



The CLEAN PAPER

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

Àwọn developer tó ní ìrí rò pé AI mú wọn yara,
sùgbón a wọn pé wọn ṣiṣẹ lóra síi — finding gidi náà,
kì í ṣe verdict lórí AI coding

Nínú randomized controlled trial METR, open-source developers 16 ṣe real tasks 246 lórí codebases tí wọn mọ dáadà, pèlú early-2025 AI tools tí a gba láàyè fún random half. Wọn retí 24% speedup; dípò bẹẹ task time ga ní 19%, lẹyìn náà wọn sì rò pé AI mú wọn yara síi ní ayíká 20%. Perception gap yẹn ni core finding. Sùgbón setting náà dín, tools jẹ snapshot ti 2025, àwọn olùkòwé sì sọ kedere pé kò fí hàn pé AI kò ran ọpọ developers lówó.

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

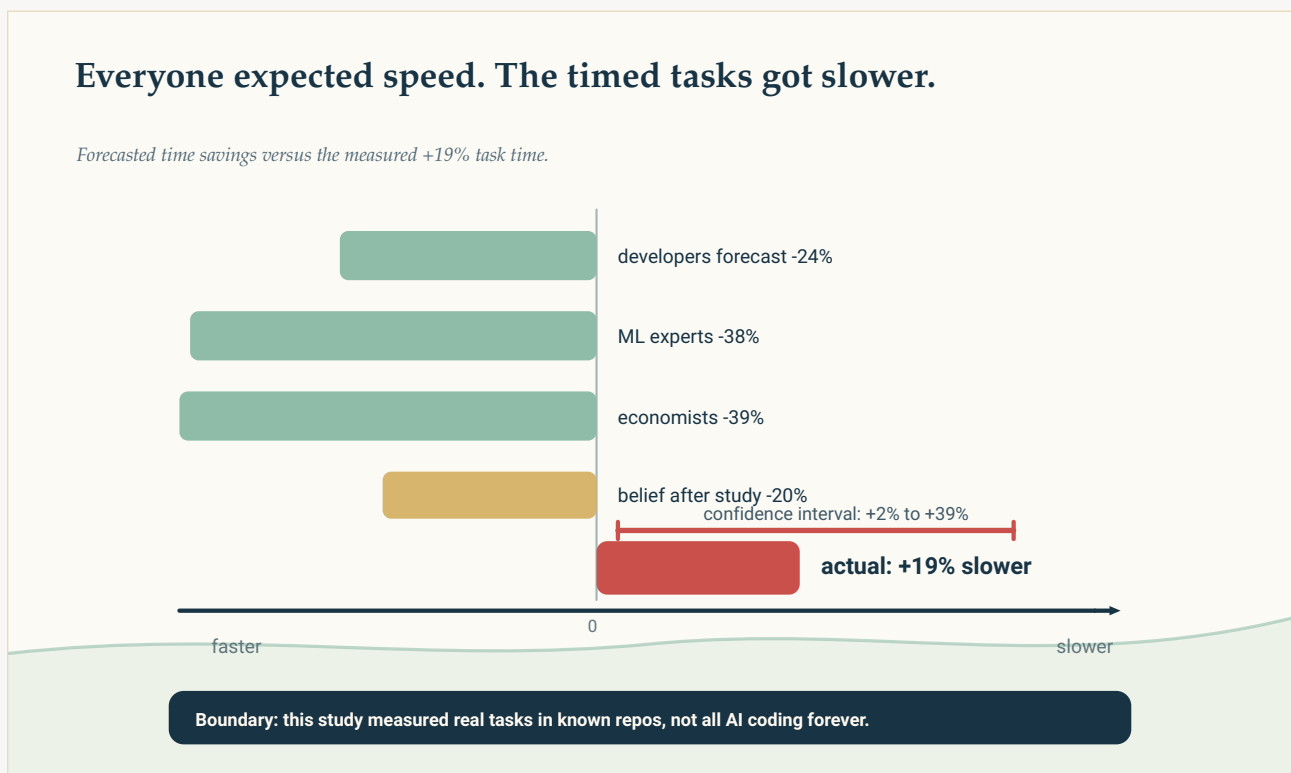
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Àwọn developer tó ní ìrírí rò pé AI mú wọn yara, sùgbón a wọn pé wọn ṣiṣẹ lórà síi — èyí ni abajade gidi, kì í ṣe ìdájò iṣẹhin lórí AI coding

