The Clean Paper

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

AI ব্যবহার করে অভিজ্ঞ ডেভেলপারদের কাজ দ্রুত মনে হয়েছিল, কিন্তু মাপে তারা ধীর হয়েছিলেন — আসল ফলটি এটাই, AI কোডিংয়ের চূড়ান্ত রায় নয়

METR-এর randomized controlled trial-এ ১৬ জন অভিজ্ঞ ওপেন-সোর্স ডেভেলপার তাদের পরিচিতি কোডবেসে ২৪৬টি বাস্তব কাজ করেছেন; অর্ধেক কাজে early-2025 AI tool ব্যবহার করা যেত। তারা ২৪% speedup আশা করেছিলেন, কিন্তু AI completion time ১৯% বাড়িয়েছে — তবুও পরে তারা মনে করেছিলেন AI তাদের প্রায় ২০% দ্রুত করেছে। এই perception gap গবেষণার সবচেয়ে শক্ত ফল; তবে sample ছোট, setting সংকীর্ণ এবং পরবর্তী tool ইতিমধ্যেই ভিন্ন ফলের ইঙ্গিত দিচ্ছে।

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

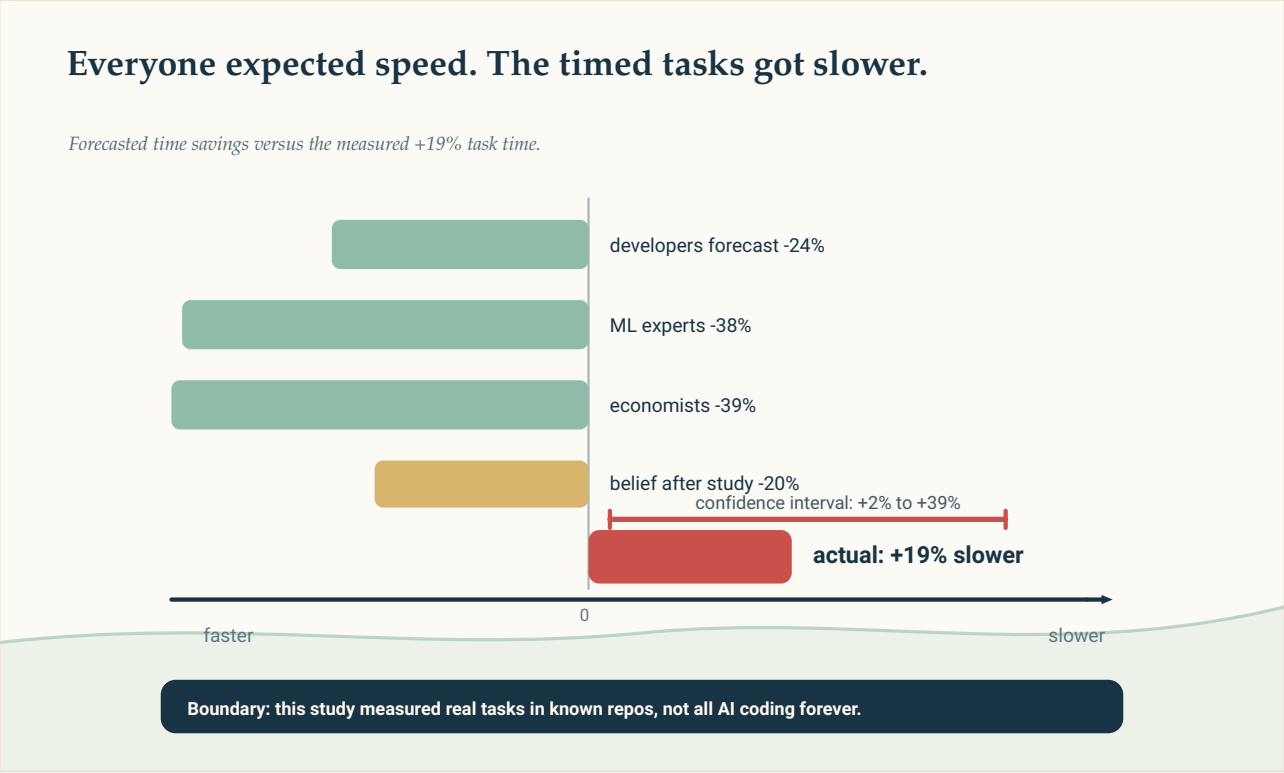
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

