The Clean Paper

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

Developers wenye uzoefu walihisi AI inawafanya wawe haraka huku wakifanya kazi polepole zaidi kwa kipimo — finding halisi, si verdict juu ya AI coding

Randomized controlled trial ya METR iliwaweka open-source developers 16 wenye uzoefu kufanya tasks 246 halisi kwenye codebases walizozijua vizuri, huku early-2025 AI tools (Cursor Pro pamoja na Claude 3.5/3.7 Sonnet) zikiruhusiwa kwa random half. Walitarajia AI ipunguze task time kwa ~24%; badala yake iliiongeza kwa 19% — na baadae bado waliamini imewafanya wawe haraka kwa ~20%. Perception gap hiyo ndiyo core robust ya study. Lakini ni developers 16, setting moja nyembamba na snapshot ya mapema 2025; waandishi wanasema wazi haisemi AI haimsaidii developer wa kawaida, na follow-up yao ya 2026 tayari inaelekea speedup, ikiwa na caveats zake.

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

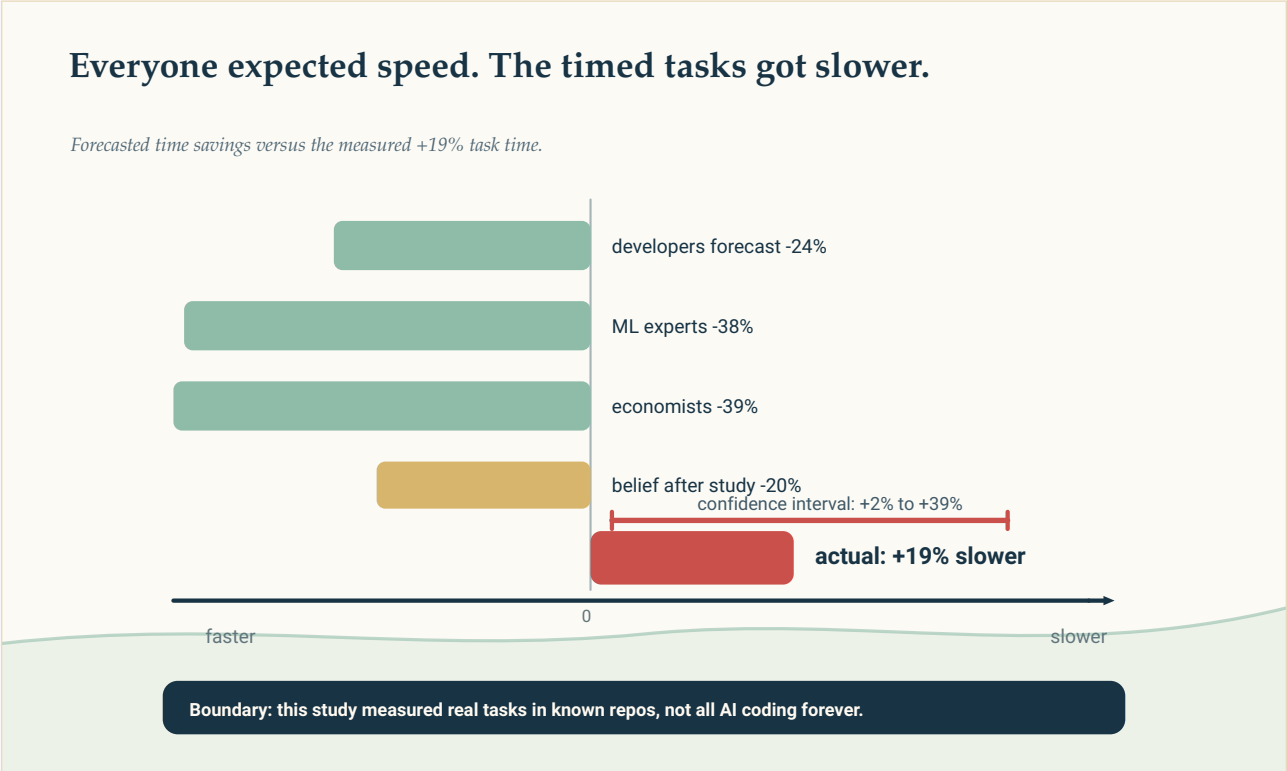
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Developers wenye uzoefu walihisi AI inawafanya wawe haraka huku wakifanya kazi polepole zaidi kwa kipimo — matokeo halisi, si hukumu juu ya AI coding

Muulize programmer mwenye uzoefu kama AI coding assistant inamharakisha na mara nyingi utapata namba: inanipunguzia muda kwa asilimia ishirini, thelathini. Muulize economist au machine-learning researcher, namba huwa kubwa zaidi. Mapema 2025 timu ya METR ilifanya jambo la polepole na ghali: ikapima. Walichukua open-chanzo developers 16 wenye uzoefu, wakawapa tasks 246 halisi kutoka codebases kubwa walizozijua vizuri, na — kwa random assignment task moja baada ya nyingine — wakati mwingine waliruhusu AI zana, wakati mwingine hawakurusu. Kisha walipima muda.

