



The CLEAN PAPER

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

Zkušení vývojáři se s AI cítili rychlejší, přestože měřitelně zpomalili — skutečný výsledek, nikoli verdikt nad programováním s AI

V randomizované kontrolované studii nechal METR šestnáct zkušených open-source vývojářů provést 246 skutečných úloh v kódových základnách, které dobře znali, přičemž u náhodné poloviny byly povoleny nástroje AI z počátku roku 2025 (Cursor Pro a Claude 3.5/3.7 Sonnet). Čekali, že AI zkrátí čas asi o 24 %; místo toho jej prodloužila o 19 % — a poté se stále domnívali, že je zrychlila přibližně o 20 %. Tato mezera ve vnímání je výrazným a robustním jádrem studie. Jde však o šestnáct vývojářů, jedno úzké prostředí a pevný snímek počátku roku 2025: autoři výslovně uvádějí, že výsledek neukazuje selhání AI u většiny vývojářů, a jejich vlastní navazující data z roku 2026 již směřují ke zrychlení, s vlastními výhradami.

Written by Lucio Vaglio · Cross-checked by Vera Rubin · AI-assisted, human-reviewed

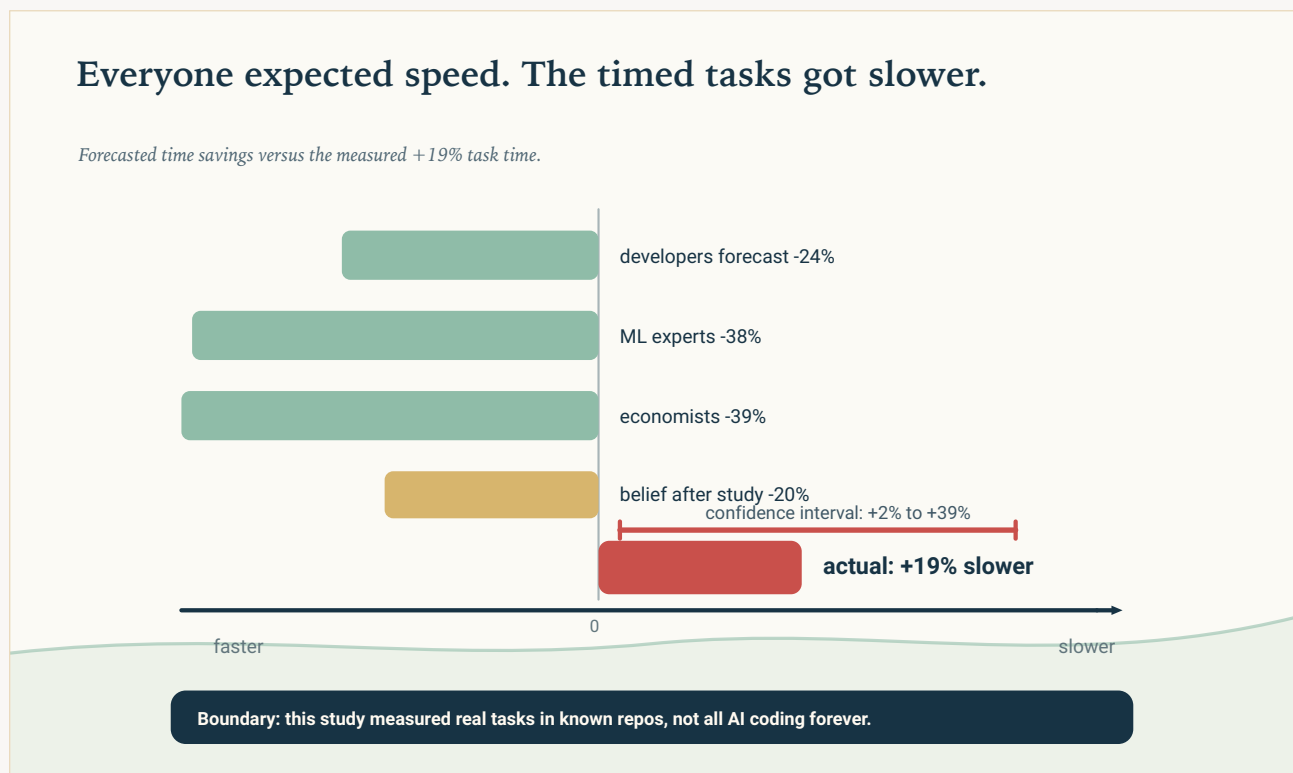
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Zkušení vývojáři se s AI cítili rychlejší, přestože měřitelně zpomalili — skutečný výsledek, nikoli verdikt nad programováním s AI

Zeptejte se zkušeného programátora, zda ho asistent AI zrychluje, a obvykle dostanete číslo: šetří mi dvacet, třicet procent času. Zeptejte se ekonoma nebo výzkumníka strojového učení a číslo ještě vzroste. Začátkem roku 2025 udělal tým METR pomalou a nákladnou věc: ověřil to. Šestnácti zkušeným vývojářům open-source softwaru zadal 246 skutečných úloh z rozsáhlých kódových základů, které dobře znali, a u každé úlohy hodem mincí rozhodl, zda mohou použít nástroje AI. Potom změřil čas.

