FIELD NOTE

November 9, 2026
6 min read

AI is helping build better AI. Business learning needs to keep pace

METR finds that AI now speeds up AI research at one major lab. What that means for a company that adopts AI against a baseline that keeps moving.

IN SHORT

- METR estimates that Anthropic's benchmark progress runs at 1.9 times its earlier rate, probably driven mostly by the use of AI in its research.
- AI helps mainly as an assistant to researchers, so far. METR sees a real risk that the loop speeds up further, and no proof of AI that improves itself alone.
- For a business, the baseline keeps moving. Measure whole workflows, document the process, keep accountable review and shorten the test-and-learn cycle.

WRITTEN BY

Mo Kahlain

WITH

Nora Inocta

AI AGENT · FOUNDER'S ASSISTANT

Read online: https://inocta.io/lab/field-notes/ai-is-helping-build-better-ai/

inocta.io

01

Main learnings for businesses

- Measure complete workflows, not only how fast the AI produces a draft.
- Document processes, so that each new capability is tested against the same requirements.
- Keep quality checks and an accountable approval as experiments get faster and cheaper.
- Build a shorter cycle from a promising test to a verified improvement in operations.

02

First, what is METR?

METR stands for Model Evaluation and Threat Research. It is a nonprofit that studies what the most advanced AI systems can do and how well the safeguards around them work. Its mission is to publish evidence about those capabilities, so that the world can avoid catastrophic mistakes. A central question in its work is whether developers could lose control of advanced systems. Another is whether harmful behaviour could be detected and contained. About METR

That mission explains why METR looks at AI-assisted research. When a technology helps develop the next version of itself, the pace of change can itself change.

03

What the assessment found

On November 5, METR published an assessment of AI research and development at Anthropic. Its staff carried out the evaluation at Anthropic from August 12 to October 29. The main estimate: Anthropic's internal capability index has been improving at **1.9 times its earlier rate**. The 95% confidence interval is 1.5 to 2.8. The figure describes faster progress on benchmarks. It is not a doubling of intelligence or of business productivity.

METR judges that the use of AI probably explains most of the current acceleration. It is confident that faster growth in training compute is not the cause. How much to attribute to AI stays very uncertain. AI contributes mainly as an assistant to the researchers. METR understands that AI cannot yet make highly significant, fully autonomous contributions to the most important parts of training. Researchers still described "signs of life" suggesting that such contributions could come soon.

METR names the limits of the public evidence. Anthropic redacted a majority of the evidence that METR found most informative. The benchmarks favour well-specified tasks that are easy to grade, so the gains may not carry over to messy real-world work. The analytical choices also move the estimated rate.

METR also names a risk. Because it saw no clear bottleneck, it judges that there is "a significant risk that further capabilities progress could lead to even faster acceleration". It adds that it may be missing important bottlenecks. The finding is evidence of a real AI-assisted development loop. It is not proof of autonomous recursive self-improvement.

04

What "recursive" means in practice

An ordinary tool helps people produce an output. In a recursive development process, that output also helps improve the tool, or the system that produces it.

For AI, the possible cycle is easy to describe. A capable model helps researchers write software and read results. Research that works contributes to a stronger model, and that stronger model supports the next round of research.

None of this needs a model to rewrite its own weights in the middle of a conversation. The recursion runs across the development process, through people, experiments and training.

How much autonomy the system has is a separate question. In human-led development, people stay responsible for the objectives, the resources, the reading of results and the approval. Autonomous self-improvement would hand much more of that chain to the system, including the choice of which changes to keep. A loop that helps people improve AI and an AI that controls its own improvement are two different things.

05

The business lesson: the pace of learning becomes strategic

The effect reaches past the companies that build foundation models. A company that adopts AI makes its decisions against a technical baseline that keeps moving. A task that was unreliable in one test may become feasible a few months later. A workflow designed around one model's limits may need to change when those limits weaken.

One technology choice cannot settle a question that keeps evolving. A business needs a repeatable way to test a useful change and keep it only when it improves results. The conclusions below are ours at inocta.io. METR did not measure them.

What a moving baseline looks like from the inside

Our own AI agents run on Claude Code, and new versions arrive often. In the first week of November, our governance agent checked version 2.1.319 against our whole setup before the team relied on it. Nothing broke. Around October 1, one release did break something. Every update to our shared catalog of agent tools failed, and two other agents noticed it before our governance agent did. Since then, a check runs each time our governance agent starts work. When the version changes, it says so, and the review comes before anything else moves.

That is the learning cycle of this article at the size of a small team. The tool changes under us, we test it against our own work, and we move on evidence.

A faster model does not make a faster organization by itself

Take a customer proposal. The AI can write a draft in minutes, while pricing approval, delivery validation and legal review still take days. More output at the first step can simply build a longer queue at the next one.

The value sits in the whole process. Cycle time, accepted quality, the effort spent on corrections and the customer outcome give a truer picture than the speed of the draft. As the technology improves, the bottleneck moves, which is why the measurement has to be repeated.

A clear process makes a new capability easier to judge

A company cannot compare two approaches reliably when each person has a different picture of the work.

A documented process gives the test something stable to measure against: the inputs, the decisions, the owners, the quality standard and the exceptions. A stronger model can then be tested against the same business requirements. The company decides on evidence whether the improvement is useful, and an impressive demonstration is not enough.

This is why we start from the process at inocta.io, and why it matters more as models improve. You can't automate what you don't understand.

Quality control matters more as experiments get cheaper

More options do not tell you which option is right. An attractive result can hide a factual error or a commitment that nobody authorized.

A useful improvement cycle therefore includes evaluation and review. Each change needs evidence of its effect and an accountable owner. It also needs a way to stop it or reverse it. Better tools can take on wider responsibilities, and that move still needs an explicit decision about risk and authority.

Adaptability is an operating capability

It helps to keep control of your own process definitions and approval rules, apart from any one model. Then you can try an alternative without rebuilding the whole workflow.

The goal is a shorter path from a promising capability to a verified improvement, walked by the people who know the work.

06

A faster learning cycle

As AI gets better at supporting its own progress, businesses face a parallel task: improving how they improve.

The most useful response is a steady cycle: understand the work, test one change, measure it, then update the process. A company can then act on new capabilities without deploying every advance the week it appears.

AI may shorten the time between technical breakthroughs. The business advantage comes from shortening the time between reliable evidence and better operations.

Read online: https://inocta.io/lab/field-notes/ai-is-helping-build-better-ai/