The
CLEAN PAPER

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

अनुभवी डेवलपर AI के साथ खुद को तेज़ समझते रहे, जबकि माप में वे धीमे निकले — और यह AI कोडिंग पर अंतिम फैसला नहीं है

METR के एक यादृच्छिकीकृत नियंत्रित परीक्षण में सोलह अनुभवी ओपन-सोर्स डेवलपरों ने उन कोडबेस पर 246 वास्तविक कार्य किए जिन्हें वे अच्छी तरह जानते थे। आधे कार्यों में, यादृच्छिक रूप से, उन्हें 2025 की शुरुआत के AI औज़ार (Cursor Pro और Claude 3.5/3.7 Sonnet) इस्तेमाल करने दिए गए। उन्हें उम्मीद थी कि AI समय लगभग 24% घटाएगा; इसके बजाय कार्य पूरा करने में 19% अधिक समय लगा— और अध्ययन के बाद भी वे मानते रहे कि AI ने उन्हें लगभग 20% तेज़ किया। अनुभव और मापन के बीच यही अंतर अध्ययन का सबसे स्पष्ट और मजबूत परिणाम है। लेकिन नमूना केवल सोलह डेवलपरों का है, परिस्थिति बहुत संकरी है और औज़ार 2025 की शुरुआत के हैं: लेखक साफ़ कहते हैं कि इससे यह निष्कर्ष नहीं निकलता कि AI अधिकांश डेवलपरों की मदद नहीं करता। उनका 2026 का अनुवर्ती अध्ययन, अपनी सीमाओं के साथ, अब तेजी की दिशा में संकेत देता है।

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

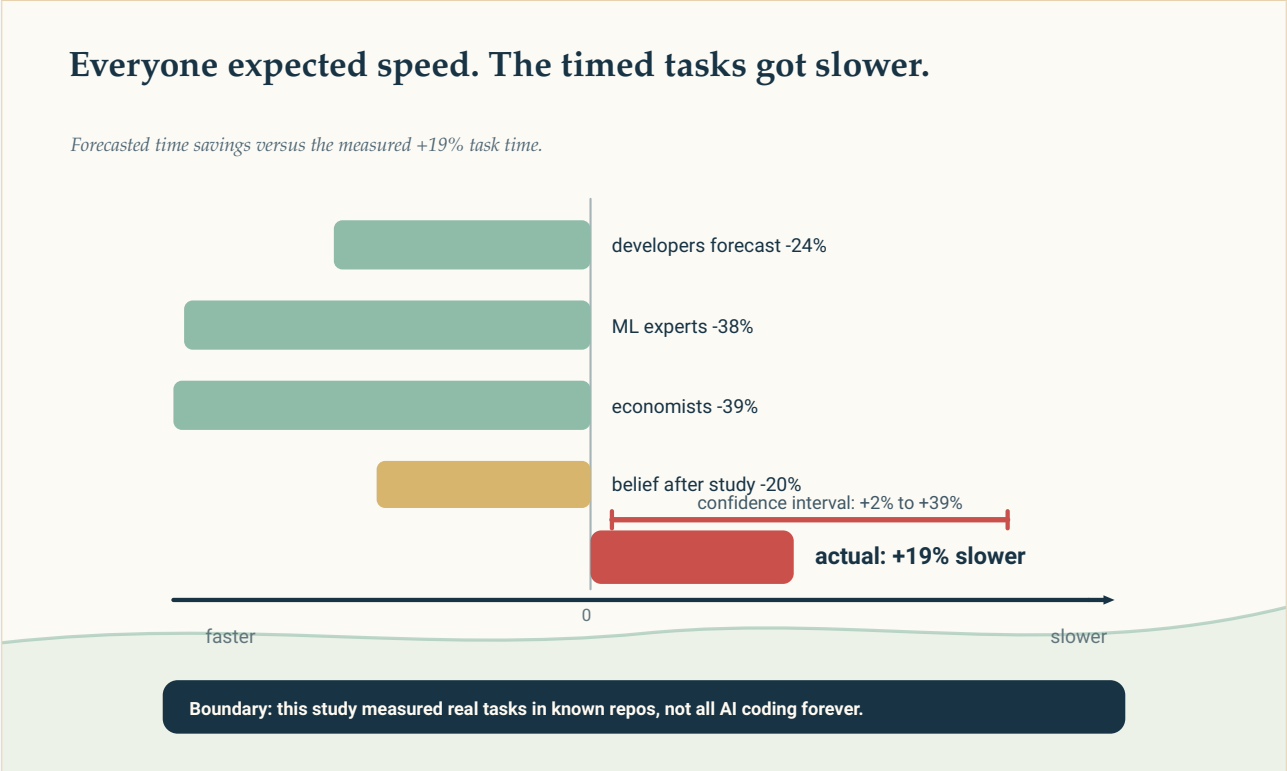
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

