



— The —
CLEAN PAPER
— —

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

ልምድ ያላቸው developers AI ሲጠቀሙ ፈጣን እንደሆኑ ተሰማቸው፤ በመለኪያ ግን ዝግ ነበሩ — ዋናው ውጤት ይህ ነው

METR ባደረገው randomized controlled trial 16 ልምድ ያላቸው open-source developers በሚያውቋቸው codebases ላይ 246 እውነተኛ tasks ሠሩ። የ2025 መጀመሪያ AI tools በrandom ግማሽ ላይ ተፈቀዱ። Developers 24% speedup ጠበቁ፤ ግን ሥራዎቹ 19% የበለጠ ጊዜ ወሰዱ — ከዚያም AI 20% እንዳፋጠናቸው አሁንም ገመቱ። ያ የperception gap ዋናው ውጤት ነው፤ ጥናቱ በAI coding ላይ አጠቃላይ ፍርድ አይደለም።

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

ልምድ ያላቸው ዴቪሎፕሞንቶች AI ሲጠቀሙ ፈጣን እንደሆኑ ተስማሞቸው፤ መለኪያው ግን ዝግ እንደነበሩ አሳየ — ዋናው ውጤት ይህ ነው፤ “AI ኮድ መጻፍ ተፈርዶበታል” ግን አይደለም

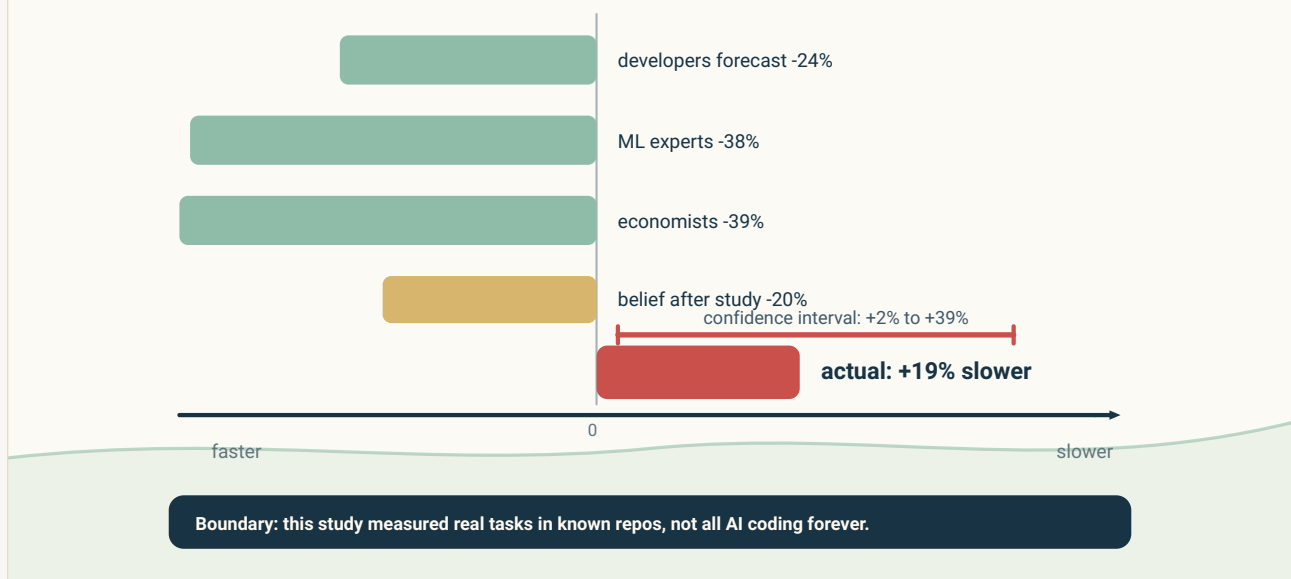
ልምድ ያለውን ፕሮግራሙር AI ኮድ መጻፍ assistant ፍጥነቱን ይጨምርለታል ወይ ብለህ ብትጠይቀው ብዙውን ጊዜ ቁጥር ይሰጥሃል፤ 20%፣ 30% ጊዜ ይቆጥብልኛል። ኢኮኖሚስት ወይም machine-learning ተመራማሪን ብትጠይቅ ቁጥሩ ይበልጥ ይጨምራል። በ2025 መጀመሪያ METR ቡድን ቀላል አይደለም ግን ትክክለኛውን ነገር አደረገ፤ መለካት። 16 ልምድ ያላቸው ያልተፈታ-ምንጭ ገንቢዎችን ወስደው በጥሩ ሁኔታ ከሚያውቋቸው ትልልቅ codebases 246 እውነተኛ ሥራዎችን ሰጧቸው። እያንዳንዱ ሥራ በየዘፈቀደ ምርጫ AI መጠቀም የሚፈቀድበት ወይም የማይፈቀድበት ሆነ፤ ከዚያ የሥራውን ጊዜ ለኩ።

