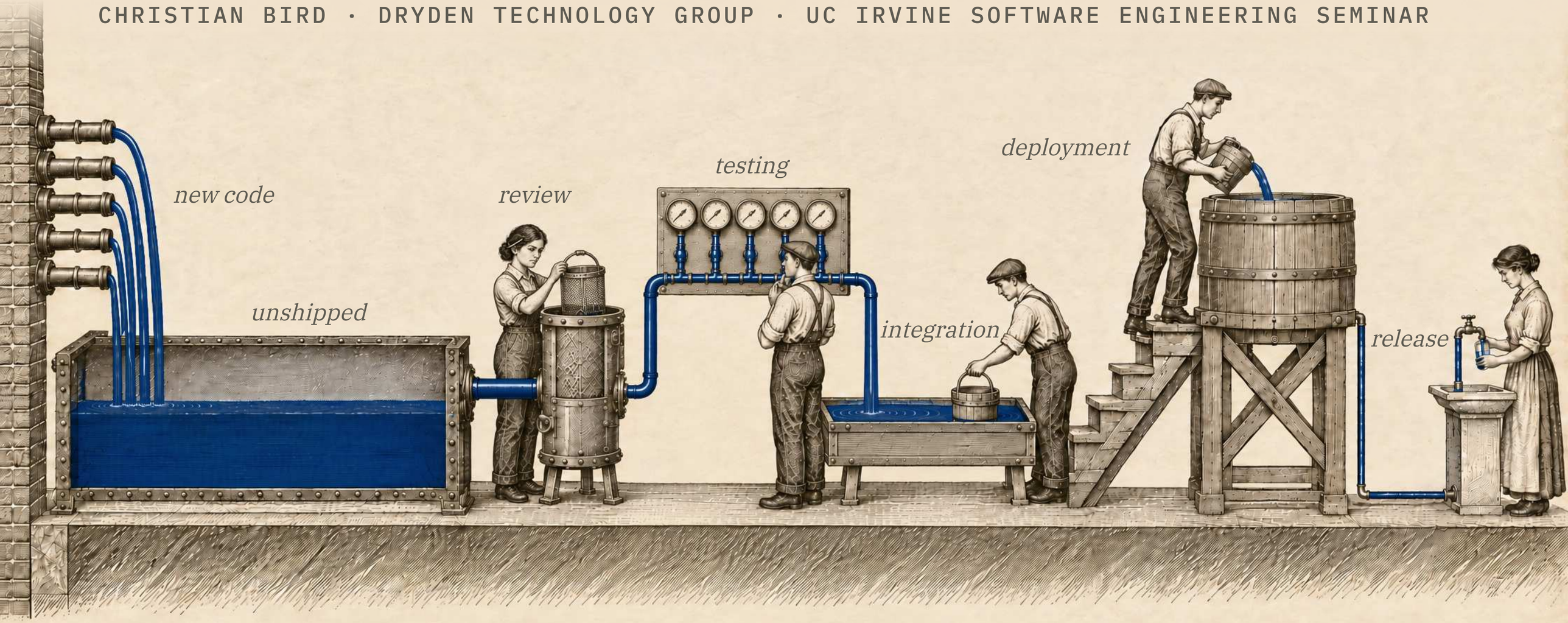
Why aren’t we shipping faster now?

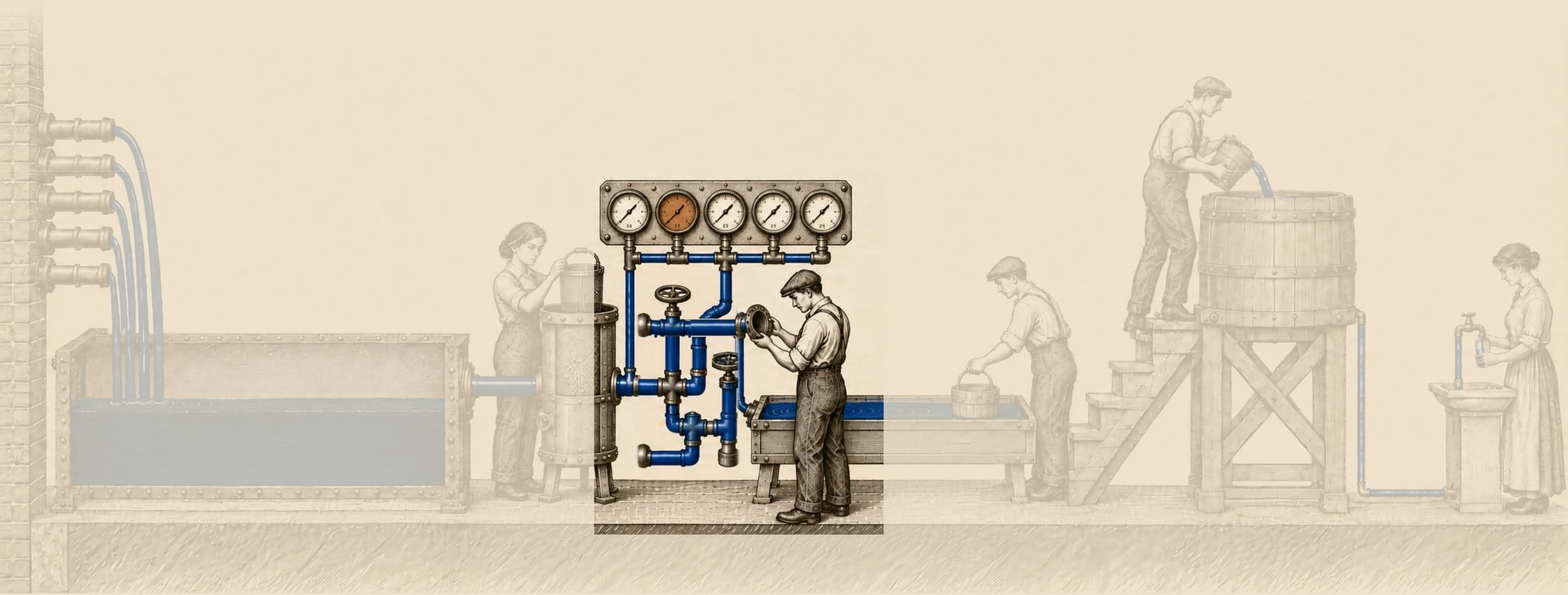
AI, activity, and outcomes in software engineering

CHRISTIAN BIRD · DRYDEN TECHNOLOGY GROUP · UC IRVINE SOFTWARE ENGINEERING SEMINAR



THE ENGINEER

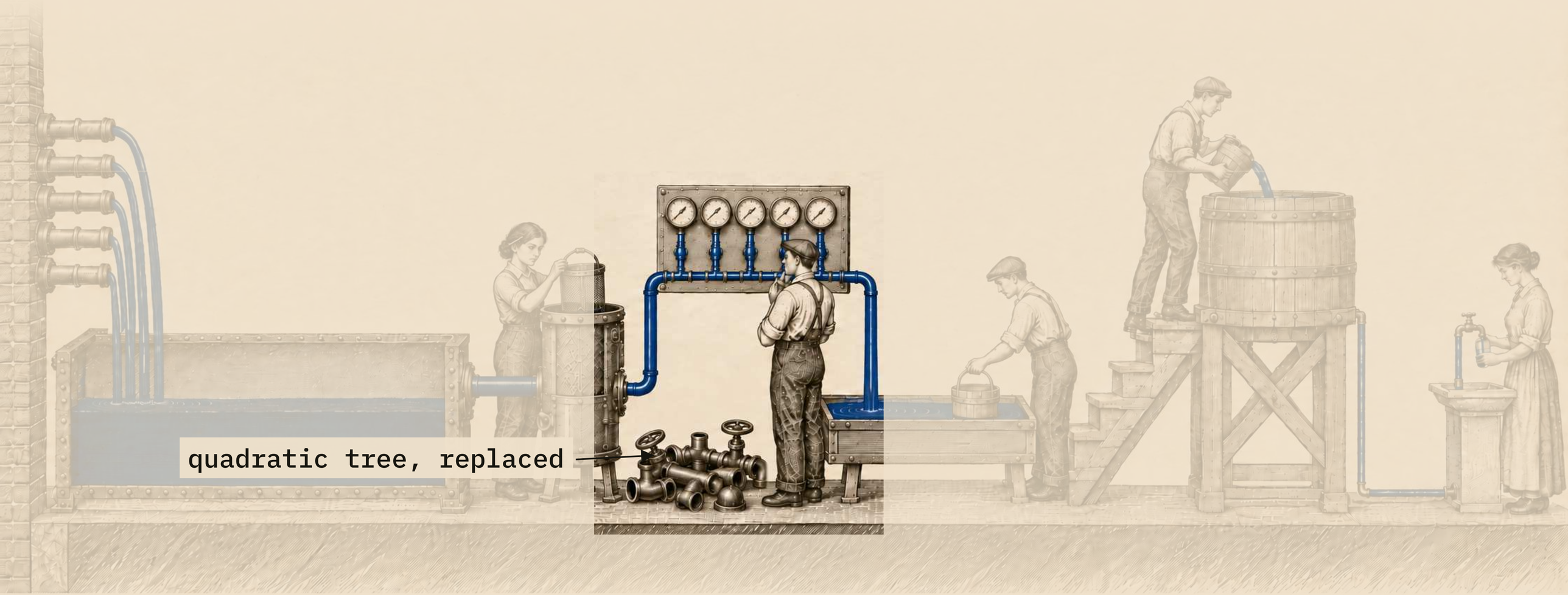
A years-old performance bug, fixed in a day.



Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

THE ENGINEER

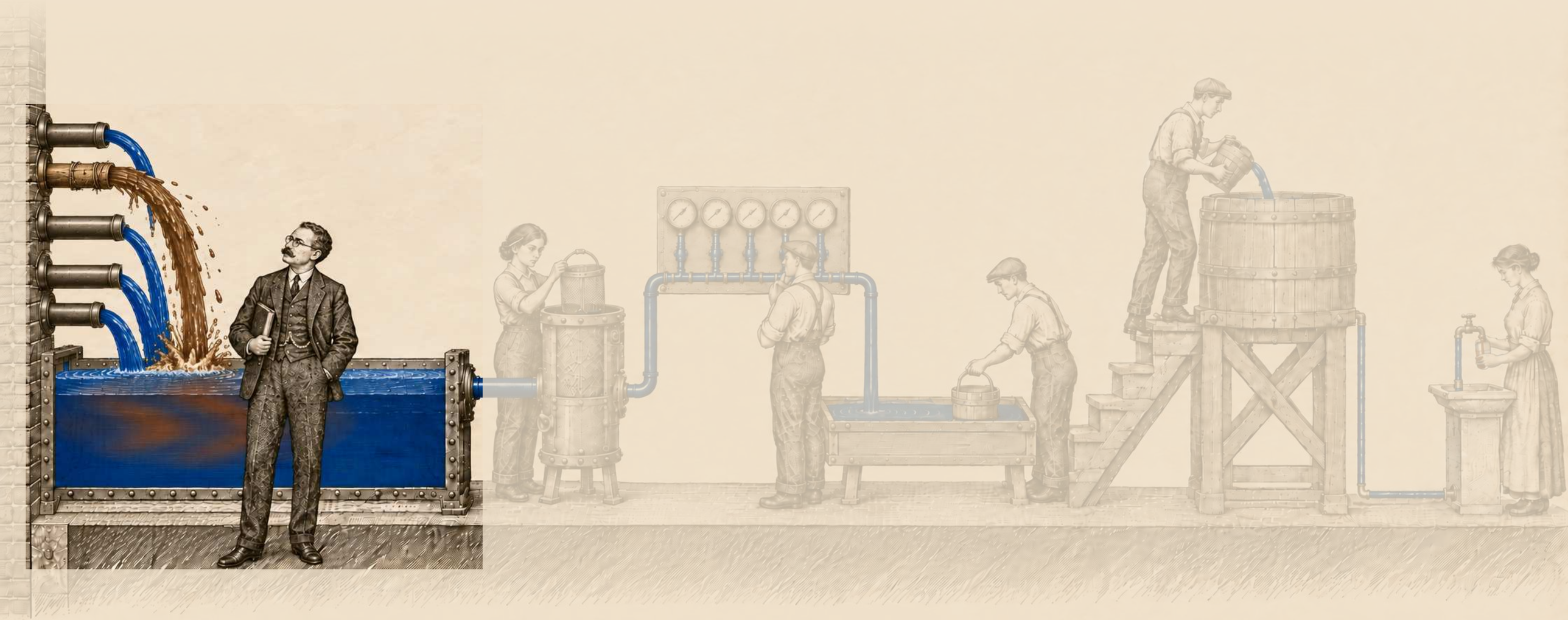
A years-old performance bug, fixed in a day.



Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

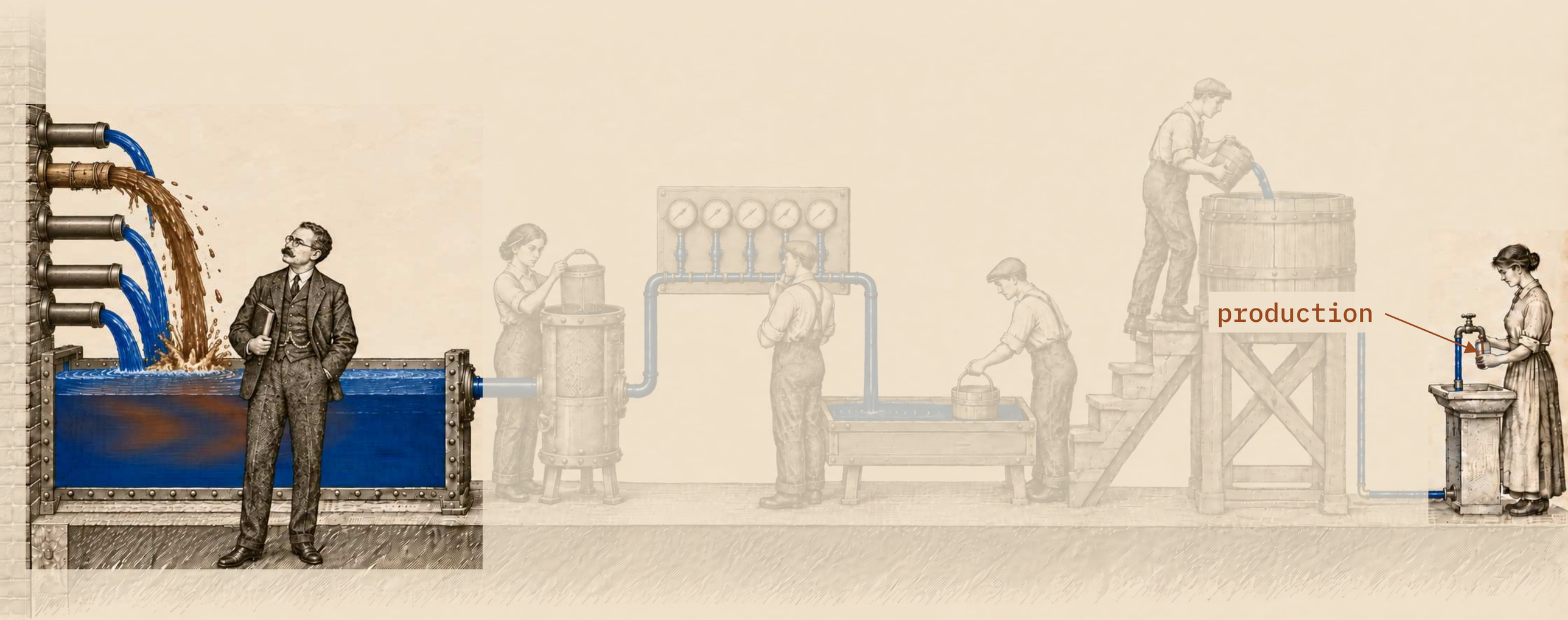
THE PRODUCT MANAGER

A long-standing bug, fixed himself. It broke production.



THE PRODUCT MANAGER

A long-standing bug, fixed himself. It broke production.



THE DEMO

What the demo shows



Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

THE DEMO

“Why is it so hard for you guys to do the same thing?”

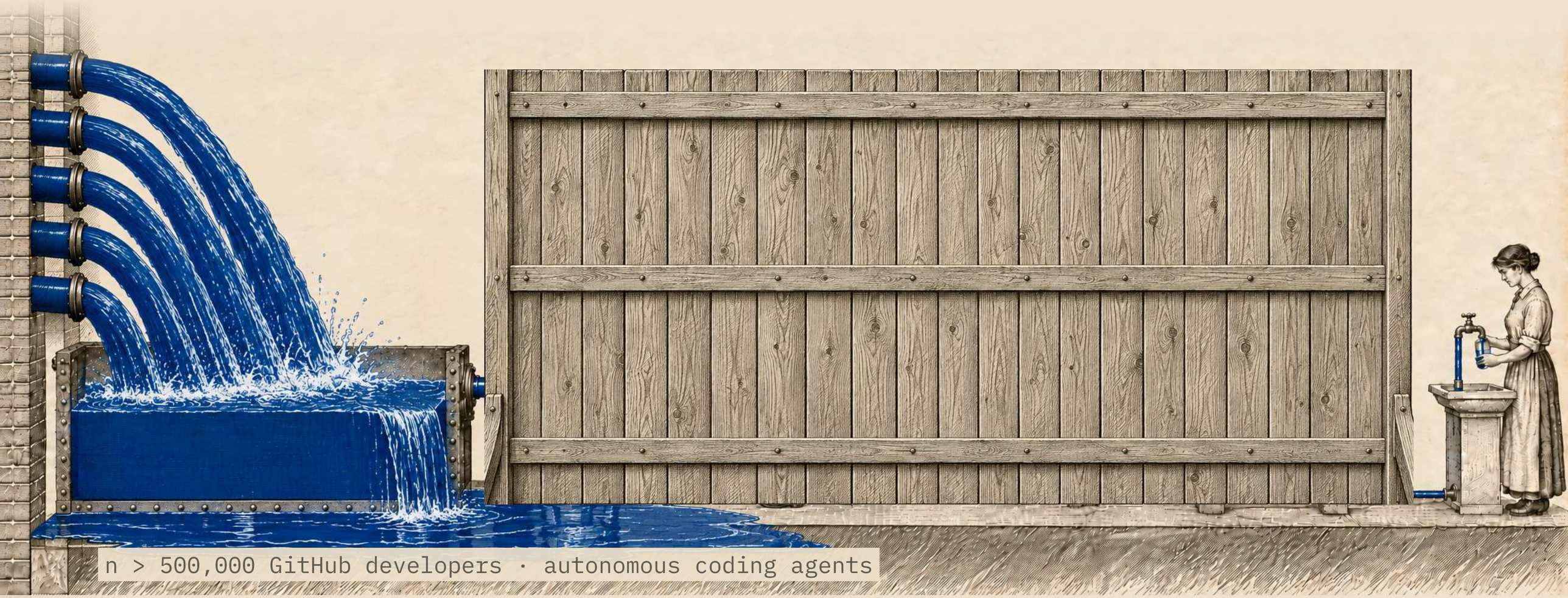
LEADERSHIP, TO OTHER TEAMS · AS P6 REPORTED IT



30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

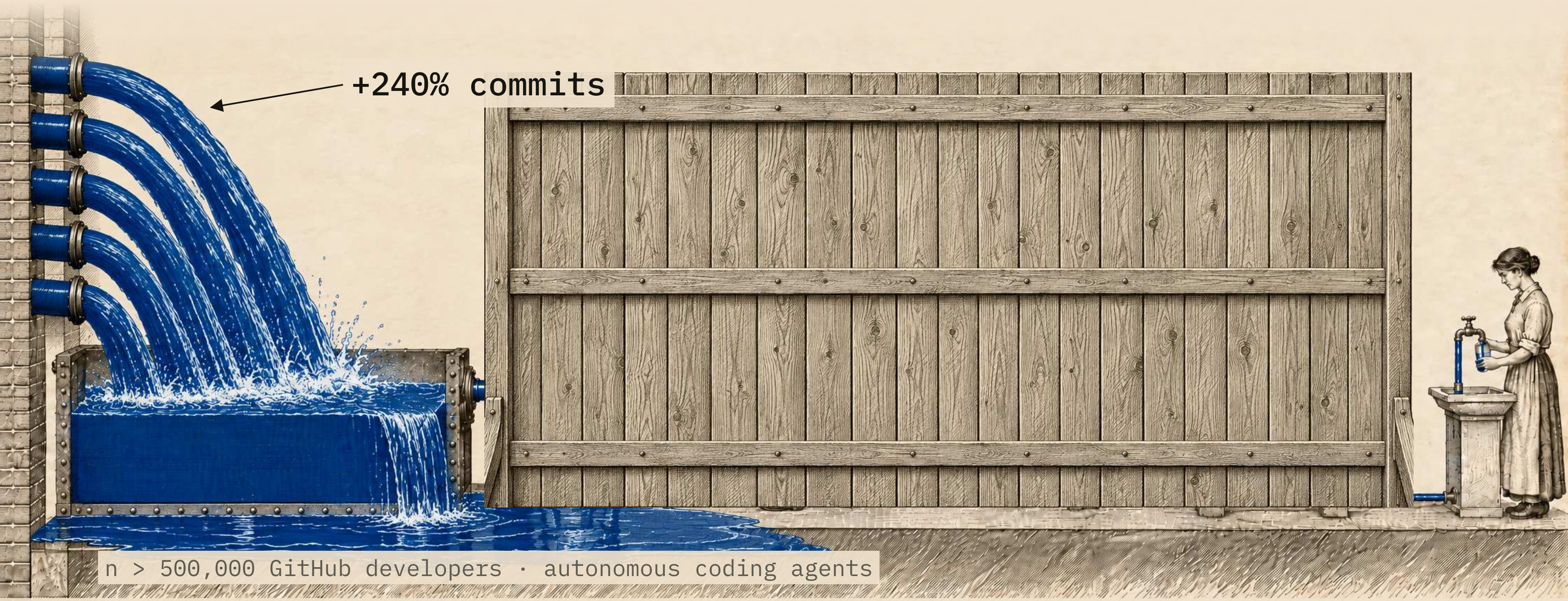
THE PUZZLE



n > 500,000 GitHub developers · autonomous coding agents

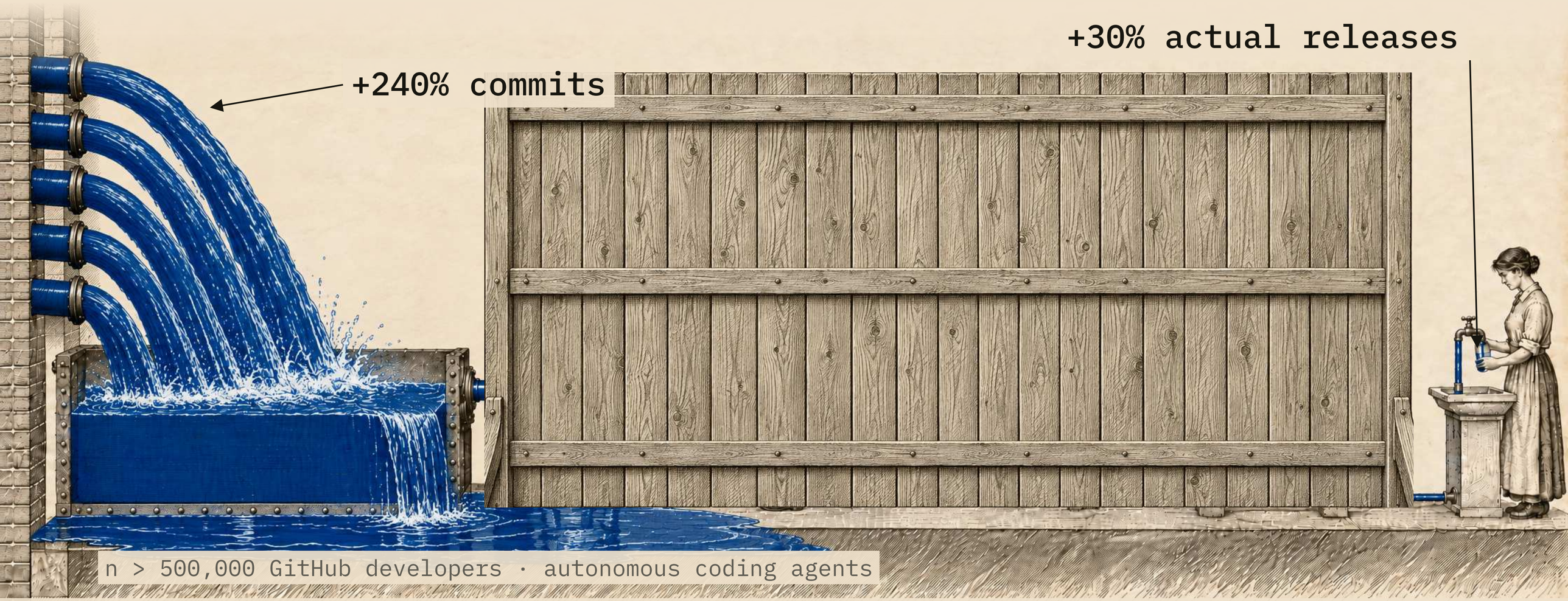
Demirer, Musolff, Yang. *Writing Code vs. Shipping Code: Productivity Effects Across Generations of AI Coding Tools*. NBER Working Paper 35275, 2026.

