



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

Kokeneet kehittäjät tunsivat olevansa tekoälyn kanssa nopeampia, vaikka työ hidastui mitattavasti — todellinen tulos ja mitä se ei kerro tekoälykoodauksesta

METR:n satunnaistetussa kontrolloidussa kokeessa kuusitoista kokenutta avoimen lähdekoodin kehittäjää teki 246 todellista tehtävää hyvin tuntemissaan koodikannoissa. Satunnaisesti puolessa tehtävistä sai käyttää vuoden 2025 alun tekoälytyökaluja (Cursor Pro sekä Claude 3.5/3.7 Sonnet). Kehittäjät odottivat tekoälyn lyhentävän tehtävääikää noin 24 %, mutta se pidensi sitä 19 % — ja kokeen jälkeenkin he uskoivat tekoälyn nopeuttaneen työtä noin 20 %. Tämä koetun ja mitatun tuottavuuden välinen kuilu on tutkimuksen terävin ja vankin tulos. Aineisto koostuu kuitenkin vain kuudestatoista kehittäjästä, yhdestä kapeasta asetelmasta ja vuoden 2025 alun työkalutilanteesta. Kirjoittajat korostavat, ettei tutkimus osoita tekoälyn olevan useimmille kehittäjille hyödytön, ja heidän oma vuoden 2026 jatkotutkimuksensa viittaa jo nopeutumiseen — omine varauksineen.

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

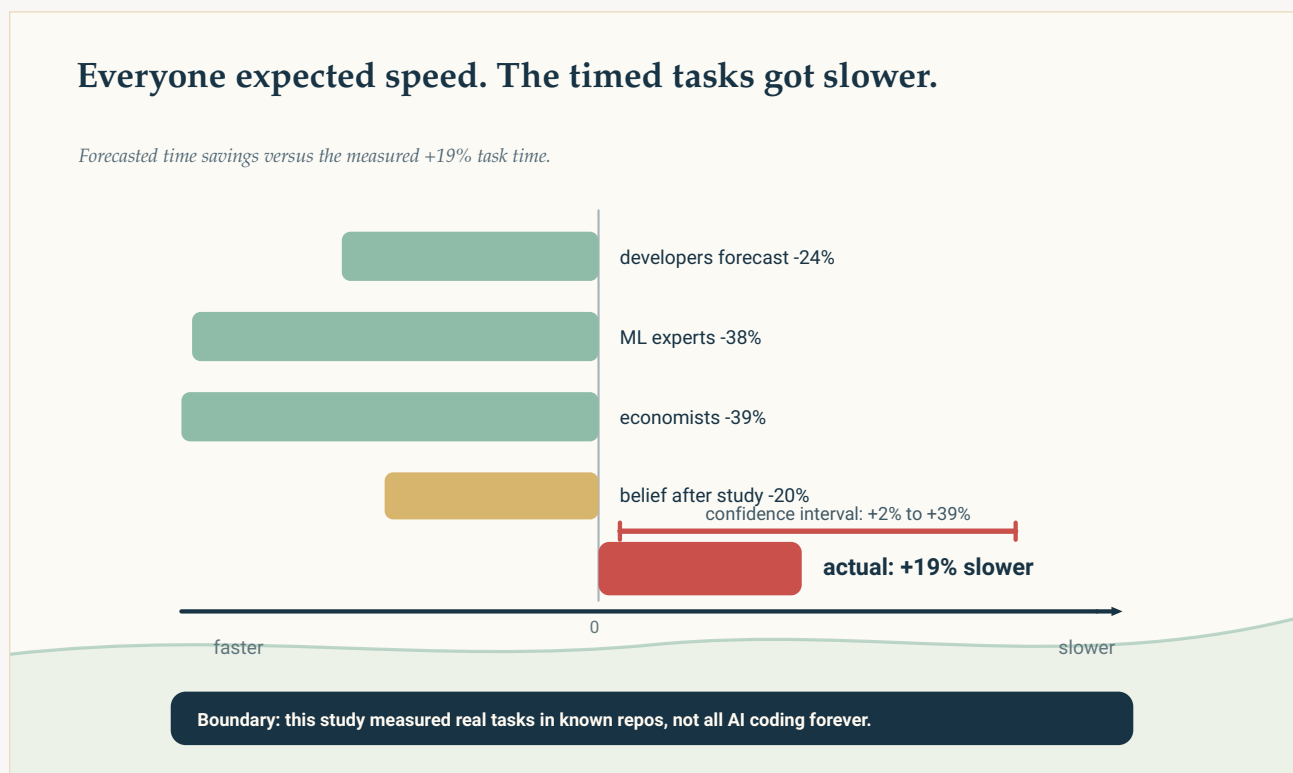
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Kokeneet kehittäjät tunsivat olevansa tekoälyn kanssa nopeampia, vaikka työ hidastui mitattavasti — todellinen tulos ja mitä se ei kerro tekoälykoodauksesta

