AI Dev Live · 11 November 2026

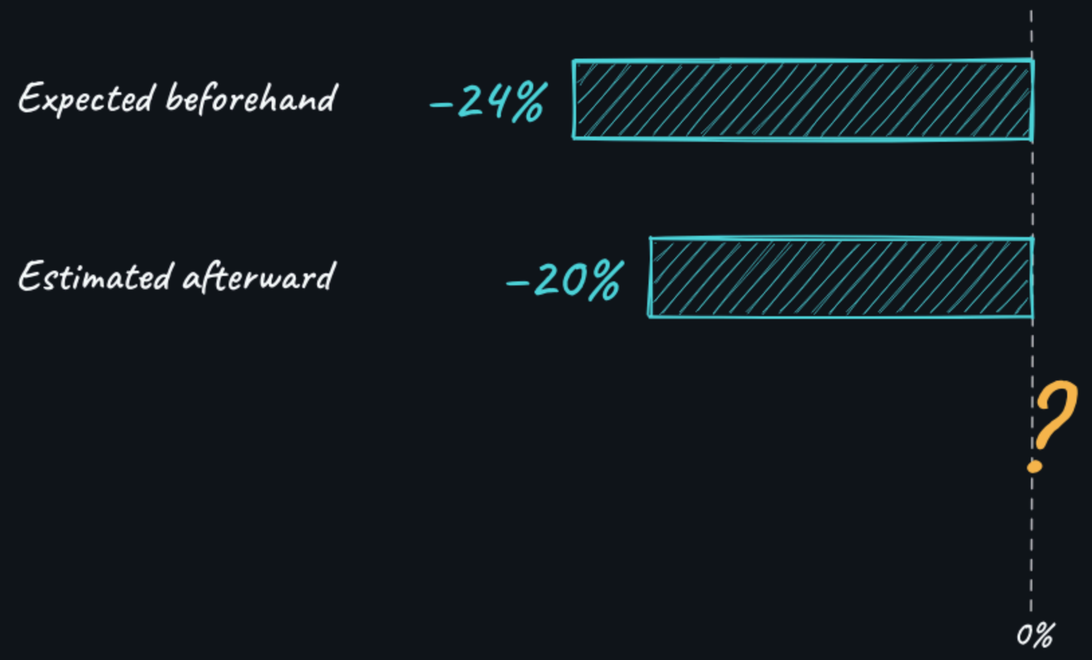
AI made development faster



How can we tell if that is true?

What did the measurement find?

METR · change in completion time

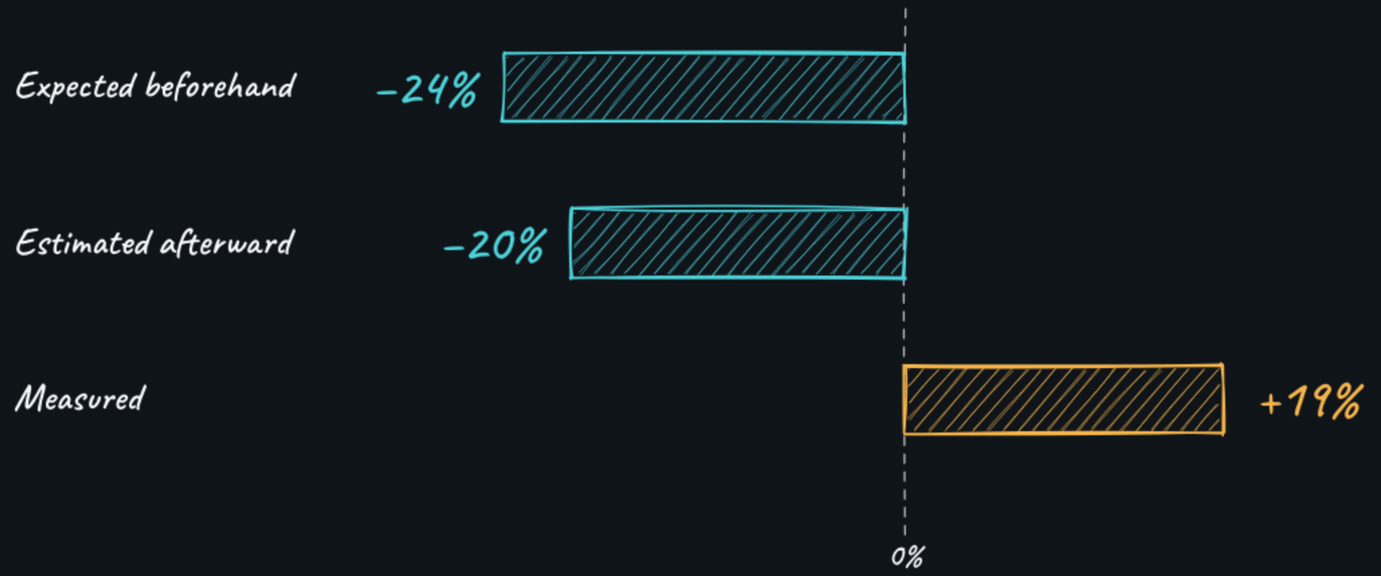


Expectation and self-report. Measurement comes next.

