



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

Искусни програмери се чувствуваа побрзи со AI додека мерливо работеа побавно — вистинскиот наод и пресудата што ова не е

Во рандомизирано контролирано испитување на METR, шеснаесет искусни развивачи на софтвер со отворен код извршувале 246 вистински задачи на кодни бази што добро ги познавале, а на случајна половина им биле дозволени AI-алатки од почетокот на 2025 година (Cursor Pro и Claude 3.5/3.7 Sonnet). Очекувале AI да го скрати времето за околу 24%; наместо тоа, времето за завршување пораснало за 19%, а потоа тие и натаму верувале дека се забрзале за околу 20%. Токму тој јаз меѓу чувството и мерењето е најсилниот дел од студијата. Но станува збор за шеснаесет програмери, тесна средина и фиксиран момент; авторите јасно велат дека ова не покажува дека AI не им помага на повеќето програмери, а нивното следење од 2026 година веќе навестува забрзување, со свои ограничувања.

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

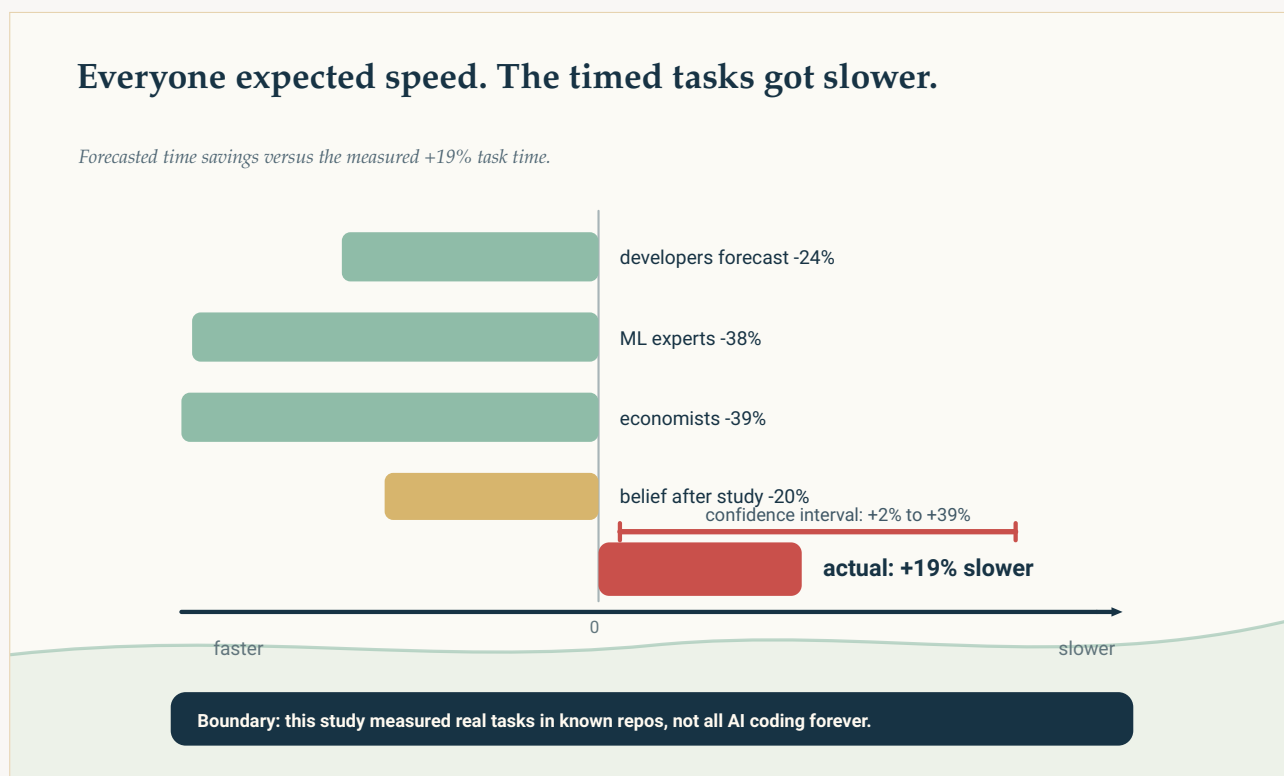
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Искусни програмери се чувствуваа побрзи со AI, а мерливо работеа побавно — тоа е вистинскиот наод, не пресуда за програмирањето со AI

Прашајте искусен програмер дали AI-помошник за кодирање му ја забрзува работата и најчесто ќе добиете бројка: ми штеди дваесет, триесет проценти. Прашајте економист или истражувач по машинско учење и бројката обично станува уште поголема. На почетокот на 2025 година тим од METR го направил бавниот и скап тест: проверил. Земале шеснаесет искусни развивачи на софтвер со отворен код, им дале 246 вистински задачи од големи кодни бази што добро ги познавале и — со случаен избор, задача по задача — или им дозволувале да користат AI-алатки или не. Потоа го мереле времето.