Bèèrè lówọ programmer tó ní ìrírí bóyá AI coding assistant ní mú ṣẹ rẹ yara, ó sáà maa fún ọ ní nọmbà kan: ó ní fipamọ ogún, oḡbọn percent àkókò mí. Bèèrè lówọ economist tàbí machine-learning olùwádíí, nọmbà náà maa ní ga síi. Ní ìbèrẹ 2025, ẹgbẹ kan ní METR ṣe ohun tó lórà, tó sì gbowó: wọn ṣàyèwò gidi. Wọn mú open-source developers méréndínlógún tó ní ìrírí, wọn fún wọn ní gidi ṣẹ-ṣiṣe 246 láti codebases nílá tí wọn mò dáadà, wọn sì lo coin flip fún ṣẹ-ṣiṣe kọkan láti pinnu bóyá wọn lè lo AI irinṣẹ tàbí rara. Lẹyìn náà wọn wọn àkókò ṣẹ náà.

Àwọn developers tí sọ tẹlẹ pé AI yóò dín ṣẹ-ṣiṣe àkókò wọn kù ní ayíká 24%. Ohun tó ṣẹlẹ jẹ idakeji: ṣẹ-ṣiṣe tí wọn ṣe pèlú AI gba **19% àkókò síi**. Apá tó yẹ ká dúró lórí ní pé — lẹyìn tí wọn parí — àwọn developers kan náà sì rò pé AI tí mú wọn yara síi ní ayíká 20%. Wọn lórà síi, sùgbón wọn ní ìmòlára pé wọn yara síi; àláfò láàárín nọmbà méjì yẹn ni ohun tó fanimọra jù nínú ìwádíí náà.

Èyí jẹ abajade gidi tí a wọn pèlú ṣọra. Sùgbón ó tún jẹ developers méréndínlógún, lórí repositories tí wọn mò gan-an, pèlú irinṣẹ tí ìbèrẹ 2025 — kì í ṣe gbolohun tó rọrùn pé “AI ní mú developers lórà.” dátà tuntun láti ẹgbẹ kan náà tí ní tọka sí ìhà mǐràn.



Gbogbo ẹgbẹ sọ tẹlẹ pé AI yóò mú ṣẹ yara — developers –24%, ML experts –38%, economists –39%, àti ìgbàgbọ developers lẹyìn ìwádíí –20%. Stopwatch rí +19% lórà síi, pèlú interval +2% sí +39%. Àláfò láàárín ohun tí wọn ní ìmòlára rẹ àti ohun tí a wọn ní àwárí náà.

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.

Ohun tí iwádíí náà wọ́n — expert open-source developers 16, gidi isẹ́-ṣiṣe 246, repos tí wọ́n mọ̀, irinṣẹ́ ibèrẹ̀ 2025, randomized — àti ohun tí kò wọ́n: kì í ṣe ọ̀pọ̀ jù lọ developers, kì í ṣe domains miiran, kì í ṣe fixed law, kì í ṣe peer-reviewed.

Original diagram — The Clean Paper CC BY 4.0

Ohun tí a wọ́n, àti ohun tí randomized trial fún ọ

Randomized controlled idánwò (RCT) ni medicine máa ní lò láti yà gidi ipa kúrò nínú ohun tí a kan ní retí. Níhìn-ín, isẹ́-ṣiṣe koṣo nínú 246 ni a fi random assignment sí “AI allowed” tàbí “AI not allowed,” kí ní apapọ̀, iyàtò systematic kan soṣo láàárín egbé méjì jẹ AI funra rẹ. Èyí ló jẹ kí a lè sọ pé AI *fa* iyípadà nínú àkókò, dípò ká kan rí pé àwọn èniyàn tó yan AI máa ní yara tàbí lóra fún àwọn idí miiran. Ó ṣe pàtàkì nítorí èrí tí a sàbà fi ní sọ productivity gain ti AI coding — self-reports àti benchmark scores — kò lè ṣe èyí: benchmark kì í ṣe gidi isẹ́, self-report sì lè jẹ aṣiṣe pèlú igboya, bí iwádíí yí ṣe fi hàn. Àwọn developers nínú iwádíí yí kì í ṣe beginners tó ní gbàyànjú toy tuntun. Wọ́n jẹ contributors tó ti pé lórí open-source projects nílá, mature, tí wọ́n mọ̀ dáadáa, wọ́n sì ti ní diẹ experience pèlú irinṣẹ́ náà.

