Your AI ROI is missing a term.

Dashboards count the time AI saves producing the work. Few count the time it takes to check it.

$$ROI = \frac{G - C - \boxed{?}}{C}$$

THE VERIFICATION BILL

review time
rework
what slips through

1.06M
merged pull requests, Oct 2024 to Feb 2026

470
organization-owned open-source repositories

1.7×
as long to merge when the change is delegated to AI

≈1.8×
as often sent back by reviewers

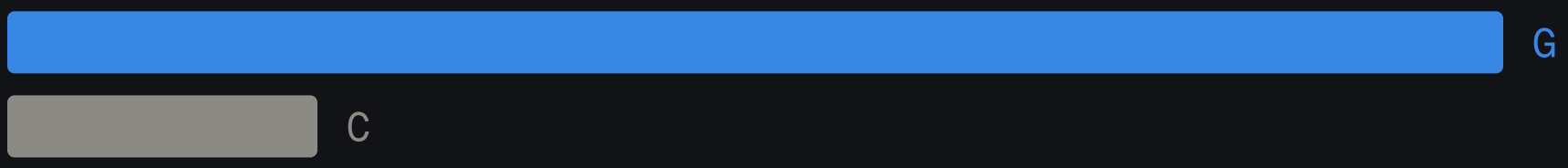
The version in the business case.

$$ROI_{dash} = \frac{G - C}{C}$$

G **Gain**
Time saved producing the work, valued at loaded cost.

C **Cost**
Licences, tokens, enablement, rollout.

WHAT THE DECK SHOWS · ILLUSTRATIVE



It assumes the output is finished the moment it is generated.

An AI agent's change is not done when it is written. It is done when someone has checked it.

Add the verification bill.

$$\mathrm{ROI}_{\mathrm{true}} = \frac{G - C - \underbrace{[V(\phi) + r \cdot R + (1-\phi) \cdot E]}_{\text{THE VERIFICATION BILL}}}{C}$$

$V(\phi)$ — **Verification**
The wait for a first human review, plus the review itself. Rises with ϕ.

$r \cdot R$ — **Rework**
Share of changes sent back, times the cost of each fix cycle.

