



The CLEAN PAPER

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

Kogenud arendajad tundsid end AI-ga kiiremana, kuigi mõõdetavalt töötasid aeglasemalt — see on tegelik tulemus, mitte otsus AI-kodeerimise kohta

METR-i juhuslikustatud kontrollitud katses tegid 16 kogenud avatud lähtekoodiga tarkvara arendajat 246 päris ülesannet koodibaasides, mida nad hästi tundsid; juhusliku poole ülesannete puhul tohtis kasutada 2025. aasta alguse AI-tööriistu (Cursor Pro koos Claude 3.5/3.7 Sonnetiga). Nad ootasid, et AI vähendab ülesandele kuluvat aega umbes 24%; selle asemel suurenes lõpetamisaeg 19% — ja pärast katset uskusid nad endiselt, et AI oli neid umbes 20% kiirendanud. See tajuerinevus on uuringu selge ja robustne tuum. Kuid tegemist oli 16 arendaja, ühe kitsa olukorra ja 2025. aasta alguse hetkepildiga: autorid rõhutavad, et see ei näita, et AI ei aitaks enamikku arendajaid, ning nende enda 2026. aasta jätku-uuring kaldub juba kiirenemise poole, ehkki oma piirangutega.

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

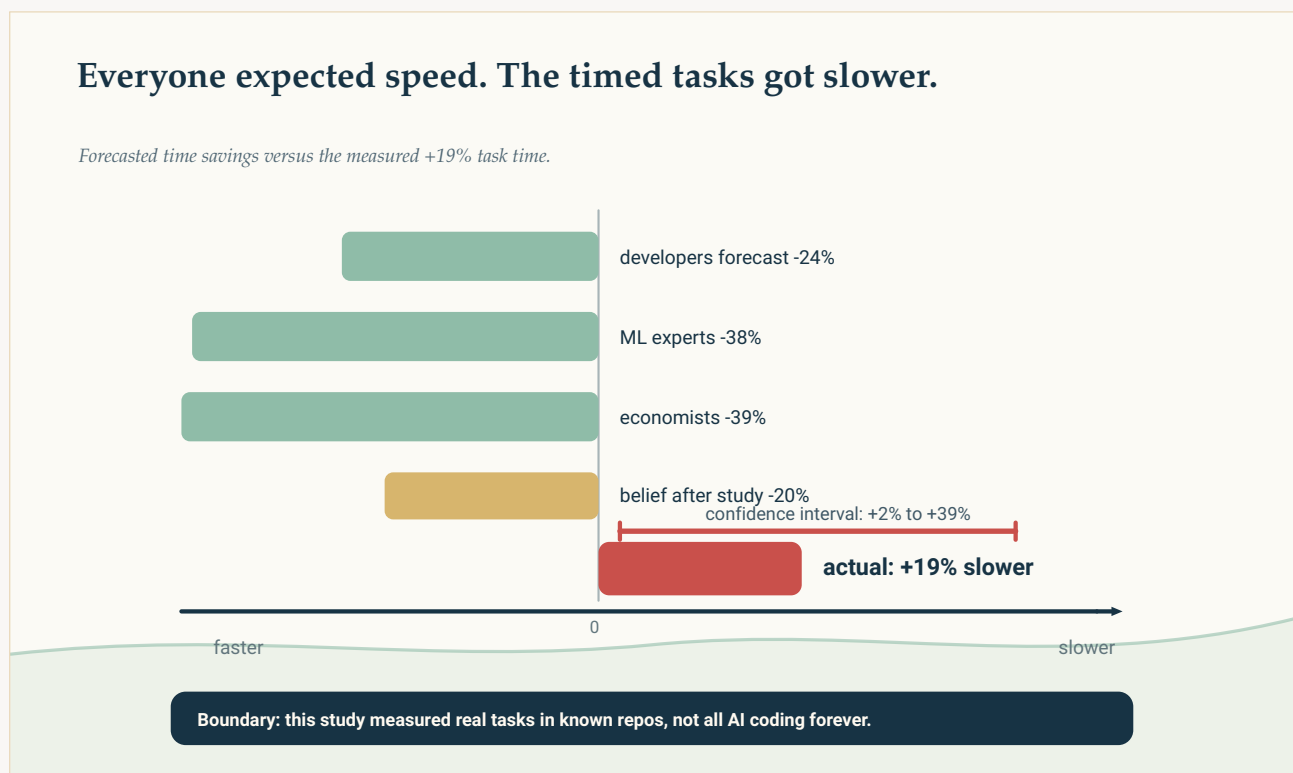
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Kogenud arendajad tundsid end tehisintellektiga kiiremana, kuigi mõõdetavalt töötasid aeglasemalt — see on tegelik tulemus, mitte otsus AI-kodeerimise kohta

