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

Təcrübəli proqramçılar AI ilə özlərini daha sürətli hiss etdilər, amma ölçülə bilən şəkildə daha yavaş işlədilər — əsl nəticə budur, AI ilə kodlaşdırma barədə hökm isə deyil

METR-in randomizə olunmuş nəzarətli sınağında on altı təcrübəli açıq mənbə proqramçısı yaxşı tanıdıqları kod bazalarında 246 real tapşırıq yerinə yetirdi; tapşırıqların təsadüfi yarısında 2025-ci ilin əvvəlindəki AI alətlərinə (Cursor Pro və Claude 3.5/3.7 Sonnet) icazə verilmişdi. Onlar AI-nin tapşırıq vaxtını təxminən 24% azaldacağını gözləyirdilər; əvəzində tamamlanma vaxtı 19% artdı — və sonradan da AI-nin onları təxminən 20% sürətləndirdiyinə inanırdılar. Bu qavrayış fərqi tədqiqatın ən kəskin və dayanıqlı nüvəsidir. Amma bu, on altı proqramçı, dar bir şərait və 2025-ci ilin əvvəlinin sabit görüntüsüdür: müəlliflər bunun AI-nin proqramçıların çoxuna kömək etmədiyini göstərmədiyini açıq deyirlər və onların öz 2026 davam tədqiqatı artıq müəyyən şərtlərlə sürətlənməyə işarə edir.

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

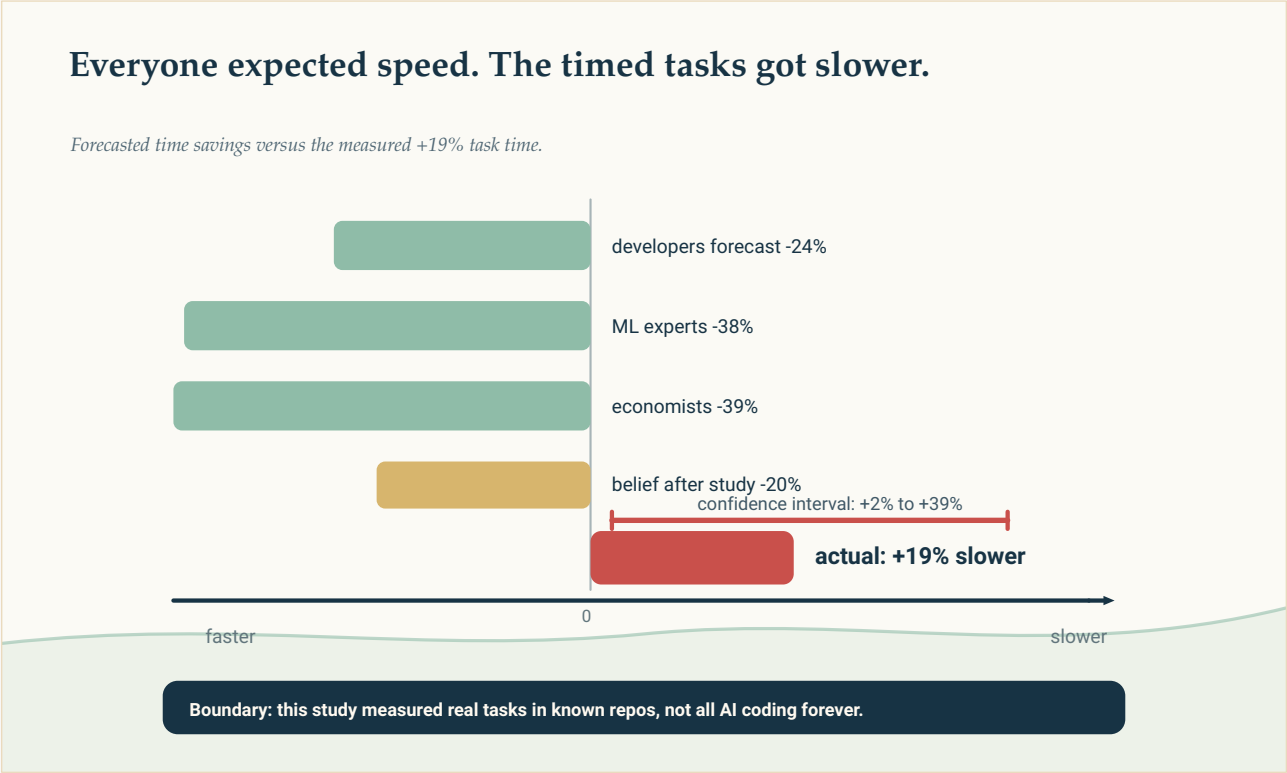
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Təcrübəli proqramçılar AI ilə özlərini daha sürətli hiss etdilər, amma ölçülə bilən şəkildə daha yavaş işlədilər — əsl nəticə budur, AI ilə kodlaşdırma barədə hökm isə deyil

