



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

تجربہ کار ڈویلپرز کو اے آئی کے ساتھ کام زیادہ تیز محسوس ہوا، مگر پیمائش میں وہ واقعی سست تھے — اصل نتیجہ کیا ہے، اور اے آئی کوڈنگ پر یہ کوئی آخری فیصلہ کیوں نہیں

ایک تصادفی کنٹرول شدہ ٹرائل میں METR نے سولہ تجربہ کار اوپن سورس ڈویلپرز سے 246 حقیقی کام ایسے *codebases* پر مکمل کروائے جنہیں وہ اچھی طرح جانتے تھے۔ آدھے کام میں تصادفی طور پر 2025 کے شروع کے اے آئی ٹولز — *Pro Cursor* کے ساتھ *Sonnet 7.5/3.3 Claude* — استعمال کرنے کی اجازت تھی۔ ڈویلپرز کو توقع تھی کہ اے آئی کام وقت تقریباً 24% کم کرے گا؛ اس کے بجائے *completion* وقت 19% بڑھ گیا، اور کام ختم ہونے کے بعد بھی انہیں لگا رہا کہ اے آئی نے انہیں تقریباً 20% تیز کیا۔ یہی *perception* خلا مطالعہ کا سب سے واضح اور مضبوط نکتہ ہے۔ لیکن یہ صرف سولہ ڈویلپرز، ایک محدود *setting* اور ابتدائی-2025 ٹولز کا *snapshot* ہے؛ مصنفین واضح ہیں کہ یہ نتیجہ زیادہ تر ڈویلپرز پر اے آئی کی ناکامی ثابت نہیں کرتا، اور ان کا اپنا 2026 *caveats follow-up* کے ساتھ *speedup* کی طرف اشارہ کرتا ہے۔ یہ ایک سنجیدہ پیمائش ہے — اے آئی *coding* پر آخری فیصلہ نہیں۔

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

تجربہ کار ڈویلپرز کو اے آئی کے ساتھ کام زیادہ تیز محسوس ہوا، مگر پیمائش میں وہ واقعی سست تھے – اصل نتیجہ کیا ہے، اور اے آئی کوڈنگ پر یہ کوئی آخری فیصلہ کیوں نہیں

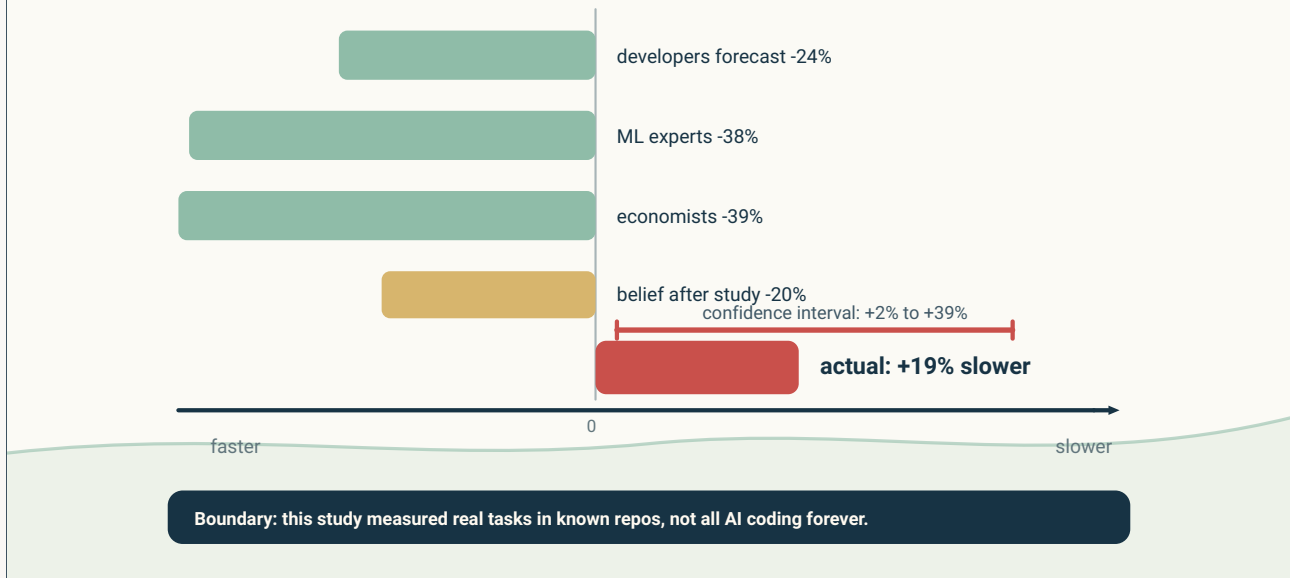
کسی تجربہ کار پروگرامر سے پوچھیں کہ اے آئی کوڈنگ اسسٹنٹ اسے کتنا تیز کرتا ہے تو عموماً جواب کسی فیصد میں آئے گا: بیس، تیس فیصد وقت بچ جاتا ہے۔ کسی ماہرِ معاشیات یا مشین لرننگ کے محقق سے پوچھیں تو اندازہ اور بڑا ہو سکتا ہے۔ 2025 کے آغاز میں METR کی ایک ٹیم نے مہنگا اور سست مگر زیادہ قابلِ اعتماد راستہ اختیار کیا: انہوں نے واقعی پیمائش کی۔ سولہ تجربہ کار اوپن سورس ڈویلپرز کو ان بڑے کوڈ بیسز سے 246 حقیقی کام دیے گئے جنہیں وہ اچھی طرح جانتے تھے، اور ہر کام کے لیے قرعہ اندازی سے طے کیا گیا کہ اے آئی ٹولز استعمال کرنے کی اجازت ہوگی یا نہیں۔ پھر کام مکمل ہونے کا وقت ناپا گیا۔

