داشتند AI زمان کار را حدود ۲۴٪ کم کند؛ در عوض زمان تکمیل ۱۹٪ بیشتر شد — و پس از آن هنوز باور داشتند AI حدود ۲۰٪ سریع‌ترشان کرده است. همین شکاف ادراک هسته محکم مطالعه است. اما این فقط شانزده توسعه‌دهنده، یک محیط محدود و عکس فوری اوایل ۲۰۲۵ است؛ نویسندگان صریح‌اند که نتیجه نشان نمی‌دهد AI به بیشتر توسعه‌دهندگان کمک نمی‌کند و پیگیری ۲۰۲۶ خودشان نیز، با احتیاط‌های خودش، به سمت افزایش سرعت متمایل است.

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

برنامه‌نویسان با تجربه با AI احساس می‌کردند سریع‌ترند، اما اندازه‌گیری نشان داد کندتر شده‌اند — یافته واقعی همین است، نه حکم نهایی درباره کدنویسی با AI

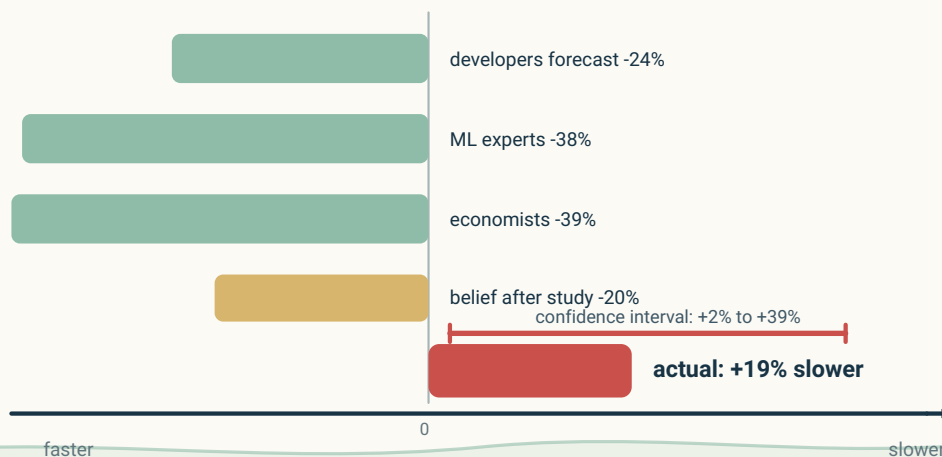
اگر از یک برنامه‌نویس با تجربه پرسید دستیار کدنویسی مبتنی بر AI چقدر سرعتش را بالا می‌برد، معمولاً عددی می‌شنوید: بیست، سی درصد در وقت صرفه‌جویی می‌کند. اگر از یک اقتصاددان یا پژوهشگر یادگیری ماشین پرسید، عدد معمولاً بزرگ‌تر می‌شود. اوایل ۲۰۲۵ تیمی در METR کار کند و پرهزینه را انجام داد و واقعاً اندازه گرفت. شانزده توسعه‌دهنده با تجربه متن‌باز را گرفتند، ۲۴۶ کار واقعی از کدیس‌های بزرگی که آن‌ها خوب می‌شناختند به دستشان دادند و — با قرعه، برای هر کار جداگانه — یا اجازه استفاده از ابزارهای AI دادند یا ندادند. سپس زمان کار را اندازه گرفتند.

توسعه‌دهندگان پیش‌بینی کرده بودند AI زمان انجام کار را حدود ۲۴٪ کاهش می‌دهد. برعکس شد: کارهایی که با AI انجام شدند ۱۹٪ بیشتر طول کشیدند. و بخش واقعاً قابل تأمل اینجا است — همان توسعه‌دهندگان پس از پایان کار هنوز باور داشتند AI آن‌ها را حدود ۲۰٪ سریع‌تر کرده است. کندتر بودند و احساس می‌کردند سریع‌ترند؛ فاصله این دو عدد جالب‌ترین یافته مطالعه است.