METR · Becker et al. (2025)

Perceived speedup needs a separate check

METR · change in completion time



All three numbers describe time; negative means faster

METR · Becker et al. (2025)

METR's result comes from familiar repositories

METR · February–June 2025

+19%

completion time

95% CI: +2…+39%

16

experienced developers

246

real tasks

Cursor Pro · Claude 3.5/3.7 Sonnet · familiar projects

METR · Becker et al. (2025)

CLI agents: more merged PRs

Microsoft · January–April 2026

Counts PRs, not time or quality

+24%

merged PRs per engineer per day

95% interval: +14.5…+33.7%

−1.1%

at a placebo launch date

interval: −10.6…+8.6%

115

days after launch

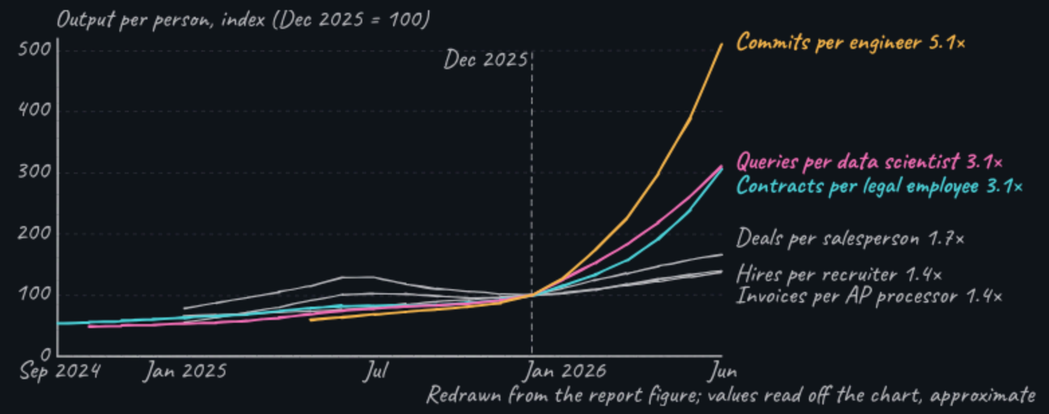
Claude Code and Copilot CLI · early adopters

Microsoft · Murphy-Hill et al. (2026)

Commits per engineer rose at Anthropic

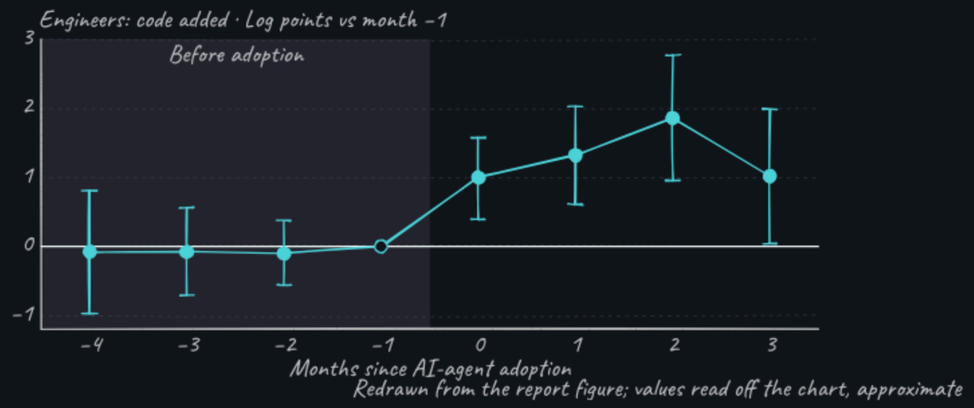
Anthropic · October 2026

Fig. 3 · Output per employee



By June 2026, commits per engineer reached 5.1 times their December 2025 level.

Fig. 4 · Code added after adoption



192 engineers adopted agents and 74 did not. Adoption was nonrandom.

Anthropic · employment report (2026) My review · RIMS #40

Each result applies to its own conditions

Compare conditions, not headlines

METR	Microsoft	Anthropic
Time per task	Merged PRs per day	Output per employee
Own repository	Azure DevOps PRs	Multiple functions
Cursor · 2025	CLI agents · 2026	Agents · 2026

Transfer the check, not the percentage

METR · Becker et al. (2025) Microsoft · Murphy-Hill et al. (2026) Anthropic · employment report (2026)

Architecture · teams · AI4SDLC

Alexander Polomodov

Technology leader with experience as a CTO and Technical Fellow at a large fintech



polomodov.tech

Research Insights Made Simple

Connecting research with engineering practice.