ڈویلپرز نے پہلے پیش گوئی کی تھی کہ اے آئی سے ہر کام میں تقریباً 24% وقت بچے گا۔ ہوا اس کے الٹ: اے آئی کے ساتھ کیے گئے کاموں میں 19% زیادہ وقت لگا۔ اور اس مطالعے کا شاید سب سے دلچسپ حصہ یہ ہے کہ کام ختم ہونے کے بعد بھی انہی ڈویلپرز کو لگا رہا کہ اے آئی نے انہیں تقریباً 20% تیز کر دیا تھا۔ وہ پیمائش میں سست تھے، مگر انہیں خود کو تیز محسوس ہوا؛ ان دونوں نمبروں کے درمیان فاصلہ ہی اس مطالعے کا سب سے اہم نتیجہ ہے۔

یہ ایک حقیقی اور احتیاط سے ناپا گیا نتیجہ ہے۔ لیکن یہ سولہ ڈویلپرز کا نتیجہ ہے، ایسے repositories پر جنہیں وہ بہت گہرائی سے جانتے تھے، اور 2025 کے شروع کے ٹولز کے ساتھ۔ اسے سیدھے جملے ”اے آئی ڈویلپرز کو سست کرتا ہے“ میں بدل دینا درست نہیں۔ اسی ٹیم کے بعد کے اعداد و شمار پہلے ہی دوسری سمت اشارہ کرتے ہیں۔

Everyone expected speed. The timed tasks got slower.

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



سب نے توقع کی تھی کہ اے آئی کام تیز کرے گا – ڈویلپرز – 24%، مشین لرننگ ماہرین – 38%، ماہرینِ معاشیات – 39%، اور کام کے بعد خود ڈویلپرز کا اندازہ – 20%۔ اصل stopwatch نے +19% یعنی سست کام دکھایا، جس کا وقفہ تقریباً 2% سے +39% تھا۔ محسوس کی

گئی اور ناپی گئی رفتار کے درمیان یہی فرق اصل نتیجہ ہے۔

Original diagram — The Clean Paper CC BY 0.4

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.