THE PUZZLE



Demirer, Musolff, Yang. *Writing Code vs. Shipping Code: Productivity Effects Across Generations of AI Coding Tools*. NBER Working Paper 35275, 2026.

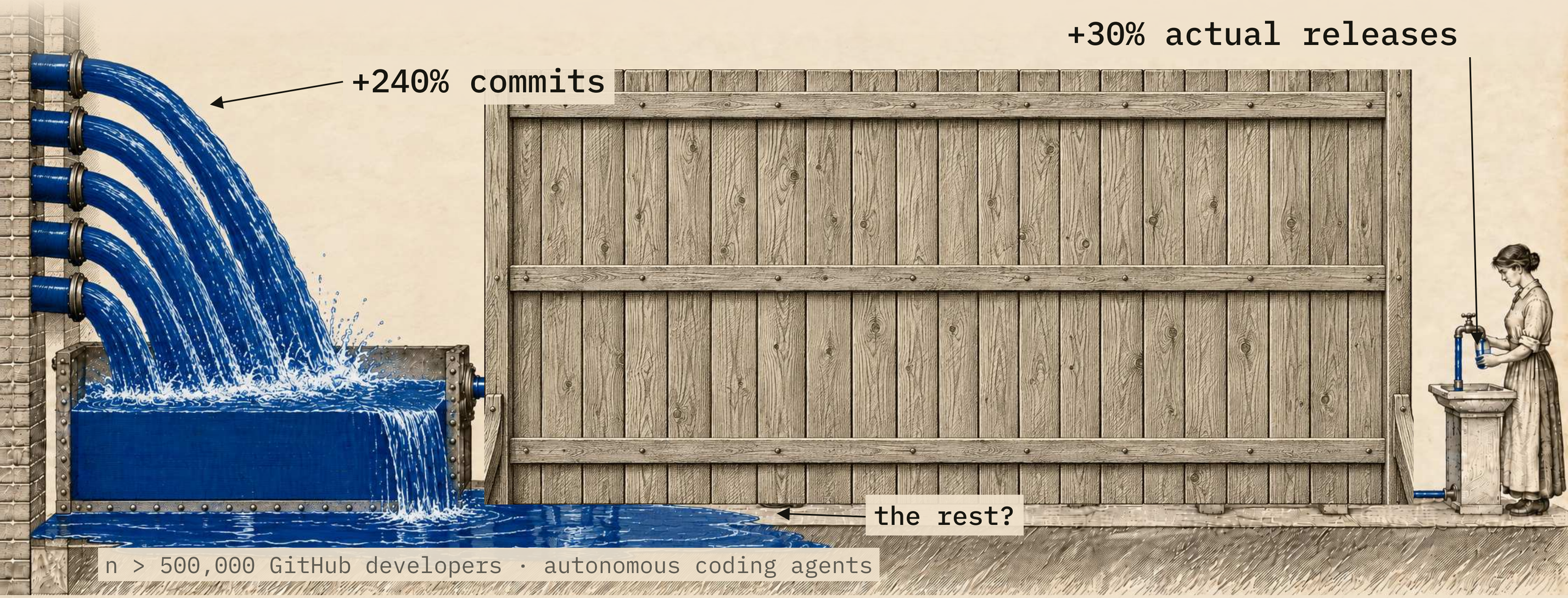
THE PUZZLE



Demirer, Musolff, Yang. *Writing Code vs. Shipping Code: Productivity Effects Across Generations of AI Coding Tools*. NBER Working Paper 35275, 2026.

THE PUZZLE

Where did the rest go?



Demirer, Musolff, Yang. *Writing Code vs. Shipping Code: Productivity Effects Across Generations of AI Coding Tools*. NBER Working Paper 35275, 2026.

PART 1 · WHAT GOT STARTED

What is different about how you work now?



Faster
45.0%

Further
36.9%

Afresh
12.4%

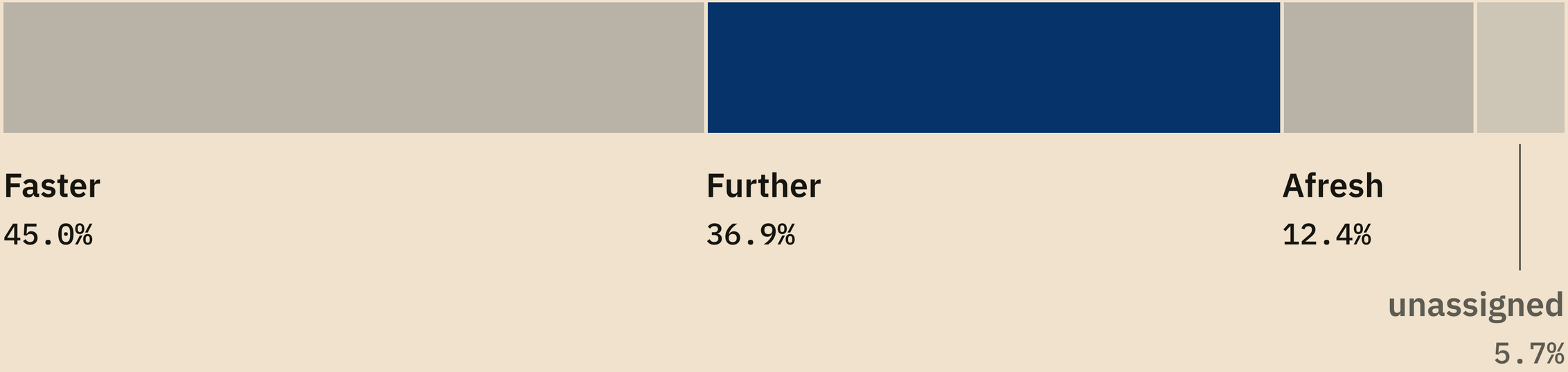
unassigned
5.7%

n = 331 developers · Aug 2026

Choudhuri, Butler, Bird, Badea, Houck, Sarma. *“To Boldly Go Where No One Has Gone Before”: What Has AI Let Developers Unlock?* In submission, 2026.

PART 1 · WHAT GOT STARTED

What is different about how you work now?

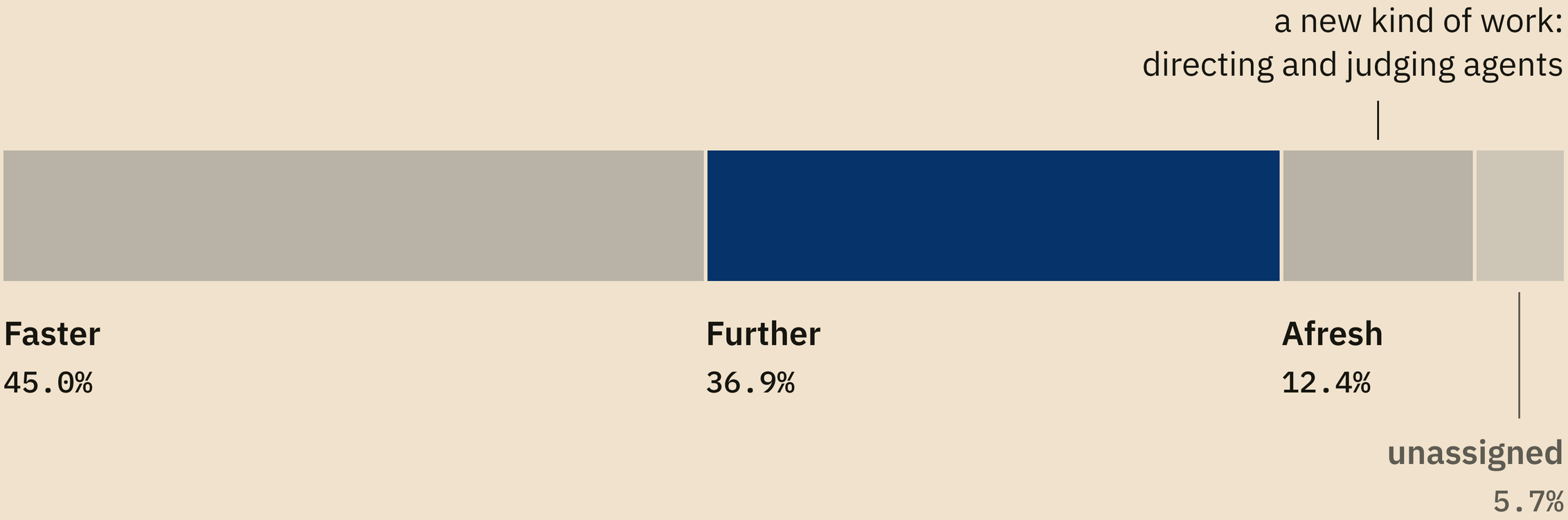


n = 331 developers · Aug 2026

Choudhuri, Butler, Bird, Badea, Houck, Sarma. *“To Boldly Go Where No One Has Gone Before”: What Has AI Let Developers Unlock?* In submission, 2026.

PART 1 · WHAT GOT STARTED

What is different about how you work now?



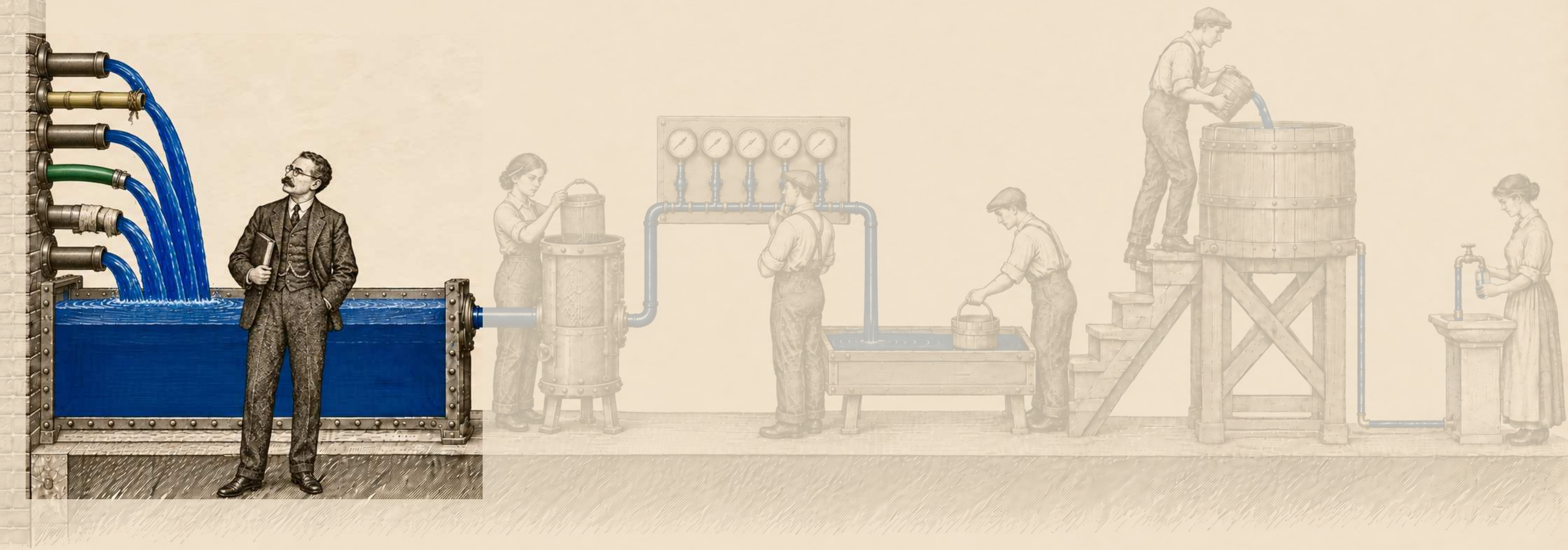
n = 331 developers · Aug 2026

Choudhuri, Butler, Bird, Badea, Houck, Sarma. *“To Boldly Go Where No One Has Gone Before”: What Has AI Let Developers Unlock?* In submission, 2026.

PART 1 · WHAT GOT STARTED

“Everyone is a programmer now — you just have to say something to the computer.”