ዴቪሎፕሞንቶች AI የተግባር ጊዜውን በ24% ገደማ እንደሚቀንስ ተንብየው ነበር። ውጤቱ ተቃራኒ ነበር፤ AI በተፈቀደባቸው ሥራዎች 19% የበለጠ ጊዜ ወሰዱ። ከሁሉ የሚያስደስተው ግን ከጨረሱ በኋላ ተመሳሳይ ዴቪሎፕሞንቶች AI በ20% ገደማ እንዳፋጠናቸው አሁንም ያምኑ ነበር። በመለኪያ ዝግ ነበር፤ በስሜት ግን ፈጣን እንደሆኑ ተስማሞቸው። በእነዚህ ሁለት ቁጥሮች መካከል ያለው ርቀት የጥናቱ በጣም ጠቃሚ ክፍል ነው።

ይህ እውነተኛና በጥንቃቄ የተለካ ውጤት ነው። ግን 16 ዴቪሎፕሞንቶች፣ በጥልቀት በሚያውቋቸው repositories፣ የ2025 መጀመሪያ መሣሪያዎችን ተጠቅመው የተገኘ ነው — እና “AI ገንቢዎችን ያዘገያል” የሚል አጠቃላይ ሕግ አይደለም። የተመሳሳይ ቡድን በኋላ ያሰባሰበው ዳታ ቀድሞውኑ ወደ ተቃራኒው አቅጣጫ ይጠቁማል።

Everyone expected speed. The timed tasks got slower.

Forecasted time savings versus the measured +19% task time.



ሁሉም AI ሥራውን እንደሚያፋጥን ተንብየዋል፤ ገንቢዎች -24%፣ ML ባለሙያዎች -38%፣ economists -39%፣ እና ዴቪሎፕሞንቶች ከመከራው በኋላ ያላቸው ግምት -20%። የሰዓት መለኪያው ግን +19% ዝግ መሆኑን አሳየ፤ ከ+2% እስከ +39% የሚደርስ interval ጋር። በተሰማውና በተለካው መካከል ያለው ልዩነት ዋናው ውጤት ነው።

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.

ጥናቱ የሚለካው፤ 16 ልምድ ያላቸው ያልተፈታ-ምንጭ ገንቢዎች፤ 246 እውነተኛ ሥራዎች፤ የሚያውቋቸው repositories፤ የ2025 መጀመሪያ መሣሪያዎች፤ በዘፈቀደ የተመደበ ንድፍ። የማይወክለው፤ አብዛኞቹ ገንቢዎች፤ ሌሎች መስኮች፤ የማይለወጥ ሕግ፤ ወይም በባለሙያዎች የተገመገመ ውጤት።

Original diagram — The Clean Paper CC BY 4.0

ይህ ምንን ለካ፤ በዘፈቀደ የተመደበ ሙከራ ምን ይጨምራል?