مطالعہ کیا ناپتا ہے — 16 ماہر اوپن سورس ڈویلپرز، 246 حقیقی کام، مانوس، 2025 repositories کے شروع کے اے آئی ٹولز، اور تصادفی تعین — اور کیا نہیں ناپتا: زیادہ تر ڈویلپرز نہیں، دوسرے شعبے نہیں، کوئی مستقل قانون نہیں، اور یہ ابھی ہم مرتبہ جائزہ شدہ بھی نہیں۔

Original diagram — The Clean Paper CC BY 0.4

یہاں کیا ناپا گیا، اور تصادفی ٹرائل کیوں اہم ہے

اثر حقیقی کہ ہیں کرتے استعمال لیے اس شعبے کئی سمیت طب جسے ہے طریقہ وہ (RCT) ٹرائل شدہ کنٹرول تصادفی گیا: رکھا میں حالتوں دو پر طور تصادفی کو کام ہر سے میں 246 یہاں سکے۔ جا کیا الگ سے اتفاق یا انتخاب امید، محض کو جاتا رہ کا آئی اے صرف فرق منظم درمیان کے گروہوں دونوں اوسطاً طرح اس بغیر۔ کے آئی اے یا اجازت کی آئی اے لوگ جو کہ گیا دیکھا یہ صرف کہ نہ، کی پیدا تبدیلی میں وقت نے آئی اے کہ ہے ہوتا ممکن کہنا یہ سے وجہ اسی ہے۔ ہیں۔ سست یا تیز ہی پہلے سے وجہ اور کسی وہ ہیں کرتے استعمال آئی اے یہ اس لیے اہم ہے کہ اے آئی coding سے متعلق عام شواہد — خود بتائی گئی رپورٹ اور بینچ مارک اسکورز — ایسا سببی فرق نہیں دکھا سکتے۔ بینچ مارک حقیقی کام نہیں، اور، رپورٹ-sefl جیسا کہ یہ مطالعہ دکھاتا ہے، پورے اعتماد کے ساتھ غلط بھی ہو سکتا ہے۔ یہاں شامل ڈویلپرز نئے ٹول سے پہلی بار الجھنے والے مبتدی نہیں تھے؛ وہ بڑے، پختہ اوپن سورس منصوبوں کے قائم شدہ شریک کار تھے، اپنے کوڈ بیس کو اچھی طرح جانتے تھے اور اے آئی ٹولز کا کچھ سابق تجربہ بھی رکھتے تھے۔

مصنفین نے کیا کیا

- METR کے Joel، Nate Becker، Barnes Beth Rush اور Rein David نے ایک تصادفی کنٹرول شدہ ٹرائل چلایا۔ سولہ تجربہ کار اوپن سورس ڈویلپرز نے ایسے بڑے repositories پر کام کیا جن میں وہ باقاعدگی سے contribute کرتے تھے؛ مخصوص منصوبوں پر ان کا اوسط تجربہ تقریباً پانچ سال تھا۔
- 246 حقیقی کام استعمال کیے گئے — خرابی، fixes، خصوصیات اور refactors جو انہی منصوبوں کے trackers issue سے لیے گئے تھے۔ ہر کام کو تصادفی طور پر "اے آئی allowed" یا "اے آئی disallowed" حالت میں رکھا گیا۔
- "اے آئی allowed" کا مطلب تھا 2025 کے شروع کا Pro Cursor tooling کے ساتھ Sonnet 7.5/3.3 Claude۔ بنیادی نتیجہ ہر کام کا اصل completion وقت تھا۔ ٹیم نے پہلے کے forecasts، بعد کے تخمینے، اور economics اور مشین لرننگ ماہرین کی پیش گوئیاں بھی جمع کیں۔