این نتیجه‌ای واقعی و با دقت اندازه‌گیری شده است. اما در عین حال درباره شانزده توسعه‌دهنده است، روی مخزن‌هایی که عمیقاً می‌شناختند، با ابزارهای اوایل ۲۰۲۵ — و معادل جمله تحت AI «توسعه‌دهندگان را کندتر می‌کند» نیست. داده‌های بعدی همان گروه نیز از همین حالا جهت دیگری را نشان می‌دهد.

Everyone expected speed. The timed tasks got slower.

Forecasted time savings versus the measured +19% task time.



Boundary: this study measured real tasks in known repos, not all AI coding forever.

همه پیش‌بینی کرده بودند AI کار را سریع‌تر کند — توسعه‌دهندگان — ۲۴٪، متخصصان یادگیری ماشین — ۳۸٪، اقتصاددانان — ۳۹٪، و حتی باور خود توسعه‌دهندگان پس از آزمایش — ۲۰٪. زمان‌سنج نتیجه +۱۹٪ کندتر را نشان داد، با بازه +۲٪ تا +۳۹٪. فاصله میان احساس و اندازه‌گیری، خود یافته است.

What this AI-productivity study measured

THIS IS

- 16 experienced open-source developers
- 246 real tasks
- repositories they knew
- randomized and timed
- early-2025 AI tools

THIS IS NOT

- not most developers
- not every domain
- not a fixed law
- not peer-reviewed
- not a verdict on future tools

Boundary: useful evidence about one setting, not a universal law of AI work.

این مطالعه چه چیزی را اندازه می گیرد — ۱۶ توسعه دهنده خبره متن باز، ۲۴۶ کار واقعی، مخزن های آشنا، ابزارهای اوایل ۲۰۲۵، تخصیص تصادفی — و چه چیزی را اندازه نمی گیرد: نه بیشتر توسعه دهندگان، نه حوزه های دیگر، نه یک قانون ثابت و نه نتیجه ای دآوری شده.

Original diagram — The Clean Paper CC BY 0.4

این مطالعه چه چیزی را اندازه گرفت و یک کارآزمایی تصادفی چه مزیتی دارد

می کند. استفاده انتظار و امید از واقعی اثر کردن جدا برای پزشکی که است ابزاری همان (RCT) کنترل شده تصادفی کارآزمایی تفاوت تنها میانگین، در بنابرین یافت؛ تخصیص غیرمجاز «AI» یا مجاز «AI» حالت به تصادفی به طور کار ۲۴۶ از یک هر اینجا کسانی شویم متوجه فقط اینکه نه شد، زمان تغییر سبب AI بگویم می دهد اجازه همین است. AI خود گروه دو میان نظام مند کدنویسی بهره وری افزایش درباره معمول شواهد چون است، مهم این کند ترند. یا سریع تر دیگری دلایل به می روند AI سراغ که این که همان طور خود اظهاری، و نیست واقعی کار بنچمارک نمی کنند: کاری چنین — بنچمارک امتیاز و خود اظهاری — AI با دست و پنجه جدید اسباب بازی با که نبودند تازه کارانی اینجا توسعه دهندگان باشد. اشتباه اطمینان با می تواند می دهد، نشان مطالعه قبلی تجربه بعضی و می شناختند خوب که بودند متن باز بالغ و بزرگ پروژه های تثبیت شده مشارکت کنندگان آن ها کنند. نرم داشتند. نیز را ابزارها این از استفاده

نویسندگان چه کردند

- یک کارآزمایی تصادفی کنترل شده اجرا کردند (METR: David و Barnes Beth Rush، Nate Becker، Joel Rein) شازده توسعه دهنده با تجربه متن باز، هر کدام روی مخزن بزرگی کار کردند که مرتب در آن مشارکت داشتند و خوب می شناختند — به طور میانگین پنج سال سابقه روی همان پروژه خاص.
- از ۲۴۶ کار واقعی استفاده کردند — رفع باگ، قابلیت جدید و بازاریابی کد از tracker issue همان پروژه ها. هر کار به طور تصادفی به «AI» مجاز یا «AI» غیرمجاز تخصیص یافت.

