



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

Izkušeni razvijalci so se z AI počutili hitrejši, merljivo pa so delali počasneje — pravi rezultat in sodba o programiranju z AI, ki je ta študija ne daje

V randomiziranem kontroliranem preskusu METR je šestnajst izkušenih odprtokodnih razvijalcev opravilo 246 resničnih nalog v kodnih bazah, ki so jih dobro poznali; pri naključni polovici nalog so smeli uporabljati orodja AI iz začetka leta 2025 (Cursor Pro ter Claude 3.5/3.7 Sonnet). Pričakovali so, da bo AI čas naloge skrajšala za približno 24 %; namesto tega se je čas dokončanja podaljšal za 19 % — po koncu pa so še vedno verjeli, da jih je AI pospešila za približno 20 %. Prav ta razkorak med občutkom in meritvijo je najostrejši in najbolj robusten rezultat študije. Toda gre za šestnajst razvijalcev, eno ozko okolje in posnetek orodij iz začetka leta 2025: avtorji izrecno poudarjajo, da rezultat ne kaže, da AI večini razvijalcev ne pomaga, njihov lastni nadaljnji poskus iz leta 2026 pa že nakazuje pospešek, z lastnimi omejitvami.

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

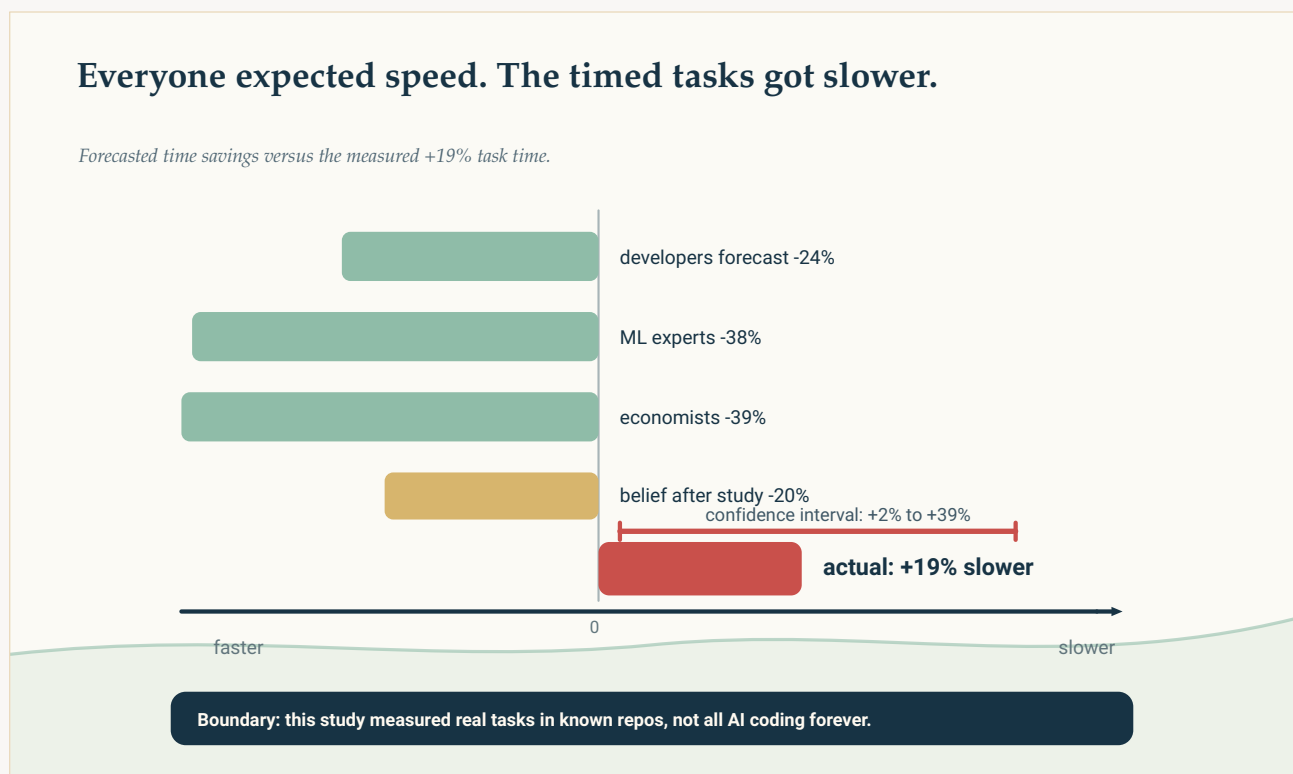
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Izkušeni razvijalci so se z AI počutili hitrejši, merljivo pa so delali počasneje — pravi rezultat in sodba o programiranju z AI, ki je ta študija ne daje

