Healthy AI First, Episode 1: It Feels Faster

Sources and evidence for this episode

Sixteen experienced developers expect AI to make them almost a quarter faster. Measured, they are slower, and afterwards they still believe they were faster. The opening episode explains why our gut feeling is a poor measuring instrument when we work with AI, what other studies say against it (large field experiments, industry reports, surveys), what working with AI can cost (more checking, fatigue, less learning) and where the series goes over the next seven episodes.

All links were checked on 17 October 2026.

What we mention in this episode

The Experiment, The First Twist, The Second Twist

- **Claim:** Experienced open-source developers expected AI to save them about 24% of their time, measured they took about 19% longer, and afterwards they estimated they had saved about 20% of their time.
  Source: Becker, J., Rush, N., Barnes, E. & Rein, D. (METR, 2025): Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity. Preprint.
  Link: https://arxiv.org/abs/2507.09089 (summary: https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/)
  Strength and limits: Randomized study with real tasks (16 people, 246 tasks), so strong evidence for exactly this situation; but a small group, experienced people in their own projects and tools from early 2025.

The Experiment Breaks Apart

- **Claim:** The follow-up study with 57 developers and more than 800 tasks did not produce a reliable number, because many no longer wanted to work without AI and 30 to 50% of the developers said they had not submitted some tasks for that reason.
  Source: Becker, J., Rush, N., Cunningham, T., Rein, D. & Mahamud, K. (METR, 24 Feb 2026): We are Changing our Developer Productivity Experiment Design.
  Link: https://metr.org/blog/2026-02-24-uplift-update/
  Strength and limits: The authors themselves call their estimate an unreliable signal (selection, hard-to-measure time with parallel agents). Early evidence, blog report.

The Other Side

- **Claim:** In three randomized field experiments at Microsoft, Accenture and another large company, developers who actually used the AI assistant completed about a quarter more tasks per week (+26.08%); less experienced developers benefited more.
  Source: Cui, K. Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S. & Salz, T. (2026): The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers. Management Science.
  Link: https://doi.org/10.1287/mnsc.2025.00535 (working paper: https://doi.org/10.2139/ssrn.4945566)

Strength and limits: Large randomized field study (4,867 people), strong evidence; the outcome is completed pull requests, not quality. The effect of usage is estimated with additional statistical assumptions, and the individual experiments vary a lot.

- **Claim:** Google's DORA reports describe AI as an amplifier: where more AI is used, more is delivered, but stability suffers; well-organized teams gain, weak structures show up faster.
  Source: DORA and Google Cloud (2025): State of AI-assisted Software Development; DORA and Google Cloud (2024): Accelerate State of DevOps Report.
  Link: https://research.google/pubs/dora-2025-state-of-ai-assisted-software-development-report/ · https://dora.dev/insights/dora-2025-year-in-review/ · https://services.google.com/fh/files/misc/2024_final_dora_report.pdf
  Strength and limits: Large surveys (almost 5,000 people in 2025); they show associations, not causes. Self-report and self-selection.

- **Claim:** In spring 2026, several hundred technical workers typically rated the value of their work with AI at 1.4 to 2 times.
  Source: Becker, J. (METR, 11 May 2026): Measuring the Self-Reported Impact of Early-2026 AI on Technical Worker Productivity.
  Link: https://metr.org/blog/2026-05-11-ai-usage-survey/
  Strength and limits: 349 respondents, 87 of them software developers; convenience sample and self-assessment, exactly the kind of measurement that was off in the first experiment.

What It Costs

- **Claim:** A repeated survey (October 2024 and April 2025) shows a shift from writing to checking; the share of people whose developer experience had worsened in at least one area rose from 14% to 27%.
  Source: Vella, A. & Blincoe, K. (2026): The Impact of AI Coding Assistants on Software Engineering: A Longitudinal Study. Preprint.
  Link: https://arxiv.org/abs/2605.23135
  Strength and limits: Preprint, 95 repeat respondents, self-report. Clarification to the episode: the 14% to 27% refers to worsened developer experience in at least one dimension, not to an overall negative rating of AI; perceived productivity stayed stable.

- **Claim:** In a lab study, participants with AI were faster (22% less time) and found the tasks less demanding; still, stress and fatigue rose across the tasks; verification load was statistically associated with this rise (mediation, not proof of cause).
  Source: Fan, G.-R., Liu, D.-D., Pan, L.-H. & Zhang, R. (2026): When Help Hurts: Verification Load and Fatigue with AI Coding Assistants. CHI 2026.
  Link: https://doi.org/10.1145/3772318.3791176
  Strength and limits: 60 people, three short Python tasks in the lab; says nothing about full working days. The link via checking effort is a statistical mediation, not a proven mechanism.

- **Claim:** People who learned an unfamiliar programming library with AI scored 50% on average in the comprehension test afterwards, without AI 67%; those who used AI to ask questions and get explanations understood more.
  Source: Shen, J. H. & Tamkin, A. (Anthropic, 2026): How AI Impacts Skill Formation. Preprint.
  Link: https://arxiv.org/abs/2601.20245 (summary: https://www.anthropic.com/research/AI-assistance-coding-skills)

Strength and limits: Randomized experiment with 52 mostly junior developers; tested right afterwards, no statement about long-term learning. Preprint by an AI vendor.

Tool tip

This episode has no tool tip. The suggestion to try only needs a pen or a notes file: before a task, estimate how long it will take with AI, afterwards write down the real time (including prompting, waiting, checking and fixing). Episode 2 shows how to read real times from Git and GitHub.

Further reading

- Maier, S., Gunzenhäuser, M., Schweisthal, J., Schneider, M. & Feuerriegel, S. (2026): A meta-analysis of the effect of generative AI on productivity and learning in programming. Preprint, 23 studies. Moderate positive effect on productivity with large variation, no significant effect on learning. https://arxiv.org/abs/2605.04779
- Basis of the series: Healthy AI First Software Development, version 2 (1 October 2026), chapters 1 and 2.

Note

Healthy AI First is AI-researched and AI-generated, narrated in Björn's voice, recreated with AI, and reviewed by him.

Daniel and the project lead in this episode are fictional characters used as illustrations. Across the series, the same applies to Miriam, Jonas and Hanna.