JENSEN HUANG, NVIDIA · 2023

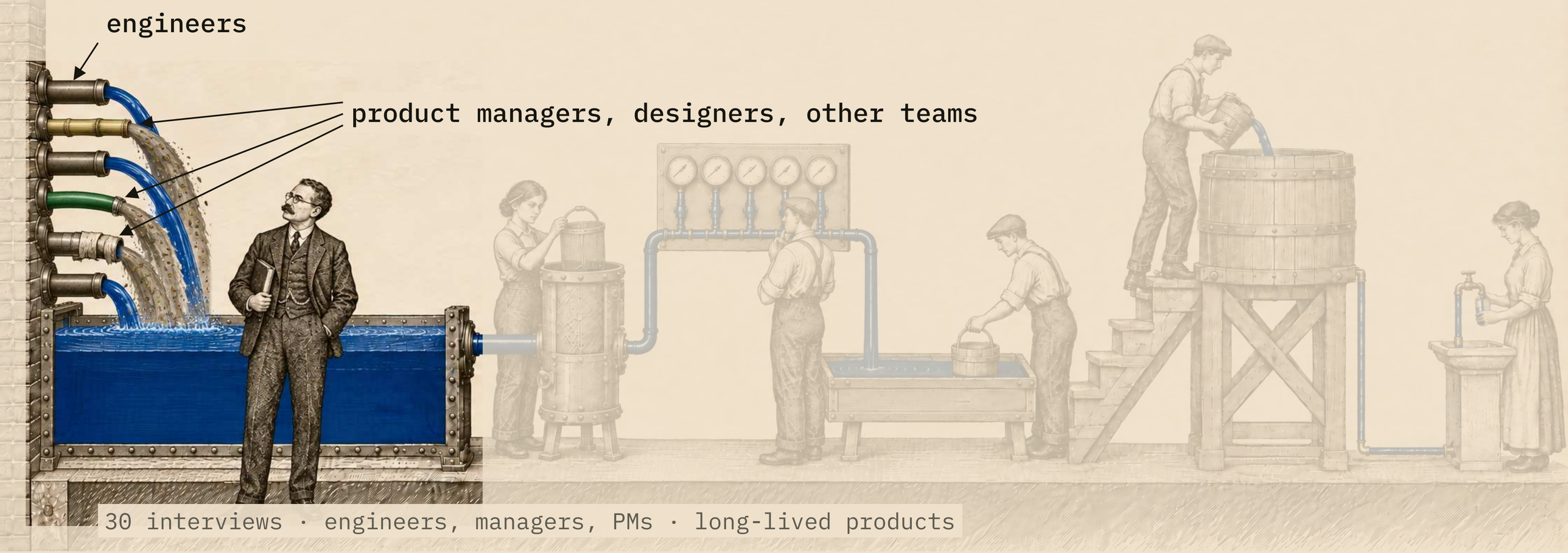


Jensen Huang (Nvidia), Computex keynote, Taipei, 29 May 2023, as reported by Reuters.

PART 1 · WHAT GOT STARTED

“I see code that is under my management being changed by people I’ve never heard of before, that I was not included on the PR for.”

P10 · OWNS A DECADES-OLD PRODUCT

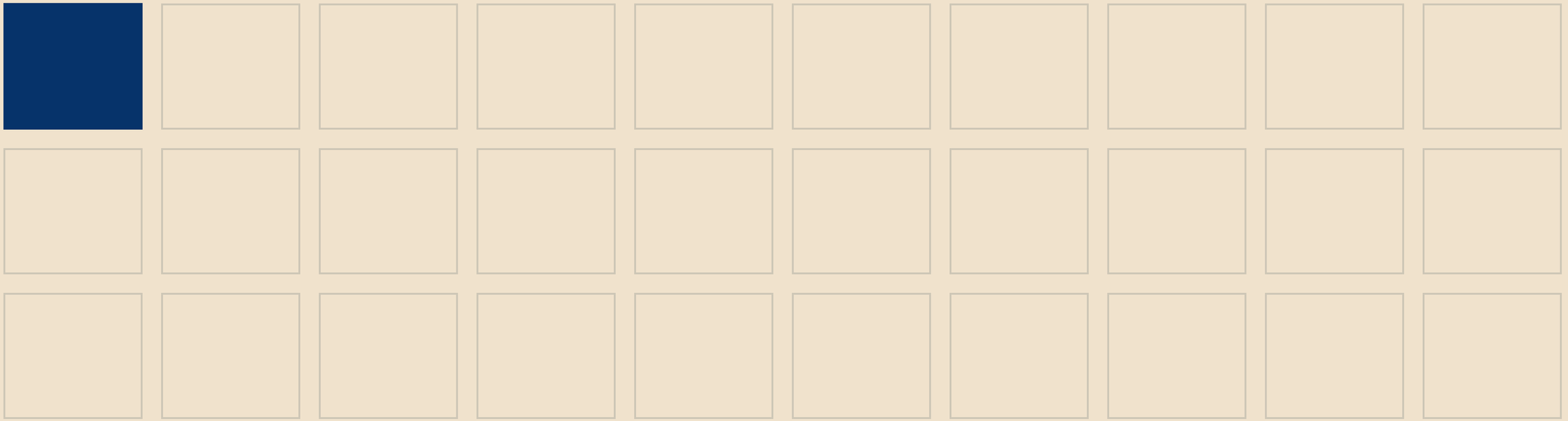


30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

> “If you can do the code in one day, you’re still talking 30 days before it actually gets deployed.”
>
> P18

code



30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

> “If you can do the code in one day, you’re still talking 30 days before it actually gets deployed.”
>
> P18



30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

> “If you can do the code in one day, you’re still talking 30 days before it actually gets deployed.”
>
> P18



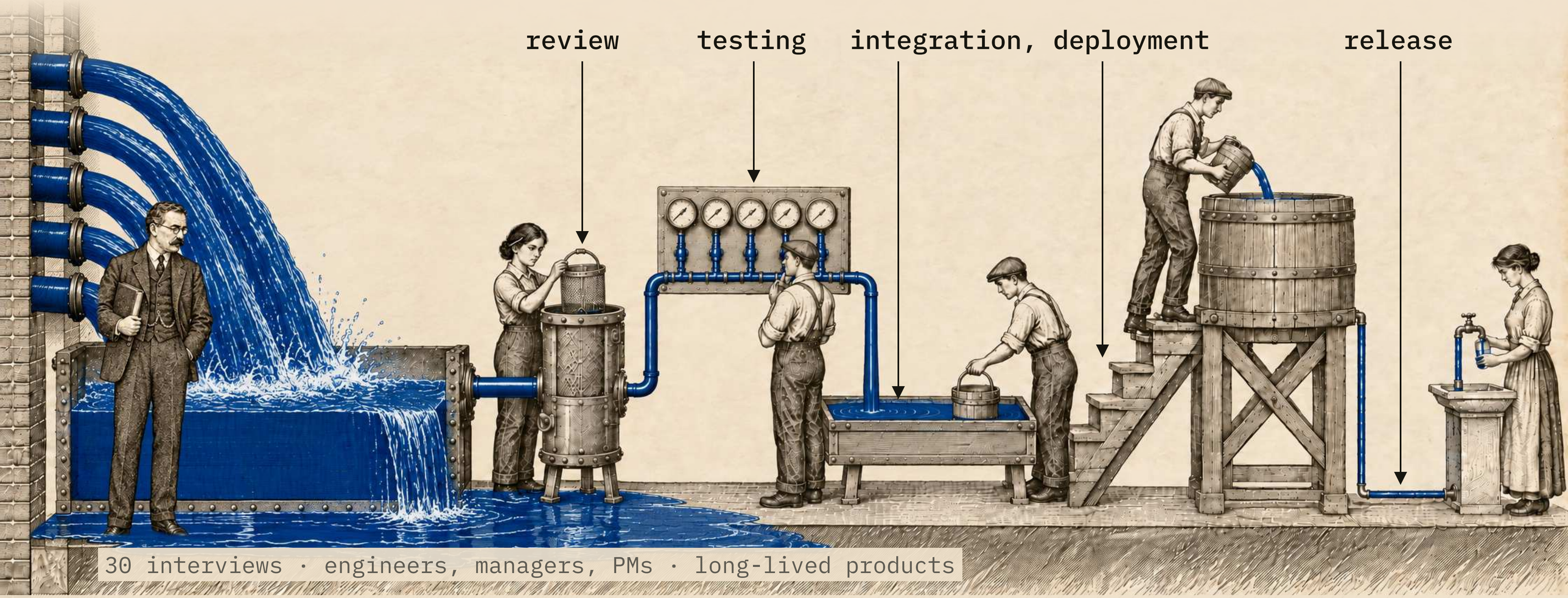
> “If we used to take a year to get a feature done that we can now do in a month, in some senses, a year’s worth of collaboration has to happen in that month.”
>
> A PARTICIPANT · FROM YOONHA CHA’S INTERVIEWS

30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

PART 2 · WHAT DID NOT KEEP UP

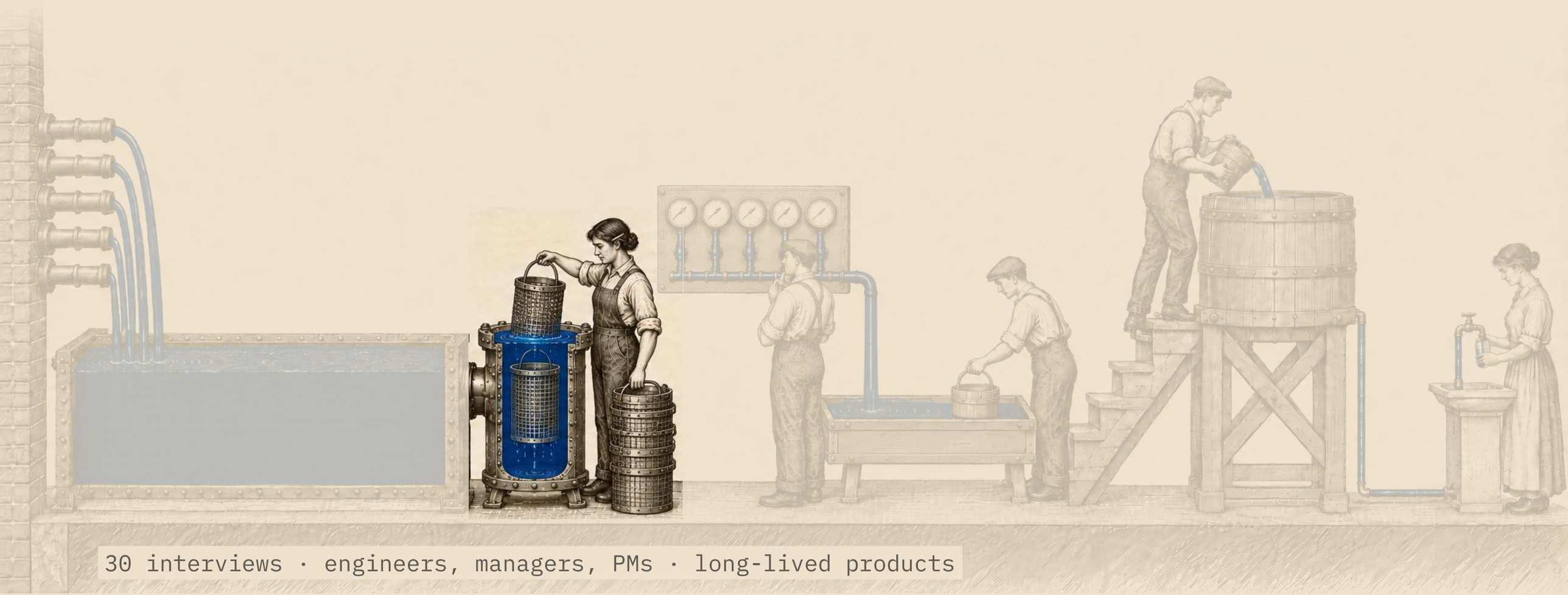
Shipping: everything between a diff and a customer



Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

PART 2 · WHAT DID NOT KEEP UP

“the people qualified to do such [shipping] work did not multiply alongside it”



30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

PART 2 · WHAT DID NOT KEEP UP

“the people qualified to do such [shipping] work did not multiply alongside it”

> “The pace at which we have to review is fast outpacing our ability to review.”
>
> SENIOR ENGINEER · P4



30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

PART 2 · WHAT DID NOT KEEP UP

“the people qualified to do such [shipping] work did not multiply alongside it”



> “The pace at which we have to review is fast outpacing our ability to review.”

> “if I’m not reviewing, I feel like maybe I’m not a team player… I’m reviewing as fast as I can and my stuff’s slipping, right?”

