



The CLEAN PAPER

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

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

U randomizovanom kontrolisanom 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) dozvoljene na nasumično odabranoj polovini. Očekivali su da će AI skratiti vreme rada za oko 24%; umesto toga, vreme završetka poraslo je za 19% — a posle su i dalje verovali da ih je AI ubrzao za oko 20%. Taj jaz između percepcije i merjenja najčvršći je deo studije. Ali reč 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 sopstveni nastavak iz 2026. već naginje prema ubrzanju, uz sopstvena ograničenja.

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

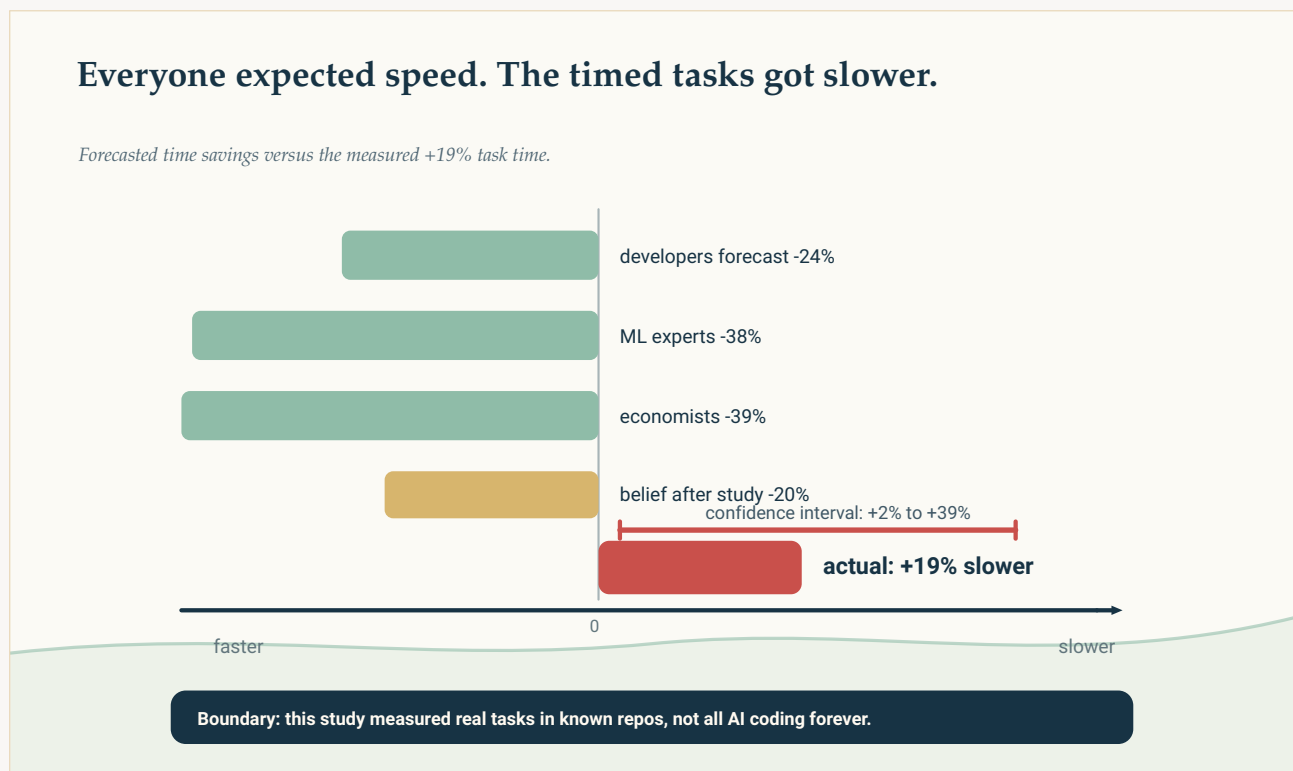
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

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

Ako pitate iskusnog programera ubrzava li ga AI pomoćnik za programiranje, obično ćete dobiti broj: uštedi mi dvadeset, trideset posto vremena. Ako pitate ekonomistu ili istraživača mašinskog učenja, broj postane još veći. Početkom 2025. godine tim METR-a uradio je ono sporo i skupo: proverio 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 dozvolili upotrebu AI alata ili je zabranili. Zatim su merili vreme rada.

Developeri su unapred procenili da će im AI skratiti vreme potrebno za zadatak za oko 24%. Dogodilo se suprotno: zadaci rađeni uz AI trajali su **19% duže**. A ovo je deo nad kojim vredi zastati — nakon završetka rada isti su developeri i dalje verovali da ih je AI ubrzao, za oko 20%. Bili su sporiji, osećali su se bržima, a razmak između ta dva broja najzanimljiviji je deo studije.

To je stvaran i pažljivo izmeren rezultat. Ali reč je o šesnaest developera, na repozitorijumima 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 smeru.



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

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.

Šta je studija merila — 16 stručnih developera otvorenog koda, 246 stvarnih zadataka, poznate repozitorijume, alate s početka 2025. i randomizaciju — i šta nije: ne većinu developera, ne druge područja, ne nepromenljiv zakon i ne recenzirani rad.