Küsige kogenud programmeerijalt, kas AI-kodeerimisabiline teeb tema töö kiiremaks, ja tavaliselt saate vastuseks arvu: säästab kakskümmend, kolmkümmend protsenti. Küsige majandusteadlaselt või masinõppeuurijalt ja arv kipub kasvama. 2025. aasta alguses tegi METR-i meeskond selle aeglase ja kalli asja ära ning kontrollis. Nad võtsid kuusteist kogenud avatud lähtekoodiga tarkvara arendajat, andsid neile 246 päris ülesannet suurtest koodibaasidest, mida nad hästi tundsid, ning lubasid iga ülesande puhul juhusliku loosiga kas AI-tööriistu kasutada või mitte. Siis mõõdeti tööle kulunud aeg.

Arendajad ennustasid, et AI vähendab ülesandele kuluvat aega umbes 24%. Tulemus oli vastupidine: AI-ga tehtud ülesanded võtsid **19% rohkem aega**. Ja siin on osa, mille juures tasub peatuda — pärast katse lõppu uskusid needsamad arendajad ikka veel, et AI oli neid umbes 20% kiirendanud. Nad olid aeglasemad, kuid tundsid end kiiremana, ning just nende kahe arvu vahe on uuringu kõige huvitavam tulemus.

See on reaalne ja hoolikalt mõõdetud leid. Samas puudutab see kuuttest arendajat, kes töötasid neile väga tuttavates repositooriumides 2025. aasta alguse tööriistadega — ning see ei võrdu lihtsa väitega „AI teeb arendajad aeglasemaks“. Sama meeskonna hilisemad andmed viitavad juba vastupidisele suunale.



Kõik ennustasid, et AI muudab töö kiiremaks — arendajad –24%, masinõppeeksperdid –38%, majandusteadlased –39% ning arendajate enda tagantjärele hinnang –20%. Stopper näitas aga +19% aeglasemat tööd, vahemikuga +2% kuni +39%. Uuringu põhitulemus ongi vahe tajutu ja mõõdetu vahel.

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.

Mida uuring mõõtis — 16 kogenud avatud lähtekoodiga tarkvara arendajat, 246 päris ülesannet, tuttavad repositooriumid, 2025. aasta alguse tööriistad, juhuslik jaotus — ja mida see ei mõõtnud: mitte enamikku arendajaid, mitte teisi valdkondi, mitte püsivat seaduspära ega eelretsenseeritud lõplikku otsust.

Original diagram — The Clean Paper CC BY 4.0

Mida siin mõõdeti ja mida juhuslikustatud katse annab

Juhuslikustatud kontrollitud katse (RCT) on meetod, mida meditsiin kasutab tegeliku mõju eristamiseks soovitud või näilisest mõjust. Siin määrati iga 246 ülesandest juhuslikult kas „AI lubatud“ või „AI keelatud“ tingimusse, nii et keskmiselt oli kahe rühma ainus süstemaatiline erinevus AI enda kasutamine. Just see võimaldab öelda, et AI *põhjustas* ajakulu muutuse, mitte lihtsalt märgata, et inimesed, kes AI poole pöörduvad, on muudel põhjustel kiiremad või aeglasemad. See on oluline, sest tavapärane tõendus AI-kodeerimise tootlikkuse kohta — enesehinnangud ja võrdlustestide tulemused — ei suuda seda teha: võrdlustest ei ole päris töö ning enesehinnang võib, nagu see uuring näitab, olla enesekindlalt vale. Siinsed arendajad ei olnud algajad, kes uue mänuasjaga kohmetult eksperimenteerisid. Nad olid suurte ja küpsete avatud lähtekoodiga projektide pikaajalised panustajad, tundsid oma projekte hästi ning neil oli ka varasem kogemus nende tööriistadega.