کیا ملا

- اے آئی کے ساتھ کام میں 19% زیادہ وقت لگا۔ تیز نہیں، سست۔ 95% اعتماد وقفہ تقریباً +2% سے +39% تک ہے، اس لیے سمت کافی مضبوط ہے چاہے درست قد میں غیر یقینی باقی رہے۔
- سب نے الٹی پیش گوئی کی تھی۔ ڈویلپرز نے 24%، speedup، مشین لرننگ ماہرین نے تقریباً 38% اور ماہرینِ معاشیات نے تقریباً 39% speedup کی توقع کی تھی۔ تینوں گروہوں کو لگا اے آئی کافی وقت بچائے گا؛ اصل وقت نے کچھ اضافی لاگت دکھائی۔
- ادراک کا فرق۔ کام ختم ہونے کے بعد، اگرچہ وہ ناپے گئے وقت میں سست نکلے، پھر بھی ڈویلپرز نے اندازہ لگایا کہ اے آئی نے انہیں تقریباً 20% تیز کیا۔ محسوس شدہ رفتار اور clock کے نتیجے کے درمیان تقریباً 40 فیصد نقاط کا فرق تھا۔
- ممکنہ وجوہات، ثابت شدہ وجوہات نہیں۔ مصنفین کچھ عوامل پیش کرتے ہیں جو slowdown سمجھا سکتے ہیں: یہ ڈویلپرز اپنے codebases کو بہت گہرائی سے جانتے تھے، اس لیے assistant کے لیے اضافی فائدہ محدود ہو سکتا ہے؛ بالغ projects میں معیار کے سخت مگر اکثر غیر تحریری تقاضے ہوتے ہیں؛ repositories بڑے اور سیاق-بہاری ہیں؛ اور، prompting اے آئی output کا جائزہ اور اصلاحات بھی وقت لیتے ہیں۔ مصنفین ان وجوہات کو possibilities کے طور پر پیش کرتے ہیں، فیصلہ کن وضاحت کے طور پر نہیں۔

یہ کیا نہیں دکھاتا

- یہ نہیں دکھاتا کہ اے آئی زیادہ تر ڈویلپرز کو تیز نہیں کرتا۔ مصنفین صاف لکھتے ہیں کہ سولہ ماہرین جو اپنے کوڈ کو ازبر جانتے ہوں، اوسط ڈویلپر اور اوسط کام کی نمائندگی نہیں کرتے۔
- یہ نہیں دکھاتا کہ اے آئی بے کار ہے یا دوسرے حالات میں لوگوں کو سست کرتا ہے — مثلاً نئے کوڈ بیس میں آنے والے لوگ، greenfield، projects، غیر مانوس programming زبانیں، یا بالکل دوسرے شعبے۔
- یہ ٹولز کو ہمیشہ کے لیے ایک ہی حالت میں freeze نہیں کرتا۔ یہ 2025 کے شروع کا Cursor اور Sonnet 7.5/3.3 Claude ہے؛ مصنفین خود کہتے ہیں کہ بہتر ٹولز یا بہتر usage اسی ترتیب میں بھی نتیجہ بدل سکتے ہیں۔
- یہ ایک پری پرنٹ ہے، جولائی 2025 میں پوسٹ ہوا اور ابھی ہم مرتبہ جائزہ شدہ نہیں۔ مصنفین یہ بھی مانتے ہیں کہ تمام artifacts experimental مکمل طور پر خارج نہیں کیے جا سکتے، اگرچہ مختلف تجزیے میں بنیادی نتیجہ قائم رہا۔
- یہ اس تسلی بخش مفروضے کو بھی ثابت نہیں کرتا کہ شاید ڈویلپرز نے کوئی اور فائدہ حاصل کیا — زیادہ سیکھا، زیادہ خوش

رہے، یا بہتر کوڈ لکھا۔ اس مطالعے میں جس چیز کو براہ راست ناپا گیا، یعنی محسوس شدہ، speedup ڈیٹا اسی کے خلاف جاتا ہے۔