በዘፈቀደ የተመደበ እና ቁጥጥር የተደረገ ሙከራ (RCT) በሕክምና እውነተኛ ተጽእኖን ከተስፋ ለመለየት የሚጠቀሙበት ዘዴ ነው። እዚህ 246 ሥራዎች እያንዳንዳቸው AI የሚፈቀድበት ወይም የማይፈቀድበት ቡድን በየዘፈቀደ ተመደቡ። ስለዚህ በአማካይ በሁለቱ ቡድኖች መካከል ያለው የተደራጀ ልዩነት AI ብቻ መሆን ይጠበቃል። ይህ የጊዜ ለውጡን AI እንዳስከተለው ለማለት ያስችላል፤ AI የሚጠቀሙ ሰዎች በሌላ ምክንያት ፈጣን ወይም ዝግ እንደሆኑ ብቻ ከማየት ይለያል። ይህ አስፈላጊ ነው፤ ምክንያቱም ስለ AI ኮድ መጻፍ የሚቀርበው ብዙ ማስረጃ — self-reports እና መመዘኛ scores — ይህን ማድረግ አይችልም። መመዘኛ እውነተኛ ሥራ አይደለም፤ self-reportም ይህ ጥናት እንዳሳየው በእርግጠኝነት ስህተት ሊሆን ይችላል። እዚህ ያሉት ገንቢዎች አዲስ መሣሪያን የሚሞክሩ ጀማሪዎች አልነበሩም፤ በትልልቅና የበሰሉ ያልተፈታ-ምንጭ ፕሮጀክቶች ውስጥ የተረጋጉ contributors ነበሩ፤ እና ከAI መሣሪያዎች ጋር አንዳንድ ቀድሞ ልምድ ነበራቸው።

ደራሲዎቹ ምን አደረጉ?

- በዘፈቀደ የተመደበ እና ቁጥጥር የተደረገ ሙከራ አካሄዱ (METR፡- Joel Becker፤ Nate Rush፤ Beth Barnes፤ David Rein)። 16 ልምድ ያላቸው ያልተፈታ-ምንጭ ገንቢዎች፤ እያንዳንዳቸው በመደበኛነት በሚያበረክቱበት እና

በጥሩ ሁኔታ በሚያውቁት ትልቅ repository ላይ ሠሩ፤ በእነዚህ የተወሰኑ ፕሮጀክቶች ላይ አማካይ ልምዳቸው 5 ዓመት ነበር።

- 246 እውነተኛ ሥራዎች ተጠቀሙ፤ bug fixes፣ features እና refactors ከፕሮጀክቶችዎቹ issue trackers የተወሰዱ። እያንዳንዱ ተግባር “AI allowed” ወይም “AI disallowed” በየዘፈቀደ ተመደበ።
- “AI allowed” ማለት የ2025 መጀመሪያ tooling፡- Cursor Pro ከClaude 3.5/3.7 Sonnet ጋር ማለት ነበር። ዋናው መለኪያ የእያንዳንዱ ተግባር እውነተኛ completion ጊዜ ነበር። ከዚያ በተጨማሪ ከገንቢዎች በፊት የተሰጡ forecasts፣ ከሥራው በኋላ የተሰጡ estimates፣ እና heconomics እና machine-learning ባለሙያዎች የመጡ ትንበያዎች ተሰበሰቡ።

ምን አገኙ?

- AI ሲፈቀድ ሥራዎቹ 19% የበለጠ ጊዜ ወሰዱ። ፈጣን ሳይሆን ዝግ። 95% የመተማመኛ ክልል በግምት +2% እስከ +39% ነው፤ ስለዚህ ትክክለኛው መጠን ያልተወሰነ ቢሆንም አቅጣጫው ጠንካራ ነው።
- ሁሉም ተቃራኒውን ተንብየው ነበር። ገንቢዎች 24% speedup፣ machine-learning ባለሙያዎች 38% ገደማ፣ economists 39% ገደማ ተንብየዋል። ሶስቱም AI ብዙ ጊዜ እንደሚቆጥብ ጠበቁ፣ ሰዓቱ ግን ተጨማሪ ጊዜ እንደወሰደ አሳየ።
- የግንዛቤ gap። ሥራውን በዝግታ ካጠናቀቁ በኋላም ገንቢዎች AI በ20% ገደማ እንዳፋጠናቸው ገመቱ። በተሰማውና በተለካው መካከል 40 percentage points ገደማ ልዩነት ነበር።
- ሊያብራሩ የሚችሉ ምክንያቶች ተዘርዝረዋል፤ ግን አልተረጋገጡም። እነዚህ ገንቢዎች codebasesዎቻቸውን በጥልቀት ስለሚያውቁ assistant የሚጨምረው ዋጋ ያነሰ ሊሆን ይችላል፤ የበሰሉ ፕሮጀክቶች ከፍተኛና ብዙ ጊዜ በግልጽ ያልተጻፉ quality standards አላቸው፤ repositories ትልቅና ሞዴል የማያውቀው ብዙ context አላቸው፤ እና እውነተኛ ጊዜ በprompting፣ በAI output ግምገማና ማስተካከያ ላይ ይወጣል። እነዚህ ምልክቶች ናቸው፤ የተረጋገጡ ምክንያቶች አይደሉም።