অভিজ্ঞ ডেভেলপাররা AI ব্যবহার করে নিজেকে দ্রুত মনে করছিলেন—কিন্তু মাপা সময়ে তারা ধীর ছিলেন; আসল ফল এটাই, AI কোডিংয়ে চূড়ান্ত রায় নয়

একজন অভিজ্ঞ programmer-কে জিজ্ঞেস করলে AI coding assistant তাকে দ্রুত করে কি না, প্রায়ই একটি সংখ্যা শোনা যায়: 20%, 30% সময় বাঁচে। অর্থনীতিবিদ বা machine-learning গবেষকদের অনুমান আরও বড় হতে পারে। 2025 সালের শুরুতে METR-এর একটি দল এই প্রশ্নে ধীর ও ব্যয়বহুল কাজটি করল: **সরাসরি মেপে দেখল**। 16 জন অভিজ্ঞ open-source developer-কে তাদের খুব পরিচিত বড় codebase থেকে **246টি বাস্তব কাজ** দেওয়া হয় এবং প্রতিটি কাজ আলাদা করে randomize করে ঠিক করা হয় কখন AI tool ব্যবহার করা যাবে, কখন যাবে না। তারপর কাজ শেষ করতে কত সময় লাগে তা মাপা হয়।

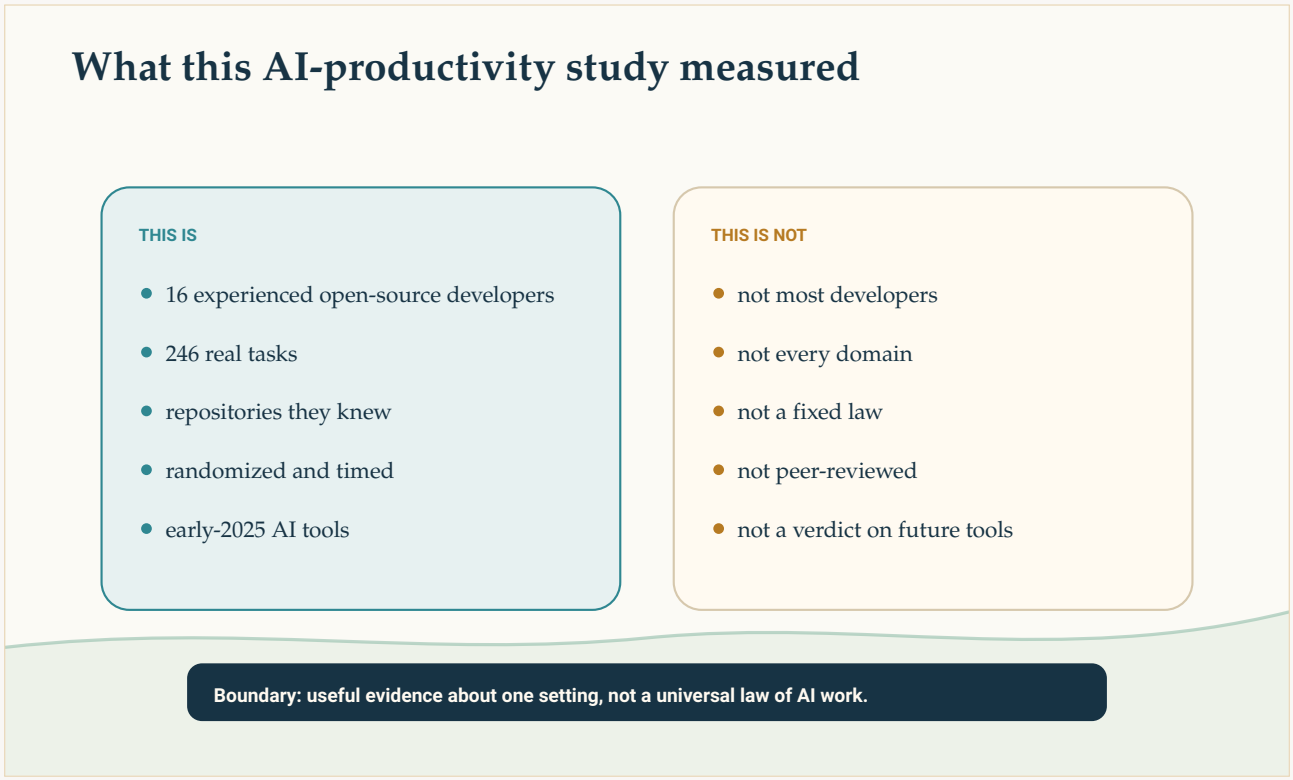
ডেভেলপাররা আগেই অনুমান করেছিলেন AI কাজ সময় প্রায় 24% কমাবে। ঘটল উল্টোটা: AI-সহ কাজ শেষ করতে **19% বেশি সময়** লাগল। আরও আকর্ষণীয় অংশ হলো—কাজ শেষ করার পরও একই ডেভেলপাররা বিশ্বাস করছিলেন AI তাদের প্রায় 20% দ্রুত করেছে। তারা মাপা সময়ে ধীর, অনুভবে দ্রুত। এই দুই সংখ্যার দূরত্বই গবেষণাটির সবচেয়ে গুরুত্বপূর্ণ ফল।