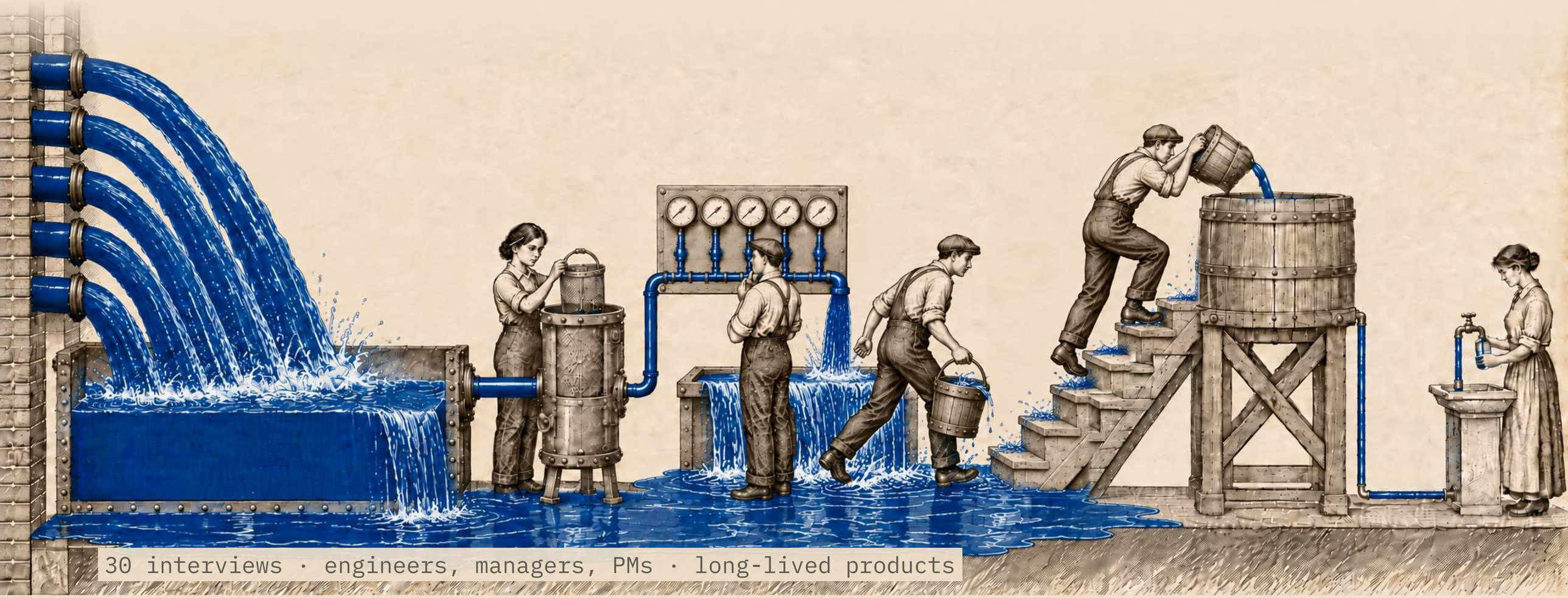
SENIOR ENGINEER · P4

30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

PART 2 · WHAT DID NOT KEEP UP

Expectations rose, and work got dropped



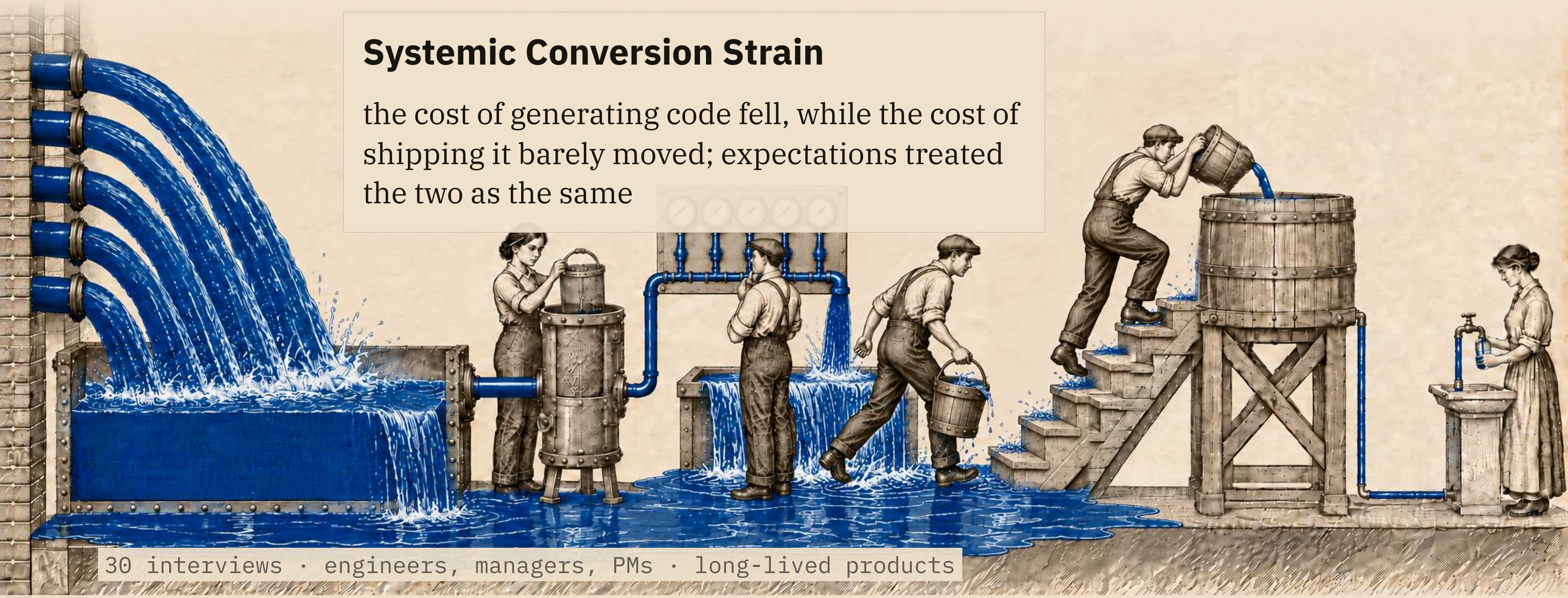
Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

PART 2 · WHAT DID NOT KEEP UP

Expectations rose, and work got dropped

Systemic Conversion Strain

the cost of generating code fell, while the cost of shipping it barely moved; expectations treated the two as the same

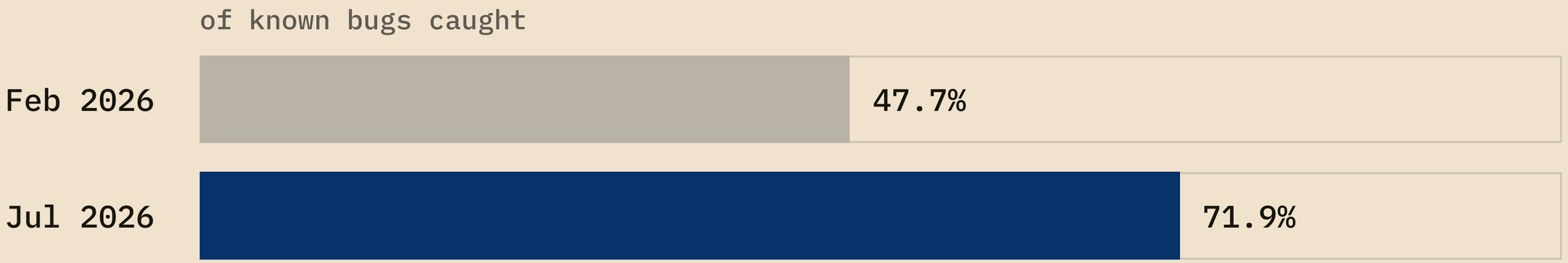


30 interviews · engineers, managers, PMs · long-lived products

Cha, Badea, Baym, Hadley, Bird, Butler, Murphy-Hill. *The Activity-Outcome Gap: Why AI Does Not Solve Software Engineering*. In submission, 2026.

PART 3 · WON’T BETTER MODELS FIX THIS?

47.7% → 71.9% in five months

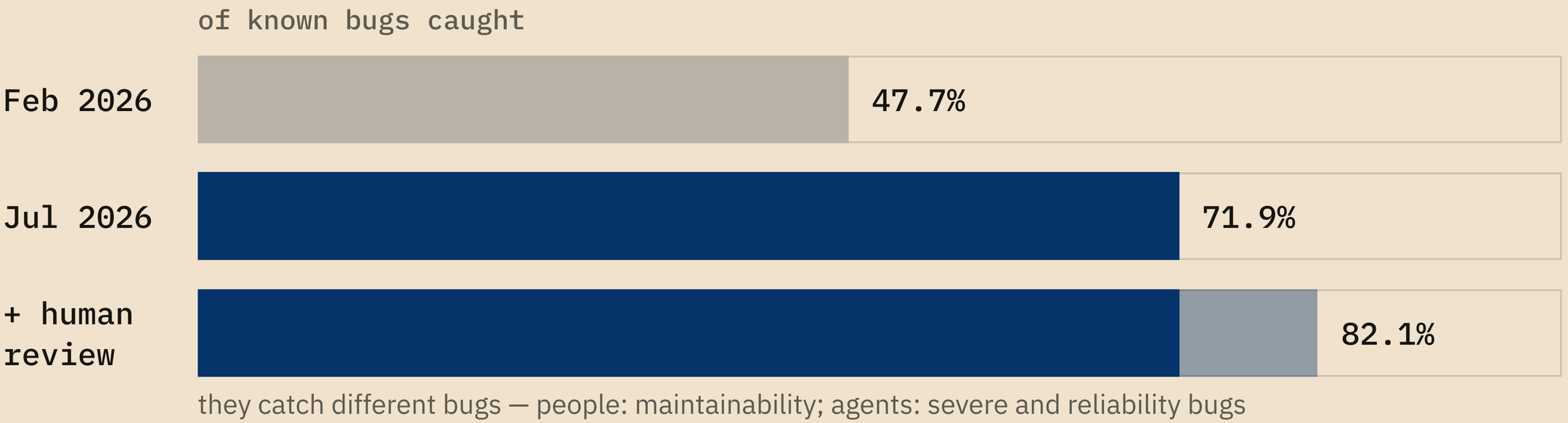


real bugs from real pull requests · Feb–Jul 2026

Yang, He, Cheng, Butler, Murphy-Hill, Badea, Bird. *(When) Should a Software Factory Be Lights-Off? Benchmarking Code Review Agents on Bug Detection*. In submission, 2026.

PART 3 · WON’T BETTER MODELS FIX THIS?

47.7% → 71.9% in five months



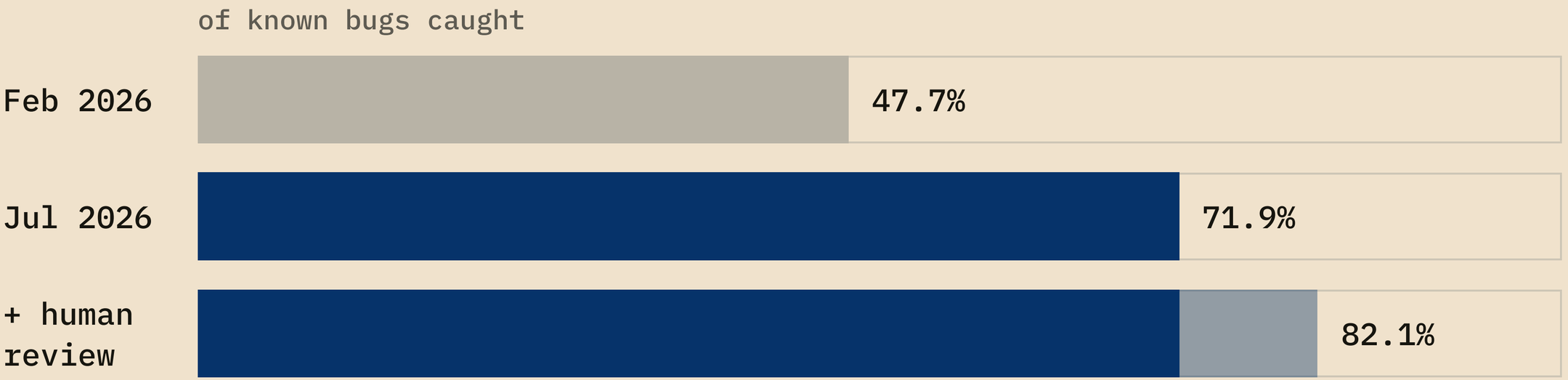
they catch different bugs — people: maintainability; agents: severe and reliability bugs

real bugs from real pull requests · Feb–Jul 2026

Yang, He, Cheng, Butler, Murphy-Hill, Badea, Bird. *(When) Should a Software Factory Be Lights-Off? Benchmarking Code Review Agents on Bug Detection*. In submission, 2026.

