The Clean Paper

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

Iskusni programeri osjećali su da su brži uz AI, iako su mjerljivo radili sporije — to je pravi nalaz, a ne presuda o programiranju uz AI

U randomiziranom kontroliranom ispitivanju METR je dao šesnaestorici iskusnih developera otvorenog koda 246 stvarnih zadataka na bazama koda koje dobro poznaju, uz alate AI-ja s početka 2025. (Cursor Pro i Claude 3.5/3.7 Sonnet) dopuštene na nasumično odabranoj polovici. Očekivali su da će AI skratiti vrijeme rada za oko 24%; umjesto toga, vrijeme završetka poraslo je za 19% — a poslije su i dalje vjerovali da ih je AI ubrzao za oko 20%. Taj jaz između percepcije i mjerenja najčvršći je dio studije. Ali riječ je o šesnaest developera, jednom uskom okruženju i fiksnoj snimci s početka 2025.: autori izričito navode da to ne pokazuje kako AI ne pomaže većini developera, a njihov vlastiti nastavak iz 2026. već naginje prema ubrzanju, uz vlastita ograničenja.

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

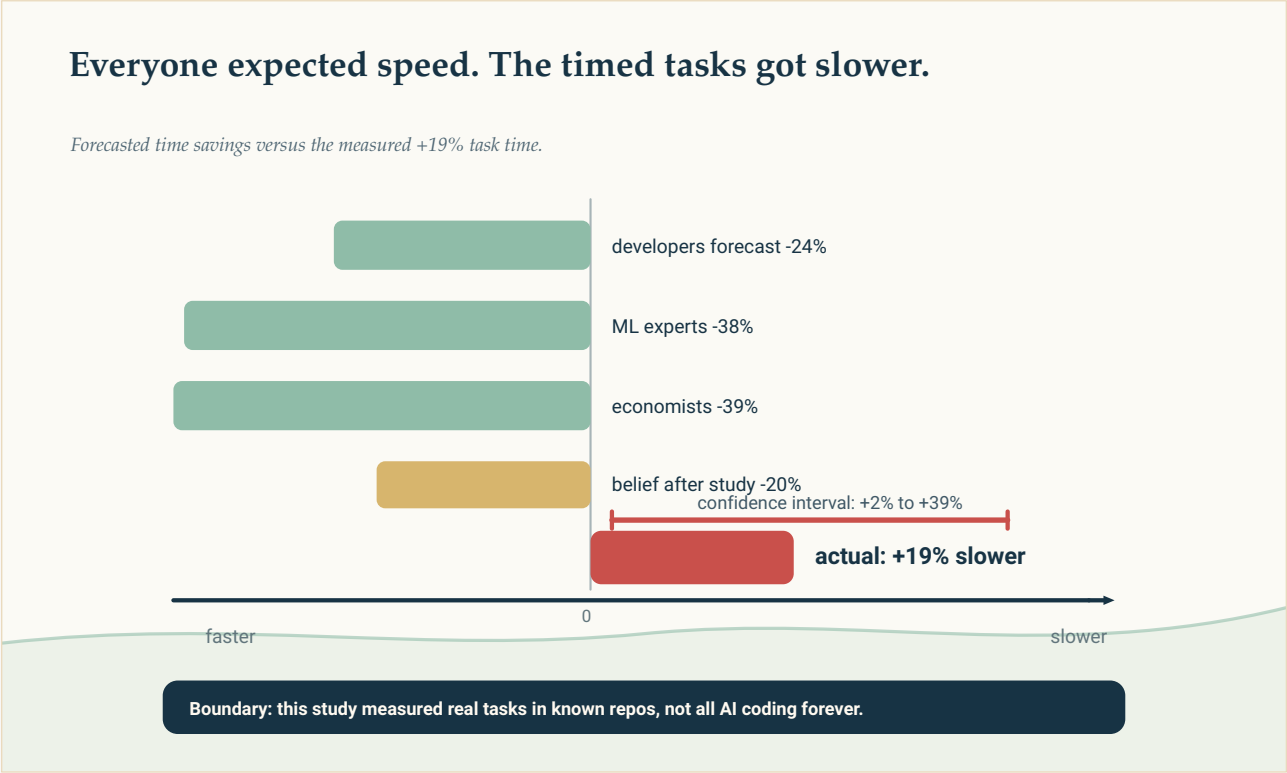
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Iskusni programeri osjećali su da su brži uz AI, iako su mjerljivo radili sporije — to je pravi nalaz, a ne presuda o programiranju uz AI

Pitate li iskusnog programera ubrzava li ga AI pomoćnik za programiranje, obično ćete dobiti broj: uštedi mi dvadeset, trideset posto vremena. Pitate li ekonomista ili istraživača strojnog učenja, broj postane još veći. Početkom 2025. tim METR-a napravio je ono sporo i skupo: provjerio je. Uključili su šesnaest iskusnih developera otvorenog koda, dali im 246 stvarnih zadataka iz velikih baza koda koje su dobro poznavali i — bacanjem novčića za svaki pojedini zadatak — ili dopustili uporabu AI alata ili je zabranili. Zatim su mjerili vrijeme rada.