अनुभवी डेवलपर AI के साथ खुद को तेज़ समझते रहे, जबकि मापने पर वे धीमे निकले — असली निष्कर्ष यही है, AI कोडिंग पर अंतिम फैसला नहीं

किसी अनुभवी प्रोग्रामर से पूछिए कि AI कोडिंग सहायक उसे कितना तेज़ बनाता है, तो अक्सर कोई संख्या सुनने को मिलेगी — बीस प्रतिशत समय बचाता है, तीस प्रतिशत बचाता है। किसी अर्थशास्त्री या मशीन-लर्निंग शोधकर्ता से पूछिए, तो अनुमान और बड़ा हो सकता है। 2025 की शुरुआत में METR की एक टीम ने धीमा और महँगा रास्ता चुना। उसने इसे वास्तव में मापा। सोलह अनुभवी ओपन-सोर्स डेवलपर लिए गए, उन्हें बड़े और परिचित कोडबेस से 246 वास्तविक काम दिए गए, और हर काम के लिए सिक्का उछालने जैसी यादृच्छिक व्यवस्था से तय किया गया कि AI औज़ार इस्तेमाल करने की अनुमति होगी या नहीं। फिर समय दर्ज किया गया।

डेवलपरों का अनुमान था कि AI उनके काम का समय लगभग 24% घटा देगा। हुआ उलटा। AI के साथ किए गए कामों में **19% अधिक समय** लगा। और अध्ययन का सबसे दिलचस्प हिस्सा यह है कि काम पूरा करने के बाद भी वे डेवलपर मानते रहे कि AI ने उन्हें लगभग 20% तेज़ किया था। वे वास्तव में धीमे थे, लेकिन उन्हें लगा कि वे तेज़ हैं — इन दोनों संख्याओं के बीच की दूरी इस अध्ययन की सबसे महत्वपूर्ण बात है।

यह वास्तविकता और साधनों से मापा गया परिणाम है। लेकिन यह केवल सोलह डेवलपरों का परिणाम है, उन रिपॉज़िटरी पर जिन्हें वे बहुत अच्छी तरह जानते थे, और 2025 की शुरुआत के औज़ारों के साथ। इसे सपाट वाक्य "AI डेवलपरों को धीमा बनाता है" में बदलना सही नहीं होगा। उसी टीम के बाद के आँकड़े पहले ही दूसरी दिशा का संकेत देते हैं।



सभी समूहों ने अनुमान लगाया था कि AI काम तेज़ करेगा — डेवलपर −24%, मशीन-लर्निंग विशेषज्ञ −38%, अर्थशास्त्री −39%, और काम के बाद डेवलपरों की अपनी धारणा −20%। वास्तविक समय-माप ने +19% धीमापन दिखाया, जिसका अंतराल +2% से +39% था। महसूस हुई गति और मापी गई गति के बीच का अंतर ही मुख्य निष्कर्ष है।

Original diagram — The Clean Paper CC BY 4.0

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.