Original diagram — The Clean Paper CC BY 4.0

Šta je ovde mereno i zašto randomizovano ispitivanje toliko vredi

Randomizovano kontrolisano ispitivanje (RCT) alat je kojim medicina razlikuje stvaran efekat od onoga kojem se samo nadamo. Ovde je svaki od 246 zadataka nasumično dodeljen uslovu u kojem je AI bio dozvoljen ili zabranjen, tako da je u proseku jedina sistematska razlika između dveju grupa upravo AI. Zbog toga se može reći da je AI *uzrokovao* promenu vremena, umesto da se samo primeti 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 — samoprocene i rezultati na benčmarkovima — to ne mogu pokazati: benčmark nije stvaran posao, a samoprocena, kako pokazuje ova studija, može biti samouvereno pogrešna. Developeri ovde nisu bili početnici koji se muče s novom igračkom. Bili su etablirani saradnici na velikim, zrelim projektima otvorenog koda koje su dobro poznavali, uz nešto prethodnog iskustva s tim alatima.

Šta su autori uradili

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

kojima redovno doprinose i koje dobro poznaju — u proseku su na tim konkretnim projektima radili pet godina.

- Koristili su **246 stvarnih zadataka** — ispravke grešaka, nove funkcije i refaktoriranja preuzete iz sopstvenih sistema za praćenje problema tih projekata. Svaki je zadatak nasumično dodeljen uslovu „AI dozvoljen” ili „AI zabranjen”.
- „AI dozvoljen” značilo je **alate s početka 2025.: Cursor Pro s modelima Claude 3.5/3.7 Sonnet**. Glavna mera bilo je stvarno vreme potrebno za dovršetak svakog zadatka. Tim je uz to prikupio predviđanja developera pre rada i njihove procene posle, kao i predviđanja stručnjaka iz ekonomije i mašinskog učenja.

Šta su otkrili

- **Uz AI zadaci su trajali 19% duže.** Ne kraće — duže. Interval pouzdanosti od 95% proteže se otprilike od +2% do +39%, pa je smer efekta prilično čvrst čak i ako njegova tačna veličina nije.
- **Svi su predvideli suprotno.** Developeri su očekivali ubrzanje od 24%; stručnjaci za mašinsko učenje oko 38%; ekonomisti oko 39%. Sve tri grupe očekivale su veliku uštedu vremena; merenje je pokazalo da je AI nešto vremena koštao.
- **Jaz u percepciji.** Nakon što su posao završili sporije, developeri su i dalje procenili da ih je AI ubrzao za oko 20% — razlika od približno 40 procentnih bodova između osećaja i onoga što je pokazao sat.
- **Moguća objašnjenja, odagnuta ali ne dokazana.** Autori navode faktore koji bi mogli objasniti usporavanje: ovi developeri duboko poznaju sopstvene baze koda pa pomoćnik ima manje toga za dodati; zreli projekti imaju visoke i često implicitne standarde kvaliteta; repozitorijumi su veliki i puni konteksta koji model nema; a stvarno se vreme troši i na pisanje upita te pregledavanje i ispravljanje AI izlaza. Autori ih predstavljaju kao tragove, ne kao konačne odgovore.

Šta ovo ne pokazuje

- **Ne** pokazuje da AI ne ubrzava većinu developera. Autori to kažu direktno: šesnaest stručnjaka na kodu koji poznaju gotovo napamet nije prosečan developer na prosečnom zadatku.
- **Ne** pokazuje da je AI beskoristan niti da usporava ljude u drugim okolnostima — novopridošle u repozitorijum, 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 upotrebe mogu promeniti rezultat čak i u potpuno istom okruženju.
- Reč je o **preprintu** (objavljenom u julu 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 utešnu pretpostavku da su developeri možda dobili nešto drugo — više naučili, bili zadovoljniji ili pisali kvalitetniji kod. Jedino što je ovde izmereno, osećaj ubrzanja, upravo je ono čemu podaci protivreče.

Koliko su dokazi jaki

- **Dizajn je neuobičajeno pošten.** Randomizacija, stvarni zadaci, stvarni repozitorijumi i stvarno merenje vremena veliki su korak napred u odnosu na samoprocene i skale benčmarkova na kojima počiva većina tvrdnji o AI programiranju. Usporavanje od 19% ostalo je prisutno i u proverama robusnosti koje su proveli autori.
- **Najprenosiviji je nalaz jaz u percepciji.** Stručna intuicija o sopstvenom ubrzanju uz AI pogrešila je za otprilike 40 procentnih bodova, i to u optimističnom smeru. 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 sopstveni nastavak iz februara 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 ko je pristao učestvovati (developeri su sve češće odbijali raditi bez AI-ja, naknada je pala, a odabir zadataka se promenio). Pošteno čitanje je da se slika menja, a čak se i ta promena izveštava bez navijanja.

Zašto je to važno

Većina rasprave o AI-ju i programiranju počiva na demonstracijama i dojmovima: glatko snimljen video ekrana, samouverena tvrdnja, pa jednako samouvereno kolutanje očima s druge strane. Ono što je retko i zbog čega ovu studiju vredi pročitati je činjenica da je neko proveo dosadan eksperiment — nasumično dodeli uslove, meri stvaran rad, pa pitaj ljude kako je prošlo. Odgovor je neprijatan 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 smeru. Najtrajnija pouka ujedno je najmanja i najljudskija: ljudi koji su radili osećali su se brže dok su merljivo bili sporiji. „Oseća se brže” nije dokaz da jeste. Izmerite.

Kratak rezime

U randomizovanom kontrolisanom 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 dozvoljeni na nasumično odabranoj polovini zadataka. Developeri su očekivali da će AI skratiti vreme rada za oko 24%; umesto toga, vreme završetka poraslo je za 19% — a nakon toga su i dalje verovali da ih je AI ubrzao za oko 20%. Taj jaz između percepcije i merenja oštro je i robusno centar studije. Ali reč 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 sopstveni nastavak iz 2026. već naginje prema ubrzanju, uz sopstvena ograničenja. Pažljivo merenje koje vredi 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/>