PART 3 · WON’T BETTER MODELS FIX THIS?

47.7% → 71.9% in five months



they catch different bugs — people: maintainability; agents: severe and reliability bugs

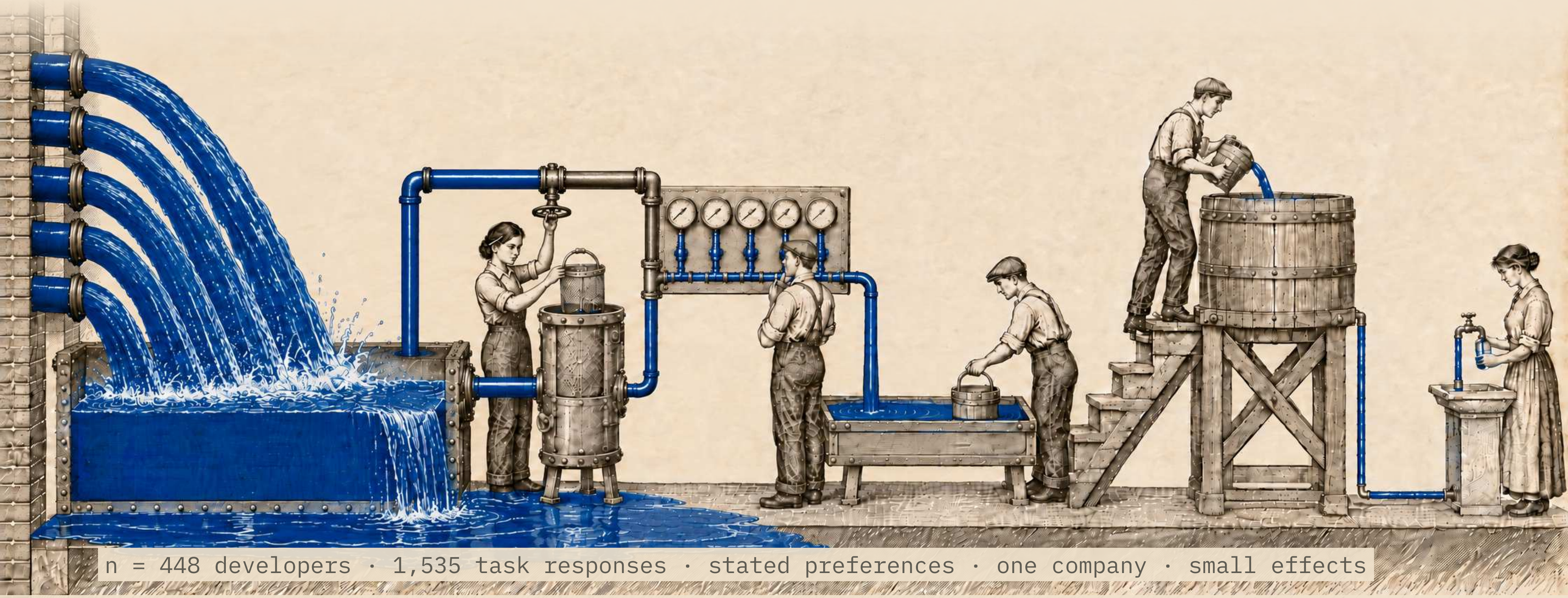
lights-off only for narrowly scoped, low-consequence PRs

real bugs from real pull requests · Feb–Jul 2026

Yang, He, Cheng, Butler, Murphy-Hill, Badea, Bird. *(When) Should a Software Factory Be Lights-Off? Benchmarking Code Review Agents on Bug Detection*. In submission, 2026.

WHY IS THERE STILL A PERSON AT THE GATE?

3 in 4 tasks: AI’s work waits for the developer’s OK

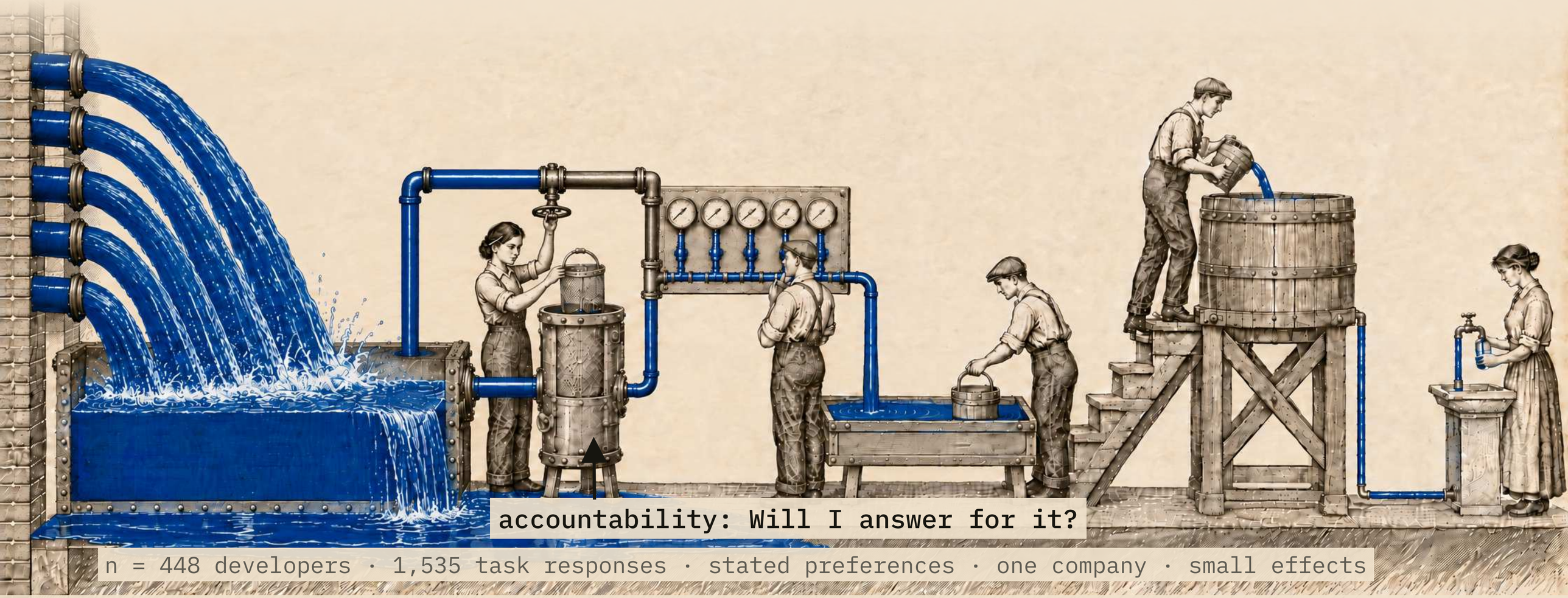


n = 448 developers · 1,535 task responses · stated preferences · one company · small effects

Choudhuri, Bird, Badea, Gerosa, Sarma. *You Shall Not Pass! Where and Why Developers Draw the Line on AI Autonomy*. arXiv:2607.00533, 2026.

WHY IS THERE STILL A PERSON AT THE GATE?

3 in 4 tasks: AI’s work waits for the developer’s OK

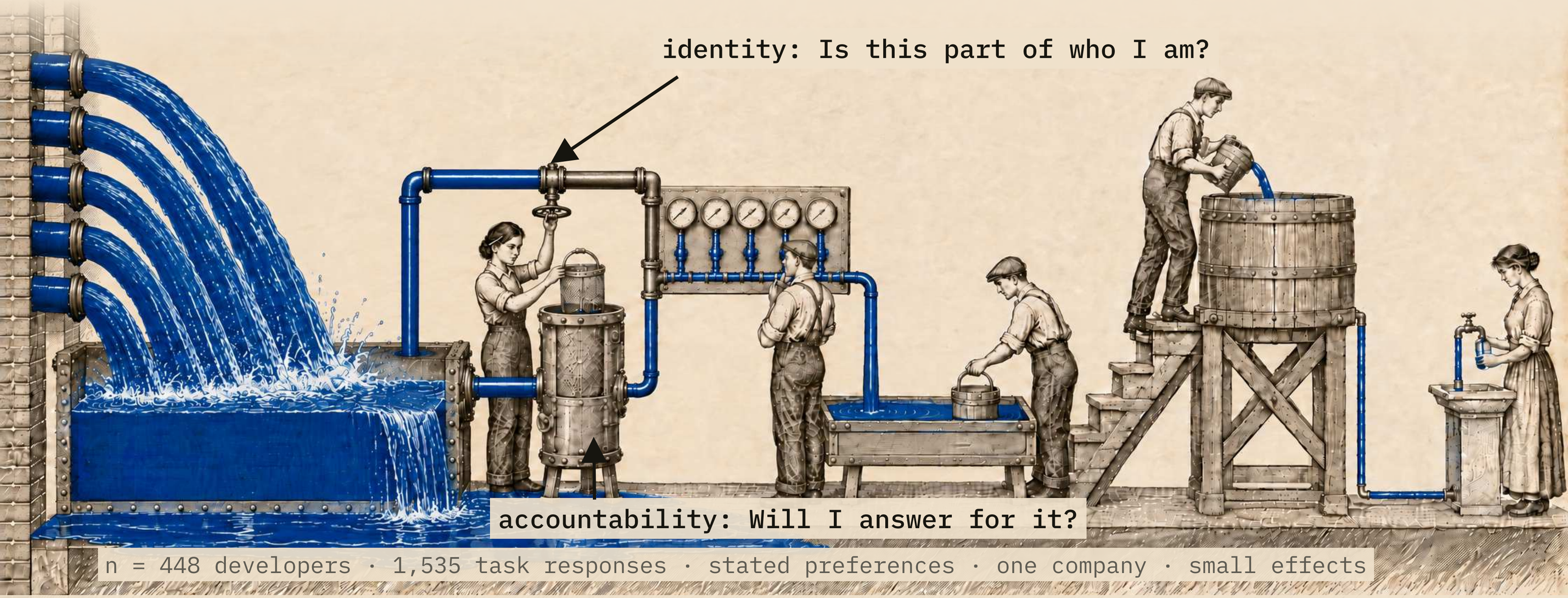


n = 448 developers · 1,535 task responses · stated preferences · one company · small effects

Choudhuri, Bird, Badea, Gerosa, Sarma. *You Shall Not Pass! Where and Why Developers Draw the Line on AI Autonomy*. arXiv:2607.00533, 2026.

WHY IS THERE STILL A PERSON AT THE GATE?

3 in 4 tasks: AI’s work waits for the developer’s OK

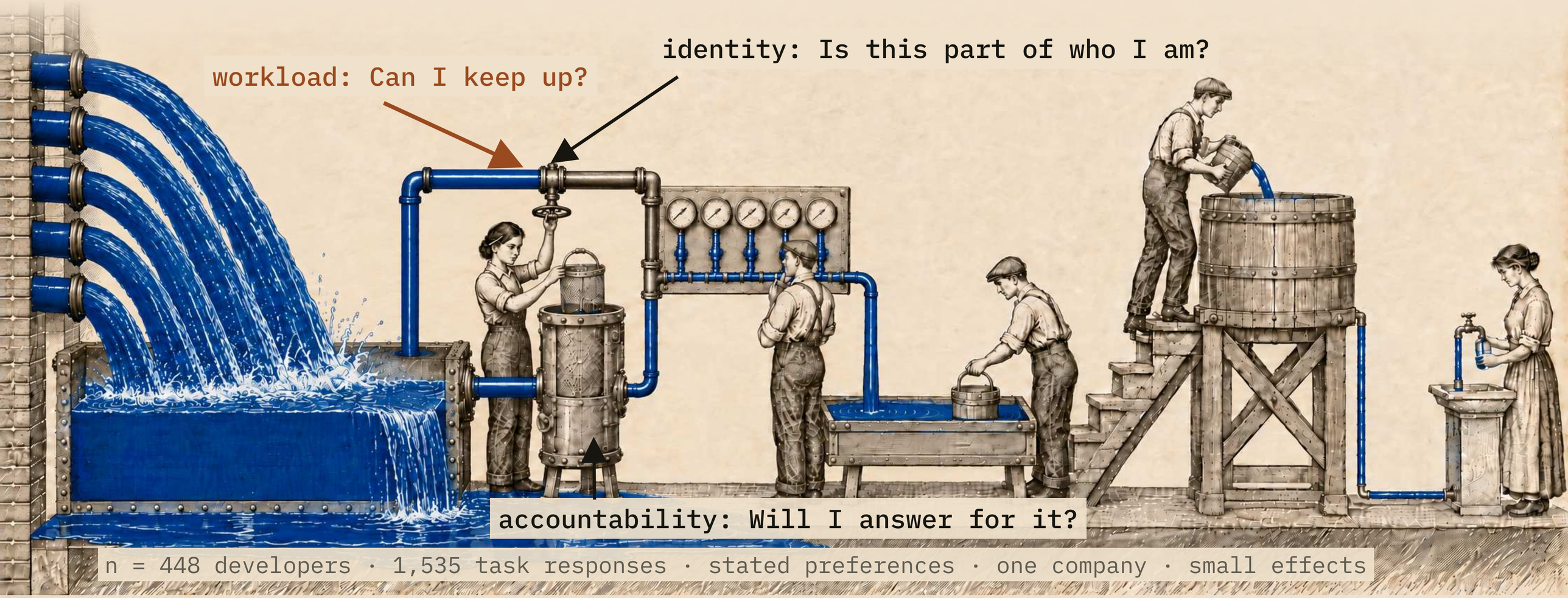


n = 448 developers · 1,535 task responses · stated preferences · one company · small effects

Choudhuri, Bird, Badea, Gerosa, Sarma. *You Shall Not Pass! Where and Why Developers Draw the Line on AI Autonomy.* arXiv:2607.00533, 2026.