Focus: Productivity · architecture · AI4SDLC

Two levels of measurement

Discovery + delivery

company level
customer value

Delivery only

company level
usage, impact, cost

Use-case economics

review, migrations,
prototypes

Choose the scope before choosing metrics

Company-wide measurement

1A

Discovery + delivery · the path to a customer outcome

From hypothesis to customer value

Company · discovery + delivery



Link the hypothesis, release and outcome for each initiative

How to measure the end-to-end outcome

Company · illustrative activation example

Activation

share of new customers first success in 7 days

End-to-end time

from hypothesis to verified outcome

Total cost

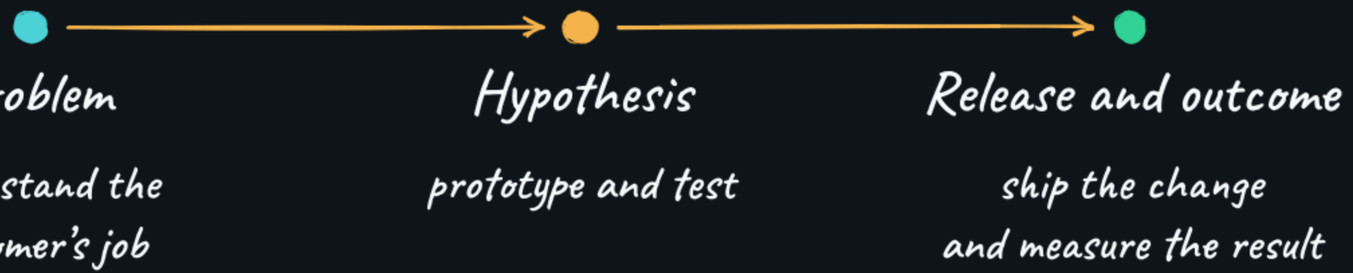
team effort and AI including failed attempts