• AI «مجاز» یعنی ابزارهای اوایل ۲۰۲۵: Pro Cursor با Sonnet 7.5/3.3 Claude معیار اصلی زمان واقعی تکمیل هر کار بود. در کنار آن، تیم پیش‌بینی‌های قبل از کار و برآوردهای پس از آن را از خود توسعه‌دهندگان، و نیز پیش‌بینی‌های متخصصان اقتصاد و یادگیری ماشین جمع‌آوری کرد.

چه پیدا کردند

- با AI، کارها ۱۹٪ بیشتر طول کشیدند. نه سریع‌تر — کندتر. فاصله اطمینان ۹۵٪ تقریباً از ۲٪+ تا ۳۹٪+ است، بنابراین جهت اثر نسبتاً محکم است هرچند اندازه دقیق آن نه.
- همه برعکسش را پیش‌بینی کرده بودند. توسعه‌دهندگان ۲۴٪ افزایش سرعت پیش‌بینی کردند؛ متخصصان یادگیری ماشین حدود ۳۸٪ و اقتصاددانان حدود ۳۹٪. هر سه گروه انتظار صرفه‌جویی قابل‌توجه در زمان داشتند؛ زمان‌سنج نشان داد مقداری زمان اضافه شد.
- شکاف ادراک. توسعه‌دهندگان پس از انجام کار، با وجود اینکه کندتر شده بودند، هنوز برآورد کردند AI حدود ۲۰٪ سرعتشان را بالا برده است — تقریباً ۴۰ واحد درصد فاصله میان احساس و چیزی که ساعت ثبت کرد.
- علت‌های نامزد، نه علت‌های اثبات‌شده. نویسندگان چند عامل ممکن برای توضیح کندی را کنار هم می‌گذارند: این توسعه‌دهندگان کدبیس خود را عمیقاً می‌شناسند، پس دستیار فضای کمتری برای افزودن ارزش دارد؛ پروژه‌های بالغ استانداردهای کیفی بالا و اغلب ضمنی دارند؛ مخزن‌ها بزرگ و پر از زمینه‌ای هستند که مدل در اختیار ندارد؛ و زمان واقعی صرف نوشتن prompt، سپس بازبینی و اصلاح خروجی AI می‌شود. مقاله این‌ها را سرخ می‌داند، نه حکم قطعی.

این مقاله چه چیزی را اثبات نمی‌کند

- نشان نمی‌دهد AI نمی‌تواند بیشتر توسعه‌دهندگان را سریع‌تر کند. نویسندگان صریحاً همین را می‌گویند: شانزده متخصص روی کدی که از بر هستند نمایندۀ توسعه‌دهندۀ متوسط روی کار متوسط نیستند.
- نشان نمی‌دهد AI بی‌فایده است یا در محیط‌های دیگر هم آدم‌ها را کند می‌کند — تازه‌واردان به یک کدبیس، پروژه‌های از صفر، زبان‌های ناآشنا یا حوزه‌های کاملاً متفاوت.
- ابزارها را در زمان منجمد نمی‌کند. این Cursor و Sonnet 7.5/3.3 Claude اوایل ۲۰۲۵ است؛ نویسندگان صریحاً می‌گویند که ابزارهای بهتر یا شیوۀ بهتر استفاده از همین ابزارها می‌تواند حتی در همین محیط دقیقاً نتیجه را تغییر دهد.
- این یک پیش‌چاپ است (منتشر شده در ژوئیه ۲۰۲۵ و هنوز داوری نشده)، و نویسندگان می‌گویند نمی‌توانند تمام مصنوعات آزمایشی را کاملاً رد کنند — هرچند نتیجه در تحلیل‌های مختلفشان پابرجا ماند.
- و اجازه نمی‌دهد خیالمان را راحت کنیم که شاید توسعه‌دهندگان از راه دیگری سود برده‌اند — بیشتر یاد گرفته‌اند، خوشحال‌تر بوده‌اند یا کد بهتری نوشته‌اند. چیزی که اینجا مستقیماً با داده تناقض دارد همان احساس افزایش سرعت است.

قدرت شواهد چقدر است؟

- طراحی به‌طور غیرمعمولی صادقانه است. تصادفی‌سازی، کار واقعی، مخزن واقعی و زمان‌گیری واقعی — جهشی بزرگ نسبت به خوداظهاری و جدول‌های بنچمارک که بیشتر ادعاهای بهره‌وری AI به آن‌ها متکی‌اند. کندی ۱۹٪ از آزمون‌های پایداری نویسندگان جان سالم به در برد.