Če izkušenega programerja vprašate, ali ga pomočnik za programiranje z umetno inteligenco pospeši, boste pogosto dobili številko: prihrani mi dvajset, trideset odstotkov časa. Če vprašate ekonomista ali raziskovalca strojnega učenja, je številka navadno še višja. Na začetku leta 2025 je ekipa METR naredila počasno in drago stvar: preverila je. Šestnajstim izkušenim odprtokodnim razvijalcem je dala 246 resničnih nalog iz velikih kodnih baz, ki so jih dobro poznali, nato pa je pri vsaki nalogi z naključnim izborom določila, ali smejo uporabljati orodja AI ali ne. Nato so izmerili čas dela.

Razvijalci so vnaprej napovedali, da bo AI čas nalog skrajšala za približno 24 %. Zgodilo se je nasprotno: naloge z dovoljeno AI so trajale **19 % dlje**. In tu je del, pri katerem se je vredno ustaviti — po koncu so isti razvijalci še vedno verjeli, da jih je AI pospešila za približno 20 %. Bili so počasnejši, počutili pa so se hitrejši. Razdalja med tema številkami je najzanimivejši del študije.

To je resničen, skrbno izmerjen rezultat. Hkrati gre za šestnajst razvijalcev, ki zelo dobro poznajo svoje repozitorije in uporabljajo orodja iz začetka leta 2025 — zato rezultat ni preprost stavek »AI razvijalce upočasni«. Poznejši podatki iste ekipe že kažejo v drugo smer.



Vsi so napovedali, da bo AI delo pospešila — razvijalci za –24 %, strokovnjaki za strojno učenje za –38 %, ekonomisti za –39 %, razvijalci pa so po koncu še vedno menili, da so bili za –20 % hitrejši. Merjenje časa je pokazalo +19 % počasnejše delo, z intervalom od +2 % do +39 %. Razkorak med občutkom in meritvijo je glavni rezultat.

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.

Kaj študija meri — 16 izkušenih odprtokodnih razvijalcev, 246 resničnih nalog, znani repozitoriji, orodja iz začetka leta 2025 in naključna razporeditev — in česa ne: ne večine razvijalcev, ne drugih področij, ne nespremenljivega zakona in ne sodbe o prihodnjih orodjih. Študija tudi še ni recenzirana.

Original diagram — The Clean Paper CC BY 4.0

Kaj je bilo izmerjeno in kaj omogoča randomiziran preskus

Randomiziran kontroliran preskus (RCT) je orodje, s katerim medicina ločuje resničen učinek od učinka, za katerega le upamo, da obstaja. V tej študiji je bila vsaka od 246 nalog naključno dodeljena v skupino, kjer je bila AI dovoljena, ali v skupino, kjer ni bila. Zato je v povprečju edina sistematična razlika med skupinama prav uporaba AI. To omogoča sklep, da je AI *povzročila* spremembo časa, namesto da bi zgolj opazili, da ljudje, ki posežejo po AI, zaradi drugih razlogov delajo hitreje ali počasneje. To je pomembno, ker običajni dokazi o koristih AI pri programiranju — samoocene in rezultati na benchmarkih — tega ne morejo pokazati: benchmark ni resnično delo, samoocena pa je lahko, kot kaže ta študija, zelo samozavestno napačna. Razvijalci v študiji niso bili začetniki, ki se prvič spoprijemajo z novim orodjem. Bili so uveljavljeni sodelavci velikih in zrelih odprtokodnih projektov, ki so jih dobro poznali, ter so imeli nekaj predhodnih izkušenj z uporabo teh orodij.

