



The CLEAN PAPER

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

Deneyimli geliştiriciler AI ile daha hızlı hissetti ama ölçüldüğünde daha yavaş çalıştı — gerçek bulgu ve bunun AI kodlama hakkında olmadığı hüküm

METR'nin randomize kontrollü çalışmasında on altı deneyimli açık kaynak geliştiricisi, iyi bildikleri kod tabanlarında 246 gerçek görev yaptı; görevlerin rastgele yarısında 2025 başı AI araçlarına (Cursor Pro + Claude 3.5/3.7 Sonnet) izin verildi. AI'nın görev süresini yaklaşık %24 azaltmasını beklediler; bunun yerine süre %19 arttı — ve sonrasında bile AI'nın kendilerini yaklaşık %20 hızlandırdığına inanıyorlardı. Bu algı boşluğu çalışmanın keskin ve sağlam çekirdeği. Ama bu on altı geliştirici, dar bir ortam ve 2025 başına ait sabit bir anlık görüntü: yazarlar bunun AI'nın geliştiricilerin çoğuna yardım etmediğini göstermediğini açıkça söylüyor ve kendi 2026 takipleri de, kendi uyarılarıyla birlikte, şimdiden hızlanma yönüne eğiliyor.

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

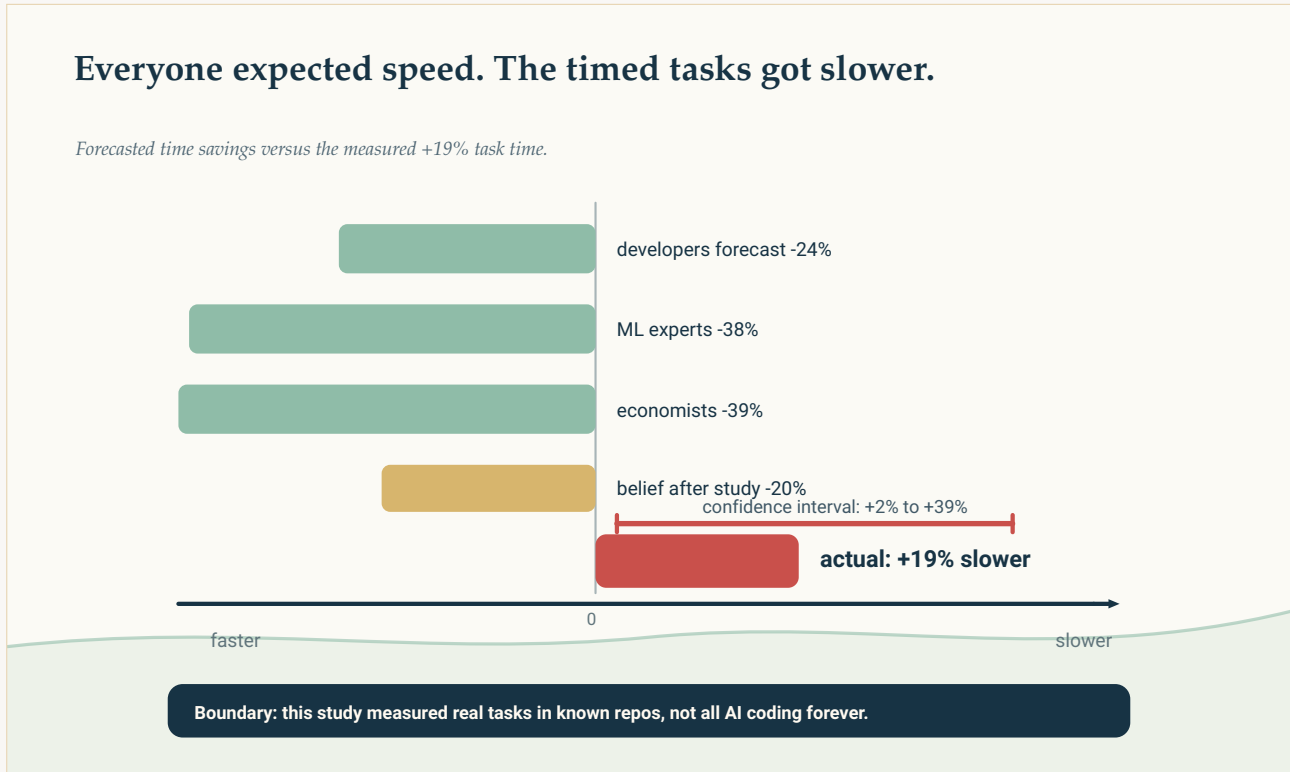
Paper: *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*, Joel Becker, Nate Rush, Beth Barnes, David Rein (METR), arXiv:2507.09089 [cs.AI] (preprint) · DOI: 10.48550/arXiv.2507.09089