Програмерите предвиделе дека AI ќе им го намали времето за задачите за околу 24%. Се случило спротивното: задачите извршени со AI траеле **19% подолго**. А најинтересниот дел е што, откако завршиле, истите програмери и натаму верувале дека AI ги забрзал за околу 20%. Работеле побавно, чувствуваа дека работат побрзо, а растојанието меѓу тие две бројки е најинтересниот резултат во студијата.

Ова е вистински и внимателно измерен резултат. Но се однесува на шеснаесет програмери, на репозиториуми што ги познаваат во детали и на алатки од почетокот на 2025 година — и не значи едноставно „AI ги прави програмерите побавни“. Подоцнежните податоци од истиот тим веќе укажуваат во спротивна насока.



Сите предвиделе дека AI ќе ја забрза работата — програмерите –24%, експертите по машинско учење

—38%, економистите —39%, а и проценката на самите програмери по завршената работа била —20%. Мерењето покажало +19% побавно, со интервал од +2% до +39%. Разликата меѓу почувствуваното и измереното е клучниот наод.

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.

Што мери студијата — 16 искусни развивачи на софтвер со отворен код, 246 вистински задачи, познати репозиториуми, алатки од почетокот на 2025 година и случајна распределба — и што не мери: не ги претставува повеќето програмери, ниту други области, не утврдува непроменлив закон и не е рецензирана студија.

Original diagram — The Clean Paper CC BY 4.0

Што точно е измерено и што ви дава рандомизиран експеримент

Рандомизирано контролирано испитување (RCT) е алатката што медицината ја користи за да разликува вистински ефект од посакуван ефект. Тука секоја од 246-те задачи била случајно распределена во група во која AI бил дозволен или во група во која не бил, така што во просек единствената систематска разлика меѓу двете групи била токму употребата на AI. Тоа овозможува да се каже дека AI ја *предизвикал* промената во времето, наместо само да се забележи дека луѓето што посегнуваат по AI по некоја друга причина се побрзи или побавни. Ова е важно затоа што вообичаените докази за зголемена продуктивност со AI — самоизвештаи и резултати на бенчмаркови — не можат да го покажат тоа: бенчмарк не е вистинска работа, а самоизвештај, како што покажува оваа студија, може со голема самоувереност да биде погрешен. Учесниците не биле почетници

што се снаоѓаат со нова играчка. Биле долгогодишни соработници на големи, зрели проекти со отворен код што добро ги познавале и веќе имале одредено искуство со алатките.

Што направиле авторите

- Спровеле **рандомизирано контролирано испитување** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Шеснаесет искусни развивачи на софтвер со отворен код работеле секој на голем репозиториум на кој редовно придонесуваат и кој добро го познаваат — во просек околу пет години работа на конкретните проекти.
- Користеле **246 вистински задачи** — поправки на грешки, нови функции и рефакторирања земени од системите за пријавување задачи на самите проекти. Секоја задача по случаен избор била означена како „AI дозволен“ или „AI недозволен“.
- „AI дозволен“ значело **алатки од почетокот на 2025 година: Cursor Pro со Claude 3.5/3.7 Sonnet**. Главната мерка била реалното време потребно за завршување на задачата. Тимот собирал и прогнози од програмерите пред задачите, нивни проценки потоа, како и предвидувања од експерти по економија и машинско учење.

Што откриле

- **Со AI задачите траеле 19% подолго.** Не побрзо — побавно. Интервалот на доверба од 95% се протега приближно од +2% до +39%, па насоката на ефектот е прилично јасна иако точната големина е понеизвесна.
- **Сите го предвиделе спротивното.** Програмерите очекувале забрзување од 24%; експертите по машинско учење околу 38%; економистите околу 39%. Сите три групи очекувале AI да заштеди многу време, а мерењето покажало дека одземал време.
- **Јаз меѓу перцепцијата и мерката.** Иако работеле побавно, по завршувањето програмерите и натаму проценувале дека AI ги забрзал за околу 20% — разлика од приближно 40 процентни поени меѓу тоа како се чувствувале и што покажал часовникот.
- **Можни причини, разгледани, но не и докажани.** Авторите наведуваат неколку фактори што би можеле да го објаснат забавувањето: овие програмери длабоко ги познаваат своите кодни бази, па помошникот има помалку простор да додаде вредност; зрелите проекти имаат високи и често непишани стандарди за квалитет; репозиториумите се големи и содржат многу контекст што моделот го нема; и реално време се троши на формулирање барања, а потоа на проверка и исправка на AI-излезот. Авторите ги нудат како можни насоки, не како конечни објаснувања.