Təcrübəli proqramçıdan AI kodlaşdırma köməkçisinin onu sürətləndirib-sürətləndirmədiyini soruşsanız, adətən bir rəqəm eşidərsiniz: mənə iyirmi, otuz faiz vaxt qazandırır. Eyni sualı iqtisadçıya və ya maşın öyrənməsi tədqiqatçısına versəniz, rəqəm daha da böyüyür. 2025-ci ilin əvvəlində METR komandası yavaş və bahalı olan işi görüb bunu yoxladı. Onlar böyük və yaxşı tanıdıqları kod bazalarında işləyən on altı təcrübəli açıq mənbə proqramçısına 246 real tapşırıq verdilər və hər tapşırıq üzrə sikkə atışı ilə ya AI alətlərindən istifadəyə icazə verdilər, ya da vermədilər. Sonra işi vaxtla ölçdülər.

Proqramçılar AI-nin tapşırıq vaxtını təxminən 24% azaldacağını əvvəlcədən proqnozlaşdırmışdılar. Əksinə oldu: AI ilə görülən tapşırıqlar **19% daha uzun çəkdi**. Ən diqqətəlayiq hissə isə budur — iş bitəndən sonra həmin proqramçılar yenə də AI-nin onları təxminən 20% sürətləndirdiyinə inanırdılar. Onlar daha yavaş işləyirdilər, amma özlərini daha sürətli hiss edirdilər; bu iki rəqəm arasındakı məsafə tədqiqatın ən maraqlı nəticəsidir.

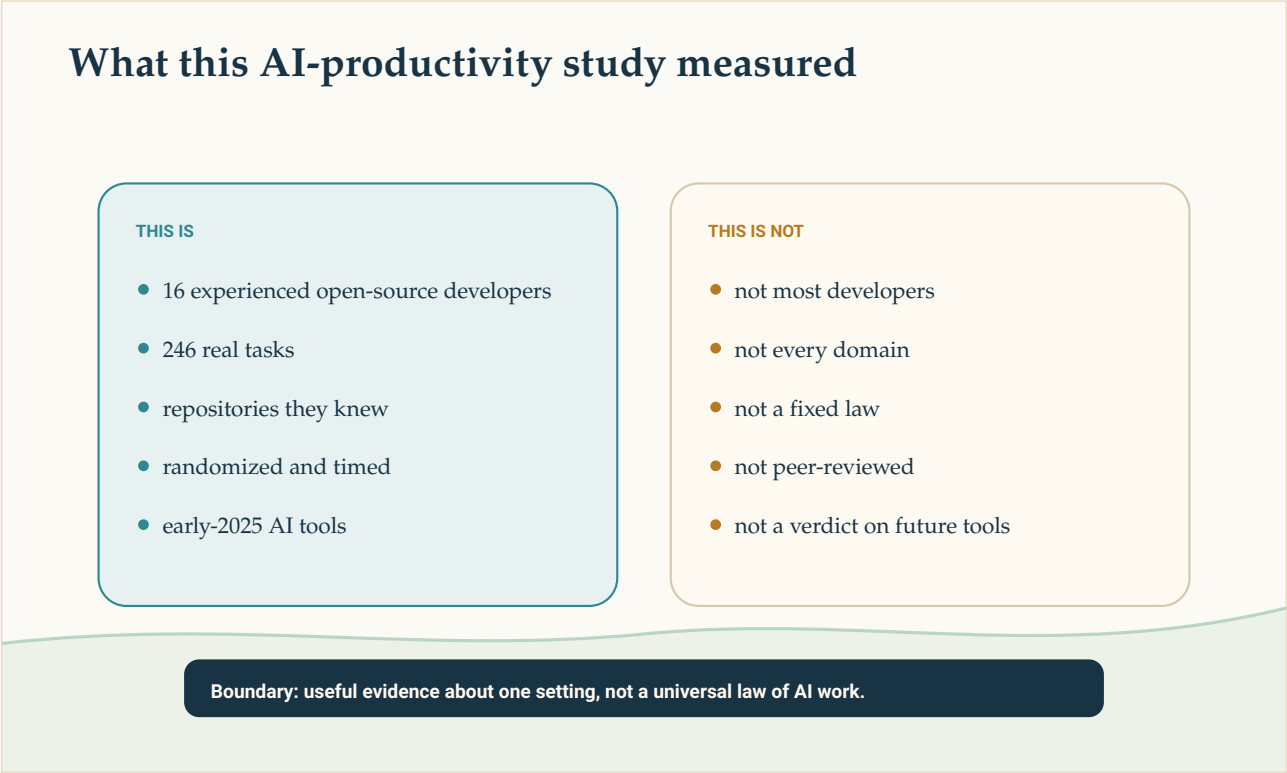
Bu, real və diqqətlə ölçülmüş nəticədir. Amma söhbət öz kod bazalarını çox yaxşı tanıyan on altı proqramçıdan, 2025-ci ilin əvvəlində mövcud olan alətlərdən gedir — və nəticə sadəcə "AI proqramçıları yavaşıdır" cümləsi deyil. Elə həmin komandanın daha sonrakı məlumatları artıq əks istiqamətə işarə edir.