Deneyimli geliştiriciler yapay zekâ ile daha hızlı hissetti, ölçüldüğünde daha yavaş çalıştı — gerçek bulgu ve bunun AI kodlama hakkında olmadığı hüküm

Deneyimli bir programcıya yapay zekâ kodlama asistanının onu hızlandırıp hızlandırmadığını sorun, genellikle bir sayı verir: bana yüzde yirmi, otuz kazandırıyor. Bir ekonomiste veya makine öğrenmesi araştırmacısına sorun, sayı daha da büyür. 2025 başında METR ekibi yavaş ve pahalı olan şeyi yaptı: kontrol etti. İyi bildikleri büyük kod tabanlarında çalışan on altı deneyimli açık kaynak geliştiricisine 246 gerçek görev verdi ve her görevi yazı-tura gibi rastgele biçimde ya AI araçlarına izin verilen ya da verilmeyen koşula atadı. Sonra işi kronometreyle ölçtü.

Geliştiriciler AI'nın görev süresini yaklaşık %24 azaltacağını tahmin etmişti. Tersi oldu: AI ile yapılan görevler **%19 daha uzun** sürdü. Ve üzerinde durulmaya değer kısım şu — iş bittikten sonra aynı geliştiriciler AI'nın onları yaklaşık %20 hızlandırdığına hâlâ inanıyordu. Daha yavaşlardı ama daha hızlı hissettiler; bu iki sayı arasındaki mesafe çalışmanın en ilginç bulgusu.

Bu gerçek, dikkatle ölçülmüş bir sonuç. Aynı zamanda iyi bildikleri depolarda, 2025 başındaki araçlarla çalışan on altı geliştiriciden ibaret — ve “AI geliştiricileri yavaşlatır” diye düz bir cümle değil. Aynı ekibin daha sonraki verileri şimdiden diğer yöne işaret ediyor.



Herkes AI'nın işi hızlandıracağını tahmin etti — geliştiriciler — %24, ML uzmanları — %38, ekonomistler — %39 ve geliştiricilerin iş bittikten sonraki kendi inancı — %20. Kronometre +%19 daha yavaş buldu; aralık +%2 ile +%39. Hissedilen ile ölçülen arasındaki boşluk bulgudur.

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.

Çalışmanın ölçtüğü: 16 uzman açık kaynak geliştiricisi, 246 gerçek görev, tanıdık depolar, 2025 başı araçları, rastgele atama. Ölçmediği: geliştiricilerin çoğu, diğer alanlar, değişmez bir yasa veya hakemli bir sonuç.

Original diagram — The Clean Paper CC BY 4.0

Bu neyi ölçtü ve randomize çalışma size ne kazandırır?

Randomize kontrollü çalışma (RCT), tıbbın gerçek bir etkiyi umut edilen etkiden ayırmak için kullandığı araçtır. Burada 246 görevin her biri rastgele AI'a izin verilen veya verilmeyen koşula atandı; böylece ortalama olarak iki yığın arasındaki tek sistematik fark AI'ın kendisi oluyor. Bu, AI'ı kullanan kişilerin başka nedenlerle tesadüfen hızlı veya yavaş olduğunu görmek yerine, zaman değişimine AI'ın *neden olduğunu* söylemenizi sağlar. Önemli, çünkü AI kodlama kazançları için olağan kanıtlar — kişisel bildirimler ve benchmark skorları — bunu yapamaz: benchmark gerçek iş değildir; öz bildirim ise bu çalışmanın gösterdiği gibi kendinden emin biçimde yanlış olabilir. Buradaki geliştiriciler yeni bir oyuncakla boğuşan acemiler değildi. İyi bildikleri büyük ve uygun açık kaynak projelerine düzenli katkı yapan, araçlarla önceden biraz deneyimli kişilerdi.