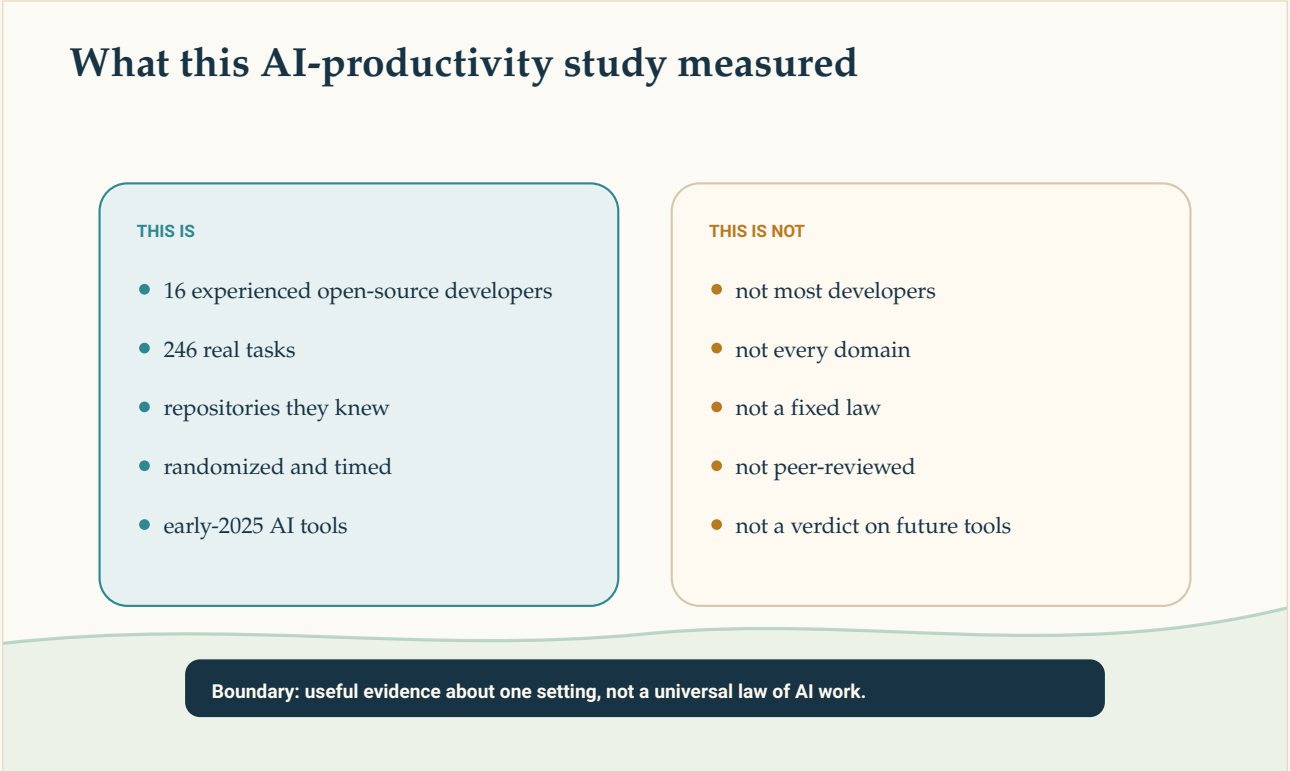
Developeri su unaprijed procijenili da će im AI skratiti vrijeme potrebno za zadatak za oko 24%. Dogodilo se suprotno: zadaci rađeni uz AI trajali su **19% dulje**. A ovo je dio nad kojim vrijedi zastati — nakon završetka rada isti su developeri i dalje vjerovali da ih je AI ubrzao, za oko 20%. Bili su sporiji, osjećali su se bržima, a razmak između ta dva broja najzanimljiviji je dio studije.