Hamı AI-nin işi sürətləndirəcəyini proqnozlaşdırmışdı — proqramçılar −24%, ML mütəxəssisləri −38%,

iqtisadçılar −39%, proqramçıların işdən sonrakı öz qiymətləndirməsi isə −20%. Xronometr isə +19% yavaşıma göstərdi; interval +2%-dən +39%-ə qədər idi. Hiss edilənlə ölçülən arasındakı fərq əsas nəticədir.

Original diagram — The Clean Paper CC BY 4.0



Tədqiqatın ölçdüyü şey — 16 təcrübəli açıq mənbə proqramçısı, 246 real tapşırıq, tanış repozitoriyalar, 2025-ci ilin əvvəlinin alətləri, randomizasiya — və ölçmədiyi şeylər: proqramçıların əksəriyyəti deyil, digər sahələr deyil, dəyişməz qanun deyil, peer-review-dan keçmiş nəticə deyil.

Original diagram — The Clean Paper CC BY 4.0

Bu nəyi ölçdü və randomizə olunmuş sınaq bizə nə qazandırır

Randomizə olunmuş nəzarətli sınaq (RCT) tibbin real təsiri ümidverici görünən təsəvvürdən ayırmaq üçün istifadə etdiyi üsuldur. Burada 246 tapşırığın hər biri təsadüfi şəkildə AI-yə icazə verilən və ya verilməyən qrupa bölündü; buna görə orta hesabla iki qrup arasındakı yeganə sistematik fərq AI-nin özü idi. Məhz bu, başqa səbəblərdən AI-yə əl atan insanların daha sürətli və ya daha yavaş olmasını sadəcə müşahidə etmək əvəzinə, vaxtdakı dəyişikliyə AI-nin *səbəb olduğunu* deməyə imkan verir. Bu vacibdir, çünki AI ilə kodlaşdırmada məhsuldarlıq artımı barədə adi sübutlar — özünüqiymətləndirmə və benchmark nəticələri — bunu edə bilmir: benchmark real iş deyil, özünüqiymətləndirmə isə bu tədqiqatın göstərdiyi kimi, böyük inamla səhv ola bilər. Buradakı proqramçılar yeni oyuncağı əlləşdirən təcrübəsiz insanlar deyildilər. Onlar yaxşı tanıdıqları iri və yetkin açıq mənbə layihələrinin uzunmüddətli iştirakçıları idilər və alətlərdən müəyyən əvvəlki istifadə təcrübələri də vardı.