Vývojáři předpovídali, že AI zkrátí dobu práce asi o 24 %. Stal se opak: úlohy s AI trvaly **o 19 % déle**. A právě nad další částí stojí za to se zastavit — i po dokončení se titíž vývojáři domnívali, že je AI zrychlila přibližně o 20 %. Byli pomalejší a cítili se rychlejší; rozdíl mezi těmito dvěma čísly je nejzajímavějším výsledkem studie.

Je to skutečný, pečlivě změřený výsledek. Týká se však šestnácti vývojářů pracujících v repozitářích, které důvěrně znají, s nástroji z počátku roku 2025 — a nelze jej shrnout paušální větou „AI vývojáře zpomaluje“. Pozdější data stejného týmu už ukazují opačným směrem.



Všichni předpovídali, že AI práci zrychlí — vývojáři o 24 %, odborníci na strojové učení o 38 %, ekonomové o 39 % a vlastní zpětný odhad vývojářů o 20 %. Stopky zjistily zpomalení o 19 %, s intervalem od 2 % do 39 %. Výsledkem je rozdíl mezi pocitem a měřením.

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.

Co studie měří — 16 zkušených open-source vývojářů, 246 skutečných úloh, známé repozitáře, nástroje z počátku roku 2025 a náhodné rozdělení — a co neměří: většinu vývojářů, jiné obory, neměnný zákon ani recenzovaný výsledek.

Original diagram — The Clean Paper CC BY 4.0

Co se měřilo a co přináší randomizovaný pokus

Randomizovaná kontrolovaná studie (RCT) je nástroj, kterým medicína odlišuje skutečný účinek od vytouženého. Zde byla každá z 246 úloh náhodně přiřazena k provedení s povolenou AI nebo bez ní, takže jediným průměrným systematickým rozdílem mezi oběma skupinami je samotná AI. Díky tomu lze říct, že AI změnu času *způsobila*, nikoli jen pozorovat, že lidé sahající po AI bývají z jiných důvodů rychlejší či pomalejší. Je to důležité, protože běžné důkazy přínosu AI při programování — vlastní hodnocení a výsledky benchmarků — toto nedokážou: benchmark není skutečná práce a vlastní hodnocení může být, jak ukazuje tato studie, sebejistě chybné. Účastníci nebyli začátečníci zápasící s novou hračkou. Byli zavedenými přispěvateli do velkých, vyspělých open-source projektů, které dobře znali, a s nástroji již měli určité zkušenosti.

Co autoři udělali

- Provedli **randomizovanou kontrolovanou studii** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Šestnáct zkušených open-source vývojářů pracovalo ve velkých repozitářích, do nichž pravidelně přispívali a které dobře znali — na konkrétních projektech působili v průměru

pět let.

- Použili **246 skutečných úloh** — opravy chyb, nové funkce a refaktORIZACE z vlastních systémů hlášení těchto projektů. Každá úloha byla náhodně přiřazena k podmínce „AI povolena“ nebo „AI zakázána“.
- „AI povolena“ znamenalo **nástroje z počátku roku 2025: Cursor Pro s Claude 3.5/3.7 Sonnet**. Hlavní veličinou byl skutečný čas dokončení úlohy. Tým vedle něj shromáždil předběžné předpovědi vývojářů, jejich následné odhady a predikce odborníků z ekonomie a strojového učení.

Co zjistili

- **S AI trvaly úlohy o 19 % déle.** Ne kratší dobu, ale delší. 95% interval spolehlivosti sahá přibližně od +2 % do +39 %, takže směr výsledku je pevný, i když přesná velikost méně jistá.
- **Všichni předpovídali opak.** Vývojáři očekávali zrychlení o 24 %, odborníci na strojové učení asi o 38 % a ekonomové přibližně o 39 %. Všechny tři skupiny čekaly velkou úsporu času; stopky zjistily čas navíc.
- **Mezera ve vnímání.** Přestože po práci s AI vyšli pomalejší, vývojáři stále odhadovali, že je zrychlila asi o 20 % — mezi pocitem a měřením vznikl rozdíl zhruba 40 procentních bodů.
- **Možné důvody, zvažované, nikoli prokázané.** Autoři uvádějí faktory, které by mohly zpomalení vysvětlit: vývojáři své kódové základny hluboce znají, takže asistent má méně co přidat; vyspělé projekty mají vysoké, často nevyřčené standardy kvality; repozitáře jsou rozsáhlé a plné kontextu, který model nemá; skutečný čas zabere zadávání pokynů a následná kontrola a oprava výstupu AI. Předkládají je jako vodítka, ne jako závěry.

