



The CLEAN PAPER

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

Skúsení vývojári sa s AI cítili rýchlejší, hoci merateľne spomalili — skutočný výsledok, nie verdikt nad programovaním s AI

V randomizovanej kontrolovanej štúdií nechal METR šesťnásť skúsených open-source vývojárov vykonať 246 skutočných úloh v kódových základniach, ktoré dobre poznali, pričom pri náhodnej polovici boli povolené nástroje AI zo začiatku roka 2025 (Cursor Pro a Claude 3.5/3.7 Sonnet). Čakali, že AI skráti čas asi o 24 %; namiesto toho ho predĺžila o 19 % — a potom sa stále domnievali, že ich zrýchlila približne o 20 %. Táto medzera vo vnímaní je ostrým a robustným jadrom štúdie. Ide však o šesťnásť vývojárov, jedno úzke prostredie a pevnú snímku začiatku roka 2025: autori výslovne uvádzajú, že výsledok neukazuje zlyhanie AI pri väčšine vývojárov, a ich vlastné nadväzujúce údaje z roku 2026 už smerujú k zrýchleniu, s vlastnými výhradami.

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

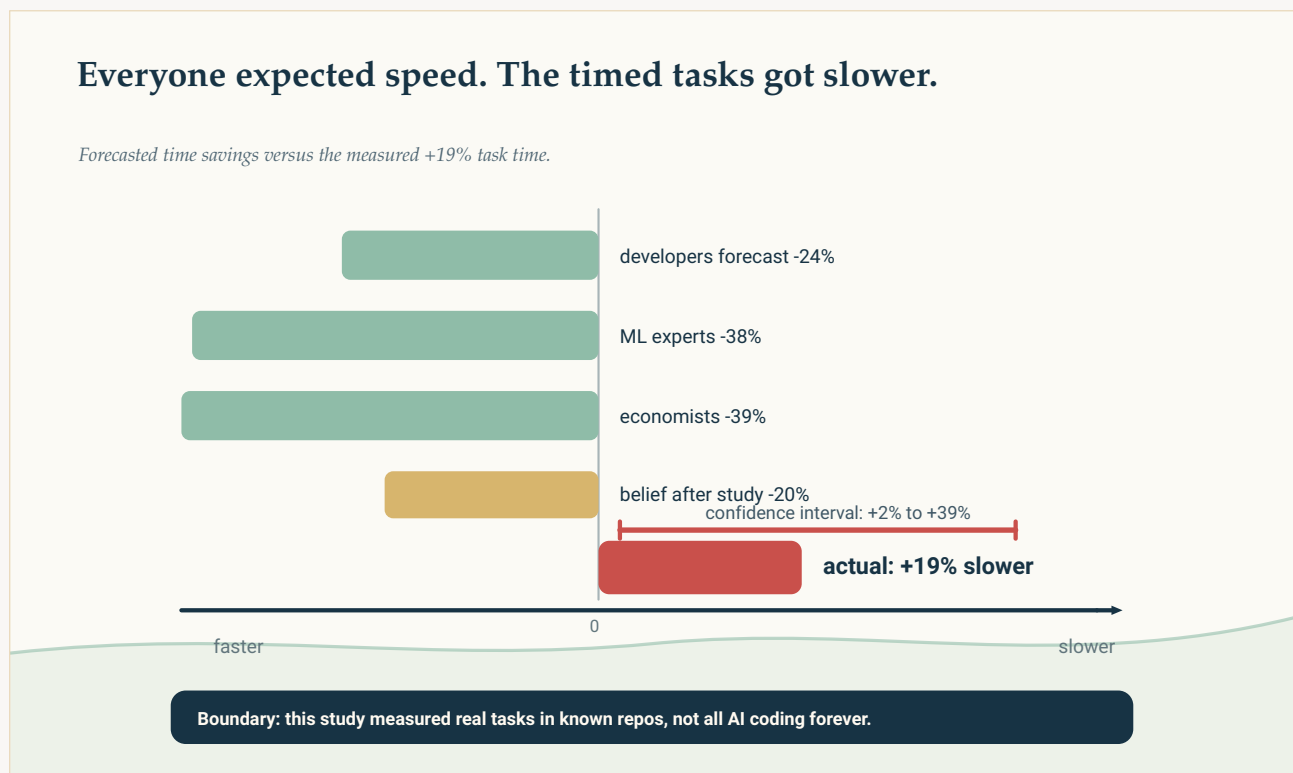
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Skúsení vývojári sa s AI cítili rýchlejší, hoci merateľne spomalili — skutočný výsledok, nie verdikt nad programovaním s AI

Spýtajte sa skúseného programátora, či ho asistent AI zrýchľuje, a zvyčajne dostanete číslo: šetrí mi dvadsať, tridsať percent času. Spýtajte sa ekonóma alebo výskumníka strojového učenia a číslo ešte vzrastie. Začiatkom roka 2025 urobil tím METR pomalú a nákladnú vec: overil to. Šestnástim skúseným vývojárom open-source softvéru zadal 246 skutočných úloh z rozsiahlych kódových základní, ktoré dobre poznali, a pri každej úlohe hodom mince rozhodol, či môžu použiť nástroje AI. Potom zmeral čas.