Müəlliflər nə etdilər

- **Randomizə olunmuş nəzarətli sınaq** keçirdilər (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). On altı təcrübəli açıq mənbə proqramçısı müntəzəm töhfə verdiyi və yaxşı tanıdığı böyük repozitoriyada işləyirdi — konkret layihələr üzrə orta təcrübə beş il idi.
- Həmin layihələrin öz issue tracker-lərindən götürülmüş **246 real tapşırıqdan** — səhv düzəlişləri, yeni funksiyalar və refaktorinqlərdən — istifadə etdilər. Hər tapşırıq təsadüfi olaraq "AI-yə icazə var" və ya "AI-yə icazə yoxdur" şəraitinə təyin edildi.
- "AI-yə icazə var" **2025-ci ilin əvvəlinin alətləri: Claude 3.5/3.7 Sonnet ilə Cursor Pro** demək idi. Əsas ölçü hər tapşırığın faktiki tamamlanma vaxtı idi. Komanda bununla yanaşı əvvəlcədən proqnozları (proqramçılardan), işdən sonrakı qiymətləndirmələri, həmçinin iqtisadiyyat və maşın öyrənməsi mütəxəssislərinin proqnozlarını topladı.

Nə tapdılar

- **AI ilə tapşırıqlar 19% daha uzun çəkdi.** Daha sürətli yox — daha yavaş. 95% etibar intervalı təxminən +2%-dən +39%-ə qədər uzanır; yəni dəqiq ölçü qeyri-müəyyən olsa da istiqamət kifayət qədər möhkəmdir.
- **Hamı əksini proqnozlaşdırmışdı.** Proqramçılar 24% sürətlənmə, maşın öyrənməsi mütəxəssisləri təxminən 38%, iqtisadçılar isə təxminən 39% gözləyirdi. Hər üç qrup AI-nin xeyli vaxt qazandıracağını düşünürdü; xronometr isə bir qədər vaxt itirdiyini göstərdi.
- **Qavrayış fərqi.** İş daha uzun çəkdiyi halda, proqramçılar yenə də AI-nin onları təxminən 20% sürətləndirdiyini hesab etdilər — hiss etdikləri ilə saatın göstərdiyi arasında təxminən 40 faiz-bəndlik fərq.
- **Mümkün səbəblər, amma sübut edilmiş hökmlər deyil.** Müəlliflər yavaşımanı izah edə biləcək amilləri sıralayırlar: bu proqramçılar öz kod bazalarını dərindən tanıyırlar, ona görə köməkçinin əlavə edə biləcəyi dəyər azdır; yetkin layihələrdə yüksək və çox vaxt yazılmamış keyfiyyət standartları var; repozitoriyalar böyükdür və modelin malik olmadığı çoxlu kontekst daşıyır; prompt yazmağa, sonra AI çıxışını nəzərdən keçirməyə və düzəltməyə real vaxt gedir. Bunlar yekun izah kimi yox, araşdırılmalı ehtimallar kimi təqdim olunur.

Bu nəyi göstərmir

- Bu, AI-nin proqramçıların əksəriyyətini sürətləndirmədiyini **göstərmir**. Müəlliflər bunu birbaşa deyirlər: əzbər bildikləri kod üzərində işləyən on altı ekspert orta proqramçını və orta tapşırığı təmsil etmir.
- AI-nin faydasız olduğunu və ya başqa şəraitlərdə — kod bazasına yeni gələnlərdə, sıfırdan layihələrdə, tanış olmayan dillərdə və ya tam başqa sahələrdə — insanları yavaşıtdığını

göstərmir.
- Alətləri zamanın bir nöqtəsində **dondurmur**. Bu, 2025-ci ilin əvvəlindəki Cursor və Claude 3.5/3.7 Sonnet-dir; müəlliflər daha yaxşı alətlərin və ya eyni alətlərdən daha yaxşı istifadə üsullarının hətta bu eyni şəraitdə nəticəni dəyişə biləcəyini açıq deyirlər.
- Bu bir **preprintdir** (2025-ci ilin iyulunda yayımlanıb, hələ peer-review-dan keçməyib) və müəlliflər eksperimental artefaktları tam istisna edə bilmədiklərini qeyd edirlər — hərçənd nəticə müxtəlif analizlərdə davamlı qalıb.
- Bu nəticə "bəlkə proqramçılar başqa cür qazandılar — daha çox öyrəndilər, daha xoşbəxt oldular, daha keyfiyyətli kod yazdılar" kimi rahatladıcı oxunuşa **əsas vermir**. Burada ölçülən məqamlardan biri, hiss edilən sürətlənmə, məhz məlumatların təkzib etdiyi şeydir.

