The
CLEAN PAPER

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

Gogaggun masu haɓaka software sun ji kamar AI ta sa su sauri, alhali ma'auni ya nuna sun yi jinkiri — wannan ne ainihin sakamakon, ba hukunci kan AI coding gaba ɗaya ba

A randomized controlled trial na METR ya sa gogaggun masu haɓaka open-source 16 su yi ayyuka na gaske 246 a codebases da suka sani sosai, tare da kayan AI na farkon 2025 a rabin ayyukan da aka rarraba bazuwar. Sun yi tsammanin AI za ta rage lokaci da kusan 24%; maimakon haka lokacin kammalawa ya ƙaru da 19%, kuma bayan gwajin har yanzu sun yi imanin AI ta hanzarta su da kusan 20%. Wannan tazarar fahimta ita ce babbar zuciyar binciken, amma mutane 16 ne, yanayi ɗaya, da hoton kayan aikin farkon 2025 — ba hukunci kan duk AI coding ba.

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

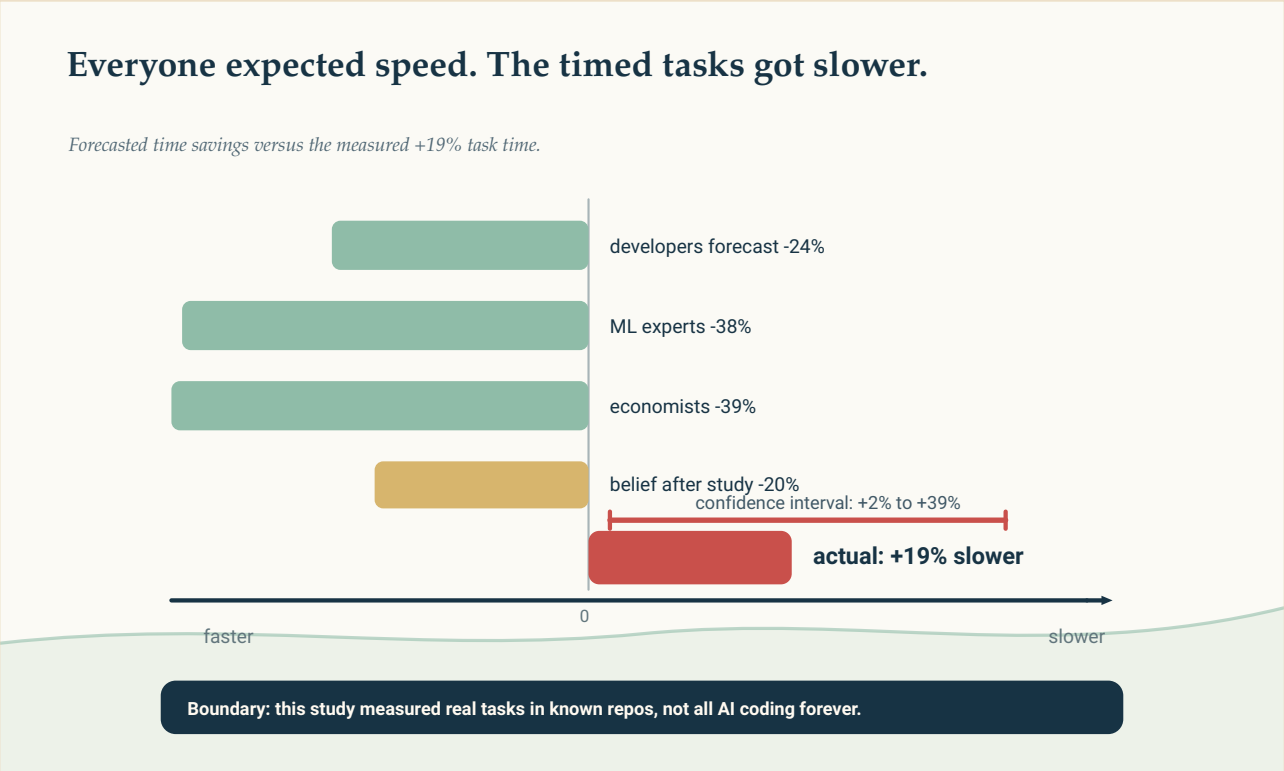
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Gogaggun masu haɓaka software sun ji kamar AI ta sa su sauri, alhali ma'auni ya nuna sun yi jinkiri — wannan ne ainihin sakamakon, ba hukunci kan AI coding gaba ɗaya ba

