



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

Έμπειροι προγραμματιστές ένιωθαν ταχύτεροι με AI ενώ μετρήθηκαν πιο αργοί — το πραγματικό εύρημα και η ετυμηγορία που δεν είναι για το AI coding

Σε τυχαιοποιημένη ελεγχόμενη δοκιμή της METR, δεκαέξι έμπειροι προγραμματιστές ανοιχτού κώδικα εκτέλεσαν 246 πραγματικές εργασίες σε βάσεις κώδικα που γνώριζαν καλά, με εργαλεία AI των αρχών του 2025 (Cursor Pro μαζί με Claude 3.5/3.7 Sonnet) να επιτρέπονται τυχαία στις μισές. Περίμεναν ότι η AI θα μείωνε τον χρόνο κατά περίπου 24%· αντί γι' αυτό, αύξησε τον χρόνο ολοκλήρωσης κατά 19% — και παρ' όλα αυτά μετά πίστευαν ακόμη ότι τους είχε επιταχύνει κατά περίπου 20%. Αυτό το χάσμα αντίληψης είναι ο πιο καθαρός και ανθεκτικός πυρήνας της μελέτης. Όμως πρόκειται για δεκαέξι προγραμματιστές, ένα στενό πλαίσιο και ένα στιγμιότυπο των αρχών του 2025: οι συγγραφείς τονίζουν ότι δεν δείχνει πως η AI δεν βοηθά τους περισσότερους προγραμματιστές, και η δική τους συνέχεια του 2026 ήδη γέρνει προς επιτάχυνση, με τις δικές της επιφυλάξεις.

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

Έμπειροι προγραμματιστές ένιωθαν ταχύτεροι με AI ενώ μετρήθηκαν πιο αργοί — το πραγματικό εύρημα και η ετυμολογία που δεν είναι για το AI coding

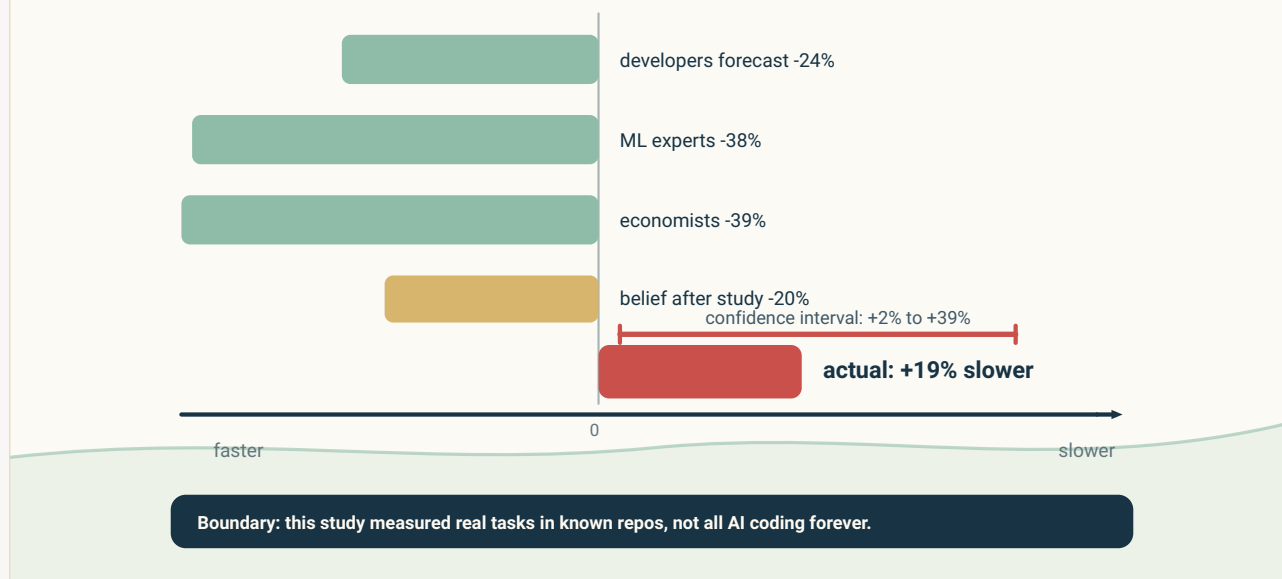
Ρωτήστε έναν έμπειρο προγραμματιστή αν ένας βοηθός προγραμματισμού με AI τον κάνει γρηγορότερο και συνήθως θα σας δώσει έναν αριθμό: μου γλιτώνει είκοσι, τριάντα τοις εκατό. Ρωτήστε έναν οικονομολόγο ή έναν ερευνητή μηχανικής μάθησης και ο αριθμός μεγαλώνει. Στις αρχές του 2025, μια ομάδα στη METR έκανε το αργό και ακριβό πράγμα: το μέτρησε. Πήρε δεκαέξι έμπειρους προγραμματιστές ανοιχτού κώδικα, τους έδωσε 246 πραγματικές εργασίες από μεγάλες βάσεις κώδικα που γνώριζαν καλά και — με τυχαία ανάθεση, εργασία προς εργασία — είτε τους επέτρεπε να χρησιμοποιούν εργαλεία AI είτε όχι. Ύστερα χρονομέτρησε τη δουλειά.