Што не покажува ова

- **Не** покажува дека AI не ги забрзува повеќето програмери. Авторите тоа го кажуваат директно: шеснаесет експерти што работат на код што го знаат речиси напамет не се просечен програмер на просечна задача.
- **Не** покажува дека AI е бескорисен или дека ги забавува луѓето во други ситуации — на пример новодојденци во кодна база, нови проекти, непознати програмски јазици или други области.
- **Не** ги замрзнува алатките во времето. Ова се Cursor и Claude 3.5/3.7 Sonnet од почетокот на 2025 година; авторите експлицитно велат дека подобри алатки, или подобар начин на употреба на истите, може да го променат резултатот дури и во токму оваа средина.
- Станува збор за **препринт** (објавен во јули 2025 година, сè уште нерецензиран), а авторите наведуваат дека не можат целосно да исклучат експериментални артефакти — иако резултатот останал стабилен низ нивните анализи.
- **Не** дозволува утешно толкување дека програмерите сигурно добиле нешто друго — научиле повеќе, биле позадоволни или напишале подобар код. Единствената таква работа што е измерена тука, почувствуваното забрзување, е токму она што податоците го побиваат.

Колку се убедливи доказите?

- **Дизајнот е невообичаено директен.** Случајна распределба, вистински задачи, вистински репозиториуми и реално мерење на времето — голем чекор напред во однос на самоизвештаите и табелите со бенчмаркови врз кои се потпираат многу тврдења за AI-кодирање. Забавувањето од 19% ги поминало проверките на робусност на авторите.
- **Најпреносливиот наод е јазот во перцепцијата.** Интуицијата на експертите за сопственото забрзување со AI била погрешна за околу 40 процентни поени, и тоа во оптимистичка насока. Тоа е предупредување за *секое* самоизвестено зголемување на продуктивноста со AI — вклучувајќи ги и бројките од оваа студија.
- **Ова е снимка во еден момент, не тренд.** Следењето на METR од февруари 2026 година, со сличен тип програмери и понови алатки, укажува на забрзување — многу грубо околу –18% за програмерите што повторно учествувале и –4% за новите — но авторите го означуваат тоа како слаб доказ, искривен од тоа кој бил подготвен да учествува (програмерите сè почесто одбивале да работат без AI, надоместокот се намалил, а изборот на задачи се променил). Најчесното читање е дека сликата се менува, а и самата промена е прикажана внимателно, без да се преувеличува.

Зошто е важно

Голем дел од расправата за AI и програмирањето се потпира на демонстрации и впечаток: впечатливо видео, самоуверено тврдење, потсмешлив контрааргумент. Ретко е некој да го направи здодевниот експеримент — да рандомизира, да мери вистинска работа и потоа да ги праша луѓето како поминало. Одговорот е непријатен за двата табора. Ја пробива приказната дека AI без исклучок ги турбозабрзува експертите на тешка, добро позната кодна база. Но ја пробива и уредната спротивна приказна — „докажано е дека AI ги прави програмерите побавни“ — затоа што поновите податоци од истиот тим веќе се наклонети во друга насока. Најтрајната лекција е и најмалата и најчовечката: луѓето што ја извршувале работата се чувствувале побрзи додека мерливо биле побавни. „Ми се чини побрзо“ не е доказ дека е побрзо. Измерете го.

Кратко резиме

Во рандомизирано контролирано испитување, METR им дал на шеснаесет искусни развивачи на софтвер со отворен код 246 вистински задачи на кодни бази што добро ги познавале; на случајно избрана половина од задачите биле дозволени AI-алатки од почетокот на 2025 година (Cursor Pro и Claude 3.5/3.7 Sonnet). Програмерите очекувале AI да им го намали времето за околу 24%; наместо тоа, времето за завршување пораснало за 19% — а потоа тие и натаму верувале дека AI ги забрзал за околу 20%. Тој јаз меѓу перцепцијата и мерката е најостриот и најробустен дел од студијата. Но се работи за шеснаесет програмери, тесна средина и фиксиран момент од почетокот на 2025 година; авторите експлицитно велат дека тоа не покажува дека AI не им помага на повеќето програмери, а нивното следење од 2026 година веќе навестува забрзување, со свои ограничувања. Внимателно мерење што заслужува сериозно внимание — не пресуда за програмирањето со AI.

Извори

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/>