Kysy kokeneelta ohjelmoijalta, nopeuttaako tekoälyavustaja hänen työtään, ja saat usein numeron: säästän sillä kaksikymmentä, kolmekymmentä prosenttia. Kysy ekonomistilta tai koneoppimistutkijalta, ja luku yleensä kasvaa. Vuoden 2025 alussa METR:n tutkijaryhmä teki hitaamman ja kalliimman asian: se mittasi. Kuusitoista kokenutta avoimen lähdekoodin kehittäjää sai 246 todellista tehtävää suurista, hyvin tuntemistaan koodikannoista. Tehtävä kerrallaan arpa ratkaisi, saiko tekoälytyökaluja käyttää vai ei. Sitten tutkijat mittasivat ajan.

Kehittäjät olivat ennustaneet tekoälyn lyhentävän tehtäväaikaa noin 24 %. Kävi päinvastoin: tekoälyn kanssa tehty tehtävät veivät **19 % enemmän aikaa**. Ja tässä on tuloksen osa, jota kannattaa hetki ajatella — tehtävien jälkeen samat kehittäjät uskoivat yhä tekoälyn nopeuttaneen heitä noin 20 %. He olivat hitaampia mutta tunsivat olevansa nopeampia, ja näiden kahden luvun välinen etäisyys on tutkimuksen kiinnostavin havainto.

Tulos on todellinen ja huolellisesti mitattu. Mutta kyse on myös kuudestatoista kehittäjästä, heidän läpikotaisin tuntemistaan repositorioista ja vuoden 2025 alun työkaluista — eikä tulos ole sama asia kuin suora väite ”tekoäly hidastaa kehittäjiä”. Saman ryhmän myöhemmät tiedot osoittavat jo toiseen suuntaan.



Kaikki ennustivat tekoälyn nopeuttavan työtä — kehittäjät –24 %, koneoppimisen asiantuntijat –38 %, ekonomistit –39 % ja kehittäjien oma jälkikäteisarvio –20 %. Sekuntikello löysi +19 % hitaamman työskentelyn, luottamusvälin ollessa +2...+39 %. Koetun ja mitatun välinen kuilu on tässä varsinainen tulos.

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.

Mitä tutkimus mittasi — 16 kokenutta avoimen lähdekoodin kehittäjää, 246 todellista tehtävää, tutut repositoriot, vuoden 2025 alun työkalut, satunnaistettu asetelma — ja mitä se ei mittaa: ei useimpia kehittäjiä, ei muita aloja, ei pysyvää lakia eikä vertaisarvioitua lopullista vastausta.

Original diagram — The Clean Paper CC BY 4.0

Mitä tässä mitattiin ja mitä satunnaistaminen antaa

Satunnaistettu kontrolloitu koe (RCT) on lääketieteessä käytetty työkalu, jolla erotetaan todellinen vaikutus toiveikkaasta tulkinnasta. Tässä jokainen 246 tehtävästä määrättiin satunnaisesti tehtäväksi joko tekoäly sallittuna tai ilman tekoälyä. Siksi kahden tehtäväjoukon ainoa järjestelmällinen ero on keskimäärin itse tekoälyn käyttö. Juuri tämä mahdollistaa väitteen, että tekoäly *aiheutti* ajan muutoksen, sen sijaan että vain huomattaisiin tekoölyyn tarttuvien ihmisten olevan muista syistä nopeampia tai hitaampia. Tämä on tärkeää, koska tavallisin näyttö tekoälykoodauksen hyödyistä — itsearviot ja vertailutestien tulokset — ei pysty tähän: vertailutesti ei ole oikeaa työtä, ja itsearvio voi tämän tutkimuksen mukaan olla itsevarmasti väärässä. Osallistujat eivät olleet aloittelijoita uuden lelun kanssa. He olivat vakiintuneita kontribuuttoreita suuriin, kypsiin avoimen lähdekoodin projekteihin, jotka he tunsivat hyvin, ja heillä oli myös aiempaa kokemusta työkaluista.