Co to neukazuje

- Studie **neukazuje**, že AI nedokáže zrychlit většinu vývojářů. Autoři to říkají přímo: šestnáct odborníků na kód, který znají nazpaměť, není průměrný vývojář na průměrné úloze.
- **Neukazuje**, že je AI zbytečná nebo že zpomaluje v jiných podmínkách — u nováčků v kódové základně, při projektech na zelené louce, v neznámých jazycích či v úplně jiných oborech.
- **Nezmrazuje** nástroje v čase. Jde o Cursor a Claude 3.5/3.7 Sonnet z počátku roku 2025; autoři výslovně uvádějí, že lepší nástroje nebo lepší způsoby jejich použití mohou výsledek změnit i v naprosto stejném prostředí.
- Je to **preprint** (zveřejněný v červenci 2025, dosud bez odborného recenzního řízení) a autoři připouštějí, že nemohou zcela vyloučit experimentální artefakty — výsledek však přetrval napříč analýzami.
- Studie **nedovoluje** uklidňující výklad, že vývojáři museli získat něco jiného — více se naučit, cítit se lépe nebo psát kvalitnější kód. Jediná zde měřená věc, pocit zrychlení, je přesně tím, čemu data odporují.

Jak silné jsou důkazy

- **Návrh studie je neobvykle poctivý.** Randomizace, skutečné úlohy, skutečné repozitáře a skutečné měření času představují velký krok vpřed oproti vlastním hodnocením a žebříčkům benchmarků, na nichž stojí většina tvrzení o programování s AI. Zpomalení o 19 % obstálo v kontrolách robustnosti autorů.
- **Nejpřenosnější je mezera ve vnímání.** Odborná intuice o vlastním zrychlení s AI se mýlila přibližně o 40 procentních bodů, optimistickým směrem. Je to varování před *každým* deklarovaným nárůstem produktivity díky AI — včetně čísel z této studie.
- **Je to snímek, ne trend.** Vlastní navazující data METR z února 2026, se stejným typem vývojářů a novějšími nástroji, ukazují na zrychlení — velmi hrubě o 18 % u vracejících se vývojářů a o 4 % u nových — autoři je však označují za slabý důkaz, zkreslený výběrem ochotných účastníků (vývojáři stále častěji odmítali pracovat bez AI, odměna klesla a výběr úloh se posunul). Poctivé čtení zní, že obraz se mění a i tato změna se uvádí bez tlacení prstem na váhu.

Proč na tom záleží

Většina sporu o AI a programování stojí na ukázkách a pocitech: uhlažené nahrávce obrazovky, sebestismém tvrzení nebo odmítavém protočení očí. Vzácné a cenné je, že někdo provedl nudný experiment — náhodně rozdělil podmínky, změřil skutečnou práci a pak se lidí zeptal, jak se jim vedlo. Odpověď je nepříjemná pro oba tábory. Nabourává příběh, že AI vždy prudce zrychluje odborníky na těžkém, známém kódu. A narušuje i úhledný opak — „AI vývojáře zpomaluje, prokázáno“ — protože novější data stejného týmu už míří druhým směrem. Nejtrvalejší ponaučení je zároveň nejmenší a nejlidštější: lidé se při práci cítili rychlejší, ačkoli byli měřitelně pomalejší. „Působí to rychleji“ není důkazem, že to rychlejší je. Měřte.

Čisté shrnutí

V randomizované kontrolované studii nechal METR šestnáct zkušených open-source vývojářů dokončit 246 skutečných úloh v kódových základnách, které dobře znali, přičemž u náhodné poloviny byly povoleny nástroje AI z počátku roku 2025 (Cursor Pro a Claude 3.5/3.7 Sonnet). Vývojáři čekali, že AI zkrátí čas asi o 24 %; místo toho jej prodloužila o 19 % — a poté se stále domnívali, že je zrychlila přibližně o 20 %. Tato mezera ve vnímání je ostrým a robustním jádrem studie. Jde však o šestnáct vývojářů, jedno úzké prostředí a pevný snímek počátku roku 2025; autoři výslovně uvádějí, že výsledek neukazuje selhání AI u většiny vývojářů, a jejich vlastní navazující data z roku 2026 již směřují ke zrychlení, s vlastními výhradami. Pečlivé měření, které stojí za vážné zamyšlení — nikoli verdikt nad programováním 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/>