Vývojári predpovedali, že AI skráti čas práce približne o 24 %. Stal sa opak: úlohy s AI trvali o **19 % dlhšie**. A práve nad ďalšou časťou sa oplatí zastaviť — aj po dokončení sa tí istí vývojári domnievali, že ich AI zrýchľila asi o 20 %. Boli pomalší a cítili sa rýchlejší; rozdiel medzi týmito dvoma číslami je najzaujímavejším výsledkom štúdie.

Je to skutočný, starostlivo zameraný výsledok. Týka sa však šestnástich vývojárov pracujúcich v repozitároch, ktoré dôverne poznajú, s nástrojmi zo začiatku roka 2025 — a nemožno ho zhrnúť paušálnou vetou „AI vývojárov spomaľuje“. Neskôršie údaje rovnakého tímu už ukazujú na opak.



Všetci predpovedali, že AI prácu zrýchli — vývojári o 24 %, odborníci na strojové učenie o 38 %, ekonómovia o 39 % a vlastný spätný odhad vývojárov o 20 %. Stopky zistili spomalenie o 19 %, s intervalom od 2 % do 39 %. Výsledkom je rozdiel medzi pocitom a meraním.

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.

Čo štúdia meria — 16 skúsených open-source vývojárov, 246 skutočných úloh, známe repozitáre, nástroje zo začiatku roka 2025 a náhodné rozdelenie — a čo nemeria: väčšinu vývojárov, iné odbory, nemenný zákon ani recenzovaný výsledok.

Original diagram — The Clean Paper CC BY 4.0

Čo sa meralo a čo prináša randomizovaný pokus

Randomizovaná kontrolovaná štúdia (RCT) je nástroj, ktorým medicína odlišuje skutočný účinok od vytúženého. Tu bola každá z 246 úloh náhodne pridelená na vykonanie s povolenou AI alebo bez nej, takže jediným priemerným systematickým rozdielom medzi oboma skupinami je samotná AI. Vďaka tomu možno povedať, že AI zmenu času *spôsobila*, nie iba pozorovať, že ľudia siahajúci po AI bývajú z iných dôvodov rýchlejší či pomalší. Je to dôležité, pretože bežné dôkazy prínosu AI pri programovaní — vlastné hodnotenia a výsledky benchmarkov — toto nedokážu: benchmark nie je skutočná práca a vlastné hodnotenie môže byť, ako ukazuje táto štúdia, sebavedomo chybné. Účastníci neboli začiatočníci zápasiaci s novou hračkou. Boli zavedenými prispievateľmi do veľkých, vyspelých open-source projektov, ktoré dobre poznali, a s nástrojmi už mali určité skúsenosti.

Čo autori urobili

- Uskutočnili **randomizovanú kontrolovanú štúdiu** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Šestnásť skúsených open-source vývojárov pracovalo vo veľkých repozitároch, do ktorých pravidelne prispievali a ktoré dobre poznali — na konkrétnych projektoch pôsobili v

priemere päť rokov.

- Použili **246 skutočných úloh** — opravy chýb, nové funkcie a refaktorizácie z vlastných systémov hlásenia týchto projektov. Každá úloha bola náhodne pridelená k podmienke „AI povolená“ alebo „AI zakázaná“.
- „AI povolená“ znamenalo **nástroje zo začiatku roka 2025: Cursor Pro s Claude 3.5/3.7 Sonnet**. Hlavnou veličinou bol skutočný čas dokončenia úlohy. Tím popri ňom zhromaždil predbežné predpovede vývojárov, ich následné odhady a predikcie odborníkov z ekonómie a strojového učenia.

Čo zistili

- **S AI trvali úlohy o 19 % dlhšie.** Nie kratšie, ale dlhšie. 95-percentný interval spoľahlivosti siaha približne od +2 % do +39 %, takže smer výsledku je pevný, aj keď presná veľkosť menej istá.
- **Všetci predpovedali opak.** Vývojári očakávali zrýchlenie o 24 %, odborníci na strojové učenie asi o 38 % a ekonómovia približne o 39 %. Všetky tri skupiny čakali veľkú úsporu času; stopky namerali čas navyše.
- **Medzera vo vnímaní.** Hoci po práci s AI vyšli pomalší, vývojári stále odhadovali, že ich zrýchlila asi o 20 % — medzi pocitom a meraním vznikol rozdiel približne 40 percentuálnych bodov.
- **Možné dôvody, zvažované, nie preukázané.** Autori uvádzajú faktory, ktoré by mohli spomalenie vysvetliť: vývojári svoje kódové základne hlboko poznajú, takže asistent má menej čo pridať; vyspelé projekty majú vysoké, často nevyslovené štandardy kvality; repozitáre sú rozsiahle a plné kontextu, ktorý model nemá; skutočný čas zaberá zadávanie pokynov a následná kontrola a oprava výstupu AI. Predkladajú ich ako vodidlá, nie závery.