Mitä tutkijat tekivät

- Toteuttivat **satunnaistetun kontrolloidun kokeen** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Kuusitoista kokenutta avoimen lähdekoodin kehittäjää työskenteli kukin suuressa repositoriossa, johon he osallistuivat säännöllisesti ja jonka he tunsivat hyvin — keskimäärin viiden vuoden kokemus juuri kyseisistä projekteista.
- Käyttivät **246 todellista tehtävää** — bugikorjauksia, ominaisuuksia ja refaktorointeja projektien omista issue-seurannoista. Jokainen tehtävä satunnaistettiin ryhmään ”tekoäly sallittu” tai ”tekoäly kielletty”.
- ”Tekoäly sallittu” tarkoitti **vuoden 2025 alun työkaluja: Cursor Pro ja Claude 3.5/3.7 Sonnet**. Ensisijainen mittari oli tehtävän todellinen valmistumisaika. Lisäksi kerättiin kehittäjien ennusteet ennen työtä ja arviot sen jälkeen sekä ekonomistien ja koneoppimisiasiantuntijoiden ennusteita.

Mitä he löysivät

- **Tekoälyn kanssa tehtävät veivät 19 % kauemmin.** Ei nopeammin vaan hitaammin. 95 prosentin luottamusväli ulottuu suunnilleen +2 prosentista +39 prosenttiin, joten suunnasta on melko vahva näyttö, vaikka vaikutuksen tarkka koko on epävarmempi.
- **Kaikki olivat ennustaneet päinvastaista.** Kehittäjät odottivat 24 prosentin nopeutusta, koneoppimisiasiantuntijat noin 38 prosenttia ja ekonomistit noin 39 prosenttia. Kaikki kolme ryhmää odottivat tekoälyn säästävän paljon aikaa; kello osoitti sen maksavan aikaa.
- **Havaitsemisen kuilu.** Vaikka kehittäjät suoriutuivat hitaammin, he arvioivat työn jälkeen tekoälyn nopeuttaneen heitä noin 20 %. Koetun ja mitatun vaikutuksen väliin jäi siis noin 40 prosenttiyksikköä.
- **Mahdollisia syitä punnittiin, ei todistettu.** Kirjoittajat nostavat esiin tekijöitä, jotka voisivat selittää hidastumista: nämä kehittäjät tuntevat omat koodikantansa niin syvästi, että avustajalla on vähemmän lisättävää; kypsissä projekteissa on korkeat ja usein implisiittiset laatuvaatimukset; repositoriot ovat suuria ja täynnä kontekstia, jota mallilla ei ole; ja todellista aikaa kuluu promptien kirjoittamiseen sekä tekoälyn tuotoksen tarkistamiseen ja korjaamiseen. Nämä esitetään johtolankoina, eivät lopullisina selityksinä.

Mitä tämä ei osoita

- Se **ei** osoita, ettei tekoäly nopeuttaisi useimpia kehittäjiä. Kirjoittajat sanovat tämän suoraan: kuusitoista oman koodinsa läpikotaisin tuntevaa asiantuntijaa eivät ole keskimääräinen kehittäjä keskimääräisessä tehtävässä.
- Se **ei** osoita tekoälyn olevan hyödytön tai hidastavan muissa tilanteissa — esimerkiksi uudelle koodikannalle tulevia kehittäjiä, alusta rakennettavaa työtä, vieraita ohjelmointikieliä tai kokonaan muita aloja.
- Se **ei** pysäytä työkaluja aikaan. Kyse on vuoden 2025 alun Cursorista ja Claude 3.5/3.7 Sonnetista.

Kirjoittajat korostavat, että paremmat työkalut tai parempi tapa käyttää näitä samoja työkaluja voivat muuttaa tulosta jopa täsmälleen tässä asetelmassa.

- Kyseessä on **preprint** (julkaistu heinäkuussa 2025, ei vielä vertaisarvioitu), ja kirjoittajat huomauttavat, etteivät he pysty täysin sulkemaan pois kokeellisia artefakteja — vaikka tulos säilyi heidän eri analyyseissään.
- Se **ei** oikeuta lohdullista tulkintaa, että kehittäjät ehkä hyötyivät jollain muulla tavalla — oppivat enemmän, viihtyivät paremmin tai kirjoittivat parempaa koodia. Tässä mitattu asia, koettu nopeutuminen, on juuri se, jonka data kumoo.