Kaj so naredili avtorji

- Izvedli so **randomiziran kontroliran preskus** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Sodelovalo je šestnajst izkušenih odprtokodnih razvijalcev, vsak na velikem repozitoriju, h kateremu redno prispeva in ga dobro pozna — v povprečju so na teh projektih delali pet let.
- Uporabili so **246 resničnih nalog** — odpravljanje napak, nove funkcije in preurejanje kode iz dejanskih sistemov za sledenje težavam teh projektov. Vsaka naloga je bila naključno dodeljena v pogoj »AI dovoljena« ali »AI ni dovoljena«.
- »AI dovoljena« je pomenilo **orodja iz začetka leta 2025: Cursor Pro s Claude 3.5/3.7 Sonnet**. Glavna meritev je bil dejanski čas dokončanja posamezne naloge. Ekipa je poleg tega zbrala napovedi razvijalcev pred delom in njihove ocene po njem ter napovedi strokovnjakov za ekonomijo in strojno učenje.

Kaj so ugotovili

- **Z AI so naloge trajale 19 % dlje.** Ne hitreje — počasneje. 95-odstotni interval zaupanja sega približno od +2 % do +39 %, zato je smer učinka precej zanesljiva, čeprav njegova natančna velikost ni.
- **Vsi so napovedali nasprotno.** Razvijalci so pričakovali 24-odstotni pospešek, strokovnjaki za strojno učenje približno 38-odstotnega, ekonomisti pa približno 39-odstotnega. Vse tri skupine so pričakovale velik prihranek časa; ura je pokazala, da ga je bilo nekaj več porabljenega.
- **Razkorak v zaznavi.** Čeprav so delo opravili počasneje, so razvijalci po koncu še vedno ocenili, da jih je AI pospešila za približno 20 % — približno 40 odstotnih točk razlike med občutkom in meritvijo.
- **Možni razlogi, ocenjeni, ne dokazani.** Avtorji navajajo dejavnike, ki bi lahko pojasnili upočasnitev: ti razvijalci svoje kodne baze poznajo zelo dobro, zato ima pomočnik manj prostora za dodano vrednost; zreli projekti imajo visoke in pogosto implicitne standarde kakovosti; repozitoriji so veliki in vsebujejo veliko konteksta, ki ga model nima; resničen čas pa porabijo tudi za pripravo navodil, pregledovanje in popravljanje izhodov AI. To so možne razlage, ne dokončni sklepi.

Česa to ne pokaže

- Rezultat **ne** kaže, da AI ne pospeši večine razvijalcev. Avtorji to povedo neposredno: šestnajst strokovnjakov na kodi, ki jo poznajo skoraj na pamet, ni povprečen razvijalec na povprečni nalogi.
- **Ne** kaže, da je AI neuporabna ali da ljudi upočasnjuje v drugih okoljih — pri novincih v kodni bazi, pri projektih od začetka, v nepoznanih jezikih ali na drugih področjih.
- **Ne** zamrzne orodij v času. Gre za Cursor in Claude 3.5/3.7 Sonnet iz začetka leta 2025; avtorji izrecno poudarjajo, da bi boljša orodja ali boljši način njihove uporabe lahko spremenili rezultat tudi v povsem istem okolju.
- Gre za **predobjavo** (objavljeno julija 2025, še ne recenzirano), avtorji pa priznavajo, da ne morejo

povsem izključiti eksperimentalnih artefaktov — čeprav je rezultat zdržal njihove preizkuse robustnosti.