Yazarlar ne yaptı?

- Bir **randomize kontrollü çalışma** yürüttüler (METR: Joel Becker, Nate Rush, Beth Barnes, David Rein). On altı deneyimli açık kaynak geliştiricisinin her biri düzenli katkı verdiği ve iyi bildiği büyük bir depoda çalıştı — belirli projelerde ortalama beş yıl deneyim.
- Projelerin kendi issue tracker'larından alınan **246 gerçek görev** kullandılar — hata düzeltmeleri,

özellikler ve refactor'lar. Her görev rastgele "AI serbest" veya "AI yasak" koşuluna atandı.

- "AI serbest", **2025 başı araçları: Claude 3.5/3.7 Sonnet ile Cursor Pro** anlamına geliyordu. Ana ölçüm görev başına gerçek tamamlanma süresiydi. Ekip ayrıca geliştiricilerin önceden tahminlerini ve sonradan değerlendirmelerini, ayrıca ekonomi ve makine öğrenmesi uzmanlarının tahminlerini topladı.

Ne buldular?

- **AI ile görevler %19 daha uzun sürdü.** Daha hızlı değil — daha yavaş. %95 güven aralığı yaklaşık +%2 ile +%39 arasında; dolayısıyla kesin büyüklük belirsiz olsa da yön sağlam.
- **Herkes tersini tahmin etmişti.** Geliştiriciler %24 hızlanma, makine öğrenmesi uzmanları yaklaşık %38, ekonomistler yaklaşık %39 öngördü. Üç grup da AI'nın büyük zaman kazandırmasını bekledi; kronometre zaman maliyeti buldu.
- **Algı boşluğu.** İşleri AI ile daha yavaş tamamladıktan sonra geliştiriciler yine de AI'nın kendilerini yaklaşık %20 hızlandırdığını tahmin etti — hissettikleri ile saatin kaydettiği arasında yaklaşık 40 puanlık fark.
- **Olası nedenler, tartılıyor ama kanıtlanmıyor.** Yazarlar yavaşlamayı açıklayabilecek faktörleri sıralıyor: bu geliştiriciler kendi kod tabanlarını çok iyi bildiği için asistanın ekleyebileceği daha az şey var; olgun projelerin yüksek ve çoğu zaman örtük kalite standartları var; depolar büyük ve modelin sahip olmadığı bağlamla dolu; gerçek zaman istem yazmaya, ardından AI çıktısını inceleyip düzeltmeye gidiyor. Bunları hüküm değil, aday açıklamalar olarak sunuyorlar.

Bunun göstermediği şeyler

- AI'nın geliştiricilerin çoğunu hızlandırmadığını **göstermez.** Yazarlar doğrudan söylüyor: ezbere bildikleri kodda çalışan on altı uzman, ortalama görevdeki ortalama geliştirici değildir.
- AI'nın yararsız olduğunu veya başka ortamlarda — kod tabanına yeni gelenlerde, greenfield işte, tanıdık olmayan dillerde veya bütünüyle başka alanlarda — insanları yavaşlattığını **göstermez.**
- Araçları zamana **dondurmaz.** Bunlar 2025 başındaki Cursor ve Claude 3.5/3.7 Sonnet; yazarlar daha iyi araçların veya mevcutları daha iyi kullanma yollarının bu kesin ortamda bile sonucu değiştirebileceğini açıkça belirtiyor.
- Bu bir **preprint** (Temmuz 2025'te yayımlandı, henüz hakemli değil) ve yazarlar deneysel artefaktları tamamen dışlayamadıklarını söylüyor — bulgu analizleri boyunca dayansa da.
- Geliştiricilerin mutlaka başka bir yoldan kazanç sağladığı — daha çok öğrendiği, daha mutlu olduğu, daha iyi kod yazdığı — rahatlatıcı yorumuna **izin vermez.** Burada ölçülen şeylerden biri olan hissedilen hızlanma, verinin tam olarak çeliştiği şey.