এটি বাস্তব এবং যত্নসহকারে মাপা ফল। কিন্তু এটি 16 জন ডেভেলপার, নিজেদের অতি পরিচিত কোডভাণ্ডার, এবং প্রাথমিক-2025 সরঞ্জামের সময়ের একটি চিত্র। তাই এটি "AI ডেভেলপারদের ধীর করে" এমন সর্বজনীন sentence নয়। একই গবেষণা দলের পরের তথ্য ইতিমধ্যেই অন্য দিকে ইঙ্গিত করছে।



সব দল AI-কে গতি বৃদ্ধি হিসেবে পূর্বাভাস করেছিল—ডেভেলপার −24%, ML বিশেষজ্ঞ −38%, অর্থনীতিবিদ −39%, এবং কাজ শেষে ডেভেলপারদের নিজের বিশ্বাস −20%। স্টপওয়াচ দেখেছে উল্টোটা: +19% ধীরতর, আস্থার ব্যবধান প্রায় +2% থেকে +39%। অনুভব আর মাপের ব্যবধানই মূল ফল।

Original diagram — The Clean Paper CC BY 4.0



গবষেণার পরধিি: 16 বশিষেজ্ঞ উন্মুক্ত-উৎস ডভেলেপার, 246 বাস্তব কাজ, পরচিতি কে াডভাণ্ডার, প্রাথমকি-2025 সরঞ্জাম, র্যান্ডমাইজড বরাদ্দ। এটি অধকিাংশ ডভেলেপাররা, অন্য domains, স্থরি সূত্র বা সহকর্মী-পর্যালে াচতি চূড়ান্ত রায় নয়।

Original diagram — The Clean Paper CC BY 4.0

এই পরীক্ষা কী মপেছে, আর randomisation কী সুবধিা দয়ে

র্যান্ডমাইজড নয়িন্ত্রতি পরীক্ষা (RCT) চকিৎিসাবজি্ঞানে hopeful সম্পর্ক আর কারণগত প্রভাব আলাদা করতে ব্যবহৃত হয়। এখানে 246 কাজ এলে ামলে াভাবে AI-অনুমে াদতি বা AI-disallowed শর্তে গছে। ফলে গড়ে দুই pile-এর পদ্ধতগিত পার্থক্য হওয়ার কথা শুধু AI প্রবশোধকিার। এই নকশা-ই "AI সময়রে পরবির্তন *সৃষ্ট* করছে" বলা সম্ভব কর—শুধু "যারা AI নয়ে তারা অন্যরকম" এমন confounding নয়।

AI কে াডংি লাভরে সাধারণ প্রমাণ—স্ব-প্রতবিদেন ও মানদণ্ড—এটি করতে পারে না। মানদণ্ড বাস্তব কে াডভাণ্ডার কাজ নয়; স্ব-প্রতবিদেন, এই গবষেণায়ই দখো যাচ্ছ, confidently ভুল হতে পার। অংশগ্রহণকারীরা novice ছলিনে না: পরণিত উন্মুক্ত-উৎস প্রকল্পরে প্রতষ্ঠিতি contributor, প্রকল্প-নর্দিষ্টি অভজি্ঞতা গড়ে প্রায় পাঁচ বছর, এবং AI সরঞ্জাম ব্যবহাররে কছিু পূর্বধারণা অভজি্ঞতা ছলি।