Ohun tí àwọn olùkòwé ṣe

- Wọ́n ṣe **randomized controlled idánwò** (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). Open-source developers méréndínlógún tó ní irírí, ọkoṣo kan ní ṣiṣe lórí repository nílá tí ó máa ní contribute sí, tí ó sì mọ̀ dáadáa — ní apapọ̀ ọdún márùn-ún lórí projects pátó wọnyí.
- Wọ́n lo **gidi isẹ́-ṣiṣe 246** — bug fixes, àbùdá àti refactors láti issue trackers ti projects náà. Task koṣo kan ni a fi random assignment sí “AI allowed” tàbí “AI disallowed.”

- “AI allowed” tùmò sí **irinṣẹ̀ ibèrẹ̀ 2025: Cursor Pro pèlú Claude 3.5/3.7 Sonnet**. Primary measure ni actual completion àkókò fún iṣẹ-ṣiṣe koṣkan. Pèlú rẹ, ẹgbẹ nàà kó forecasts (látí developers ṣáájú), iṣíro àfojúsùn (lẹyìn), àti predictions látí economics àti machine-learning experts.

Ohun tí wọn rí

- **Pèlú AI, iṣẹ-ṣiṣe gba 19% àkókò síi.** Kì í ṣe yara — wọn lóra. 95% ìgbẹkẹlẹ interval jẹ ní ayíká +2% sí +39%, nítorí nàà itọsọna nàà dájú bó tilẹ̀ jẹ pé exact iwọn kò péye.
- **Gbogbo ènìyàn sọ idakeji tẹlẹ.** Developers retí 24% speedup; machine-learning experts tó 38%; economists tó 39%. Gbogbo ẹgbẹ retí pé AI yóò fipamọ àkókò púpọ; stopwatch rí pé ó fi àkókò kún un.
- **Àláfó perception.** Lẹyìn iṣẹ nàà, bó tilẹ̀ jẹ pé wọn lóra síi, developers sì iṣíro àfojúsùn pé AI mú wọn yara síi ní ayíká 20% — àláfó tó fẹrẹ̀ jẹ 40 kókó láàárín ohun tí wọn ní imọlára rẹ àti ohun tí clock ṣe àkọsílẹ.
- **Àwọn idí tó ṣeé ṣe, tí a ṣe iwọn sùgbón tí a kò fi proof hàn.** Àwọn olùkòwé tóka sí factors tó lè ṣàlàyé slowdown: developers wònyí mọ codebases wọn jinlẹ, nítorí nàà ohun tí assistant lè fi kún kere; mature projects ní high quality standards tí a kò sàbà kọ síta; repositories nílá ni context púpọ tí àwòṣe kò ní; àkókò gidi sì í ló sí prompting, review àti correction ti AI output. Wọn fi wònyí hàn gégé bí leads, kì í ṣe verdicts.

Ohun tí èyí kò fi hàn

- Kò fi hàn pé AI kò lè mú ọpọ̀ jù lọ developers yara. Àwọn olùkòwé sọ ọ taara: experts 16 lórí code tí wọn mọ dáadàa kì í ṣe average developer lórí average iṣẹ-ṣiṣe.
- Kò fi hàn pé AI kò wúlò, tàbí pé ó máa ní fa slowdown nínú settings mǐràn — newcomers sí codebase, greenfield iṣẹ, unfamiliar èdè, tàbí fields mǐràn.
- Kò di irinṣẹ nàà mọ sí ipo kan. Èyí ni Cursor àti Claude 3.5/3.7 Sonnet ti ibèrẹ̀ 2025; àwọn olùkòwé sọ pé better irinṣẹ tàbí better ways of using them lè yí èsì padà paápàa nínú setting kan nàà.
- Ó jẹ **preprint** (July 2025, kò tī peer-reviewed), àwọn olùkòwé sì sọ pé wọn kò lè yọ gbogbo experimental artifacts kúrò, bó tilẹ̀ jẹ pé àwá rí nàà dúró nípasẹ analyses wọn.
- Kò fún wa ní àṣe látí sọ pèlú itùnù pé bóyá developers gba nńkan mū — kọ síi, ní ayọ síi, tàbí kọ code tó dára síi. Ohun kan tí a wọn nínú imọlára wọn, felt speedup, ni dátà gan-an tako.