- Rezultat **ne** dovoljuje tolažilne razlage, da so razvijalci morda nekaj pridobili na drug način — se več naučili, bili bolj zadovoljni ali napisali boljšo kodo. Edina taka stvar, ki je bila tukaj izmerjena, je občutek pospeška, in prav tega podatki zavrnejo.

Kako močni so dokazi

- **Zasnova je nenavadno neposredna.** Naključna razporeditev, resnične naloge, resnični repozitoriji in dejansko merjenje časa so velik korak naprej v primerjavi s samoocenami in lestvicami benchmarkov, na katerih temelji večina trditev o AI pri programiranju. 19-odstotna upočasnitev je preživela preverjanja robustnosti, ki so jih izvedli avtorji.
- **Najbolj prenosljiv rezultat je razkorak v zaznavi.** Intuicija strokovnjakov o lastnem pospešku z AI je bila napačna za približno 40 odstotnih točk in napaka je bila optimistična. To je opozorilo za *vsako* samoocenjeno povečanje produktivnosti zaradi AI — tudi za številke iz te študije.
- **To je posnetek trenutka, ne trend.** Lastna nadaljnja študija METR iz februarja 2026, z razvijalci podobnega tipa in novjšimi orodji, že kaže proti pospešku — zelo grobo približno -18% pri razvijalcih, ki so sodelovali že prej, in -4% pri novih — vendar avtorji te dokaze označujejo kot šibke. Rezultate izkrivlja izbor udeležencev: razvijalci so vse pogosteje zavračali delo brez AI, plačilo se je znižalo, izbor nalog pa se je spremenil. Pošteno branje je, da se slika spreminja in da je tudi to spreminjanje poročano z veliko previdnosti.

Zakaj je pomembno

Velik del razprave o AI in programiranju temelji na demonstracijah in občutkih: ugleden posnetek zaslona, samozavestna trditev, skeptičen odziv. Redko pa kdo izvede dolgočasen poskus — naloge naključno razdeli, izmeri resnično delo in nato ljudi vpraša, kako se jim je zdelo. Zaradi tega je ta študija vredna pozornosti. Odgovor je neprijeten za oba tabora. Prebije zgodbo, da AI enakomerno turbo-pospeši izkušene razvijalce pri zahtevni kodi, ki jo dobro poznajo. Hkrati pa prebije tudi urejeno nasprotno zgodbo — »dokazano je, da AI razvijalce upočasni« — saj novejši podatki iste ekipe že kažejo v drugo smer. Najbolj trajen nauk je hkrati najmanjši in najbolj človeški: ljudje, ki so opravljali delo, so se počutili hitrejši, čeprav so bili merljivo počasnejši. »Zdi se mi hitreje« ni dokaz, da je res hitreje. Izmerite.

Kratek povzetek

V randomiziranem kontroliranem preskusu METR je šestnajst izkušenih odprtokodnih razvijalcev opravilo 246 resničnih nalog na kodnih bazah, ki so jih dobro poznali; pri naključni polovici nalog so smeli uporabljati orodja AI iz začetka leta 2025 (Cursor Pro ter Claude 3.5/3.7 Sonnet). Razvijalci

so pričakovali, da bo AI čas naloge skrajšala za približno 24 %; namesto tega se je čas dokončanja podaljšal za 19 % — po koncu pa so še vedno verjeli, da jih je pospešila za približno 20 %. Prav ta razkorak med občutkom in meritvijo je najostrejši in najbolj robusten rezultat. Toda gre za šestnajst razvijalcev, eno ozko okolje in posnetek orodij iz začetka leta 2025; avtorji jasno poudarjajo, da študija ne kaže, da AI večini razvijalcev ne pomaga, njihovi lastni podatki iz leta 2026 pa že kažejo proti pospešku, z lastnimi omejitvami. Skrbna meritev, ki jo velja vzeti resno — ne pa sodba o programiranju z AI.

Viri

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/>