WHY IS THERE STILL A PERSON AT THE GATE?

3 in 4 tasks: AI’s work waits for the developer’s OK

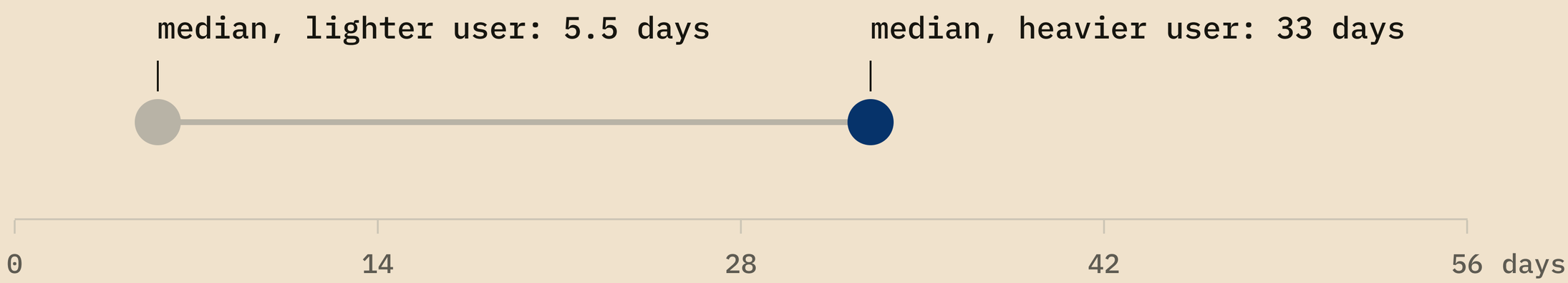


n = 448 developers · 1,535 task responses · stated preferences · one company · small effects

Choudhuri, Bird, Badea, Gerosa, Sarma. *You Shall Not Pass! Where and Why Developers Draw the Line on AI Autonomy.* arXiv:2607.00533, 2026.

NO TIME TO LEARN

27 matched pairs · same team, manager, level



27 matched pairs · 54 developers · 8-week telemetry window

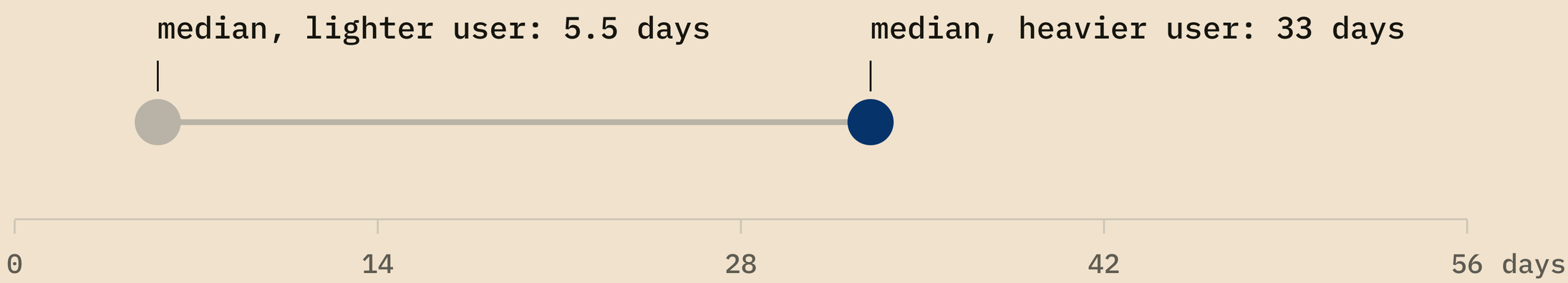
Miller, Choudhuri, Ulloa, Haniyur, DeLine, Storey, Murphy-Hill, Bird, Butler. *“Maybe We Need Some More Examples”: Individual and Team Drivers of Developer GenAI Tool Use*. ICSE 2026.

NO TIME TO LEARN

The Productivity Pressure Paradox

> developers lack the time necessary to develop the skills that would save time

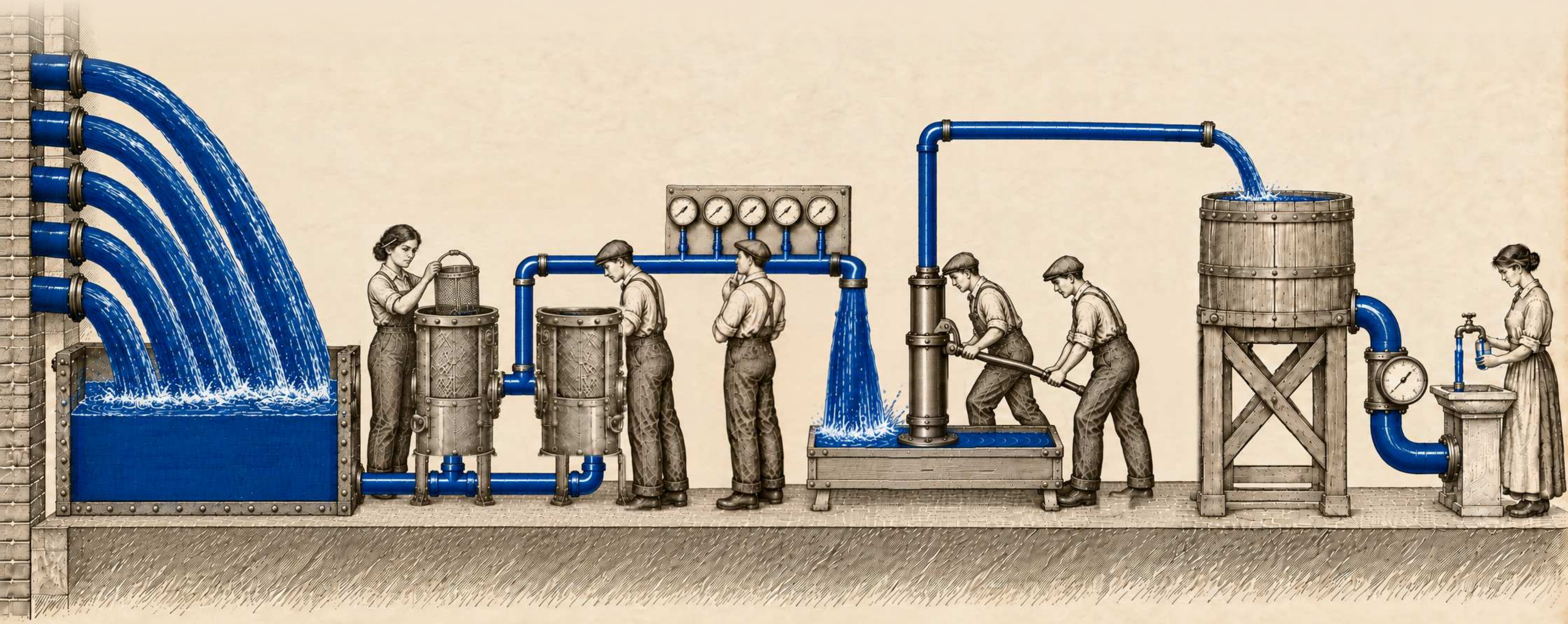
27 matched pairs · same team, manager, level



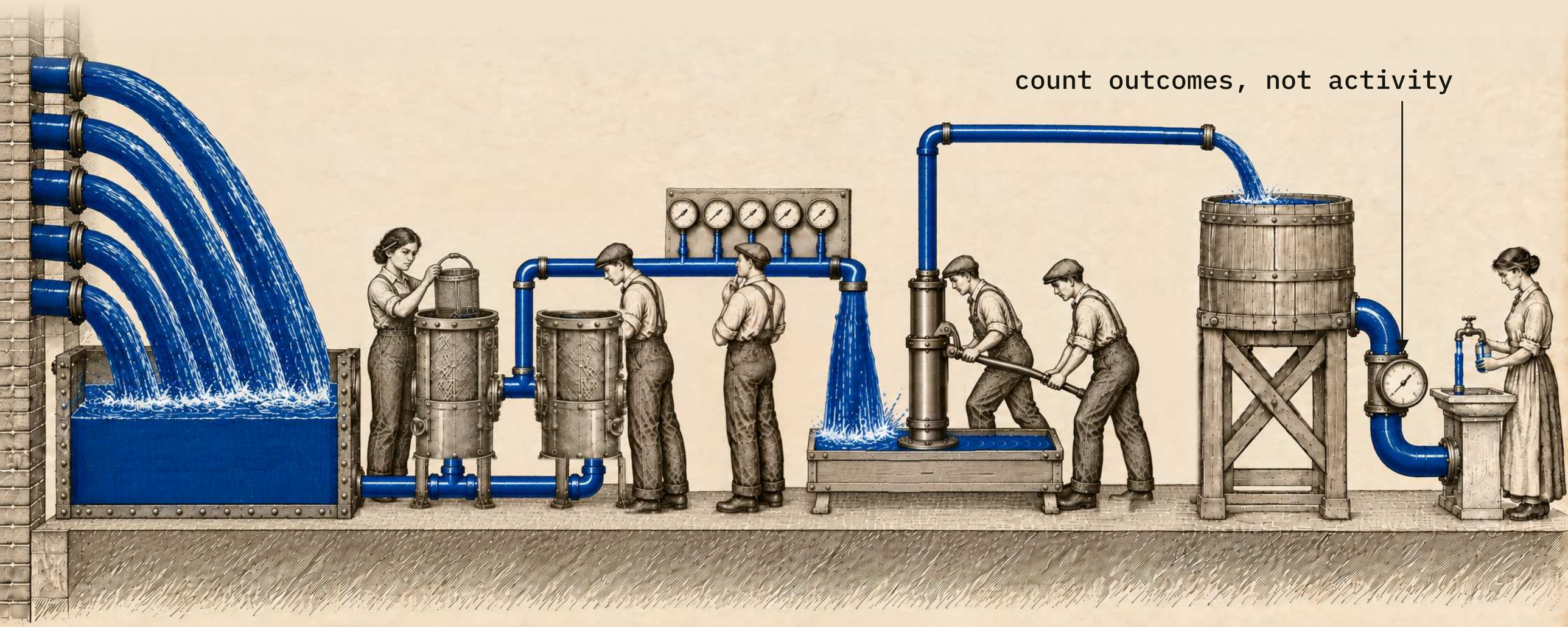
27 matched pairs · 54 developers · 8-week telemetry window

Miller, Choudhuri, Ulloa, Haniyur, DeLine, Storey, Murphy-Hill, Bird, Butler. *“Maybe We Need Some More Examples”: Individual and Team Drivers of Developer GenAI Tool Use*. ICSE 2026.