গবষেকরো কী করছেনে

- METR-এর Joel Becker, Nate Rush, Beth Barnes ও David Rein একটি **র্যান্ডমাইজড নয়িন্ত্রতি পরীক্ষা** চালান। 16 experienced উন্মুক্ত-উৎস ডভেলেপার নজিদেরে নয়িমতি অবদান রাখা বড় কে াডভাণ্ডার-তে কাজ করনে; নর্দিষ্টি প্রকল্পে গড় অভজি্ঞতা প্রায় পাঁচ বছর।
- প্রকল্প সমস্যা tracker থকেে নওেয়া **246 বাস্তব কাজ**—ত্রুটি সংশে াধন, বশৈষ্ট্যি ও refactor—ব্যবহার করা হয়। প্রতটিি কাজ এলে ামলে াভাবে "AI অনুমে াদতি" বা "AI disallowed" করা হয়।
- AI-অনুমে াদতি শর্ত ছলি **Cursor Pro + Claude 3.5/3.7 Sonnet**, অর্থাৎ প্রাথমকি-2025 tooling। প্রধান ফলাফল ছলি বাস্তব

সম্পূর্ণতা সময়। পাশাপাশি ডেভেলপারদের pre-কাজ পূর্বাভাস, পরবর্তী-কাজ অনুমান এবং অর্থনীতিবিদ/ML বিশেষজ্ঞ পূর্বাভাস সংগ্রহ করা হয়।

তাঁরা কী পেয়েছেন

- **AI ব্যবহার করলে কাজ 19% বেশি সময় নিয়েছে।** 95% আস্থার ব্যবধান প্রায় +2% থেকে +39%; অর্থাৎ সুনির্দিষ্ট প্রভাবের মাত্রা অনিশ্চিত হলেও দিক এই নমুনায় যথেষ্ট কঠিন।
- **সবাই উল্টো টা পূর্বাভাস করেছিল।** ডেভেলপার ~24% গতি বৃদ্ধি, ML বিশেষজ্ঞ ~38%, অর্থনীতিবিদ ~39% আশা করেছিলেন। বাস্তব স্টপওয়াচ কিছু গতি বৃদ্ধি নয়, slowdown দেখায়।
- **উপলব্ধি ব্যবধান:** কাজ শেষে ডেভেলপাররা এখনও মনে করেছিলেন AI তাদের ~20% দ্রুত করেছে। Felt প্রভাব ও পরিমাপিত প্রভাবের ব্যবধান প্রায় 40 শতাংশ বিন্দু।
- লেখকেরা সম্ভব ব্যাখ্যাগুলো কে অনুমান হিসেবে সাজান, রায় হিসেবে নয়: পরিচিত কো ডভাণ্ডারে বিশেষজ্ঞের own প্রেক্ষাপট সুবিধা; পরিণত প্রকল্পের implicit মান মানক; কো ডভাণ্ডারের প্রেক্ষাপট মডেলের বাইরে থাকা; prompting-এর সময়; AI আউটপুট পর্যালো চনা/সংশো ধন খরচ।

এটি কী দেখায় না

- AI **অধিকাংশ ডেভেলপারকে ধীর করে**—এটি দেখায় না। লেখকেরা নিজেরাই বলেন, 16 বিশেষজ্ঞ পরিচিত কো ডভাণ্ডার গড় ডেভেলপার/কাজের প্রতিনিধি নয়।
- Newcomer, greenfield প্রকল্প, অপরিচিত ভাষা, ছো টতর কো ডভাণ্ডার বা অন্য ক্ষেত্রে AI অকার্যকর/ধীর—এমন উপসংহার নয়।
- সরঞ্জাম মান স্থির—না। এটি প্রাথমিক-2025 Cursor এবং Claude 3.5/3.7 Sonnet। সরঞ্জাম বা usage ধরন improve করলে একই বিন্যাসেও ফল বদলাতে পারে।
- এটি **প্রিপ্রিন্ট** (জুলাই 2025), তখন সহকর্মী পর্যালো চনা সম্পন্ন ছিল না; লেখকেরা পরীক্ষামূলক কৃত্রিম প্রভাব পুরো পুরি নিয়ম বাইরে করতে পারেননি, যদিও ফল বিভিন্ন দৃঢ়তা যাচাইে টিকে ছিল।
- "ডেভেলপাররা হয়তো অন্যভাবে লাভবান হয়েছে—শিখছে, happier ছিল, ভালো কো ড লিখছে"—এই আশ্বস্তকর গল্প গবেষণা প্রমাণ করে না। পরিমাপিত ব্যক্তিনির্ভর দাবি ছিল felt গতি বৃদ্ধি, আর সেটিই তথ্যের সঙ্গে disagree করছে।

প্রমাণ কতটা শক্তিশালী?