Báwo ni èrí ṣe lágbára tó?

- **Design nàà dá lórí measurement tó dára ju ohun tí a sàbà rí lọ.** Randomized, gidi iṣẹ-ṣiṣe, gidi repositories, actual timing — ìgbésẹ̀ nílá ju self-reports àti benchmark leaderboards tí ọpọ̀ claims AI coding gbára lé. 19% slowdown nàà dúró lẹyìn robustness checks àwọn olùkòwé.
- **Finding tó rọrùn jù látí gbe lọ sí ibòmíràn ni perception gap.** Expert intuition nípa speedup AI ti ara ẹni jẹ aṣiṣe ní ayíká 40 kókó sí ihà optimistic. Èyí jẹ ikilọ̀ fún *gbogbo* self-reported productivity gain látí AI — pèlú dátà iwádíi yìí funra rẹ.
- **Snapshot ni, kì í ṣe trend.** Follow-up METR ti February 2026, pèlú irú developers kan nàà lórí irinṣẹ tuntun,

ń tóká sí speedup — ní iwón tò rò tò — 18% fún returning developers àti — 4% fún tuntun — sùgbón àwón olùkòwé pe e ní weak ẹrì, tí olukopa selection ti yí: developers ń kọ láti sísẹ̀ láìsì AI síi, pay rate dín kù, isẹ̀-sísẹ̀ selection sì skew. Itumọ̀ tò yẹ ni pé picture ń yí, wón sì ń rọ̀yìn iyípadà nàà pèlú isọra.

Kí nìdí tí ó fi sẹ̀ pàtàkì?

Ọ̀pọ̀ ariyanjiyan nípa AI àti programming gbára lé demos àti “vibes”: screen recording tò lẹ̀wa, claim tò dájú, àti eye-roll láti apa kejì. Ohun tí kò wọ̀pọ̀, tí ó sì jẹ́ kí iwádìí yí yẹ ká ka, ni pé ẹ̀nikan sẹ̀ idánwò tò boring sùgbón tò wúlò — randomize, wón gidì isẹ̀, lẹ̀yìn nàà béèrè lówò àwón ènìyàn bí ó sẹ̀ lọ. Ìdáhùn nàà kò rọ̀rùn fún camp kankan. Ó fọ̀ itan pé AI máa ní turbocharge expert developers lórí hard, familiar code ní gbogbo igba. Ó tún fọ̀ idakeji tò mò — “a ti fi ẹ̀rì múlẹ̀ pé AI ń mú developers lóra” — nítorí newer dátà ti egbẹ̀ kan nàà ti ń tẹ̀ sí ihà speedup. Èkọ̀ tò lè pé jù ni kéré, tò sì jẹ́ ti ènìyàn: àwón ènìyàn tò sẹ̀ isẹ̀ ní ìmọ̀lára pé wón yara, nígbà tí measurement fi hàn pé wón lóra. “Ó dà bí ẹ̀ni pé ó yara” kì í sẹ̀ ẹ̀rì pé ó yara. Wón un.

Àkótán kedere

Nínú randomized controlled idánwò, METR jẹ́ kí experienced open-source developers 16 sẹ̀ gidì isẹ̀-sísẹ̀ 246 lórí codebases tí wón mò dáadàá, pèlú early-2025 AI irinsẹ̀ (Cursor Pro àti Claude 3.5/3.7 Sonnet) tí a gba láàyè fún random half. Àwón developers retí pé AI yóò dín isẹ̀-sísẹ̀ àkókò kù ní ayíká 24%; dípò bẹ̀, completion àkókò ga ní 19% — lẹ̀yìn nàà wón sì rò pé AI mú wón yara síi ní ayíká 20%. Perception gap yẹn ni core àwárí tò muna jù. Sùgbón developers 16 ni, setting kan tò dín, snapshot ti ìbèrẹ̀ 2025; àwón olùkòwé sọ kedere pé kò fi hàn pé AI kò ran ọ̀pọ̀ developers lówò, newer follow-up wón sì ti ń tóká sí speedup, pèlú caveats tirẹ̀. Measurement tò tò sí ifiyèsì — kì í sẹ̀ verdict lórí AI coding.

Àwón orísun

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/>