Compare similar initiatives and customer cohorts

Code of Leadership · conversation with Vladimir Dolgov

Product engineers own work end to end

Company · discovery + delivery



Back to the basics of product development

Code of Leadership #76 · Gleb Mikheev My talk · Software development evolution

A tiny team owns the whole cycle

Handoffs

handoff count
and waiting time

Time to outcome

from framing the task
to customer impact

Sustainability

production incidents
and team workload

Team size depends on the product, risk and platform support

My talk · HSE University My talk · Software development evolution

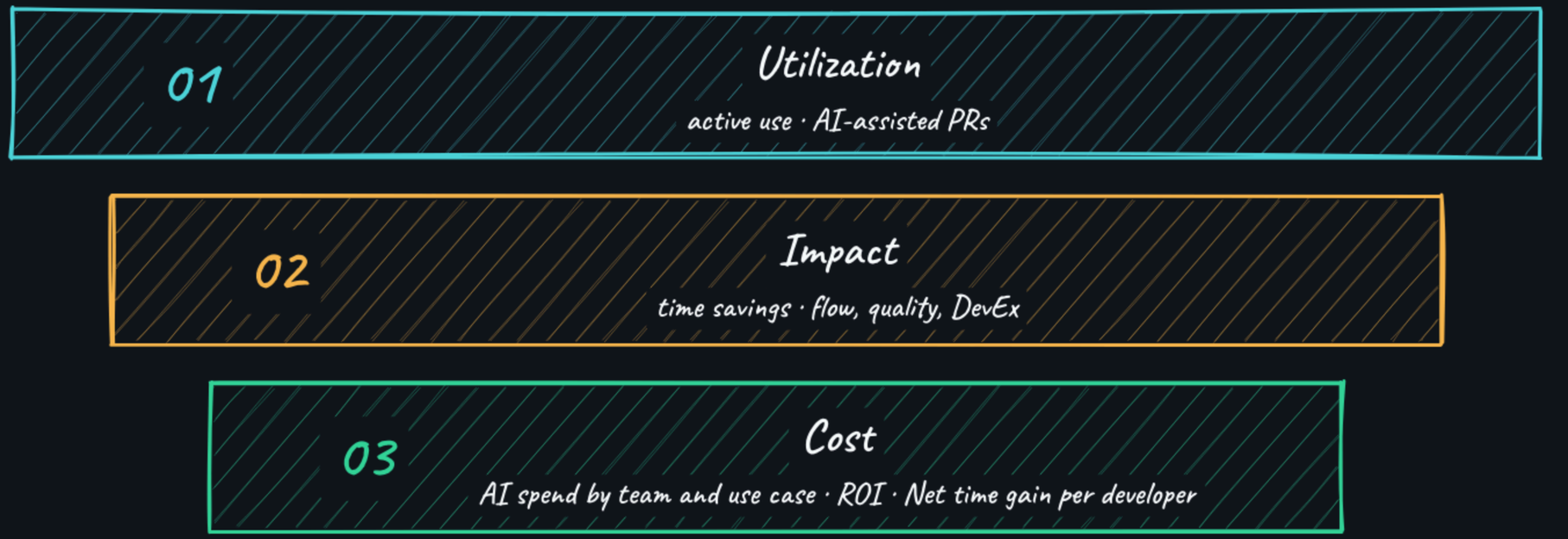
AI Dev Live · 2026

Company-wide delivery

1B

Measuring AI code assistants and agents · DX and RIMS #11

How DX measures AI in delivery



Read AI metrics alongside overall team productivity

DX · Measuring AI code assistants and agents RIMS #11 My talk · DotNext 2026

Measure adoption in the context of work

Delivery · utilization

Who	Where	How
team · experience	job · repository	frequency · mode

A licence does not describe a working practice

My talk · DotNext 2026

An adoption rate needs a denominator

Define the population

People

used / had access

Tasks

AI-assisted / eligible

Set the observation window and inclusion rules in advance

Assistance mode changes the work

Assistant

human leads the task
AI assists with operations

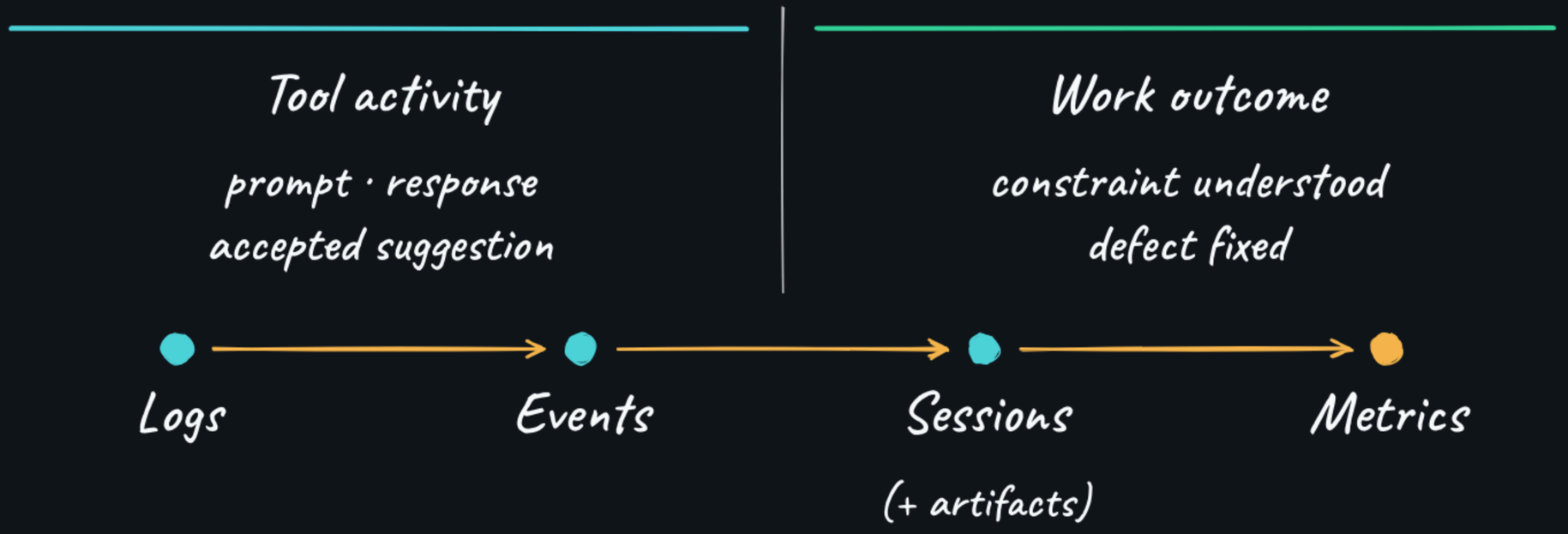
Delegation

agent executes the task
human defines and accepts

Concurrency is a separate dimension of either mode

My talk · DotNext 2026

A tool event is not an outcome



Check the outcome against the task's success criterion

InSession (2020) Google Research · Measuring Developer Goals My review of the series · RIMS #30

