



The CLEAN PAPER

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

Dezvoltatorii experimentați s-au simțit mai rapizi cu IA în timp ce lucrau măsurabil mai lent — rezultatul real, și verdictul despre coding cu IA care nu este

Un trial randomizat controlat de METR a pus șaisprezece dezvoltatori open-source experimentați să facă 246 de task-uri reale pe codebase-uri pe care le cunoșteau bine, cu instrumente IA de la începutul lui 2025 (Cursor Pro plus Claude 3.5/3.7 Sonnet) permise pe o jumătate aleatorie. Ei se așteptau ca IA să reducă timpul pe task cu aproximativ 24%; în schimb, l-a crescut cu 19% — iar după aceea încă au crezut că îi accelerase cu aproximativ 20%. Acel decalaj de percepție este nucleul ascuțit și robust al studiului. Dar sunt șaisprezece dezvoltatori, un setting îngust și o fotografie fixă de la începutul lui 2025: autorii sunt expliciti că nu arată că IA nu îi ajută pe majoritatea dezvoltatorilor, iar follow-up-ul lor din 2026 înclină deja spre o accelerare, cu propriile caveat-uri.

Written by Lucio Vaglio · Cross-checked by Vera Rubin · 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

Dezvoltatorii experimentați s-au simțit mai rapizi cu IA în timp ce lucrau măsurabil mai lent — rezultatul real, și verdictul despre coding cu IA care nu este

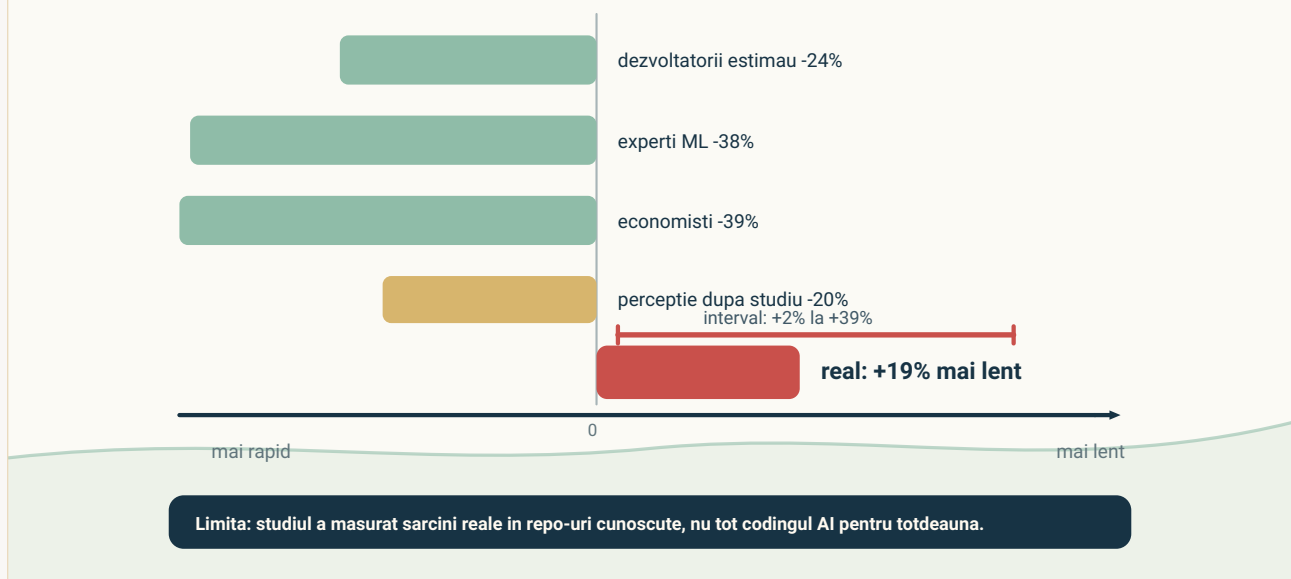
Întreabă un programator experimentat dacă un asistent de coding cu IA îl face mai rapid și de obicei vei primi un număr: îmi economisește douăzeci, treizeci la sută. Întreabă un economist, sau un cercetător în machine learning, și numărul crește. La începutul lui 2025, o echipă de la METR a făcut lucrul lent și scump: a verificat. A luat șaisprezece dezvoltatori open-source cu experiență, le-a dat 246 de task-uri reale din codebase-uri mari pe care le cunoșteau bine și — prin aruncarea unei monede, task cu task — fie le-a permis să folosească instrumente IA, fie nu. Apoi a cronometrat munca.

Dezvoltatorii estimaseră că IA le va reduce timpul de lucru cu aproximativ 24%. A făcut opusul: task-urile făcute cu IA au durat **cu 19% mai mult**. Iar aici este partea la care merită să stai — după ce au terminat, aceiași dezvoltatori încă credeau că IA îi accelerase, cu aproximativ 20%. Erau mai lenți, și se simțeau mai rapizi, iar distanța dintre acele două numere este cel mai interesant lucru din studiu.