Ka tambayi gogaggen mai shirya manhaja ko mataimakin AI na rubuta code yana hanzarta aikinsa, sau da yawa zai ba ka lamba: yana ceton ni kashi ashirin, talatin cikin ɗari. Ka tambayi masanin tattalin arziki ko mai binciken machine learning, lambar kan ƙaru. A farkon 2025, wata ƙungiya a METR ta yi abin da yake jinkiri kuma mai tsada: ta gwada wannan kai tsaye. Ta ɗauki gogaggun masu haɓaka open-source guda 16, ta ba su ayyuka na gaske 246 daga manyan codebases da suka sani sosai, sannan — ta hanyar zaɓin bazuwar kamar jefa tsabar kuɗi, aiki bayan aiki — ta ba su damar amfani da kayan AI ko ta hana su. Daga nan aka auna lokacin aikin.

Masu haɓakar sun yi hasashen cewa AI za ta rage lokacin kammala aiki da kusan 24%. Abin da ya faru ya kasance akasin haka: ayyukan da aka yi da AI sun ɗauki **19% fiye da lokaci**. Kuma ga ɓangaren da ya fi dacewa a tsaya a yi tunani a kansa — bayan sun gama, waɗannan masu haɓakar har yanzu sun yi imanin AI ta hanzarta su da kusan 20%. Sun yi jinkiri, amma sun ji kamar sun yi sauri; tazarar da ke tsakanin waɗannan lambobi biyu ita ce abin da ya fi ban sha'awa a binciken.

Wannan sakamako ne na gaske da aka auna da kulawa. Amma kuma masu haɓaka 16 ne kawai, suna aiki a repositories da suka sani sosai, da kayan aikin farkon 2025 — don haka ba ya nufin jumlar da ba ta da sharadi cewa "AI tana sa masu haɓaka software su yi jinkiri." Sabbin bayanai daga ƙungiyar ɗaya ma sun riga sun nuna yiwuwar kishiyar haka.



Kowa ya yi hasashen AI za ta hanzarta aikin — masu haɓaka −24%, masana machine learning −38%, masana tattalin arziki −39%, kuma imanin masu haɓaka bayan gwajin −20%. Agogon lokaci ya nuna +19% jinkiri, da tazara daga +2% zuwa +39%. Tazarar tsakanin abin da aka ji da abin da aka auna ita ce ainihin sakamakon.

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.

Abin da binciken ya auna — gogaggun masu haɓaka open-source 16, ayyuka na gaske 246, repositories da suka saba da su, kayan aikin farkon 2025, da rarrabawa bazuwar — da abin da bai auna ba: ba mafi yawan masu haɓaka ba, ba sauran fannoni ba, ba doka tabbatacciya ba, kuma ba peer-reviewed ba.

Original diagram — The Clean Paper CC BY 4.0

Abin da aka auna a nan, da abin da randomized gwaji yake ba ka

Randomized controlled gwaji (RCT) shi ne irin kayan aikin da magani ke amfani da shi don bambance tasiri na gaske da abin da ake fata kawai. A nan, an rarraba kowane ɗaya daga cikin ayyuka 246 bazuwar zuwa “AI an yarda” ko “AI ba a yarda ba”, don haka a matsakaici bambanci na tsari tsakanin rukunan biyu shi ne kasancewar AI kanta. Wannan ne ya ba masu bincike damar cewa AI *ta jawo* canjin lokacin, maimakon kawai su lura cewa mutanen da suka zaɓi AI sun fi sauri ko jinkiri saboda wasu dalilai. Wannan yana da muhimmanci domin shaidun da aka fi amfani da su wajen ikirarin samun ribar AI a coding — rahoton mutum game da kansa da maki na benchmark — ba sa iya nuna wannan. Benchmark ba aikin yau da kullum na gaske ba ne, kuma self-report, kamar yadda wannan binciken ya nuna, na iya zama da tabbaci amma ya kasance kuskure. Masu haɓakar a nan ba sababbi ba ne da ke rikicewa da sabon kayan wasa. Sun kasance masu ba da gudummawa na dindindin ga manyan, tsofaffin ayyukan open-source da suka sani sosai, kuma wasu suna da ƙwarewar amfani da kayan AI tun da farko.