अध्ययन ने क्यामापा— 16 अनुभवीओपन-सोर्स डेवलपर, 246 वास्तवकि काम, परचिति रपिॉज़िटरी, 2025 कीशुरुआत के औज़ार, और यदृच्छकि आवंटन — और क्यानहीं: यह अधकिांश डेवलपरोकानमूनानहीं, दूसरे क्षेत्रोकापरीक्षण नहीं, कोई स्थायीनयिम नहीं, और सहकर्मीसमीक्षति अध्ययन भीनहीं।

Original diagram — The Clean Paper CC BY 4.0

इस प्रयोग ने क्यामापा और यदृच्छकि नयिंत्रति परीक्षण से क्यालाभ मलिताहै

यदृच्छकि नयिंत्रति परीक्षण (रैंडमइज़्ड नयिंत्रति परीक्षण, RCT) वहीतरीकाहै जसिकाउपयोग चकित्सामें कसीवास्तवकि प्रभाव कोकेवल आशाजनक दखिने वाले प्रभाव से अलग करने के लएि कयिाजताहै। यहाँ246 में से हर काम कोयदृच्छकि रूप से ऐसे समूह में रखागयाजहाँAI कीअनुमतथीयानहींथी। इससे औसतन दोसमूहों के बीच व्यवस्थति फ़र्क केवल AI कारह जताहै। यहीकारण है कि शोधकर्ताकह सकते हैं कि समय में बदलाव AI के कारण हुआ, न कि केवल यह कि जोलोग AI इस्तेमाल करते हैं वे कसीऔर वजह से पहले से तेज़ यादीमें होते हैं। यह इसलएि महत्वपूर्ण है क्योंकि AI कोडंगि के लाभ के समाान्य प्रमाण — स्वयं दीगई रपिोर्ट और बेंचमार्क स्कोर — ऐसानहींकर सकते। बेंचमार्क वास्तवकि काम नहींहै, और जैसायह अध्ययन दखिताहै, स्वयं कीधारणापूरे आत्मवशि्वास के साथ गलत होसकतीहै। यहाँडेवलपर नए खलिौने से जूझते नौसखिएि नहींथे। वे बड़े, परपिक्व ओपन-सोर्स प्रोजेक्टोंके स्थापति योगदानकर्ताथे, जनिहें वे अच्छीतरह जानते थे, और कुछ कोइन औज़ारोंका पहले से अनुभव भीथा।

शोधकर्ताओं ने क्याकयिा

- **यदृच्छकि नयिंत्रति परीक्षण** चलाया(METR: Joel Becker, Nate Rush, Beth Barnes, David Rein)। इसमें सोलह अनुभवीओपन-सोर्स डेवलपर शामलि थे। हर व्यक्तिउसीबड़े रपिॉज़िटरीपर काम कर रहाथाजसिमें वह नयिमति योगदान देताथाऔर जसिे अच्छीतरह जानताथा— संबंधति प्रोजेक्ट पर औसतन पाँच साल काअनुभव।

- **246 वास्तविक काम** इस्तेमाल किए — बग सुधार, फीचर और रीफ़ैक्टरिंग, जो उन्हीं प्रोजेक्टों के इश्यू ट्रैकर से लिए गए थे। हर काम को यादृच्छिक रूप से “AI की अनुमति” या “AI की अनुमति नहीं” समूह में रखा गया।
- “AI की अनुमति” का अर्थ था **2025 की शुरुआत के औज़ार: Cursor Pro के साथ Claude 3.5/3.7 Sonnet**। मुख्य मान हर काम को पूरा करने में लगा वास्तविक समय था। साथ में टीम ने पहले से डेवलपरों के अनुमान, काम के बाद उनके आकलन, और अर्थशास्त्र तथा मशीन-लर्निंग विशेषज्ञों की भविष्यवाणियाँ भी दर्ज कीं।