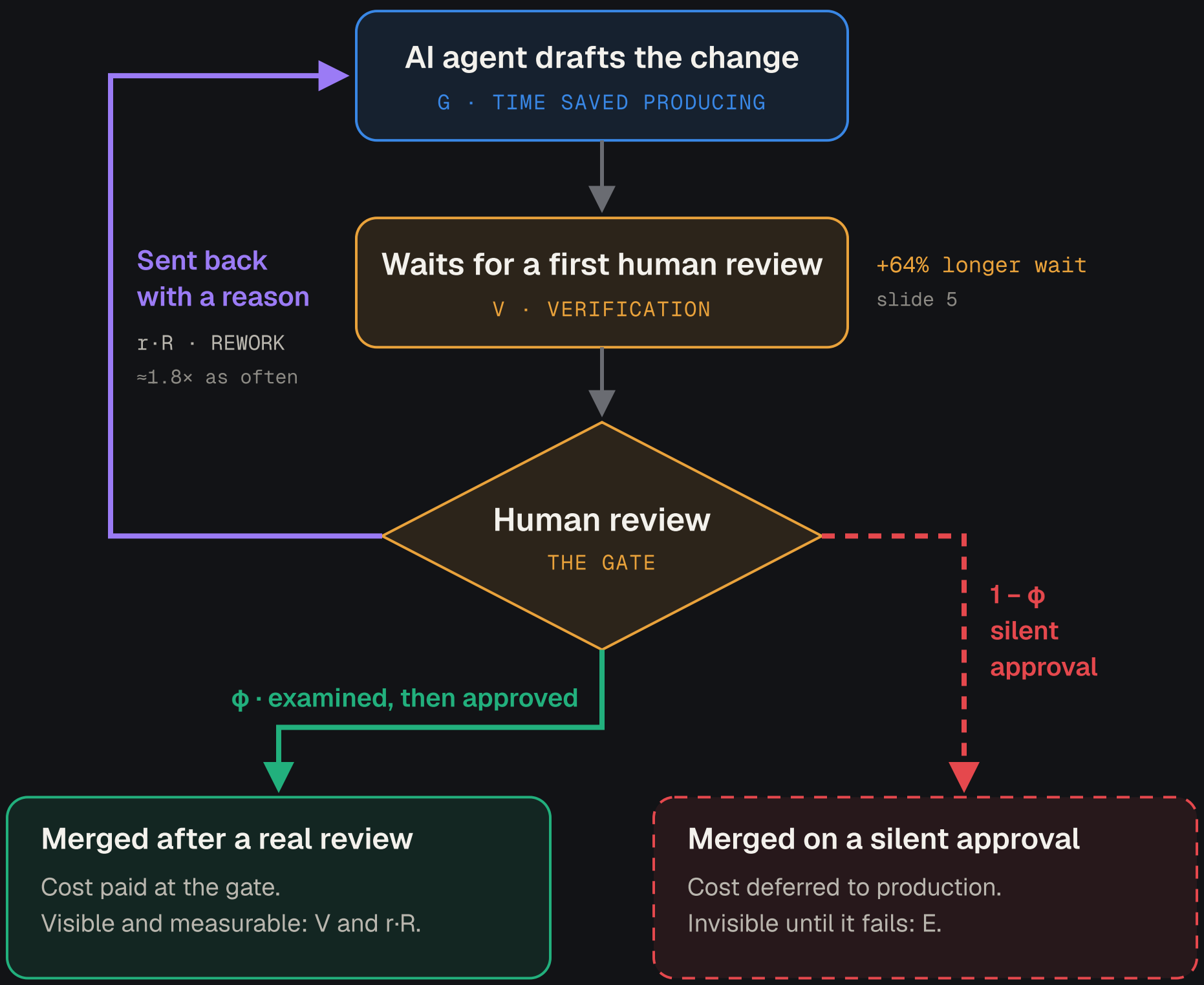
$(1-\phi) \cdot E$ — **Escapes**
Output nobody really examined, times what it costs when it fails downstream.

ϕ — **Faithful appropriation**
The share of AI output a human genuinely examines before it ships.

Lower ϕ and the visible bill shrinks. It does not disappear. It moves to E, where no dashboard looks.

Every delegated change passes one gate.

What happens there decides where the cost lands.



ϕ is faithful appropriation: the share of AI output a human genuinely examines before it ships.

An AI-delegated change vs the same developer's own, in the same repository and month.

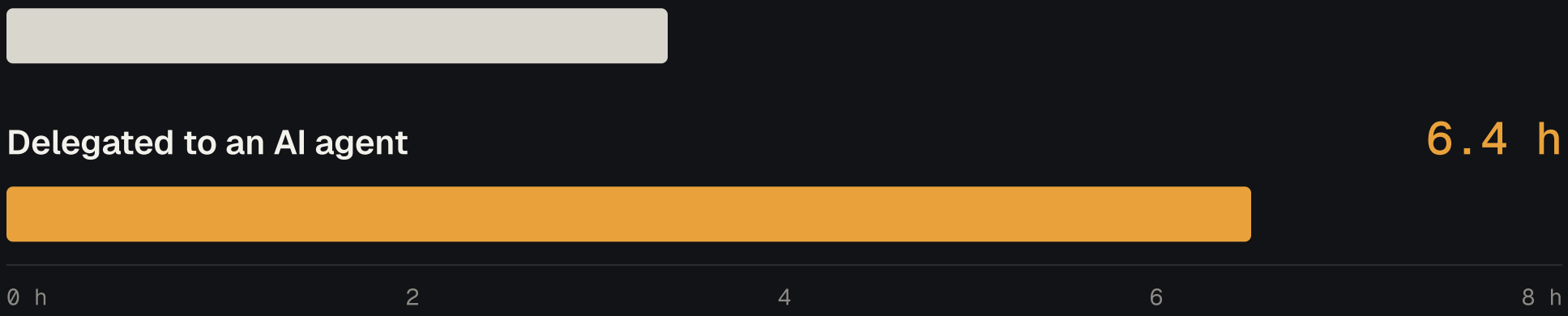
1.7×

as long to merge

+64%

longer wait for a first human review

MEDIAN WAIT FOR THE FIRST HUMAN REVIEW



HELD FIXED IN THE COMPARISON

- developer × repository × month
- task type
- complexity
- weekday × hour

Not a night-shift effect. On a business-hours clock, most of the merge-time gap remains.