Οι προγραμματιστές είχαν προβλέψει ότι η AI θα μείωνε τον χρόνο των εργασιών τους κατά περίπου 24%. Συνέβη το αντίθετο: οι εργασίες με AI χρειάστηκαν **19% περισσότερο χρόνο**. Και εδώ είναι το σημείο που αξίζει να σταθεί κανείς — όταν τελείωσαν, οι ίδιοι προγραμματιστές εξακολουθούσαν να πιστεύουν ότι η AI τους είχε επιταχύνει, κατά περίπου 20%. Ήταν πιο αργοί και ένιωθαν πιο γρήγοροι· η απόσταση ανάμεσα σε αυτούς τους δύο αριθμούς είναι το πιο ενδιαφέρον εύρημα της μελέτης.

Είναι ένα πραγματικό, προσεκτικά μετρημένο αποτέλεσμα. Είναι επίσης δεκαέξι προγραμματιστές, σε αποθετήρια που γνωρίζουν σε βάθος, με εργαλεία των αρχών του 2025 — και δεν ισοδυναμεί με τη γενική πρόταση «η AI κάνει τους προγραμματιστές πιο αργούς». Τα μεταγενέστερα δεδομένα της ίδιας ομάδας ήδη δείχνουν προς την αντίθετη κατεύθυνση.

Everyone expected speed. The timed tasks got slower.

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



Όλοι προέβλεψαν ότι η AI θα επιτάχυνε τη δουλειά — προγραμματιστές −24%, ειδικοί ML −38%, οικονομολόγοι −39% και η εκ των υστέρων πεποίθηση των ιδίων των προγραμματιστών −20%. Το χρονόμετρο βρήκε +19% πιο αργά, με διάστημα +2% έως +39%. Το χάσμα ανάμεσα στο αισθητό και στο μετρημένο είναι το εύρημα.

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.

Τι μετρά η μελέτη — 16 έμπειρους προγραμματιστές ανοιχτού κώδικα, 246 πραγματικές εργασίες, γνώριμα αποθετήρια, εργαλεία των αρχών του 2025, τυχαιοποίηση — και τι δεν μετρά: όχι τους περισσότερους προγραμματιστές, όχι άλλους τομείς, όχι έναν αμετάβλητο νόμο, όχι αξιολόγηση από ομοτίμους.