क्या मिला

- **AI के साथ कामों में 19% अधिक समय लगा।** तेज़ नहीं — धीमा। 95% विश्वास-अंतराल लगभग +2% से +39% तक है, इसलिए दिशा काफ़ी स्पष्ट है, भले ही प्रभाव का सटीक आकार नहीं।
- **सभी ने उलटा अनुमान लगाया था।** डेवलपरों ने 24% तेजी का अनुमान लगाया; मशीन-लर्निंग विशेषज्ञों ने लगभग 38%; अर्थशास्त्रियों ने लगभग 39%। तीनों समूहों को लगा था कि AI काफ़ी समय बचाएगा; घड़ी ने दिखाया कि कुछ समय बढ़ा।
- **धारणा का अंतर।** धीमे काम करने के बावजूद, काम के बाद डेवलपरों ने अनुमान लगाया कि AI ने उन्हें लगभग 20% तेज़ किया था — यानी अनुभव और वास्तविक समय-मान के बीच करीब 40 प्रतिशत-अंक का अंतर।
- **संभावित कारण, लेकिन सिद्ध कारण नहीं।** लेखक कई कारक सामने रखते हैं जो धीमेपन की व्याख्या कर सकते हैं: डेवलपर अपने कोडबेस को बहुत गहराई से जानते थे, इसलिए सहायक के पास जोड़ने को कम था; परिपक्व प्रोजेक्टों में गुणवत्ता के ऊँचे और अक्सर अनकहे मानक होते हैं; रिपॉज़िटरी बड़े हैं और उनमें बहुत-सा संदर्भ है जो मॉडल के पास नहीं होता; और वास्तविक समय प्रॉम्प्ट लिखने, AI के आउटपुट की समीक्षा करने और उसे सुधारने में जाता है। लेखक इन्हें संकेत मानते हैं, अंतिम निष्कर्ष नहीं।

यह क्या नहीं दिखाता

- यह **नहीं** दिखाता कि AI अधिकांश डेवलपरों को तेज़ नहीं करता। लेखक इसे साफ़ कहते हैं: अपने कोड को बहुत अच्छी तरह जानने वाले सोलह विशेषज्ञ औसत डेवलपर और औसत काम का प्रतिनिधित्व नहीं करते।
- यह **नहीं** दिखाता कि AI बेकार है, या दूसरी परिस्थितियों में लोगों को धीमा करता है — जैसे किसी कोडबेस में नए व्यक्ति, बिल्कुल नया प्रोजेक्ट, अपरिचित प्रोग्रामिंग भाषा या पूरी तरह अलग क्षेत्र।
- यह औज़ारों को समय में स्थिर नहीं मानता। यह 2025 की शुरुआत का Cursor और Claude 3.5/3.7 Sonnet है; लेखक स्पष्ट कहते हैं कि बेहतर औज़ार, या इन्हीं औज़ारों को बेहतर तरीक़े से इस्तेमाल करना, ठीक इसी परिस्थिति में भी नतीजा बदल सकता है।
- यह एक **प्रीप्रिंट** है (जुलाई 2025 में पोस्ट किया गया; अभी सहकर्मी-समीक्षित नहीं), और लेखक मानते हैं कि वे प्रयोग से पैदा हुए सभी कृत्रिम प्रभाव पूरी तरह खारिज नहीं कर सकते — हालाँकि अलग-अलग विश्लेषणों में मुख्य परिणाम कायम रहा।
- इससे यह सुझाव देने वाला निष्कर्ष भी नहीं निकाला जा सकता कि डेवलपरों को ज़रूर कोई दूसरा लाभ मिला होगा — उन्होंने अधिक सीखा, बेहतर महसूस किया या बेहतर कोड लिखा। यहाँ जिस चीज़ को मापा गया, “मुझे लगा मैं तेज़ था”, डेटा ठीक उसी धारणा का खंडन करता है।