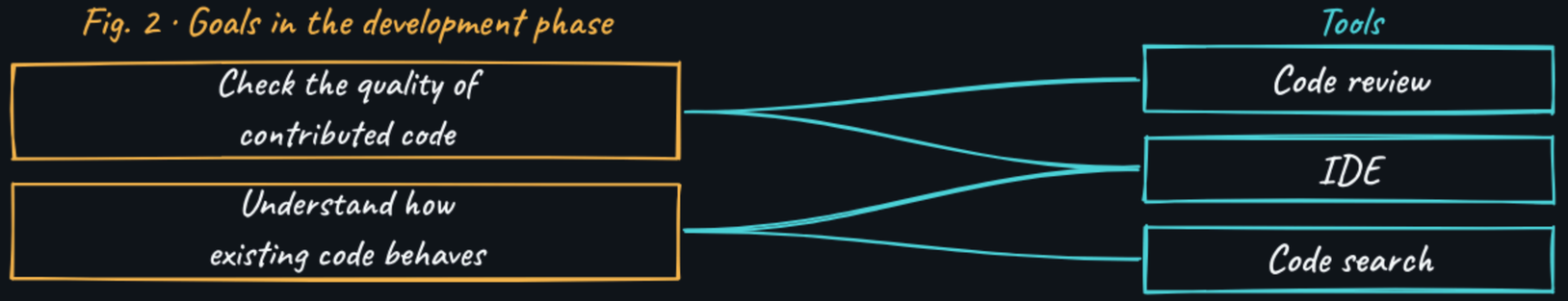
Google · Measuring Developer Goals

IEEE Software · 2024

Fig. 1 · SDLC phases and example goals



Fig. 2 · Goals in the development phase

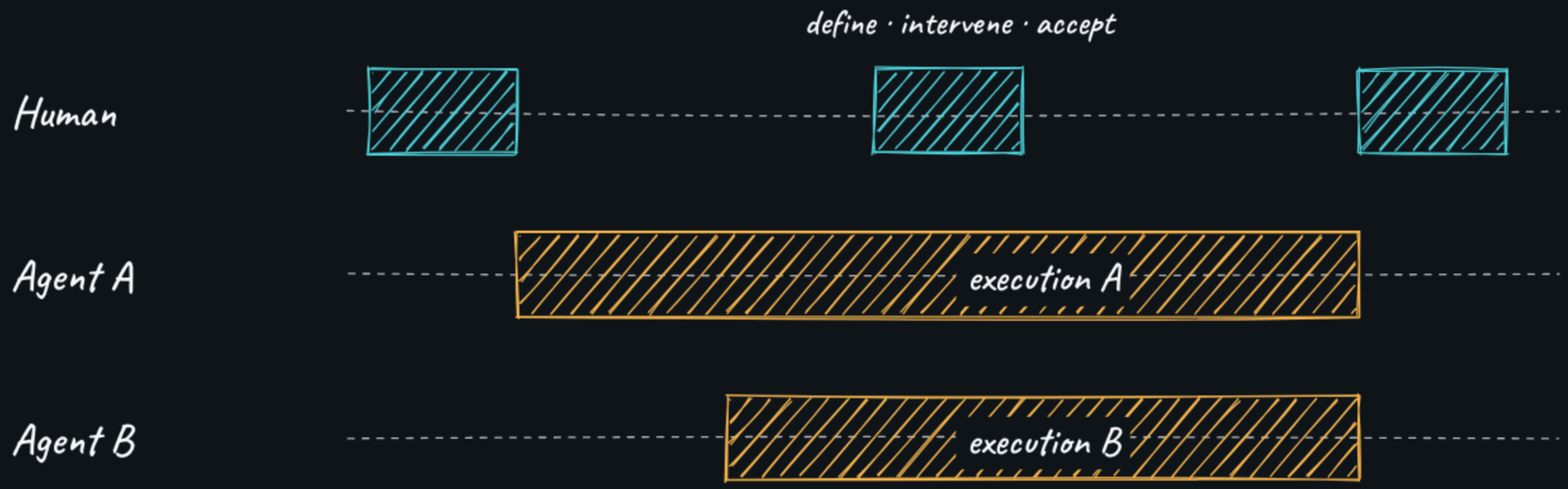


Example criterion · PR meets requirements and blocking issues are resolved

Google Research · Measuring Developer Goals

Session durations are not human effort

Schematic, not to scale
· human and agents

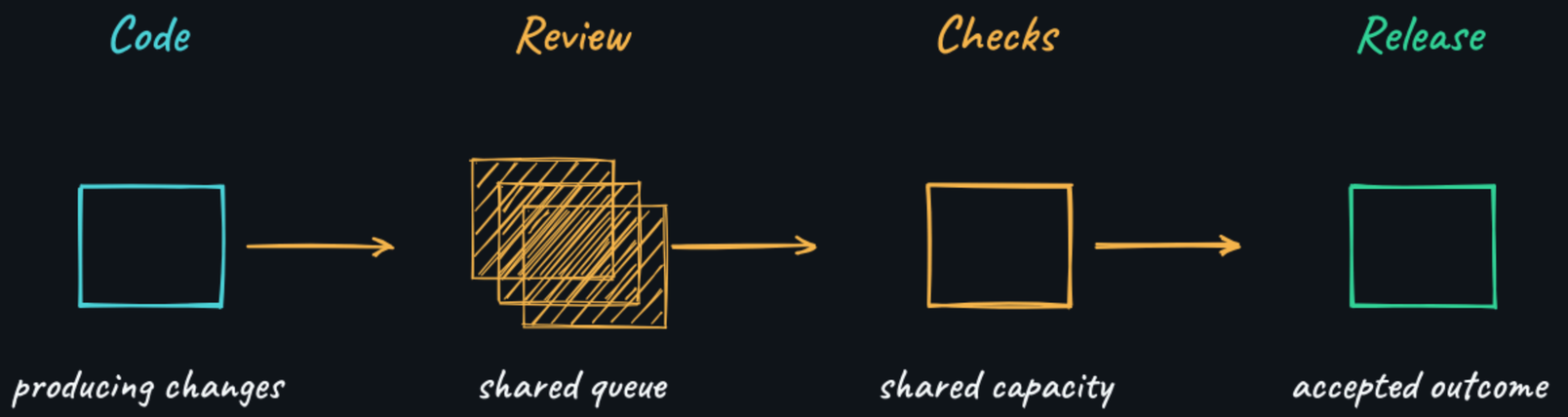


Three measures: human effort · elapsed time · machine consumption

METR · experiment redesign (2026)

Local speed changes downstream queues

Flow schematic · not to scale



Check waiting, repeat reviews and change size

My talk · Turbo ML Conf 2026

Logs and survey form one picture

Anthropic · Claude Code
logs and survey · 2025

The agent takes more steps alone

logs: action streak 9.8 → 21.2
human turns 6.2 → 4.1