Mida autorid tegid

- Viisid läbi **juhuslikustatud kontrollitud katse** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Osales kuusteist kogenud avatud lähtekoodiga tarkvara arendajat, kellest igaüks töötas suures repositooriumis, kuhu ta regulaarselt panustas ja mida hästi tundis — keskmiselt oli neil konkreetse projektiga viis aastat kogemust.
- Kasutasid **246 päris ülesannet** — vigade parandusi, uusi funktsioone ja refaktoreerimist, mis pärinesid nende projektide enda issue-tracker'itest. Iga ülesanne määrati juhuslikult tingimusse „AI lubatud“ või „AI keelatud“.
- „AI lubatud“ tähendas **2025. aasta alguse tööriistu: Cursor Pro koos Claude 3.5/3.7 Sonnetiga**. Peamine mõõdik oli tegelik ülesande lõpetamiseks kulunud aeg. Lisaks kogus meeskond arendajate prognoose enne katset ja hinnanguid pärast katset ning majandusteadlaste ja masinõppeekspertide ennustusi.

Mida nad leidsid

- **AI-ga võtsid ülesanded 19% rohkem aega.** Mitte vähem, vaid rohkem. 95% usaldusvahemik ulatub ligikaudu +2%-st +39%-ni, seega on mõju suund üsna kindel isegi siis, kui täpne suurus ei ole.
- **Kõik olid ennustanud vastupidist.** Arendajad prognoosisid 24% kiirenemist, masinõppeekspertid umbes 38% ja majandusteadlased umbes 39%. Kõik kolm rühma ootasid suurt ajavõitu; stopper näitas hoopis ajakulu kasvu.
- **Tajuerinevus.** Pärast töö tegemist ja aeglasemat tulemust hindasid arendajad endiselt, et AI oli neid umbes 20% kiirendanud — umbes 40 protsendipunkti erinevust selle vahel, mida nad tundsid, ja selle vahel, mida kell mõõtis.
- **Võimalikud põhjused, mida kaaluti, kuid ei tõestatud.** Autorid toovad välja tegureid, mis võiksid aeglustumist seletada: arendajad tunnevad oma koodibaase väga sügavalt, mistõttu assistendil on vähem lisaväärtust; küpsetel projektidel on kõrged ja sageli sõnastamata kvaliteedinõuded; repositooriumid on suured ja täis konteksti, mida mudel ei tea; ning päris aega kulub promptimisele, seejärel AI-väljundi ülevaatamisele ja parandamisele. Autorid esitavad neid võimalike seletustena, mitte lõplike järeldustena.

Mida see ei näita

- See **ei** näita, et AI ei suuda enamikku arendajaid kiirendada. Autorid ütlevad seda otsesõnu: kuusteist eksperti koodis, mida nad peast tunnevad, ei esinda keskmist arendajat ega keskmist ülesannet.
- See **ei** näita, et AI on kasutu või et see aeglustab inimesi muudes olukordades — näiteks uus-tulnukaid koodibaasis, täiesti uue projekti loomisel, võõras programmeerimiskeeles või teistes valdkondades.

- See **ei** külmuta tööriistu ajas. Jutt käib 2025. aasta alguse Cursorist ja Claude 3.5/3.7 Sonnetist; autorid rõhutavad, et paremad tööriistad või nende oskuslikum kasutamine võivad tulemust muuta isegi täpselt samas olukorras.
- Tegemist on **eelpublikatsiooniga** (avaldatud 2025. aasta juulis, eelretsenseerimata) ning autorid märgivad, et nad ei saa täielikult välistada katsega seotud artefakte — kuigi leid püsis nende erinevates analüüsides.
- See **ei** anna alust lohutavale tõlgendusele, et arendajad pidid siis vähemalt muul viisil võitma — õppima rohkem, tundma end paremini või kirjutama kvaliteetsemat koodi. Üks mõõdetud asi, tajutud kiirenemine, on just see, millele andmed vastu räägivad.