Este un rezultat real, măsurat cu grijă. Este și despre șaisprezece dezvoltatori, pe repository-uri pe care le cunosc intim, folosind instrumente de la începutul lui 2025 — și nu este propoziția plată “IA îi face pe dezvoltatori mai lenți”. Datele ulterioare ale aceleiași echipe arată deja în cealaltă direcție.

Toti așteptau viteza. Sarcinile cronometrate au devenit mai lente.

Economii de timp prognozate versus +19% măsurat în timpul sarcinilor.



Toată lumea a prezis că IA va accelera munca — dezvoltatorii —24%, experții ML —38%, economiștii —39%, iar convingerea de după lucru a dezvoltatorilor înșiși —20%. Cronometrul a găsit +19% mai lent, cu un interval de la +2% la +39%. Decalajul dintre perceput și măsurat este rezultatul.

Original diagram — The Clean Paper CC BY 4.0

Ce a masurat acest studiu despre productivitatea AI

ACESTA ESTE

- 16 dezvoltatori open-source experimentati
- 246 sarcini reale
- repo-uri pe care le cunosteau
- randomizat si cronometrat
- unelte AI inceput de 2025

ACESTA NU ESTE

- nu majoritatea dezvoltatorilor
- nu orice domeniu
- nu o lege fixa
- nu peer-reviewed
- nu un verdict despre unelte viitoare

Limita: dovada utila despre un context, nu o lege universala a muncii cu AI.

Ce măsoară studiul — 16 dezvoltatori open-source experți, 246 de task-uri reale, repository-uri familiare, instrumente de la începutul lui 2025, randomizare — și ce nu măsoară: nu majoritatea dezvoltatorilor, nu alte domenii, nu o lege fixă, nu peer-reviewed.

Original diagram — The Clean Paper CC BY 4.0

Ce a măsurat asta și ce îți cumpără un trial randomizat

Un **trial randomizat controlat (RCT)** este instrumentul pe care medicina îl folosește pentru a deosebi un efect real de unul sperat. Aici, fiecare dintre cele 246 de task-uri a fost alocat aleatoriu să fie făcut cu IA permisă sau nu, astfel încât, în medie, singura diferență sistematică dintre cele două grămezi să fie IA însăși. Asta îți permite să spui că IA a *cauzat* schimbarea de timp, în loc doar să observi că oamenii care ajung să folosească IA sunt mai rapizi sau mai lenți din alte motive. Contează pentru că dovezile obișnuite despre câștigurile din coding cu IA — self-report-uri și scoruri de benchmark — nu pot face asta: un benchmark nu este muncă reală, iar un self-report, cum arată acest studiu, poate fi greșit cu încredere. Dezvoltatorii de aici nu erau începători care băjbâiau cu o jucărie nouă. Erau contributory stabili la proiecte open-source mari, mature, pe care le cunoșteau bine, cu o anumită experiență anterioară în folosirea instrumentelor.

Ce au făcut autorii

- Au rulat un **trial randomizat controlat** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Şaisprezece dezvoltatori open-source experimentaţi, fiecare lucrând pe un repository mare la care contribuie regulat şi pe care îl cunoaşte bine — în medie de cinci ani pe proiectele respective.
- Au folosit **246 de task-uri reale** — bug fix-uri, feature-uri şi refactorizări luate din issue tracker-ele propriilor proiecte. Fiecare task a fost alocat aleatoriu la “IA permisă” sau “IA interzisă”.
- “IA permisă” însemna **tooling de la începutul lui 2025: Cursor Pro cu Claude 3.5/3.7 Sonnet**. Măsura principală a fost timpul real de finalizare pe task. În paralel, echipa a colectat predicţii (de la dezvoltatori înainte) şi estimări (de la ei după), plus predicţii de la experţi în economie şi machine learning.

Ce au găsit

- **Cu IA, task-urile au durat cu 19% mai mult.** Nu mai repede — mai lent. Intervalul de încredere de 95% merge de la aproximativ +2% la +39%, deci direcţia este solidă chiar dacă mărimea exactă nu este.
- **Toată lumea prezisese opusul.** Dezvoltatorii estimaseră o accelerare de 24%; experţii în machine learning, aproximativ 38%; economişti, aproximativ 39%. Toate cele trei grupuri se aşteptau ca IA să economisească mult timp; cronometrul a găsit că a costat timp.
- **Decalajul de percepţie.** După ce au făcut munca şi au ieşit mai lenţi, dezvoltatorii încă estimau că IA îi accelerase cu aproximativ 20% — un decalaj de aproximativ 40 de puncte între ce simţeau şi ce înregistra ceasul.
- **Motive candidate, cântărite mai degrabă decât dovedite.** Autorii aliniază factori care ar putea explica încetinirea: aceşti dezvoltatori îşi cunosc codebase-urile foarte bine, deci un asistent are mai puţin de adăugat; proiectele mature poartă standarde de calitate înalte, adesea implicite; repository-urile sunt mari şi pline de context pe care un model nu îl are; şi timp real intră în prompting, apoi în revizuirea şi corectarea outputului IA. Le prezintă ca piste, nu ca verdicte.