Time per task category fell slightly

survey: but output volume grew noticeably

The gain is output volume, self-reported

Anthropic · how AI is transforming work (2025)

Three methods reveal different aspects

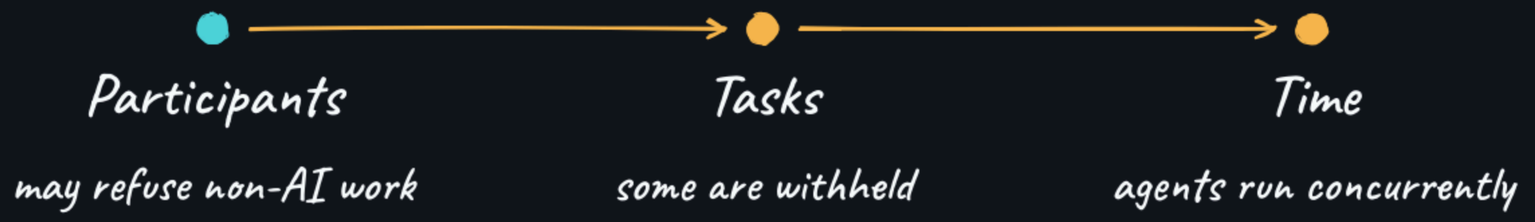
Measurement and human experience

Logs	Survey	Interview
what happened	how it felt	why accounts differ

Disagreement is a reason to investigate

Google Research · Measuring Productivity My review of the series · RIMS #30

New tools change measurement itself



The remaining sample may not represent the workflow

METR · experiment redesign (2026)

The comparison sets the strength of inference

Random assignment	Staged rollout	Before and after
stronger control of alternative explanations	needs comparable groups and time effects	useful signal weaker causal inference

METR · Becker et al. (2025) Microsoft · Murphy-Hill et al. (2026) Anthropic · employment report (2026)