شواہد کتنے مضبوط ہیں

- ڈیزائن غیر معمولی حد تک دیانت دار ہے۔ بے ترتیب تقسیم، حقیقی کام، حقیقی repositories اور حقیقی وقت بندی — اے آئی coding کے اکثر دعووں کی بنیاد بننے والے خود بتائی گئی رپورٹ اور بینچ مارک leaderboards کے مقابلے میں بڑا قدم ہے۔ 19% slowdown مصنفین کے مضبوطی جانچیں میں بھی قائم رہا۔
- سب سے زیادہ عمومی سبق ادراک خلا ہے۔ اپنے اے آئی speedup کے بارے میں ماہرین کی وجدان تقریباً 40 نقاط سے غلط نکلی، اور غلطی امید والی سمت میں تھی۔ یہ اے آئی پیداواری صلاحیت کے ہر رپورٹ-sefl شدہ دعوے کے لیے احتیاط کا اشارہ ہے — خود اس مطالعے کے بعد میں خود بتائی گئی رپورٹ سمیت۔
- یہ snapshot ہے، trend نہیں۔ METR کا فروری 2026 فالو آپ، اسی طرح کے ڈویلپرز مگر نئے ٹولز کے ساتھ، speedup کی طرف اشارہ کرتا ہے — returning ڈویلپرز کے لیے بہت کھردرا طور پر — 18% اور نئے شرکاء کے لیے — 4% — مگر مصنفین اسے کمزور شواہد کہتے ہیں کیونکہ participation میں انتخاب bias بڑھ گیا تھا: زیادہ ڈویلپرز اے آئی کے بغیر کام کرنے سے انکار کر رہے تھے، pay شرح کم ہوا، اور کام کا انتخاب بدل گیا۔ درست نتیجہ یہ ہے کہ تصویر بدل رہی ہے، اور خود اس تبدیلی کا اندازہ بھی احتیاط سے لینا چاہیے۔

یہ کیوں اہم ہے

اے آئی اور programming پر زیادہ تر بحث demos اور ”feel“ پر چلتی ہے: ایک شاندار اسکرین recording، ایک پراعتماد دعویٰ، پھر کسی دوسرے کی طنز یہ نفی۔ یہاں اہم بات یہ ہے کہ کسی نے boring مگر ضروری تجربہ کیا — randomize، حقیقی کام کا وقت ناپا، پھر لوگوں سے پوچھا کہ انہیں کیسا لگا۔ جواب دونوں camps کے لیے بے آرام ہے۔ یہ اس کہانی کو کمزور کرتا ہے کہ اے آئی مشکل، مانوس کوڈ پر ماہر ڈویلپرز کو ہمیشہ turbocharge کرتا ہے۔ لیکن یہ اس tidy مخالف کہانی کو بھی ثابت نہیں کرتا کہ ”اے آئی ڈویلپرز کو سست کرتا ہے، بات ختم“، کیونکہ اسی ٹیم کا نیا ڈیٹا پہلے ہی دوسری سمت جھک رہا ہے۔ سب سے پائیدار سبق چھوٹا اور انسانی ہے: لوگوں کو کام تیز محسوس ہوا جبکہ پیمائش میں وہ سست تھے۔ ”مجھے تیز لگا“ اس بات کا ثبوت نہیں کہ واقعی تیز تھا۔ اسے ناپیں۔

صاف خلاصہ

ایک تصادفی کنٹرول شدہ ٹرائل میں METR نے سولہ تجربہ کار اوپن سورس ڈویلپرز سے 246 حقیقی کام ایسے codebases پر مکمل کروائے جنہیں وہ اچھی طرح جانتے تھے۔ آدھے کام میں تصادفی طور پر 2025 کے شروع کے اے آئی ٹولز — Pro Cursor — کے ساتھ Sonnet 7.5/3.3 Claude — استعمال کرنے کی اجازت تھی۔ ڈویلپرز کو توقع تھی کہ اے آئی کام وقت تقریباً 24% کم کرے گا؛ اس کے بجائے completion وقت 19% بڑھ گیا، اور کام ختم ہونے کے بعد بھی انہیں لگا رہا کہ اے آئی نے انہیں تقریباً 20% تیز کیا۔ یہی ادراک خلا مطالعے کا سب سے واضح اور مضبوط نکتہ ہے۔ لیکن یہ صرف سولہ ڈویلپرز، ایک محدود ترتیب اور ابتدائی-2025 ٹولز کا snapshot ہے؛ مصنفین واضح ہیں کہ یہ نتیجہ زیادہ تر ڈویلپرز پر اے آئی کی ناکامی ثابت نہیں کرتا، اور ان کا

اپنا 2026 فالو آپ caveats کے ساتھ speedup کی طرف اشارہ کرتا ہے۔ یہ ایک سنجیدہ پیمائش ہے — اے آئی coding پر آخری فیصلہ نہیں۔

ماخذ

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-uplift METR, <https://metr.org/blog/2026-02-24-uplift-update/>