নকশা অস্বাভাবিকভাবে শক্তিশালী। Randomization, বাস্তব সমস্যা, পরিচিত কো ডভাণ্ডার এবং বাস্তব সময়বিন্যাস স্ব-প্রতিবেদন/মানদণ্ড-ভারী গবেষণা-সাহিত্যের তুলনায় বড় upgrade। 19% slowdown লেখকদের দৃঢ়তা বিশ্লেষণেও থাকে।

সবচেয়ে transferable সিদ্ধান্ত উপলব্ধি ব্যবধান। বিশেষজ্ঞ নিজস্ব উৎপাদনশীলতা পরিবর্তন সম্পর্কে ~40 শতাংশ বিন্দু optimisticভাবে ভুল ছিল। AI উৎপাদনশীলতার স্ব-প্রতিবেদন শুনলে এই ফল সতর্ক করে: মানুষ নিজের speed পরিবর্তন খুব বিশ্বাসযো গ্যভাবে অনুমান নাও করতে পারে।

এটি একটি সময়বিন্দুর ছবি, দীর্ঘমেয়াদি প্রবণতা নয়। METR-এর ফেব্রুয়ারি 2026 follow-up-এ একই ধরনের developer কিন্তু নতুনতর tool ব্যবহার করে কিছুটা গতি বৃদ্ধির সংকেত দিয়েছেন—ফিরে আসা developer-দের ক্ষেত্রে আনুমানিক −18%, নতুনদের ক্ষেত্রে −4%। লেখকেরা একে দুর্বল প্রমাণ বলেন, কারণ অংশগ্রহণকারী নির্বাচনের ধরন বদলেছে: developer-রা AI ছাড়া কাজ করতে ক্রমে অনিচ্ছুক

হয়েছে, পারশ্রিমকিরে হার কমছে এবং বছে নওেয়া কাজরে ধরনওে পক্ষপাত থাকতে পার।ে তাই সতর্ক পাঠ হল ো: প্রভাব বদলাচ্ছ হতে পার,ে কন্িতু সইে পরবির্তনটওি অনশ্চিতি।

কনে এটি গুরুত্বপূর্ণ

AI coding নয়িে আলে োচনা প্রায়ই demo, ব্যক্তগিত অনুভূতি ও anecdote-এ আটকে যায়। এই গবষেণার মূল্য হল ো বরিক্তকির কন্িতু শক্ত পদ্ধতি: randomize করা, বাস্তব কাজরে সময় মাপা, তারপর অংশগ্রহণকারীদরে নজিদেরে অভজ্ঞিতাও জজ্ঞিসে করা। ফলটি দুই পক্ষরে সহজ গল্পই ভঙেে দয়ে। “AI সব বশিষেজ্ঞকে বহুগুণ দ্রুত কর”ে এখানে ভুল; আবার “AI coding প্রমাণতিভাবইে ধীর” বলাও অতরিঞ্জতি, কারণ পরবর্তী তথ্য ইতমিধ্যে অন্য দকিরে সংকতে দচ্ছি।ে

সবচয়েে স্থায়ী শক্ষিা খুব মানব: মানুষ পরমিাপতিভাবে ধীর ছলি, কন্িতু নজিকেে দ্রুত মনে করছ।ে **“দ্রুত মনে হচ্ছ”ে প্রমাণ নয় যে সত্যইি দ্রুত। পরমিাপ করুন।**

সংক্ষপ্তি সারাংশ

METR-এর র্যান্ডমাইজড পরীক্ষায় 16 experienced উন্মুক্ত-উৎস ডভেলেপার নজিদেরে পরচিতি কে াডভাণ্ডারে 246 বাস্তব কাজ করনে। এলে োমলে ো অর্ধকেে প্রাথমকি-2025 AI সরঞ্জাম (Cursor Pro + Claude 3.5/3.7 Sonnet) অনুমে োদতি ছলি। ডভেলেপাররা ~24% গতবৃিদ্ধি আশা করলওে AI শর্তে কাজ সম্পূর্ণতা সময় **19% বড়েছে;ে** কাজ শষেে তারা তবুও ~20% গতবৃিদ্ধি হয়ছেে বলে মনে করছেনে। উপলব্ধি ব্যবধান গবষেণাটরি সবচয়েে শক্তশিালী শক্ষিা। কন্িতু নমুনা ছ োট, বন্যিাস সংকীর্ণ, সরঞ্জাম সময়রে একটি চত্রি কালনর্িধারতি, এবং লখেকদরে 2026 অনুসরণকালীন newer সরঞ্জামে গতবৃিদ্ধরি দকিে দুর্বল সংকতে দয়ে। এটি সতর্ক কারণগত পরমিাপ—AI কে াডংিরে চূড়ান্ত রায় নয়।

উৎস

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/