Čo to neukazuje

- Štúdia **neukazuje**, že AI nedokáže zrýchliť väčšinu vývojárov. Autori to hovoria priamo: šesťnásť odborníkov na kóde, ktorý poznajú naspamäť, nie je priemerný vývojár pri priemernej úlohe.
- **Neukazuje**, že je AI zbytočná alebo že spomaľuje v iných podmienkach — pri nováčikoch v kódovej základni, projektoch na zelenej lúke, neznámych jazykoch či v celkom iných odboroch.
- **Nezmrazuje** nástroje v čase. Ide o Cursor a Claude 3.5/3.7 Sonnet zo začiatku roka 2025; autori výslovne uvádzajú, že lepšie nástroje alebo lepšie spôsoby ich použitia môžu výsledok zmeniť aj v úplne rovnakom prostredí.
- Je to **preprint** (zverejnený v júli 2025, zatiaľ bez odborného recenzného konania) a autori pripúšťajú, že nemôžu úplne vylúčiť experimentálne artefakty — výsledok však pretrval naprieč analýzami.
- Štúdia **neumožňuje** upokojujúci výklad, že vývojári museli získať niečo iné — viac sa naučiť, cítiť sa lepšie alebo písať kvalitnejší kód. Jediná tu meraná vec, pocit zrýchlenia, je presne tým, čomu údaje odporujú.

Aké silné sú dôkazy

- **Návrh štúdie je nezvyčajne poctivý.** Randomizácia, skutočné úlohy, skutočné repozitáre a skutočné meranie času predstavujú veľký krok vpred oproti vlastným hodnoteniam a rebríčkom benchmarkov, na ktorých stojí väčšina tvrdení o programovaní s AI. Spomalenie o 19 % obstálo v kontrolách robustnosti autorov.
- **Najprenosnejšia je medzera vo vnímaní.** Odborná intuícia o vlastnom zrýchlení s AI sa mýlila približne o 40 percentuálnych bodov, optimistickým smerom. Je to varovanie pred *každým* deklarovaným nárastom produktivity vďaka AI — vrátane čísel z tejto štúdie.
- **Je to momentka, nie trend.** Vlastné nadväzujúce údaje METR z februára 2026, s rovnakým typom vývojárov a novšími nástrojmi, ukazujú na zrýchlenie — veľmi hrubo o 18 % pri vracajúcich sa vývojároch a o 4 % pri nových — autori ich však označujú za slabý dôkaz, skreslený výberom ochotných účastníkov (vývojári čoraz častejšie odmietali pracovať bez AI, odmena klesla a výber úloh sa posunul). Poctivý záver je, že obraz sa mení a aj táto zmena sa uvádza bez prikrášľovania výsledku.

Prečo na tom záleží

Väčšina sporu o AI a programovanie stojí na ukážkach a pocitoch: uhladenej nahrávke obrazovky, sebavedomom tvrdení alebo odmietavom prevrátení očí. Vzácné a cenné je, že niekto vykonal nudný experiment — náhodne rozdelil podmienky, zmeral skutočnú prácu a potom sa ľudí spýtal, ako sa im darilo. Odpoveď je nepríjemná pre oba tábory. Narúša príbeh, že AI vždy prudko zrýchľuje odborníkov na ťažkom, známom kóde. A narúša aj úhladný opak — „AI vývojárov spomaľuje, dokázané“ — pretože novšie údaje rovnakého tímu už mieria druhým smerom. Najtrvalejšie ponaučenie je zároveň najmenšie a najľudskejšie: ľudia sa pri práci cítili rýchlejší, hoci boli merateľne pomalší. „Pôsobí to rýchlejšie“ nie je dôkazom, že to rýchlejšie je. Merajte.

Stručné zhrnutie

V randomizovanej kontrolovanej štúdii nechal METR šesťnásť skúsených open-source vývojárov dokončiť 246 skutočných úloh v kódových základniach, ktoré dobre poznali, pričom pri náhodnej polovici boli povolené nástroje AI zo začiatku roka 2025 (Cursor Pro a Claude 3.5/3.7 Sonnet). Vývojári čakali, že AI skráti čas asi o 24 %; namiesto toho ho predĺžila o 19 % — a potom sa stále domnievali, že ich zrýchľila približne o 20 %. Táto medzera vo vnímaní je ostrým a robustným jadrom štúdie. Ide však o šesťnásť vývojárov, jedno úzke prostredie a pevnú snímku začiatku roka 2025; autori výslovne uvádzajú, že výsledok neukazuje zlyhanie AI pri väčšine vývojárov, a ich vlastné nadväzujúce údaje z roku 2026 už smerujú k zrýchleniu, s vlastnými výhradami. Starostlivé meranie, ktoré stojí za vážne zváženie — nie verdikt nad programovaním s AI.

Zdroje

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/>