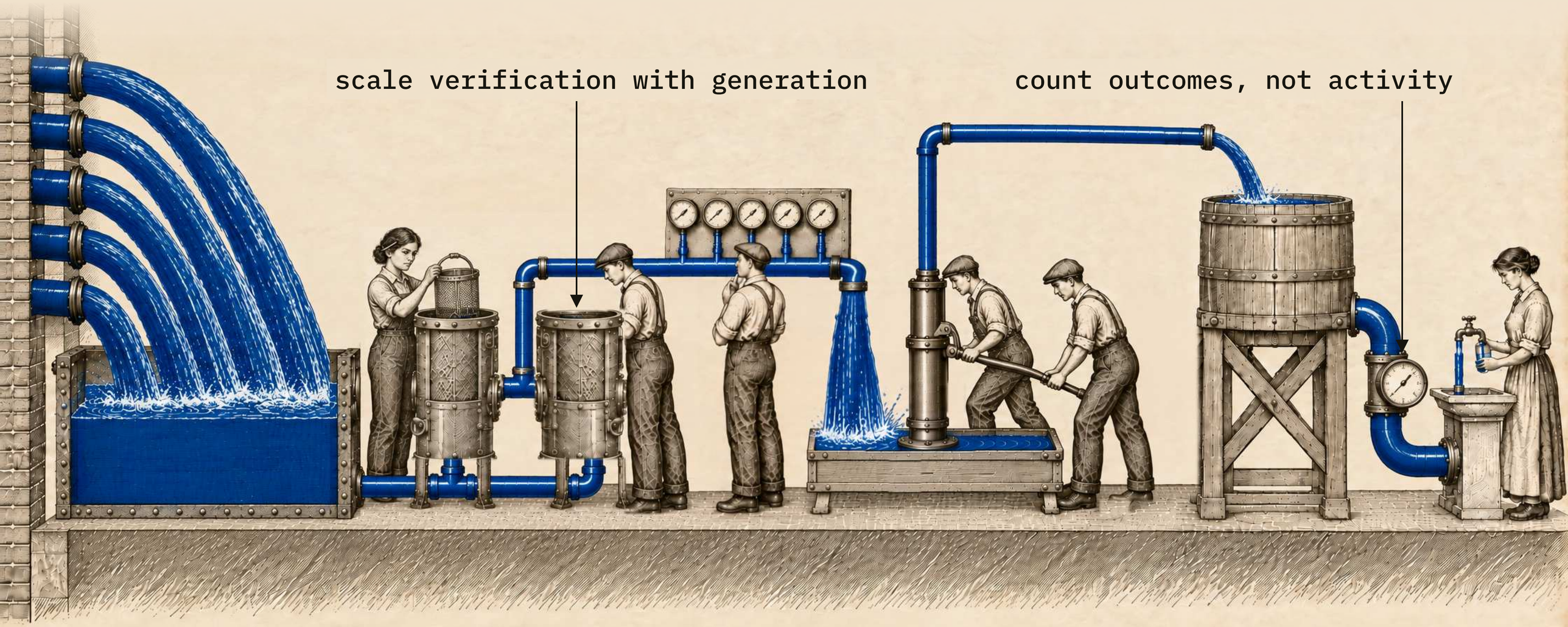
SO WHAT DO WE DO?



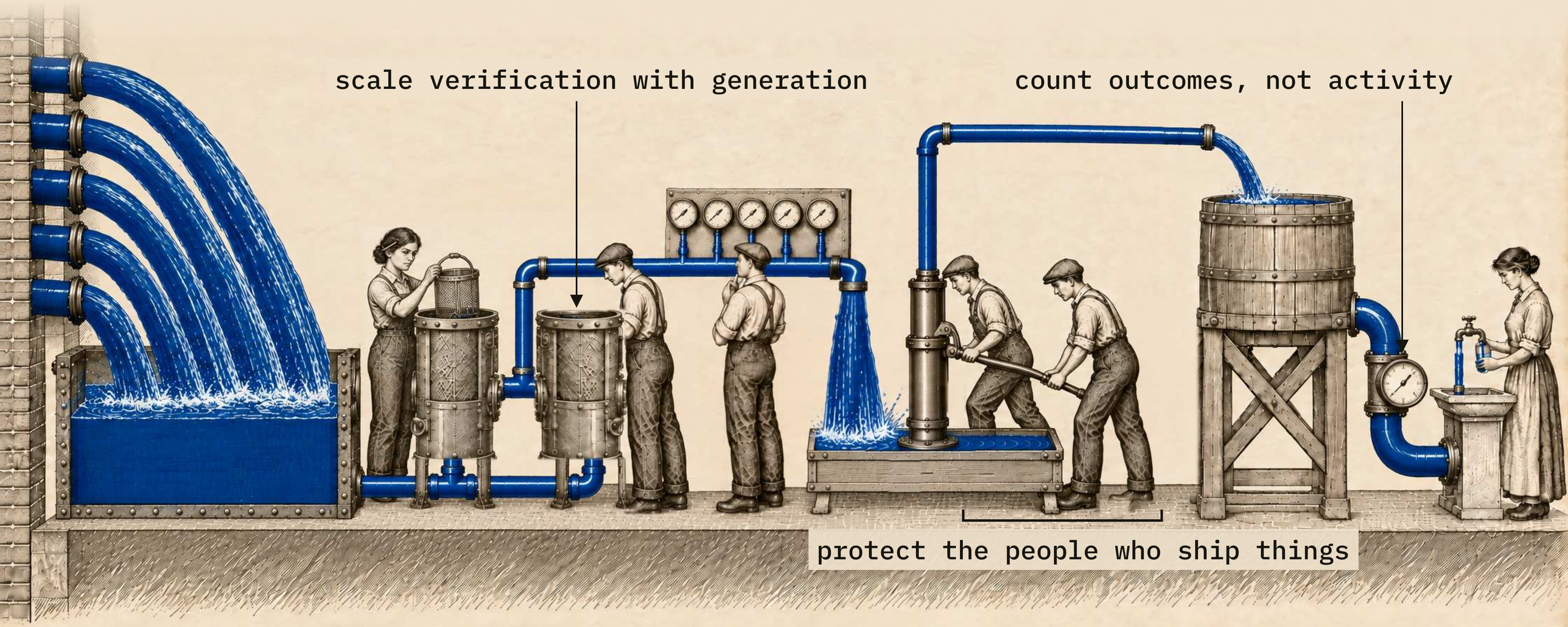
SO WHAT DO WE DO?



SO WHAT DO WE DO?

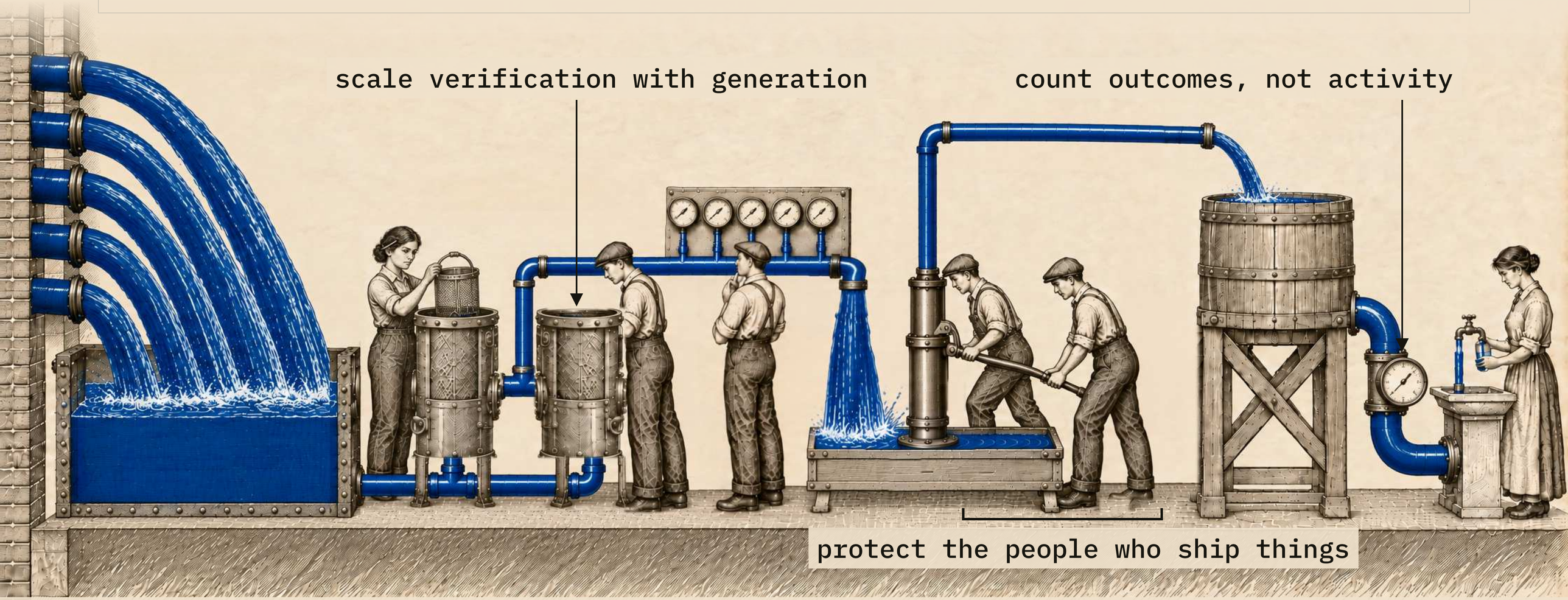


SO WHAT DO WE DO?



SO WHAT DO WE DO?

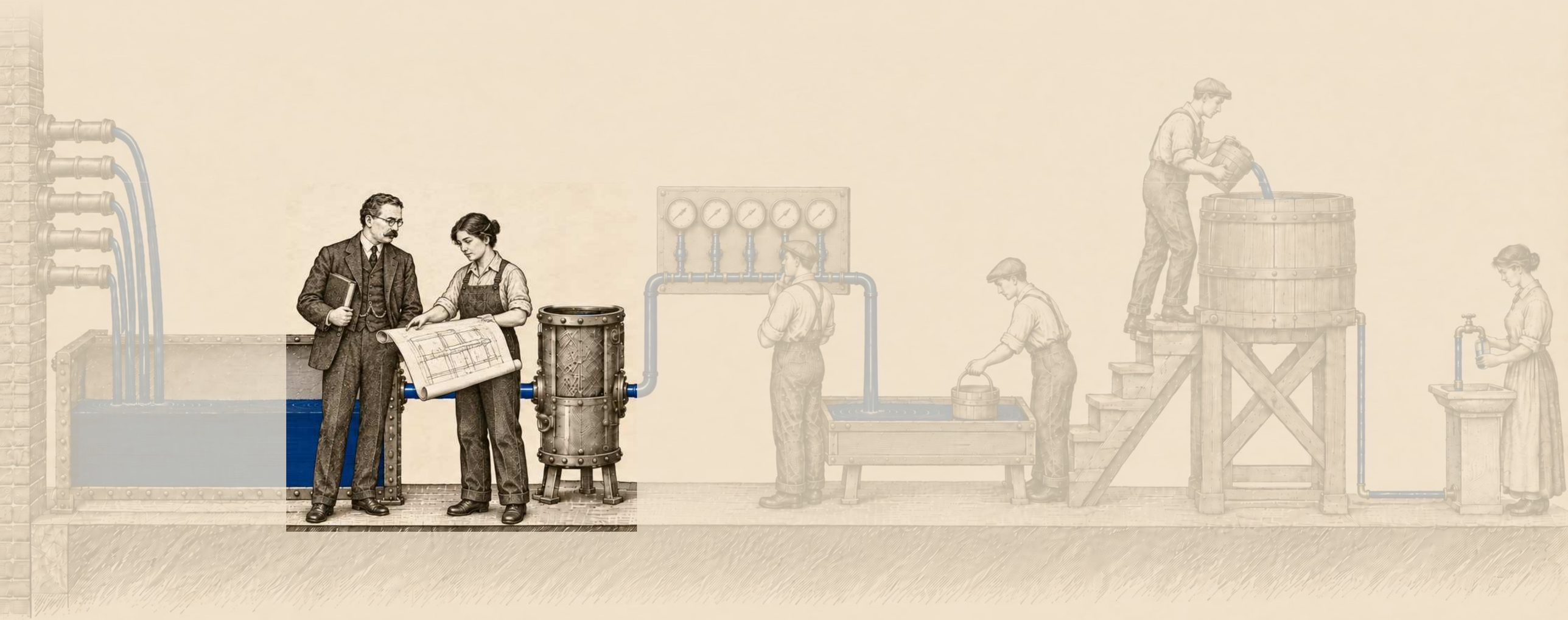
> “We have a choice to set these boundaries deliberately, through how roles, workflows, and jobs are designed, or let tool defaults reset them with each release.”



Choudhuri, Bird, Badea, Gerosa, Sarma. *You Shall Not Pass! Where and Why Developers Draw the Line on AI Autonomy*. arXiv:2607.00533, 2026.

THE PRODUCT MANAGER

Afterwards, a new respect for his engineers.



WHY AREN’T WE SHIPPING FASTER NOW?

More work is starting, by more people.

The people who ship it did not multiply, and they keep the sign-off on purpose.

We count what goes in, not what comes out.