Developers walikuwa wametabiri AI ingepunguza task time kwa takribani 24%. Ilifanya kinyume: tasks zilizofanywa na AI zilichukua **19% muda zaidi**. Na hii ndiyo sehemu inayostahili kukaa nayo — baada ya kumaliza, developers wale wale bado waliamini AI ilikuwa imewafanya wawe haraka kwa takribani 20%. Walikuwa polepole zaidi, lakini walihisi haraka zaidi, na umbali kati ya namba hizo mbili ndiyo sehemu ya kuvutia zaidi ya utafiti.

Haya ni matokeo halisi, yaliyopimwa kwa uangalifu. Pia ni developers 16, kwenye repositories wanazozijua kwa undani, wakitumia zana za mapema 2025 — na si sentensi tambarare "AI huwafanya developers wawe polepole." Data ya baadaye kutoka timu ileile tayari inaelekea upande mwingine.



Kila kundi lilitabiri AI ingeharakisha kazi — developers −24%, ML experts −38%, economists −39%, na belief ya developers wenyewe baada ya utafiti −20%. Stopwatch ilipata +19% polepole zaidi, na interval ya +2% hadi

+39%. pengo kati ya kilichohisiwa na kilichopimwa ndiyo matokeo.

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.

Kile utafiti unapima ni finyu lakini halisi: watengenezaji 16 wenye uzoefu wa miradi ya chanzo huria, kazi 246 za kweli katika hazina za msimbo walizozijua, na zana za AI za mapema 2025 zilizogawiwa kwa nasibu kwa nusu ya kazi. Haimaanishi watengenezaji wote, taaluma nyingine au sheria isiyobadilika kuhusu AI, na wakati wa utafiti matokeo hayakuwa bado yamepitia mapitio ya rika.

Original diagram — The Clean Paper CC BY 4.0

Hii ilipima nini, na jaribio la nasibu inakupa nini

Randomized controlled jaribio (RCT) ni zana inayotumika kutenganisha athari halisi na matumaini. Hapa, kila moja ya tasks 246 iliwekwa kwa random kuwa “AI allowed” au “AI disallowed,” hivyo kwa wastani tofauti ya kimfumo kati ya piles mbili ni AI yenyewe. Hilo ndilo linaloruhusu kusema AI *ilisababisha* mabadiliko ya muda, badala ya kugundua tu kwamba watu wanaochagua AI kwa sababu nyingine huwa haraka au polepole zaidi. Ni muhimu kwa sababu ushahidi wa kawaida wa gains za AI coding — self-reports na alama za vipimo linganishi — hauwezi kufanya hivi: kipimo linganishi si kazi halisi, na self-report, kama utafiti huu inavyoonyesha, inaweza kuwa wrong kwa confidence. Developers hapa hawakuwa novices wakihangaika na toy mpya. Walikuwa contributors waliozoea projects kubwa na mature za open chanzo walizozijua vizuri, wakiwa na experience fulani ya zana hizi tayari.

Waandishi walifanya nini

- Walifanya **randomized controlled jaribio** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Open-chanzo developers 16 wenye uzoefu walifanya kazi kwenye repository kubwa ambayo kila mmoja alikuwa contributor wa kawaida na aliijua vizuri — wastani wa miaka mitano kwenye project husika.
- Walitumia **tasks 246 halisi** — bug fixes, features na refactors kutoka issue trackers za projects hizo. Kila task iliwekwa kwa random kuwa "AI allowed" au "AI disallowed."
- "AI allowed" ilimaanisha **tooling ya mapema 2025: Cursor Pro pamoja na Claude 3.5/3.7 Sonnet.** Primary measure ilikuwa actual completion time kwa task. Pia walikusanya forecasts za developers kabla, estimates zao baada, na predictions kutoka economics na machine-learning experts.

Walichokuta

- **Kwa AI, tasks zilichukua 19% muda zaidi.** Si haraka — polepole. 95% kipindi cha kujiamini ni takribani +2% hadi +39%, hivyo direction ni imara hata kama exact size siyo.
- **Kila mtu alikuwa ametabiri kinyume.** Developers walitabiri speedup ya 24%; ML experts takribani 38%; economists takribani 39%. Makundi yote yalitarajia AI kuokoa muda mwingi; stopwatch ilipata iliongeza muda.
- **Perception pengo.** Baada ya kufanya kazi na kuibuka polepole zaidi, developers bado walikadiria AI ilikuwa imewafanya wawe haraka kwa takribani 20% — pengo karibu points 40 kati ya walichohisi na saa ilichorekodi.
- **Sababu zinazowezekana, zimepimwa kama leads si verdicts.** Waandishi wanataja mambo yanayoweza kueleza slowdown: developers hawa wanajua codebases zao kwa undani, hivyo assistant ana nafasi ndogo ya kuongeza; mature projects zina standards za quality nyingi, mara nyingi implicit; repositories ni kubwa na zina context ambayo modeli haina; na muda halisi huenda kwenye prompting, reviewing na correcting AI output. Wanaziwasilisha kama hypotheses, si hitimisho.