ይህ ጥናት የማያሳየው

- AI አብዛኞቹን ገንቢዎች አያፋጥንም ብሎ አያሳይም። ደራሲዎቹ በቀጥታ ይህን ይላሉ፤ በጣም የሚያውቁትን ኮድ ላይ የሚሠሩ 16 ባለሙያዎች አማካይ ገንቢን ወይም አማካይ ተግባርን አይወክሉም።
- AI አይጠቅምም ወይም በሌሎች ሁኔታዎች ሰዎችን ያዘገያል ብሎ አያሳይም፤ አዲስ ወደ codebase የገቡ ሰዎች፣ greenfield ሥራ፣ የማይታወቁ ቋንቋዎች፣ ወይም ሙሉ በሙሉ ሌሎች መስኮች የተለየ ውጤት ሊኖራቸው ይችላል።
- መሣሪያዎቹን በዚያ ደረጃ አያቆማቸውም። ይህ የ2025 መጀመሪያ Cursor እና Claude 3.5/3.7 Sonnet ነው። ደራሲዎቹ የተሻሉ መሣሪያዎች ወይም እነዚህን የተሻለ መጠቀም በተመሳሳይ ሁኔታ እንኳ ውጤቱን ሊቀይር እንደሚችል ይገልጻሉ።
- ይህ preprint ነው — በJuly 2025 የተለጠፈ፣ ገና በባለሙያዎች የተገመገመ ያልሆነ። ደራሲዎቹ experimental artifactsን ሙሉ በሙሉ ማስወገድ እንደማይችሉ ይጠቁማሉ፤ ምንም እንኳ ውጤቱ በrobustness analyses ውስጥ ቢቆይም።
- ገንቢዎች ምናልባት በሌላ መንገድ ጥቅም አግኝተዋል — ይበልጥ ተማሩ፣ ደስ አላቸው፣ የተሻለ ኮድ ጻፉ — ብሎ ለማረጋገጥ ፈቃድ አይሰጥም። እዚህ የተለካው ነገር የተሰማቸው speedup ነው፤ ዳታውም ያንን ስሜት ይቃረናል።

ማስረጃው ምን ያህል ጠንካራ ነው?

- የጥናቱ ንድፍ ከተለመደው በላይ ታማኝ ነው። በዘፈቀደ የተመደበ፣ እውነተኛ tasks፣ እውነተኛ repositories፣ እውነተኛ timing። ይህ ብዙ የAI ኮድ መጻፍ አባባሎች ከሚያርፉባቸው self-reports እና መመዘኛ leaderboards የበለጠ ጠንካራ ነው። 19% slowdown በደራሲዎቹ robustness checks ውስጥ ቆይቷል።
- በብዙ ሁኔታ ሊተላለፍ የሚችለው ውጤት ግንዛቤ gap ነው። ባለሙያዎች ስለራሳቸው AI speedup ያላቸው ስሜት በ40 points ገደማ በoptimistic direction ስህተት ነበር። ይህ ስለ ማንኛውም self-reported AI productivity gain ጥንቃቄ ያስፈልጋል የሚል ትምህርት ነው።
- ይህ የአንድ ጊዜ snapshot ነው፣ trend አይደለም። የMETR የFebruary 2026 follow-up በተመሳሳይ ገንቢዎች እና አዲስ መሣሪያዎች ላይ ወደ speedup ይጠቁማል — በጣም በጣም ት —18% ለreturning ገንቢዎች እና —4% ለnew ones — ግን ደራሲዎቹ ይህን ደካማ ማስረጃ ብለው ያቀርባሉ። የመሳተፍ ምርጫ ተዛብቷል፤ ገንቢዎች ያለ AI ለመሥራት እየተቃወሙ መጡ፣ የክፍያ መጠን ቀነሰ፣ የተግባር ምርጫም ተዛባ። ትክክለኛው ንባብ፤ ሁኔታው እየተለወጠ ነው፣ እና ያ ለውጥም በጥንቃቄ ነው የሚዘገበው።