Kui tugev on tõendus

- **Katsekujundus on ebatavaliselt aus.** Juhuslikustatud jaotus, päris ülesanded, päris repositooriumid, tegelik ajamõõtmine — see on suur samm edasi võrreldes enesehinnangute ja võrdlustestide edetabelitega, millel suur osa AI-kodeerimise väidetest põhineb. 19% aeglustumine pidas vastu autorite robustsuskontrollidele.
- **Kõige üldistatavam tulemus on tajuerinevus.** Ekspertide sisetunne enda AI-ga saavutatud kiirenemise kohta eksis umbes 40 protsendipunkti võrra optimistlikus suunas. See on hoiatus *kõigi* enese raporteeritud AI-tootlikkuse võitude suhtes — kaasa arvatud selle uuringu enda arvud.
- **See on hetkepilt, mitte trend.** METR-i enda 2026. aasta veebruari jätku-uuring, kus sarnased arendajad kasutasid uuemaid tööriistu, viitab juba kiirenemisele — väga ligikaudu –18% naasvate ja –4% uute arendajate puhul — kuid autorid nimetavad tõendit nõrgaks ja märgivad osalejate valikust tulenevaid moonutusi (arendajad keeldusid üha sagedamini töötamast ilma AI-ta, tasu langes ning ülesannete valik nihkus). Aus tõlgendus on, et pilt muutub, ja isegi muutust ennast raporteeritakse ettevaatlikult.

Miks see oluline on

Suur osa AI ja programmeerimise vaidlusest käib demode ja kõhutunde ümber: efektne ekraanisalvestus, enesekindel väide, vastuseks skeptiline silmapööritus. Haruldane — ja seetõttu lugemist väärt — on see, et keegi tegi ära igava katse: juhuslikusta, mõõda päris töö aega ja küsi siis inimestelt, kuidas nende arvates läks. Vastus on ebamugav mõlemale leerile. See lõhub loo, et AI annab kogenud arendajatele raskes, tuttavas koodis ühtlaselt turbokiirenduse. Ja see lõhub ka kena vastandloo „tõestatud: AI teeb arendajad aeglasemaks“, sest sama meeskonna uuemad andmed kalduvad juba teisele poole. Kõige püsivam õppetund on ka kõige väiksem ja inimlikum: inimesed tundsid end tööd tehes kiiremana, kuigi olid mõõdetavalt aeglasemad. „Tundub kiirem“ ei ole tõend, et see on kiirem. Mõõtke.

Lühidalt

Juhuslikustatud kontrollitud katses lasi METR kuueteistkümmel kogenud avatud lähtekoodiga tarkvara arendajal lahendada 246 päris ülesannet koodibaasides, mida nad hästi tundsid; juhusliku poole ülesannete puhul tohtis kasutada 2025. aasta alguse AI-tööriistu (Cursor Pro koos Claude 3.5/3.7 Sonnetiga). Arendajad ootasid, et AI vähendab ülesandele kuluvat aega umbes 24%; tegelikult suurenes lõpetamisaeg 19% — ja pärast katset uskusid nad endiselt, et AI oli neid umbes 20% kiirendanud. See tajuerinevus on uuringu selge ja robustne tuum. Kuid tegemist oli 16 arendaja, ühe kitsa olukorra ja 2025. aasta alguse hetkepildiga; autorid rõhutavad, et uuring ei näita, et AI ei aitaks enamikku arendajaid, ning nende enda 2026. aasta jätkuandmed viitavad juba kiirenemisele, kuigi ka neil on omad piirangud. Hoolikas mõõtmine, mida tasub tõsiselt võtta — mitte kohtuotsus AI-kodeerimisele.

Allikad

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/>