Abin da marubutan suka yi

- Sun gudanar da **randomized controlled gwaji** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Gogaggun masu haɓaka open-source guda 16, kowannensu yana aiki a babban repository da yake ba da gudummawa akai-akai kuma ya sani sosai — matsakaicin kusan shekaru biyar a takamaiman aikin.
- Sun yi amfani da **ayyuka na gaske 246** — gyaran bugs, ƙarin features da refactors da aka ɗauko daga issue trackers na waɗannan ayyukan. An rarraba kowane aiki bazuwar zuwa "AI allowed" ko "AI disallowed."
- "AI allowed" yana nufin **kayan aikin farkon 2025: Cursor Pro tare da Claude 3.5/3.7 Sonnet.** Babban ma'aunin shi ne ainihin lokacin kammala kowane aiki. Haka kuma ƙungiyar ta tattara hasashen masu haɓaka kafin gwajin, kimantawarsu bayan gwajin, da hasashen masana tattalin arziki da machine learning.

Abin da suka gano

- **Da AI, ayyukan sun ɗauki 19% fiye da lokaci.** Ba sauri ba — jinkiri. 95% confidence interval ɗin ya tashi daga kusan +2% zuwa +39%, don haka alkiblar sakamakon tana da ƙarfi ko da ainihin girman tasirin bai daidai sosai ba.
- **Kowa ya yi hasashen kishiyar haka.** Masu haɓaka sun yi hasashen 24% hanzari; masana machine learning kusan 38%; masana tattalin arziki kusan 39%. Duk rukunan uku sun yi tsammanin AI za ta ceci lokaci mai yawa; ma'aunin lokaci ya nuna ta ƙara shi.
- **Tazarar fahimta.** Bayan sun yi aikin kuma sun fito a hankali, masu haɓakar har yanzu sun kimanta AI ta hanzarta su da kusan 20% — tazara kusan maki 40 tsakanin abin da suka ji da abin da agogo ya nuna.
- **Dalilai masu yiwuwa, an tantance su amma ba a tabbatar da su ba.** Marubutan sun jera abubuwan da za su iya bayyana jinkirin: waɗannan masu haɓaka sun san codebases nasu sosai, don haka ƙarin abin da mataimaki zai kawo ya ragu; tsofaffin projects suna da manyan matakan inganci da yawa daga cikinsu ba a rubuta su ba; repositories suna da girma kuma suna cike da context da model bai sani ba; sannan lokaci na gaske yana shiga prompting, dubawa da gyara abin da AI ta samar. Waɗannan alamu ne, ba hukunci na ƙarshe ba.

Abin da wannan ba ya nuna

- Ba ya nuna cewa AI **ba ta hanzarta mafi yawan masu haɓaka software ba**. Marubutan sun faɗi haka kai tsaye: ƙwararru 16 a code da suka sani tamkar tafin hannunsu ba su wakiltar matsakaicin developer a matsakaicin aiki.
- Ba ya nuna AI ba ta da amfani, ko tana sa mutane su yi jinkiri a sauran yanayi — sabbin shiga codebase, sabon project daga farko, yarukan da ba a saba da su ba, ko sauran fannoni gaba ɗaya.
- Ba ya kulle kayan aikin a wuri guda. Wannan Cursor da Claude 3.5/3.7 Sonnet na farkon 2025 ne; marubutan sun bayyana cewa kayan aikin da suka fi kyau, ko hanyoyin amfani da waɗannan kayan da suka fi kyau, na iya canza sakamakon ko a wannan yanayin ɗaya.
- Wannan **preprint** ne (an wallafa a Yuli 2025, ba a yi peer review ba tukuna), kuma marubutan sun lura cewa ba za su iya kawar da yiwuwar wasu experimental artifacts gaba ɗaya ba — duk da cewa sakamakon ya tsaya a nazarce-nazarcensu daban-daban.

- Ba ya ba da lasisin cewa tabbas masu haɓakar sun samu wani amfani daban — sun koyi ƙari, sun ji daɗi, ko sun rubuta code mafi kyau. Abin da aka auna a nan, jin cewa an hanzarta su, shi ne ainihin abin da bayanan suka saɓa masa.