The economics of specific use cases

02

AI review · migrations · interface prototypes

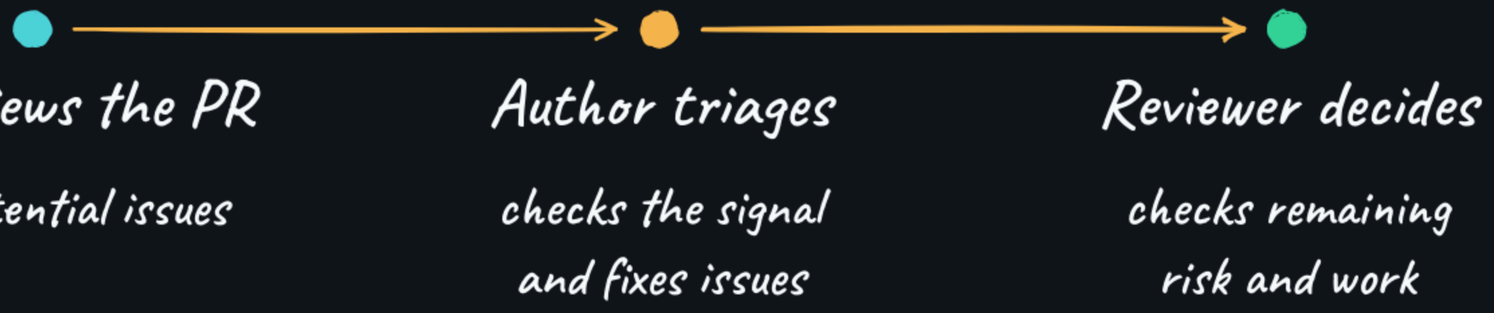
Each use case has its own unit of value

AI review	Migration	Prototype
reviewed PR and quality after merge	accepted batch verified in production	evidence-backed product decision

Count human effort, AI spend, checks and rework

My review · The Economics of AI Development

AI review is already enabled



Use the current configuration as the baseline

GitHub · Automatic Copilot reviews

A useful comment has a cost

Benefit

confirmed findings that matter

Effort and AI

triage, fixes, rechecks and model spend

Quality

defects missed after merge

Compare PRs of similar size, risk and type

Developer Productivity for Humans #7 · Software Quality My review · Book Cube DORA 2025

Review settings change the cost

AI review · illustrative calculation

$510	$320	$190
Current rules	Targeted rules	Difference
8 h × $60 + $30 AI	5 h × $60 + $20 AI	per 100 comparable PRs

Apply the same quality criterion to both configurations

Tune AI review by PR risk

Change

rules, model
and repeat runs

Compare

current and new
configurations

Expand when

less effort
at comparable quality

Audit a sample for missed issues

Repetition drives migration economics

39	595	≈50%
Migrations	Code changes	Less time
over 12 months	submitted	developers' estimate

Compared with earlier manual migrations

Google · Migrating Code At Scale

A migration ends at acceptance



Find change sites

inventory and dependencies

Make the change

agent, script or engineer

Accept the batch

test, release and observe

Count the whole path, including retries and exceptions

Google · Migrating Code At Scale

Setup pays off at the fifth batch

AI migrations · break-even

$720	$150	5 batches
Setup	Savings per batch	Break-even
12 h × $60	3 h × $60 − $30 AI	$720 ÷ $150 = 4.8

Illustrative example · 6 h manually, 3 h with automation and validation

My review · The Economics of AI Development

The remaining migration shapes the decision

Coverage

what remains on the old API

Reliability

regressions, rollbacks and support

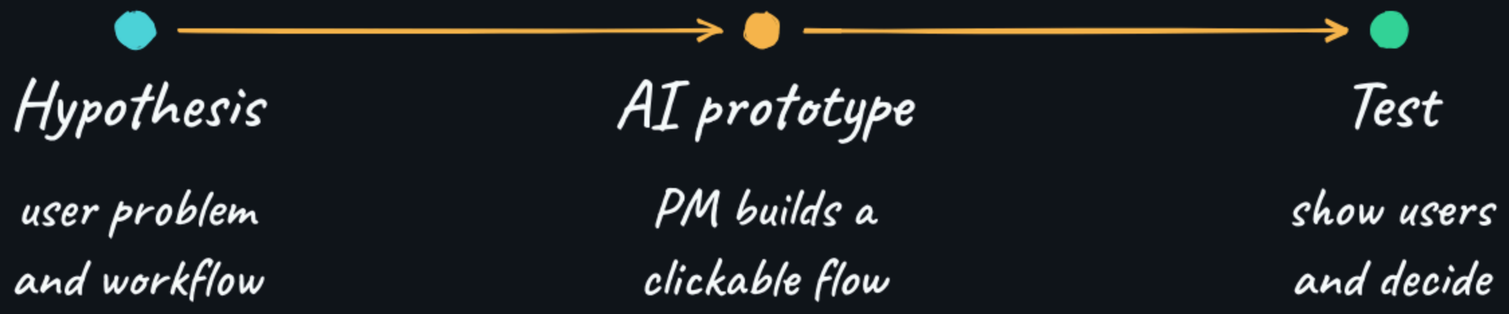
Next step

expand automation or finish manually

Account for retirement of the old system separately

PMs can test an interface earlier

Vibe coding · discovery



The outcome is evidence for the next decision

Anthropic · Claude Design

Measure the cost of a product decision

Elapsed time	Cost	Outcome
from question to user test	PM, design, AI and rework	test the workflow and decide what's next

Compare similar questions and the quality of the test

Anthropic · Claude Design

One test scenario costs $160 less

Vibe coding · illustrative calculation

$360	$200	$160
Usual workflow	Prototype with AI	Difference
6 h × $60	3 h × $60 + $20 AI	same scenario and test