- قابل انتقال ترین یافته شکاف ادراک است. شهود متخصص درباره افزایش سرعت خودش با AI حدود ۴۰ واحد درصد در جهت خوش بینانه خطا کرد. این هشدار است درباره هر ادعای خوداظهاری از افزایش بهره‌وری با AI — از جمله عددهای همین مطالعه.
- این یک عکس فوری است، نه روند. پیگیری فوری ۲۰۲۶ خود، METR، با همان نوع توسعه‌دهنده و ابزارهای جدیدتر، به سمت افزایش سرعت اشاره می‌کند — بسیار تقریبی، حدود -۱۸٪ زمان برای توسعه‌دهندگان بازگشته و -۴٪ برای افراد جدید — اما نویسندگان آن را شواهد ضعیفی می‌دانند، چون اینکه چه کسی حاضر به شرکت شد داده را منحرف می‌کند (توسعه‌دهندگان بیش از پیش حاضر نبودند بدون AI کار کنند، نرخ پرداخت پایین آمد و انتخاب کارها سوگیرانه شد). خوانش صادقانه این است که تصویر در حال حرکت است و حتی خود این حرکت هم با احتیاط گزارش شده است.

چرا مهم است

بخش زیادی از بحث AI و برنامه‌نویسی روی دمو و حس می‌چرخد: یک ویدیوی براق، یک ادعای مطمئن، یک چشم‌غره مخالف. چیزی که نادر است — و این مطالعه را ارزشمند می‌کند — این است که کسی آزمایش کسل‌کننده را انجام داده: تصادفی‌سازی، زمان‌گیری کار واقعی، سپس پرسیدن اینکه «چطور بود؟» پاسخ برای هر دو اردوگاه ناراحت‌کننده است. داستانی را سوراخ می‌کند که AI همیشه توسعه‌دهندگان خبره را روی کد سخت و آشنا توربوشارژ می‌کند. و در عین حال روایت مرتب مخالف — «ثابت شد AI توسعه‌دهندگان را کند می‌کند» — را هم خراب می‌کند، چون داده جدیدتر همان تیم از همین حالا جهت مخالف را نشان می‌دهد. پایدارترین درس کوچک‌ترین و انسانی‌ترین است: کسانی که کار را انجام می‌دادند، در حالی که اندازه‌گیری نشان می‌داد کندترند احساس می‌کردند سریع‌ترند. «حس می‌کنم سریع‌تر است» شواهدی نیست که واقعاً چنین باشد. اندازه بگیرید.

خلاصه تمیز

در یک کارآزمایی تصادفی کنترل‌شده، METR از شانزده توسعه‌دهنده باتجربه متن‌باز خواست ۲۴۶ کار واقعی را روی کدیس‌هایی که خوب می‌شناختند انجام دهند؛ برای نیمی تصادفی از کارها استفاده از ابزارهای AI اوایل ۲۰۲۵ (Cursor Pro) به همراه Claude 7.5/3.3 Sonnet مجاز بود. توسعه‌دهندگان انتظار داشتند AI زمان کار را حدود ۲۴٪ کاهش دهد؛ در عوض زمان تکمیل ۱۹٪ افزایش یافت — و پس از کار هم هنوز باور داشتند AI حدود ۲۰٪ سریع‌ترشان کرده است. این شکاف ادراک هسته تیز و مقاوم مطالعه است. اما این فقط شانزده توسعه‌دهنده، یک محیط محدود و یک عکس فوری از اوایل ۲۰۲۵ است؛ نویسندگان صریحاً می‌گویند نتیجه نشان نمی‌دهد AI به بیشتر توسعه‌دهندگان کمک نمی‌کند، و پیگیری ۲۰۲۶ خودشان از همین حالا به سمت افزایش سرعت اشاره دارد، با احتیاط‌های خودش. یک اندازه‌گیری دقیق که باید جدی گرفته شود — نه حکم نهایی درباره کدنویسی با AI.

منابع

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/>