Yaya ƙarfin shaidar yake

- **Tsarin gwajin yana da gaskiya fiye da yawancin irin wannan bincike.** Rarrabawa bazuwar, ayyuka na gaske, repositories na gaske, ainihin auna lokaci — babban mataki ne fiye da self-reports da benchmark leaderboards da yawancin ikirarin AI coding suke dogara da su. Jinkirin 19% ya ci gaba da bayyana a robustness checks na marubutan.
- **Sakamakon da ya fi sauƙin ɗauka zuwa sauran yanayi shi ne tazarar fahimta.** Kwararrun mutane sun yi kuskure wajen kimanta tasirin AI a kan saurinsu da kusan maki 40, kuma kuskuren ya kasance a bangaren optimism. Wannan gargadi ne ga *duk* wani ikirarin karin productivity da aka dogara da self-report — har da lambobin wannan binciken.
- **Hoto ne na wani lokaci, ba trend na dindindin ba.** Follow-up na METR a Fabrairu 2026, da irin waɗannan masu haɓaka amma sabbin kayan aiki, ya nuna yiwuwar hanzari — a ƙalla kusan −18% ga developers da suka dawo da −4% ga sababbi — amma marubutan sun kira wannan shaida mai rauni saboda waɗanda suka yarda su shiga sun canza (masu haɓaka da yawa sun ƙi yin aiki ba tare da AI ba, kuɗin aikin ya ragu, kuma zaɓin tasks ya karkata). Karatu mai gaskiya shi ne hoton yana canzawa, kuma ko wannan canjin an ruwaito shi da taka-tsantsan.

Me ya sa wannan yake da muhimmanci

Yawancin muhawarar AI da programming tana gudana ne da demos da ji: bidiyo mai kyau, ikirari mai ƙarfin gwiwa, sai martani mai watsi. Abin da yake da wuya — kuma ya sa wannan binciken ya cancanci karantawa — shi ne wani ya gudanar da gwajin mai gajiyarwa: rarraba bazuwar, auna aikin gaske, sannan a tambayi mutane yadda suka ji. Amsar ba ta da daɗi ga bangarorin biyu. Tana karya labarin cewa AI koyaushe tana ƙara wa ƙwararrun developers sauri a code mai wahala da suka saba da shi. Amma tana kuma karya kishiyar labarin mai sauƙi — "AI tana sa developers su yi jinkiri, an tabbatar" — domin sabbin bayanan ƙungiyar ɗaya sun riga sun karkata zuwa wani bangare. Darasin da ya fi ɗorewa shi ne mafi ƙanƙanta kuma mafi na ɗan Adam: mutanen da ke yin aikin sun ji kamar sun yi sauri alhali ma'auni ya nuna sun yi jinkiri. "Yana ji kamar ya fi sauri" ba shaida ba ce cewa hakika haka yake. Auna shi.

Taƙaitaccen bayani

A cikin randomized controlled gwaji, METR ta sa gogaggun masu haɓaka open-source 16 su kammala ayyuka na gaske 246 a codebases da suka sani sosai, tare da kayan AI na farkon 2025 (Cursor Pro da Claude 3.5/3.7 Sonnet) a yarda da su a rabin ayyukan da aka zaɓa bazuwar. Masu haɓakar sun yi tsammanin AI za ta rage lokacin aiki da kusan 24%; maimakon haka ta ƙara lokacin kammalawa da 19% — kuma bayan gwajin har yanzu sun yi imanin ta

hanzarta su da kusan 20%. Wannan tazarar fahimta ita ce babbar, ƙaƙƙarfan zuciyar binciken. Amma mutane 16 ne kawai, yanayi ɗaya mai takamaimai, da hoton kayan aikin farkon 2025; marubutan sun bayyana cewa wannan ba ya nuna AI ba ta taimaka wa mafi yawan developers, kuma follow-up nasu na 2026 ya riga ya nuna yiwuwar hanzari, tare da nasa caveats. Ma'auni mai kyau da ya cancanci a ɗauke shi da muhimmanci — ba hukunci na ƙarshe kan AI coding ba.

Majiyoyi

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/