Sübut nə qədər güclüdür

- **Dizayn qeyri-adi dərəcədə dürüstdür.** Randomizə olunmuş, real tapşırıqlar, real repozitoriyalar, faktiki vaxt ölçümü — AI kodlaşdırması barədə iddiaların çoxunun söykəndiyi özünüqiymətləndirmələr və benchmark siyahıları ilə müqayisədə böyük addımdır. 19%-lik yavaşıma müəlliflərin dayanıqlılıq yoxlamalarından keçib.
- **Ən geniş daşına bilən nəticə qavrayış fərqidir.** İnsanların öz AI sürətlənməsinə dair ekspert intuisiyası optimist istiqamətdə təxminən 40 faiz-bənd səhv çıxdı. Bu, AI məhsuldarlığı barədə *bütün* özünübildiriş iddialarına qarşı ehtiyatlılıq siqnalıdır — elə bu tədqiqatın öz rəqəmləri də daxil olmaqla.
- **Bu, trend yox, bir anlıq görüntüdür.** METR-in 2026-cı ilin fevralındakı davam tədqiqatı eyni tip proqramçılarla, daha yeni alətlərdə, sürətlənməyə işarə edir — geri qayıdan proqramçılar üçün çox təxmini −18%, yeni iştirakçılar üçün −4%. Amma müəlliflər bunu zəif sübut sayırlar, çünki kimlərin iştirak etməyə razı olduğu nəticəni təhrif edir: proqramçılar AI-siz işləməkdən getdikcə daha çox imtina edirdi, ödəniş dərəcəsi azalmışdı və tapşırıq seçimi meyllənmişdi. Dürüst oxunuş budur: mənzərə dəyişir və hətta bu dəyişiklik barədə məlumat da tərəzinin gözünə barmaq basmadan verilir.

Niyə vacibdir

AI və proqramlaşdırma barədə mübahisələrin çoxu demo və hisslər üzərində qurulur: parlaq ekran videosu, özünəinamlı iddia, qarşı tərəfdən istehzalı baxış. Nadir olan və bu işi oxumağa dəyər edən şey odur ki, kimsə darıxdırıcı eksperimenti həqiqətən aparıb — randomizə et, real işi vaxtla ölç, sonra insanlardan necə keçdiyini soruş. Cavab hər iki düşərgə üçün narahatdır. Bir tərəfdən, AI-nin mürəkkəb və tanış kod üzərində ekspert proqramçıları hər yerdə turbosürətləndirdiyi hekayəsini deşir. Digər tərəfdən, "AI proqramçıları yavaşıdır, sübut olundu" kimi səliqəli əks hökmü də deşir, çünki eyni komandanın daha yeni məlumatları artıq əks istiqamətə meyllənir. Ən davamlı dərs həm də ən kiçik və ən insani olandır: işi görən insanlar ölçülə bilən şəkildə daha yavaş olduqları halda özlərini daha sürətli hiss edirdilər. "Daha

sürətli hiss olunur” — bunun həqiqətən daha sürətli olduğuna sübut deyil. Ölçün.

Təmiz xülasə

METR randomizə olunmuş nəzarətli sınaqda on altı təcrübəli açıq mənbə proqramçısına yaxşı tanıdıqları kod bazalarında 246 real tapşırıq verdi; tapşırıqların təsadüfi yarısında 2025-ci ilin əvvəlindəki AI alətlərinə (Cursor Pro və Claude 3.5/3.7 Sonnet) icazə vardı. Proqramçılar AI-nin tapşırıq vaxtını təxminən 24% azaldacağını gözləyirdilər; əvəzində tamamlanma vaxtı 19% artdı — və sonradan da AI-nin onları təxminən 20% sürətləndirdiyinə inanırdılar. Bu qavrayış fərqi tədqiqatın ən kəskin və dayanıqlı nüvəsidir. Amma bu, on altı proqramçı, dar bir şərait və 2025-ci ilin əvvəlinin sabit görüntüsüdür; müəlliflər açıq deyirlər ki, nəticə AI-nin proqramçıların çoxuna kömək etmədiyini göstərmir və onların öz 2026 davam tədqiqatı artıq müəyyən şərtlərlə sürətlənməyə işarə edir. Ciddiyə alınmalı diqqətli ölçü — AI kodlaşdırması barədə hökm yox.

Mənbələr

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/