Original diagram — The Clean Paper CC BY 4.0

Τι μέτρησε αυτό και τι κερδίζουμε από μια τυχαιοποιημένη δοκιμή

Μια **τυχαιοποιημένη ελεγχόμενη δοκιμή (RCT)** είναι το εργαλείο που χρησιμοποιεί η ιατρική για να ξεχωρίσει ένα πραγματικό αποτέλεσμα από ένα ελπιδοφόρο φαινόμενο. Εδώ, καθεμία από τις 246 εργασίες ανατέθηκε τυχαία να γίνει είτε με AI είτε χωρίς, έτσι ώστε κατά μέσο όρο η μόνη συστηματική διαφορά ανάμεσα στις δύο ομάδες να είναι η ίδια η AI. Αυτό επιτρέπει να πούμε ότι η AI προκάλεσε τη μεταβολή στον χρόνο, αντί να παρατηρούμε απλώς ότι όσοι καταφεύγουν σε AI τυχαίνει για άλλους λόγους να είναι γρηγορότεροι ή πιο αργοί. Έχει σημασία επειδή τα συνήθη στοιχεία για τα οφέλη του AI coding — αυτοαναφορές και βαθμολογίες σε benchmarks — δεν μπορούν να το κάνουν: ένα benchmark δεν είναι πραγματική εργασία και μια αυτοαναφορά, όπως δείχνει αυτή η μελέτη, μπορεί να είναι με αυτοπεποίθηση λανθασμένη. Οι προγραμματιστές εδώ δεν ήταν αρχάριοι που πάλευαν με ένα καινούργιο παιχνίδι. Ήταν καθιερωμένοι συνεισφέροντες σε μεγάλα, ώριμα έργα ανοιχτού κώδικα που γνώριζαν καλά και είχαν ήδη κάποια εμπειρία με τα εργαλεία.

Τι έκαναν οι συγγραφείς

- Πραγματοποίησαν μια **τυχαιοποιημένη ελεγχόμενη δοκιμή (METR)**: Joel Becker, Nate Rush, Beth Barnes, David Rein). Δεκάξι έμπειροι προγραμματιστές ανοιχτού κώδικα, ο καθένας σε μεγάλο αποθετήριο στο οποίο συνεισφέρει τακτικά και γνωρίζει καλά — κατά μέσο όρο πέντε χρόνια στα συγκεκριμένα έργα.
- Χρησιμοποίησαν **246 πραγματικές εργασίες** — διορθώσεις σφαλμάτων, νέες λειτουργίες και refactoring από τα ίδια τα συστήματα παρακολούθησης θεμάτων των έργων. Κάθε εργασία ανατέθηκε τυχαία σε «AI επιτρέπεται» ή «AI δεν επιτρέπεται».
- «AI επιτρέπεται» σήμαινε **εργαλεία των αρχών του 2025: Cursor Pro με Claude 3.5/3.7 Sonnet**. Η κύρια μέτρηση ήταν ο πραγματικός χρόνος ολοκλήρωσης κάθε εργασίας. Παράλληλα, η ομάδα συνέλεξε προβλέψεις (από τους προγραμματιστές πριν) και εκτιμήσεις (από τους ίδιους μετά), καθώς και προβλέψεις από ειδικούς οικονομικών και μηχανικής μάθησης.

Τι βρήκαν

- **Με AI, οι εργασίες χρειάστηκαν 19% περισσότερο χρόνο.** Όχι λιγότερο — περισσότερο. Το διάστημα εμπιστοσύνης 95% εκτείνεται περίπου από +2% έως +39%, οπότε η κατεύθυνση του αποτελέσματος είναι σταθερή ακόμη κι αν το ακριβές μέγεθος έχει αβεβαιότητα.
- **Όλοι είχαν προβλέψει το αντίθετο.** Οι προγραμματιστές προέβλεψαν επιτάχυνση 24%, οι ειδικοί μηχανικής μάθησης περίπου 38% και οι οικονομολόγοι περίπου 39%. Και οι τρεις ομάδες περίμεναν ότι η AI θα εξοικονομούσε πολύ χρόνο· το χρονόμετρο έδειξε ότι κόστισε χρόνο.
- **Το χάσμα αντίληψης.** Αφού έκαναν τη δουλειά και βγήκαν πιο αργοί, οι προγραμματιστές εξακολουθούσαν να εκτιμούν ότι η AI τους είχε επιταχύνει κατά περίπου 20% — διαφορά περίπου 40 ποσοστιαίων μονάδων ανάμεσα σε αυτό που ένιωθαν και σε αυτό που κατέγραψε το ρολόι.
- **Πιθανές αιτίες, σταθμισμένες αλλά όχι αποδεδειγμένες.** Οι συγγραφείς παραθέτουν παράγοντες που θα μπορούσαν να εξηγούν την επιβράδυνση: αυτοί οι προγραμματιστές γνωρίζουν πολύ βαθιά τις δικές τους βάσεις κώδικα, οπότε ένας βοηθός έχει λιγότερα να προσθέσει· ώριμα έργα έχουν υψηλά, συχνά άρρητα πρότυπα ποιότητας· τα αποθετήρια είναι μεγάλα και γεμάτα συμφραζόμενα που το μοντέλο δεν έχει· και πραγματικός χρόνος ξοδεύεται στη διατύπωση prompts, έπειτα στον έλεγχο και στη διόρθωση της εξόδου της AI. Τα παρουσιάζουν ως πιθανές εξηγήσεις, όχι ως τελικές αποδείξεις.