Kile ambacho hii haionyeshi

- **Haionyeshi** kwamba AI haimharakishi developer wa kawaida. Waandishi wanasema wazi: experts 16 kwenye code wanayoijua kwa moyo si average developer kwenye average task.
- **Haionyeshi** AI haina matumizi, au kwamba huwafanya watu kuwa polepole katika settings nyingine — newcomers kwenye codebase, greenfield kazi, unfamiliar languages, au fields nyingine.
- **Haiigandishi zana.** Hizi ni Cursor na Claude 3.5/3.7 Sonnet za mapema 2025; waandishi wanaweka wazi zana bora zaidi, au matumizi bora ya hizi, yanaweza kubadilisha matokeo hata katika setting hii.
- Ni **preprint** (Julai 2025, bado haijafanyiwa mapitio ya rika), na waandishi wanasema hawawezi kuondoa kabisa experimental artifacts — ingawa matokeo ilisimama katika analyses zao.

- **Hairuhusu** tafsiri ya kutuliza kwamba developers lazima walipata faida nyingine — walijifunza zaidi, walifurahia zaidi, waliandika code bora. Kitu kimoja kilichopimwa hapa, felt speedup, ndicho data inapingana nacho.

Ushahidi una nguvu kiasi gani

- **Design ni ya uaminifu isivyo kawaida.** Randomized, real tasks, real repositories, actual timing — hatua kubwa juu ya self-reports na kipimo linganishi leaderboards ambazo madai nyingi za AI coding hutegemea. Slowdown ya 19% ilisimama katika robustness checks za waandishi.
- **matokeo inayohamishika zaidi ni perception pengo.** Expert intuition kuhusu AI speedup yake mwenyewe ilikuwa wrong kwa takribani points 40, katika direction optimistic. Hii ni caution juu ya *kila* self-reported productivity gain ya AI — pamoja na namba za utafiti huu yenyewe.
- **Ni picha ya wakati mmoja, si mwelekeo wa kudumu.** Ufuatiliaji wa METR wa Februari 2026 kwa watengenezaji wa aina ileile na zana mpya zaidi unaelekea kuonyesha kasi zaidi — makadirio ya takribani 18% kwa waliorejea na 4% kwa wapya — lakini waandishi wanaita ushahidi huo dhaifu kwa sababu uteuzi wa washiriki ulibadilika: watengenezaji wengi walikataa kufanya kazi bila AI, malipo yalipungua na aina za kazi zilizochaguliwa zikabadilika. Usomaji wa tahadhari ni kwamba hali inabadilika, na hata mabadiliko hayo yanapaswa kutafsiriwa kwa uangalifu.

Kwa nini ni muhimu

Mjadala mwingi kuhusu AI na programming unaendeshwa na demos na vibes: screen recording nzuri, dai yenye confidence, eye-roll ya upande mwingine. Kitu nadra, na kinachofanya utafiti huu istahili kusomwa, ni kwamba mtu alifanya experiment boring — randomize, time real kazi, kisha waulize watu ilikuwaje. Jibu ni uncomfortable kwa camps zote mbili. Linatoboa simulizi kwamba AI kwa uniform huwatia turbo expert developers kwenye hard, familiar code. Pia linatoboa opposite tidy simulizi — "AI huwafanya developers wawe slow, proven" — kwa sababu newer data ya team ileile tayari inaelekea upande mwingine. Somo linaloweza kudumu zaidi ni dogo na la kibinadamu: watu waliokuwa wakifanya kazi walihisi haraka huku wakipimwa polepole. "Inahisi haraka" si ushahidi kwamba ni haraka. Pima.

Muhtasari safi

Katika jaribio la nasibu lenye kundi dhibiti, METR iliwaweka watengenezaji 16 wenye uzoefu wa chanzo huria kufanya kazi 246 za kweli katika codebase walizozijua vizuri. Katika nusu ya kazi iliyochaguliwa kwa nasibu waliruhusiwa kutumia zana za AI za mapema 2025 — Cursor Pro pamoja na Claude 3.5/3.7 Sonnet. Watengenezaji walitarajia AI ipunguze muda wa kazi kwa takribani 24%; badala yake muda uliongezeka kwa 19%. Cha kuvutia zaidi, baada ya kazi bado waliamini AI ilikuwa imewaharakisha kwa karibu 20%. Pengo kati ya hisia na kipimo ndilo matokeo thabiti zaidi ya utafiti.

Lakini sampuli ilikuwa watengenezaji 16 tu, mazingira yalikuwa finyu na huu ulikuwa picha ya zana za mapema 2025. Waandishi wanasema wazi kuwa hilo halimaanishi AI haimsaidii mtengenezaji wa kawaida, na ufuatiliaji wao wa 2026 tayari unaelekea upande wa kuharakisha kazi, ukiwa na tahadhari zake. Hiki ni kipimo makini kinachostahili kuchukuliwa kwa uzito, si hukumu ya jumla kuhusu programu zinazosaidia kuandika msimbo.

Vyanzo

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/