To je stvaran i pažljivo izmjeren rezultat. Ali riječ je o šesnaest developera, na repozitorijima koje poznaju u detalje, uz alate iz početka 2025. — i to nije isto što i tvrditi: „AI usporava developere." Kasniji podaci istog tima već pokazuju u suprotnom smjeru.



Svi su predvidjeli da će AI ubrzati rad — developeri −24%, stručnjaci za strojno učenje −38%, ekonomisti −39%, a i naknadna procjena samih developera bila je −20%. Štoperica je pokazala +19% sporiji rad, uz interval od +2% do +39%. Razlika između doživljenog i izmjerenog pravi je nalaz.

Original diagram — The Clean Paper CC BY 4.0



Što je studija mjerila — 16 stručnih developera otvorenog koda, 246 stvarnih zadataka, poznate repozitorije, alate s početka 2025. i randomizaciju — i što nije: ne većinu developera, ne druga područja, ne nepromjenjiv zakon i ne recenzirani rad.

Original diagram — The Clean Paper CC BY 4.0

Što je ovdje mjereno i zašto randomizirano ispitivanje toliko vrijedi

Randomizirano kontrolirano ispitivanje (RCT) alat je kojim medicina razlikuje stvaran učinak od onoga kojem se samo nadamo. Ovdje je svaki od 246 zadataka nasumično dodijeljen uvjetu u kojem je AI bio dopušten ili zabranjen, tako da je u prosjeku jedina sustavna razlika između dviju skupina upravo AI. Zbog toga se može reći da je AI *uzrokovao* promjenu vremena, umjesto da se samo primijeti kako su ljudi koji posežu za AI-jem slučajno brži ili sporiji zbog nekog drugog razloga. To je važno jer uobičajeni dokazi za dobitke od AI programiranja — samoprocjene i rezultati na benchmarkovima — to ne mogu pokazati: benchmark nije stvaran posao, a samoprocjena, kako pokazuje ova studija, može biti samouvjereno pogrešna. Developeri ovdje nisu bili početnici koji se muče s novom igračkom. Bili su etablirani suradnici na velikim, zrelim projektima otvorenog koda koje su dobro poznavali, uz nešto prethodnog iskustva s tim alatima.

Što su autori napravili

- Proveli su **randomizirano kontrolirano ispitivanje** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Šesnaest iskusnih developera otvorenog koda radilo je na velikim repozitorijima

kojima redovito pridonose i koje dobro poznaju — u prosjeku su na tim konkretnim projektima radili pet godina.
- Koristili su **246 stvarnih zadataka** — ispravke pogrešaka, nove značajke i refaktoriranja preuzete iz vlastitih sustava za praćenje problema tih projekata. Svaki je zadatak nasumično dodijeljen uvjetu „AI dopušten” ili „AI zabranjen”.
- „AI dopušten” značilo je **alate s početka 2025.: Cursor Pro s modelima Claude 3.5/3.7 Sonnet.** Glavna mjera bilo je stvarno vrijeme potrebno za dovršetak svakog zadatka. Tim je uz to prikupio predviđanja developera prije rada i njihove procjene poslije, kao i predviđanja stručnjaka iz ekonomije i strojnog učenja.

Što su otkrili

- **Uz AI zadaci su trajali 19% dulje.** Ne kraće — dulje. Interval pouzdanosti od 95% proteže se otprilike od +2% do +39%, pa je smjer učinka prilično čvrst čak i ako njegova točna veličina nije.
- **Svi su predvidjeli suprotno.** Developeri su očekivali ubrzanje od 24%; stručnjaci za strojno učenje oko 38%; ekonomisti oko 39%. Sve tri skupine očekivale su veliku uštedu vremena; mjerenje je pokazalo da je AI nešto vremena koštao.
- **Jaz u percepciji.** Nakon što su posao završili sporije, developeri su i dalje procijenili da ih je AI ubrzao za oko 20% — razlika od približno 40 postotnih bodova između osjećaja i onoga što je pokazao sat.
- **Moguća objašnjenja, odvagnuta ali ne dokazana.** Autori navode čimbenike koji bi mogli objasniti usporavanje: ovi developeri duboko poznaju vlastite baze koda pa pomoćnik ima manje toga za dodati; zreli projekti imaju visoke i često implicitne standarde kvalitete; repozitoriji su veliki i puni konteksta koji model nema; a stvarno se vrijeme troši i na pisanje upita te pregledavanje i ispravljanje AI izlaza. Autori ih predstavljaju kao tragove, ne kao konačne odgovore.