Ce nu arată

- **Nu** arată că IA nu reuşeşte să accelereze majoritatea dezvoltatorilor. Autorii o spun direct: şaisprezece experţi pe cod pe care îl ştiu pe dinafară nu sunt dezvoltatorul mediu pe task-ul mediu.
- **Nu** arată că IA este inutilă, sau că îi încetineşte pe oameni în alte contexte — nou-veniţi într-un codebase, lucru greenfield, limbaje nefamiliare sau domenii cu totul diferite.
- **Nu** îngheaţă instrumentele în loc. Acesta este Cursor şi Claude 3.5/3.7 Sonnet de la începutul lui 2025; autorii sunt expliciti că instrumente mai bune, sau moduri mai bune de a le folosi pe acestea, ar putea schimba rezultatul chiar în exact acest setting.
- Este un **preprint** (postat în iulie 2025, încă neevaluat inter pares), iar autorii notează că nu pot exclude complet artefacte experimentale — deşi rezultatul a rezistat în analizele lor.

- **Nu** autorizează lectura liniștitoare că dezvoltatorii trebuie să fi câștigat altceva — au învățat mai mult, s-au simțit mai bine, au scris cod mai bun. Singurul lucru măsurat aici, accelerarea simțită, este exact ceea ce datele contrazic.

Cât de puternică este evidența

- **Designul este neobișnuit de onest.** Randomizat, task-uri reale, repository-uri reale, cronometrare efectivă — un pas mare peste self-report-urile și leaderboard-urile de benchmark pe care se sprijină cele mai multe afirmații despre coding cu IA. Încetinirea de 19% a supraviețuit verificărilor de robustețe ale autorilor.
- **Rezultatul cel mai portabil este decalajul de percepție.** Intuiția experților despre propria accelerare cu IA a fost greșită cu aproximativ 40 de puncte, în direcția optimistă. Este un avertisment despre *orice* câștig de productivitate cu IA raportat de utilizatori — inclusiv numerele acestui studiu.
- **Este o fotografie, nu o tendință.** Follow-up-ul METR din februarie 2026, cu același tip de dezvoltatori pe instrumente mai noi, indică o accelerare — foarte aproximativ — 18% pentru dezvoltatorii care au revenit și -4% pentru cei noi — dar autorii îl marchează ca evidență slabă, distorsionată de cine a fost dispus să participe (dezvoltatorii refuzau tot mai mult să lucreze fără IA, rata de plată a scăzut, iar selecția task-urilor s-a înclinat). Lectura onestă este că tabloul se mișcă, iar chiar și mișcarea este raportată cu degetul ținut departe de cântar.

De ce contează

Cea mai mare parte a argumentului despre IA și programare se poartă pe demo-uri și vibe-uri: o înregistrare de ecran elegantă, o afirmație sigură, o ridicare de ochi în direcția opusă. Ce este rar, și ce face studiul acesta demn de citit, este că cineva a făcut experimentul plictisitor — randomizează, cronometrează muncă reală, apoi întreabă oamenii cum a mers. Răspunsul este inconfortabil pentru ambele tabere. Sparge povestea că IA turbo-acelerează uniform dezvoltatorii experți pe cod greu și familiar. Și sparge și opusul ordonat — “IA îi face pe dezvoltatori mai lenți, dovedit” — pentru că datele mai noi ale aceleiași echipe înclină deja în cealaltă direcție. Lecția cea mai durabilă este și cea mai mică și mai umană: oamenii care făceau munca se simțeau mai rapizi în timp ce erau măsurabil mai lenți. “Se simte mai rapid” nu este dovadă că este. Măsoară.

Sinteză curată

Într-un trial randomizat controlat, METR a pus șaisprezece dezvoltatori open-source experimentați să completeze 246 de task-uri reale pe codebase-uri pe care le cunoșteau bine, cu instrumente IA de la începutul lui 2025 (Cursor Pro plus Claude 3.5/3.7 Sonnet) permise pe o jumătate aleatorie. Dezvoltatorii se așteptau ca IA să le reducă timpul pe task cu aproximativ 24%; în schimb, l-a crescut

cu 19% — iar după aceea încă au crezut că îi accelerase cu aproximativ 20%. Acel decalaj de percepție este nucleul ascuțit și robust al studiului. Dar sunt șaisprezece dezvoltatori, un setting îngust și o fotografie fixă de la începutul lui 2025; autorii sunt expliți că nu arată că IA nu îi ajută pe majoritatea dezvoltatorilor, iar follow-up-ul lor din 2026 indică deja o accelerare, cu propriile caveat-uri. O măsurare atentă care merită luată în serios — nu un verdict despre coding cu IA.

Sources

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/>