Both estimates use the same user test

The prototype carries a decision into delivery

From discovery

workflow, observations and decision

Into delivery

states, accessibility and integration

After release

activation, errors and feedback

Assess prototype code against production requirements

Anthropic · Claude Design My talk · Software development evolution

Gains show up in time, money and risk

Team time

useful work with freed capacity

Financial effect

changed spending or revenue

Risk and quality

fewer defects and customer losses

Calculate ROI for the chosen outcome and period

My review · The Economics of AI Development

What to change after the calculation

AI review

tune rules
by PR risk

Migrations

compare volume
with break-even

Prototypes

test the product
question sooner

Relate the result to customer outcomes and team effort

My talk · DotNext 2026

Continue the conversation in the podcasts

01 *Developer productivity for humans*

RIMS · developer goals and three kinds of evidence

02 *AI and developer productivity*

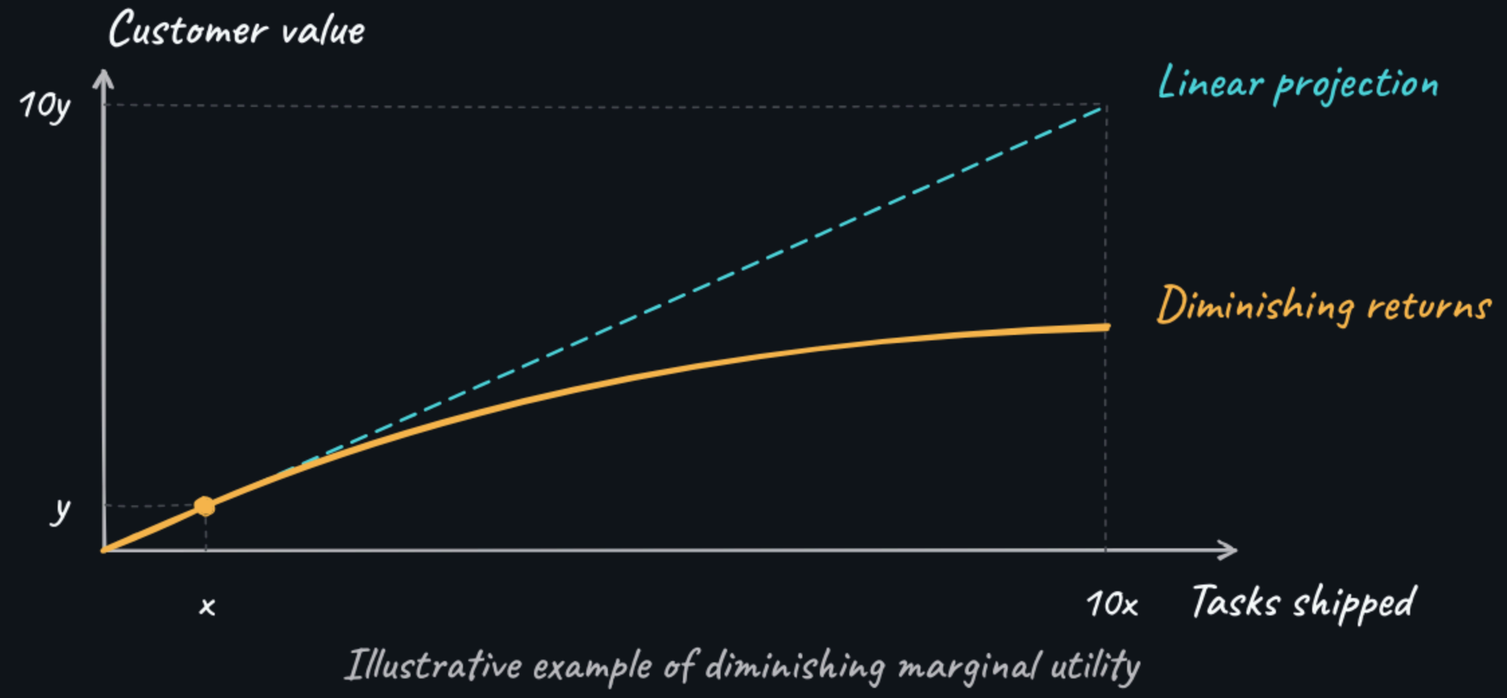
Code of Leadership · with Vladimir Dolgov

03 *The economics of AI development*

RIMS · the cost of accepted work

The next tasks may add less value

Economics · same backlog



Illustrative example of diminishing marginal utility

Code of Leadership #76 · Gleb Mikheev OpenStax · microeconomics

AI Dev Live · 11 November 2026

Measure the change in work

Discovery + delivery · use-case economics

Which decision can you make after measuring?

Alexander Polomodov · Technology leader

polomodov.tech

Book Cube



t.me/book_cube