Τι δεν δείχνει αυτό

- **Δεν δείχνει ότι η AI αποτυγχάνει να επιταχύνει τους περισσότερους προγραμματιστές.** Οι συγγραφείς το λένε ρητά: δεκαέξι ειδικοί σε κώδικα που γνωρίζουν απ' έξω δεν είναι ο μέσος προγραμματιστής στη μέση εργασία.
- **Δεν δείχνει ότι η AI είναι άχρηστη ή ότι επιβραδύνει τους ανθρώπους σε άλλα περιβάλλοντα** — νεοεισερχόμενους σε μια βάση κώδικα, έργα από το μηδέν, άγνωστες γλώσσες ή εντελώς διαφορετικούς τομείς.
- **Δεν παγώνει τα εργαλεία στον χρόνο.** Πρόκειται για Cursor και Claude 3.5/3.7 Sonnet των αρχών του 2025· οι συγγραφείς τονίζουν ότι καλύτερα εργαλεία ή καλύτεροι τρόποι χρήσης των ίδιων εργαλείων θα μπορούσαν να αλλάξουν το αποτέλεσμα ακόμη και στο ίδιο ακριβώς πλαίσιο.
- Είναι **preprint** (αναρτήθηκε τον Ιούλιο του 2025, χωρίς ακόμη αξιολόγηση από ομοτίμους), και οι συγγραφείς σημειώνουν ότι δεν μπορούν να αποκλείσουν πλήρως πειραματικές τεχνικές επιδράσεις — παρότι το αποτέλεσμα άντεξε στις αναλύσεις τους.
- **Δεν δικαιολογεί την καθυστερημένη ερμηνεία** ότι οι προγραμματιστές σίγουρα κέρδισαν κάπου αλλού — έμαθαν περισσότερα, ένιωσαν καλύτερα, έγραψαν καλύτερο κώδικα. Το μόνο σχετικό πράγμα που μετρήθηκε εδώ, η αισθητή επιτάχυνση, είναι ακριβώς αυτό που τα δεδομένα διαψεύδουν.

Πόσο ισχυρά είναι τα στοιχεία

- **Ο σχεδιασμός είναι ασυνήθιστα ειλικρινής.** Τυχαιοποίηση, πραγματικές εργασίες, πραγματικά αποθετήρια, πραγματική χρονομέτρηση — μεγάλο βήμα πάνω από τις αυτοαναφορές και τους πίνακες κατάταξης benchmarks στους οποίους βασίζονται οι περισσότερες αξιώσεις για AI coding. Η επιβράδυνση 19% άντεξε στους ελέγχους ευρωστίας των συγγραφέων.
- **Το πιο γενικεύσιμο εύρημα είναι το χάσμα αντίληψης.** Η διαίσθηση ειδικών για τη δική τους επιτάχυνση με AI έκανε λάθος κατά περίπου 40 μονάδες, προς την αισιόδοξη κατεύθυνση. Αυτό είναι προειδοποίηση για κάθε αυτοαναφερόμενο κέρδος παραγωγικότητας από AI — συμπεριλαμβανομένων των αριθμών αυτής της ίδιας μελέτης.
- **Είναι στιγμιότυπο, όχι τάση.** Η ίδια η συνέχεια της METR τον Φεβρουάριο του 2026, με παρόμοιους προγραμματιστές και νεότερα εργαλεία, δείχνει προς επιτάχυνση — πολύ χονδρικά −18% για όσους επέστρεψαν και −4% για νέους συμμετέχοντες — αλλά οι συγγραφείς τη χαρακτηρίζουν ασθενή ένδειξη, επηρεασμένη από το ποιοι δέχτηκαν να συμμετάσχουν (οι προγραμματιστές αρνούνταν όλο και περισσότερο να εργαστούν χωρίς AI, η αμοιβή μειώθηκε και η επιλογή εργασιών μετατοπίστηκε). Η έντιμη ανάγνωση είναι ότι η εικόνα αλλάζει και ακόμη και αυτή η αλλαγή αναφέρεται χωρίς να σπρώχνονται τα δεδομένα προς μια βολική κατεύθυνση.