Kanıt ne kadar güçlü?

- **Tasarım alışılmadık derecede dürüst.** Rastgele atama, gerçek görevler, gerçek depolar, gerçek süre ölçümü — AI kodlama hakkındaki iddiaların çoğunun dayandığı özbildirim ve benchmark liderlik tablolarına göre büyük ilerleme. %19 yavaşlama yazarların sağlamlık kontrollerinden geçti.
- **En taşınabilir bulgu algı boşluğu.** Uzmanların kendi AI hızlanmalarına ilişkin sezgisi iyimser yönde yaklaşık 40 puan yanlıştı. Bu, AI'dan bildirilen *her* özbildirimsel verimlilik kazancına — bu çalışmanın kendi sayıları dahil — karşı bir uyarıdır.
- **Bir anlık görüntü, trend değil.** METR'nin aynı tür geliştiricilerle daha yeni araçlar kullanan Şubat 2026 takibi hızlanma yönüne işaret ediyor — geri dönen geliştiricilerde çok kabaca —%18, yenilerde —%4 — ama yazarlar bunu zayıf kanıt olarak işaretliyor; kimlerin katılmaya razı olduğu sonucu bozuyor (geliştiriciler AI olmadan çalışmayı giderek daha çok reddetti, ödeme oranı düştü ve görev seçimi kaydı). Dürüst okuma: tablo hareket ediyor ve bu hareket bile teraziye parmak koymadan raporlanıyor.

Neden önemli?

AI ve programlama tartışmasının çoğu demo ve hislerle dönüyor: şık bir ekran kaydı, kendinden emin bir iddia, karşıdan bir göz devirmesi. Nadir olan ve bunu okumaya değer kılan, birilerinin sıkıcı deneyi yapması: rastgele ata, gerçek işi zamanla, sonra insanlara nasıl gittiğini sor. Cevap iki kamp için de rahatsız edici. AI'nın zor, tanıdık kodda uzman geliştiricileri tekdüze biçimde turbo hızlandırdığı hikâyeyi deliyor. Aynı zamanda düzenli ters sloganı da — “AI geliştiricileri yavaşlatır, kanıtlandı” — deliyor; çünkü aynı ekibin yeni verisi şimdiden öteki yöne eğiliyor. En kalıcı ders aynı zamanda en küçük ve en insani olanı: iş yapanlar ölçülebilir biçimde daha yavaşken daha hızlı hissetti. “Daha hızlı geliyor” onun hızlı olduğuna kanıt değildir. Ölçün.

Kısa özet

METR randomize kontrollü bir çalışmada, on altı deneyimli açık kaynak geliştiricisine iyi bildikleri kod tabanlarında 246 gerçek görev yaptırdı; görevlerin rastgele yarısında 2025 başı AI araçlarına (Cursor Pro + Claude 3.5/3.7 Sonnet) izin verdi. Geliştiriciler AI'nın görev süresini yaklaşık %24 azaltmasını bekliyordu; bunun yerine tamamlanma süresini %19 artırdı — ve iş bittikten sonra bile AI'nın kendilerini yaklaşık %20 hızlandırdığına inanıyorlardı. Bu algı boşluğu çalışmanın keskin ve sağlam çekirdeği. Ama bu on altı geliştirici, dar bir ortam ve 2025 başına ait sabit bir anlık görüntü; yazarlar bunun AI'nın geliştiricilerin çoğuna yardım etmediğini göstermediğini açıkça belirtiyor ve kendi 2026 takipleri de, kendi uyarılarıyla birlikte, şimdiden hızlanma yönüne işaret ediyor. Ciddiye alınmaya değer dikkatli bir ölçüm — AI kodlama hakkında bir hüküm değil.

Kaynaklar

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/>