Kuinka vahvaa näyttö on

- **Asetelma on poikkeuksellisen rehellinen.** Satunnaistaminen, todelliset tehtävät, oikeat repositorit ja todellinen ajanmittaus ovat suuri askel eteenpäin verrattuna itsearvioihin ja vertailutesteihin, joihin monet väitteet tekoälykoodauksen hyödyistä nojaavat. 19 prosentin hidastuminen säilyi kirjoittajien herkkyystarkistuksissa.
- **Yleistettävien havainto on koetun ja mitatun ero.** Asiantuntijan intuitio omasta tekoälyn tuomasta nopeutuksesta oli noin 40 prosenttiyksikköä väärässä optimistiseen suuntaan. Se on varoitus *kaikille* itse raportoiduille tekoälyn tuottavuushyödyille — myös tämän tutkimuksen omille ennakoarvioille.
- **Kyse on tilannekuvasta, ei trendistä.** METR:n oma helmikuun 2026 jatkotutkimus samantyyppisillä kehittäjillä ja uudemmilla työkaluilla viittaa nopeutumiseen — hyvin karkeasti –18 % palaavilla kehittäjillä ja –4 % uusilla — mutta kirjoittajat pitävät näyttöä heikkona, koska osallistujien valikoituminen väärästi asetelmaa (kehittäjät kieltäytyivät yhä useammin työskentelemästä ilman tekoälyä, palkkiotaso laski ja tehtävien valinta vinoutui). Rehellinen tulkinta on, että kuva muuttuu, ja muutoskin raportoidaan varovasti ilman peukaloa vaa’alla.

Miksi tällä on merkitystä

Suurin osa tekoälystä ja ohjelmoinnista käytävästä väittelystä nojaa demoihin ja tuntumaan: sulava ruutukaappaus, itsevarma väite ja vastaukseksi silmien pyörittely. Harvinaista — ja tämän tutkimuksen arvoista — on se, että joku teki tylsän kokeen: satunnaisti, mittasi oikean työn ajan ja kysyi sitten osallistujilta, miten työ sujui. Vastaus on epämukava kummallekin leirille. Se puhkaisee tarinan, jonka mukaan tekoäly kiihdyttää järjestelmällisesti asiantuntijakehittäjiä vaikeassa, tutussa koodissa. Samalla se puhkaisee yksinkertaisen vastatarinan ”tekoäly tekee kehittäjistä hitaampia, todistettu”, koska saman ryhmän uudempi aineisto kallistuu jo toiseen suuntaan. Kestävin opetus on pienin ja inhimillisin: työn tekijät tunsivat olevansa nopeampia samalla kun he olivat mitattavasti hitaampia. ”Tuntuu nopeammalta” ei ole näyttöä siitä, että se on. Mittaa se.

Tiivistelmä

METR:n satunnaistetussa kontrolloidussa kokeessa kuusitoista kokenutta avoimen lähdekoodin kehittäjää teki 246 todellista tehtävää hyvin tuntemissaan koodikannoissa. Satunnaisesti puolessa tehtävistä sai käyttää vuoden 2025 alun tekoälytyökaluja (Cursor Pro sekä Claude 3.5/3.7 Sonnet). Kehittäjät odottivat tekoälyn lyhentävän tehtäväaikaa noin 24 %, mutta se pidensi valmistumisaikaa 19 % — ja työn jälkeenkin he uskoivat tekoälyn nopeuttaneen heitä noin 20 %. Tämä koetun ja mitatun välinen kuilu on tutkimuksen terävä ja vankka ydin. Aineisto on kuitenkin vain kuusitoista kehittäjää, yksi kapea asetelma ja vuoden 2025 alun tilannekuva. Kirjoittajat korostavat, ettei tutkimus osoita tekoälyn olevan useimmille kehittäjille hyödytön, ja heidän oma vuoden 2026 jatkotutkimuksensa viittaa jo nopeutumiseen, omine varauksineen. Huolellinen mittaus, joka kannattaa ottaa vakavasti — ei lopullinen tuomio tekoälykoodauksesta.

Lähteet

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/>