Što ovo ne pokazuje

- **Ne** pokazuje da AI ne ubrzava većinu developera. Autori to kažu izravno: šesnaest stručnjaka na kodu koji poznaju gotovo napamet nije prosječan developer na prosječnom zadatku.
- **Ne** pokazuje da je AI beskoristan niti da usporava ljude u drugim okolnostima — novopridošle u repozitorij, greenfield projekte, nepoznate programske jezike ili sasvim druga područja.
- **Ne** zamrzava alate u vremenu. Ovo su Cursor i Claude 3.5/3.7 Sonnet s početka 2025.; autori izričito navode da bolji alati ili bolji načini njihove uporabe mogu promijeniti rezultat čak i u potpuno istom okruženju.
- Riječ je o **preprintu** (objavljenom u srpnju 2025., još bez recenzije), a autori napominju da ne mogu potpuno isključiti eksperimentalne artefakte — iako se nalaz održao kroz njihove različite analize.
- Studija **ne** opravdava utješnu pretpostavku da su developeri možda dobili nešto drugo — više naučili, bili zadovoljniji ili pisali kvalitetniji kod. Jedino što je ovdje izmjereno, osjećaj ubrzanja, upravo je ono čemu podaci proturječe.

Koliko su dokazi jaki

- **Dizajn je neuobičajeno pošten.** Randomizacija, stvarni zadaci, stvarni repozitoriji i stvarno mjerenje vremena veliki su korak naprijed u odnosu na samoprocjene i ljestvice benchmarkova na kojima počiva većina tvrdnji o AI programiranju. Usporavanje od 19% ostalo je prisutno i u provjerama robusnosti koje su proveli autori.
- **Najprenosiviji je nalaz jaz u percepciji.** Stručna intuicija o vlastitom ubrzanju uz AI pogriješila je za otprilike 40 postotnih bodova, i to u optimističnom smjeru. To je upozorenje za *svaku* samoprijavljenu dobit u produktivnosti od AI-ja — uključujući i brojke iz ove studije.
- **Ovo je snimka trenutka, ne trend.** METR-ov vlastiti nastavak iz veljače 2026., s istom vrstom developera ali novijim alatima, naginje prema ubrzanju — vrlo grubo oko −18% za developere koji su se vratili i −4% za nove — no autori to označavaju kao slab dokaz, izobličen time tko je pristao sudjelovati (developeri su sve češće odbijali raditi bez AI-ja, naknada je pala, a odabir zadataka se promijenio). Pošteno čitanje jest da se slika mijenja, a čak se i ta promjena izvještava bez navijanja.

Zašto je to važno

Većina rasprave o AI-ju i programiranju počiva na demonstracijama i dojmovima: glatko snimljen video zaslona, samouvjerena tvrdnja, pa jednako samouvjereno kolutanje očima s druge strane. Ono što je rijetko i zbog čega ovu studiju vrijedi pročitati jest činjenica da je netko proveo dosadan eksperiment — nasumično dodijeli uvjete, mjeri stvaran rad, pa pitaj ljude kako je prošlo. Odgovor je neugodan za oba tabora. Ruši priču da AI jednako snažno ubrzava stručne developere na teškom, dobro poznatom kodu. Ali ruši i urednu suprotnu priču — „dokazano je da AI usporava developere” — jer noviji podaci istog tima već naginju drugom smjeru. Najtrajnija pouka ujedno je najmanja i najljudskija: ljudi koji su radili osjećali su se brže dok su mjerljivo bili sporiji. „Osjeća se brže” nije dokaz da jest. Izmjerite.

Sažetak bez uljepšavanja

U randomiziranom kontroliranom ispitivanju METR je dao šesnaestorici iskusnih developera otvorenog koda 246 stvarnih zadataka na bazama koda koje dobro poznaju, pri čemu su alati AI-ja s početka 2025. (Cursor Pro i Claude 3.5/3.7 Sonnet) bili dopušteni na nasumično odabranoj polovici zadataka. Developeri su očekivali da će AI skratiti vrijeme rada za oko 24%; umjesto toga, vrijeme završetka poraslo je za 19% — a nakon toga su i dalje vjerovali da ih je AI ubrzao za oko 20%. Taj jaz između percepcije i mjerenja oštro je i robusno središte studije. Ali riječ je o šesnaest developera, jednom uskom okruženju i fiksnoj snimci s početka 2025.; autori izričito navode da to ne pokazuje kako AI ne pomaže većini developera, a njihov vlastiti nastavak iz 2026. već naginje prema ubrzanju, uz vlastita ograničenja. Pažljivo mjerenje koje vrijedi ozbiljno shvatiti — ne presuda o programiranju uz AI.

Izvori

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/