Γιατί έχει σημασία

Το μεγαλύτερο μέρος της διαμάχης για την AI και τον προγραμματισμό βασίζεται σε demos και εντυπώσεις: μια εντυπωσιακή καταγραφή οθόνης, ένας βέβαιος ισχυρισμός, ένα ειρωνικό σήκωμα των ματιών από την άλλη πλευρά. Αυτό που είναι σπάνιο — και κάνει τη μελέτη αξία ανάγνωσης — είναι ότι κάποιος έκανε το βαρετό πείραμα: τυχαιοποίησε, χρονομέτρησε πραγματική εργασία και μετά ρώτησε τους ανθρώπους πώς πήγε. Η απάντηση είναι άβολη και για τα δύο στρατόπεδα. Τρυπά την ιστορία ότι η AI κάνει ομοιόμορφα τους έμπειρους προγραμματιστές υπερταχείς σε δύσκολο, γνώριμο κώδικα. Τρυπά και το τακτοποιημένο αντίθετο — «αποδείχθηκε ότι η AI κάνει τους προγραμματιστές πιο αργούς» — επειδή τα νεότερα δεδομένα της ίδιας ομάδας ήδη γέρνουν προς την άλλη κατεύθυνση. Το πιο ανθεκτικό μάθημα είναι και το μικρότερο και πιο ανθρώπινο: όσοι έκαναν τη δουλειά ένιωθαν ταχύτεροι ενώ μετρήθηκαν πιο αργοί. Το «μου φαίνεται πιο γρήγορο» δεν είναι απόδειξη ότι είναι. Μετρήστε το.

Καθαρή σύνοψη

Σε μια τυχαιοποιημένη ελεγχόμενη δοκιμή, η METR έβαλε δεκάξι έμπειρους προγραμματιστές ανοιχτού κώδικα να ολοκληρώσουν 246 πραγματικές εργασίες σε βάσεις κώδικα που γνώριζαν καλά, με εργαλεία AI των αρχών του 2025 (Cursor Pro μαζί με Claude 3.5/3.7

Sonnet) να επιτρέπονται τυχαία στις μισές. Οι προγραμματιστές περίμεναν ότι η AI θα μείωνε τον χρόνο τους κατά περίπου 24%. Αντί γι' αυτό αύξησε τον χρόνο ολοκλήρωσης κατά 19% — και μετά εξακολουθούσαν να πιστεύουν ότι τους είχε επιταχύνει κατά περίπου 20%. Αυτό το χάσμα αντίληψης είναι ο καθαρός και ισχυρός πυρήνας της μελέτης. Αλλά πρόκειται για δεκαέξι προγραμματιστές, ένα στενό πλαίσιο και ένα σταθερό στιγμιότυπο των αρχών του 2025· οι συγγραφείς λένε ρητά ότι δεν αποδεικνύει πως η AI δεν βοηθά τους περισσότερους προγραμματιστές, και η δική τους συνέχεια του 2026 ήδη δείχνει προς επιτάχυνση, με τις δικές της επιφυλάξεις. Μια προσεκτική μέτρηση που αξίζει να ληφθεί σοβαρά — όχι ετυμηγορία για το AI coding.

Πηγές

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/>