Calendar time, not reviewer hours. This is V, the first term of the bill.

Aziz Camara · PhD research on AI coding tools

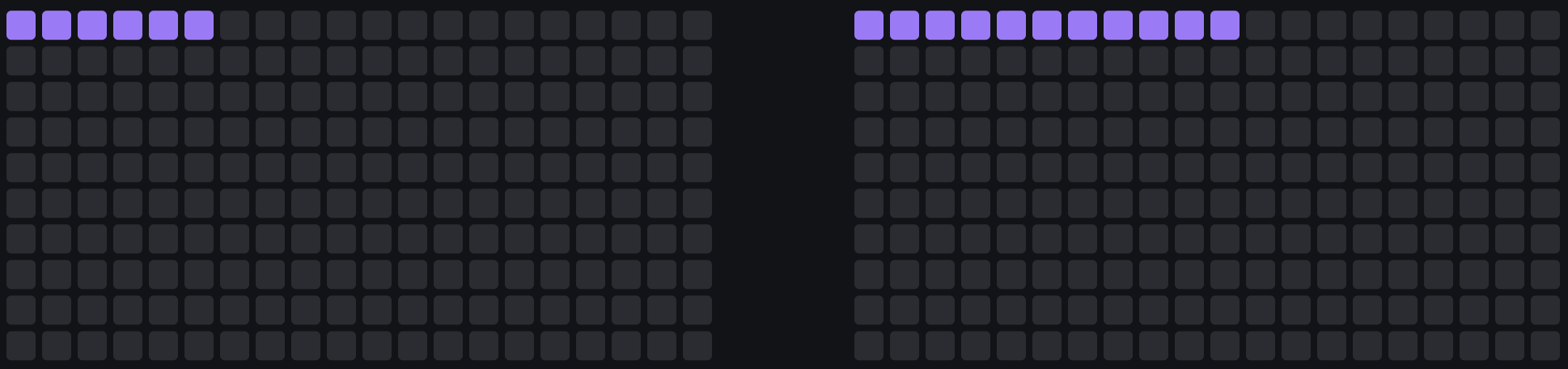
1,064,292 merged PRs · 470 repos · Oct 2024 to Feb 2026

On average, yes.

Reviewers behaved as calibrated trust predicts, not as complacency predicts.