ለምን ይጠቅማል?

ስለ AI እና programming የሚደረገው ክርክር ብዙውን ጊዜ በdemos እና በስሜት ይሄዳል፤ የሚያምር ማያ recording፣ እርግጠኛ አባባል፣ እና ተቃራኒ ተሳልቆ፣ እዚህ ያልተለመደው ነገር አንድ ሰው አሰልፏቸውን ግን ጠቃሚውን መከራ ማድረጉ ነው፤ randomize አድርግ፣ እውነተኛ ሥራን በጊዜ ለካ፣ ከዚያ ሰዎቹ እንዴት እንደተሰማቸው ጠይቅ። መልሱ ለሁለቱም ካምፖች የማይመች ነው። AI በአስቸጋሪና በታወቀ ኮድ ላይ expert ገንቢዎችን ሁልጊዜ እጅግ ያፋጥናል የሚለውን ታሪክ ይቃወማል። በተመሳሳይ ጊዜ “AI ገንቢዎችን ያዘገያል፣ ተረጋግጧል” የሚለውን ቀላል ተቃራኒ ታሪክም ይቃወማል፤ ምክንያቱም የዚያው ቡድን አዲስ ዳታ ወደ ተቃራኒ አቅጣጫ ይጠቁማል። በጣም የሚቆየው ትምህርት ቀላል ነው፤ ሰዎቹ በመለኪያ ዝግ ሲሆኑ ፈጣን እንደሆኑ ተሰማቸው። “ፈጣን ይመስላል” የሚለው ስሜት ማስረጃ አይደለም። ለካው።

ንጹህ ማጠቃለያ

በዘፈቀደ የተመደበ እና ቁጥጥር የተደረገ መከራ፣ METR 16 ልምድ ያላቸው ያልተፈታ-ምንጭ ገንቢዎች በጥሩ ሁኔታ በሚያውቋቸው codebases ላይ 246 እውነተኛ ሥራዎችን እንዲሠሩ አደረገ። በየዘፈቀደ የተመረጠ ግማሽ ሥራ ላይ የ2025 መጀመሪያ AI መሣሪያዎች — Cursor Pro + Claude 3.5/3.7 Sonnet — ተፈቀዱ። ገንቢዎች AI ተግባር ጊዜውን በ24% ገደማ እንደሚቀንስ ጠበቁ፤ በምትኩ completion ጊዜውን በ19% ጨመረ። ከሥራው በኋላ ግን AI በ20% ገደማ እንዳፋጠናቸው አሁንም ያምኑ ነበር። ይህ ግንዛቤ gap የጥናቱ ግልጽና ጠንካራ ዋና ውጤት ነው። ግን ይህ 16 ገንቢዎች፣ አንድ ጠባብ አውድ፣ እና የ2025 መጀመሪያ snapshot ነው። ደራሲዎቹ በግልጽ AI አብዛኞቹን ገንቢዎች አይረዳም ብሎ እንደማያሳይ ይገልጻሉ፣ እና የ2026 follow-up በራሱ ገደቦች ወደ speedup ይጠቁማል። በጥንቃቄ ሊወሰድ የሚገባ መለኪያ ነው — በAI ኮድ መጻፍ ላይ የመጨረሻ ፍርድ አይደለም።

ምንጮች

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/>