प्रमाण कितना मजबूत है

- **डिज़ाइन असमान्य रूप से साफ़ है।** यादृच्छिक आवंटन, वास्तविक काम, वास्तविक रिपॉज़िटरी और वास्तविक समय-माप — यह उन स्वयं-रिपोर्ट और बेंचमार्क तालिकाओं से बड़ा कदम आगे है जिन पर AI कोडिंग के बहुत से दावे टिके हैं। 19% का धीमापन लेखकों की मज़बूती जाँचों में भी बना रहा।
- **सबसे सामान्य रूप से लागू होने वाला निष्कर्ष धारणा का अंतर है।** अपनी AI-सहायता से मिली गति पर विशेषज्ञों की अंतर्ज्ञान लगभग 40 प्रतिशत-अंक से गलत थी और गलती आशावादी दिशा में थी। इसलिए AI से उत्पादकता बढ़ने की हर स्वयं-रिपोर्ट को सावधानी से लेना चाहिए — इस अध्ययन के अपने आँकड़ों को भी।
- **यह एक समय-बिंदु की तस्वीर है, रुझान नहीं।** METR का फरवरी 2026 का follow-up, इसी तरह के डेवलपरों और नए औज़ारों के साथ, तेजी की दिशा दिखाता है — लौटकर आए डेवलपरों के लिए बहुत मोटे तौर पर −18% और नए प्रतिभागियों के लिए −4% — लेकिन लेखक इसे कमजोर प्रमाण बताते हैं, क्योंकि इसमें प्रतिभागी चयन का असर है: डेवलपर AI के बिना काम करने से अधिक इनकार करने लगे, भुगतान दर घट गई, और चुने गए कामों का स्वरूप बदल गया। सावधान निष्कर्ष यह है कि तस्वीर बदल रही है, और उस बदलाव को भी बिना बढ़ा-चढ़ाकर बताया जाना चाहिए।

यह क्यों महत्वपूर्ण है

AI और प्रोग्रामिंग पर बहस का बड़ा हिस्सा डेमो और अनुभूति पर चलता है: कोई चमकदार स्क्रीन रिकॉर्डिंग, कोई आत्मविश्वास भरा दावा, और उसके जवाब में आँखें घुमाना। इस शोध को पढ़ने लायक जो बनाता है वह यह है कि किसी ने उबाऊ लेकिन ज़रूरी प्रयोग किया— कामों को यादृच्छिक बाँटा, वास्तविक समय मापा, फिर लोगों से पूछा कि अनुभव कैसा रहा। जवाब दोनों पक्षों के लिए असहज है। यह उस कहानी को तोड़ता है कि AI कठिन और परिचित कोड पर विशेषज्ञ डेवलपरों को हमेशा बहुत तेज़ कर देता है। और यह साफ़-सुथरी विपरीत कहानी— "AI डेवलपरों को धीमा करता है, साबित हो गया" — भी नहीं टिकने देता, क्योंकि उसी टीम का नया डेटा पहले ही दूसरी दिशा में झुक रहा है। सबसे टिकाऊ सबक सबसे छोटा और सबसे मानवीय है: काम करने वालों को लगा वे तेज़ थे, जबकि मापने पर वे धीमे थे। "तेज़ लगता है" इस बात का प्रमाण नहीं कि वह सच में तेज़ है। मापिए।

संक्षिप्त सार

एक यादृच्छिक नियंत्रित परीक्षण में METR ने सोलह अनुभवी ओपन-सोर्स डेवलपरों से उन कोडबेस पर 246 वास्तविक काम करवाए जिन्हें वे अच्छी तरह जानते थे। लगभग आधे कामों में यादृच्छिक रूप से 2025 की शुरुआत के AI औज़ार (Cursor Pro और Claude 3.5/3.7 Sonnet) इस्तेमाल करने की अनुमति थी। डेवलपरों को उम्मीद थी कि AI काम का समय लगभग 24% घटा देगा; इसके बजाय पूरा करने का समय 19% बढ़ गया— और बाद में भी उन्हें लगा कि AI ने उन्हें लगभग 20% तेज़ किया था। यही धारणा-अंतर अध्ययन का सबसे स्पष्ट और मजबूत केंद्र है। लेकिन यह केवल सोलह डेवलपर, एक संकीर्ण परिस्थिति और 2025 की शुरुआत की एक तस्वीर है; लेखक साफ़ कहते हैं कि इससे यह साबित नहीं होता कि AI अधिकांश डेवलपरों की मदद नहीं करता, और उनका अपना 2026 follow-up पहले ही कुछ सावधानियों के साथ तेजी की दिशा दिखाता है। यह एक ऐसा सावधानी से किया गया मापन है जिसे गंभीरता से लेना चाहिए — AI कोडिंग पर अंतिम फैसला नहीं।

स्रोत

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

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

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