≈1.8×

as often sent back
with changes requested



Each square = 1 in 200 pull requests · filled = sent back

−5.5 pts

silent approvals, same reviewer, same repository

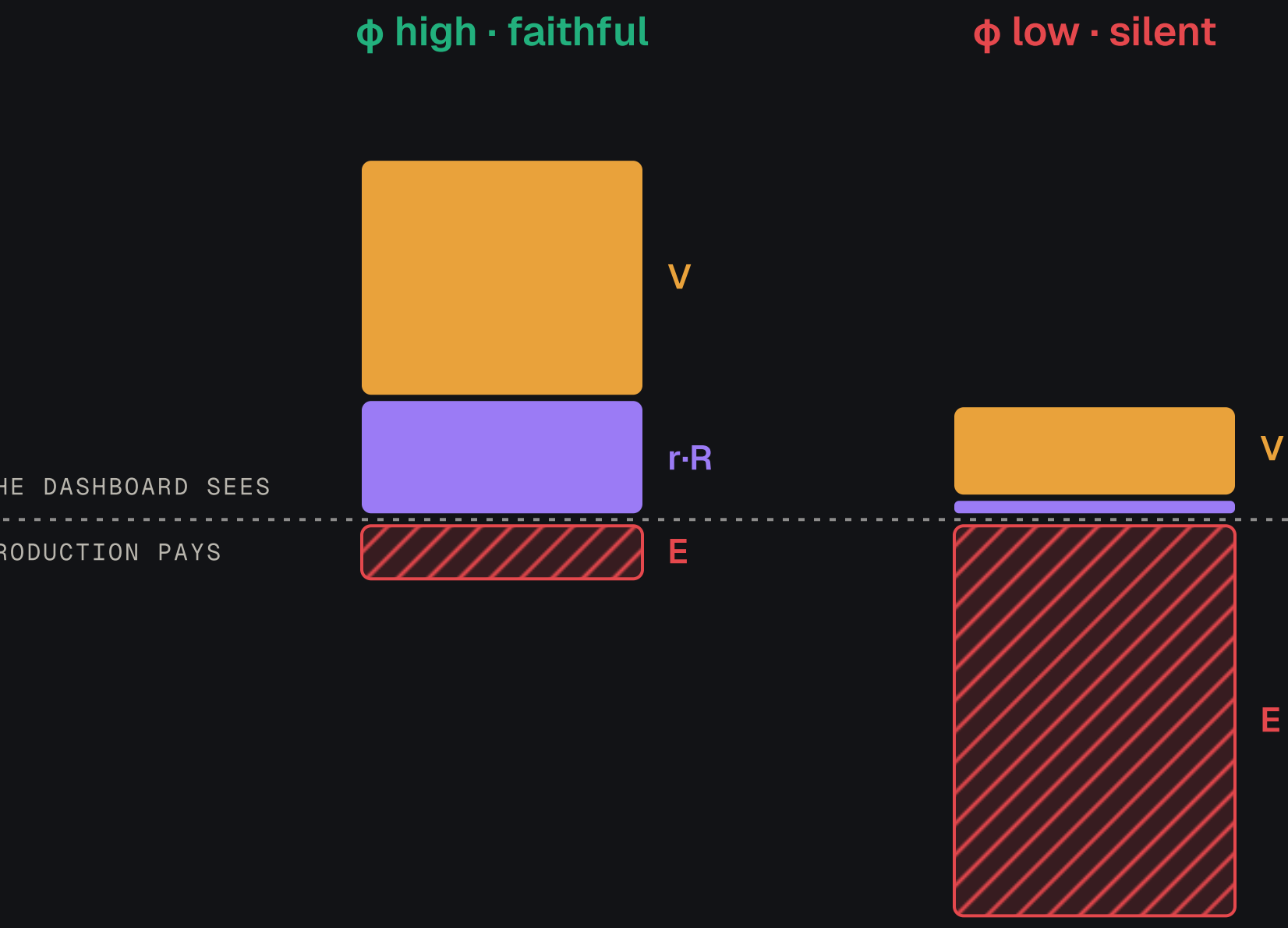
No penalty

detected in reverts after merge (equivalence not shown)

> Φ · FAITHFUL APPROPRIATION
>
> Using a tool in line with its spirit (DeSanctis & Poole, 1994). For an AI agent, that spirit is extending a developer's judgment, not replacing it. **A change sent back with a reason is faithful use. A silent approval is its absence.**

- **Early warning:** the scrutiny gap narrows as developers delegate more (an exploratory reading).

Same G. Same C.
Different bill.



Paid at the gate

ROI looks lower. It is real.

Paid in production

ROI looks higher. It is borrowed.

Cutting the checks is the fastest way to improve AI ROI on paper.

Scale delegation only as fast as you can measure it.

Compare delegated work with the same people's own work, in the same weeks.

01	Wait for the first human review	V
02	Time to merge	V
03	Share sent back	r·R
04	Share reverted or repaired after merge	E

> **Set no speed targets on delegated work.**
>
> A target on what is easy to observe rewards the reviewer who examines least.
>
> Holmström & Milgrom, 1991

WHAT MY DATA CANNOT TELL YOU

Whether the time saved producing (G) pays this bill. Production time is not in the record. **So measure it. Do not assume it.**

THE EQUATION

$$ROI_{true} = (G - C - [V(\phi) + r \cdot R + (1-\phi) \cdot E]) / C$$

Aziz Camara · PhD research, Université Paris Dauphine | PSL

GitHub data · Oct 2024 to Feb 2026