

## RESEARCH REPORT

# The AI Proficiency Gap: How Engineering Leaders Are Flying Blind on Their Biggest Investment

*Why measuring AI tool adoption is not the same as measuring AI proficiency — and why the difference costs enterprises millions.*

---

Published by **Pairbility (SameThoughts Inc.)** • February 2026 • [pairbility.com/research](https://pairbility.com/research)

Research synthesized from: *LeadDev 2025 AI Impact Report (880 engineering leaders)* • *METR Randomized Controlled Trial* • *Jellyfish (20M pull requests, 200K developers)* • *IBM Institute for Business Value* • *Cortex 2026 Engineering Report* • *DX AI Measurement Framework* • *World Economic Forum*

## . Executive Summary

---

This report synthesizes publicly available data on one of the fastest-growing blind spots in enterprise technology: the inability to measure whether AI tools are making engineering teams genuinely more productive — and whether individual engineers are developing real AI collaboration skills.

The findings are striking. While 85% of developers now use AI tools daily and organizations invest heavily in tooling subscriptions, no standardized way exists to assess whether engineers are using those tools effectively. Leaders face board-level pressure to justify AI ROI with no objective data to provide.

More alarming: controlled research shows that experienced developers using AI tools are sometimes working slower than those without them — while reporting that they feel faster. The perception-reality gap is not a minor calibration error. It is a systematic organizational blindspot.

Critically, this report finds that no established industry standard exists for measuring human-AI collaborative proficiency — not from ISO, IEEE, or any standards body. Every major research institution acknowledges the gap. No one has closed it.

## 1. Key Findings at a Glance

---

**60%**

**of engineering leaders cite lack of clear metrics as their #1 AI challenge**

*LeadDev 2025 AI Impact Report — 880 engineering leaders surveyed*

**19%**

**slower — experienced developers using AI tools took longer than peers without them in a controlled trial**

*METR Randomized Controlled Trial, July 2025 — 16 experienced open-source developers*

**73%**

**of engineering leaders say AI has changed the skills they look for when hiring — but no assessment standard exists**

*LeadDev 2025 AI Impact Report*

**77%**

**of teams formally recommend AI tools — yet the adoption process is described as ‘chaotic’ with no structured measurement**

*LeadDev 2025 AI Impact Report*

**1 in 8**

**companies are achieving both revenue gains and cost savings from AI investment — despite near-universal deployment**

*IBM Institute for Business Value*

**ZERO**

**established industry standards exist for measuring human-AI collaborative proficiency**

*ISO / IEEE / WEF landscape review, January 2026*

## 2. The Measurement Crisis

---

The single most consistent finding across all research sources is that engineering leaders are making major AI investment decisions without adequate measurement tools. The gap between AI deployment and measurable outcomes is creating organizational paralysis.

### What Gets Measured vs. What Actually Matters

Industry measurement frameworks have emerged — DORA, SPACE, DX Core 4 — but they all share a fundamental limitation: they measure adoption and output, not proficiency.

Metric Category	Who Measures It	What's Missing
Adoption rates (tool usage %)	DX, Jellyfish, Swarmia	Whether usage is effective
Lines of code generated by AI	GitHub Copilot dashboards	Whether that code is good
PR throughput / cycle time	DORA metrics, Jellyfish	Whether speed = quality
Developer satisfaction surveys	SPACE framework, DX	Objective skill measurement
Code suggestion acceptance rate	Copilot, Cursor dashboards	Whether accepted code should have been rejected

The critical gap: no existing framework measures the skill of human-AI collaboration — how well an engineer decomposes problems for AI, how critically they evaluate AI output, how strategically they manage model interactions, or how effectively they debug with AI assistance.

## 3. Engineering Leader Pain Points

### Chaotic, Uneven Adoption

Even at companies with formal AI tool programs, adoption is fragmented and deeply inconsistent. 77% of teams formally recommend AI tools, but the adoption process is described as 'chaotic' — management provides tool access but things usually stop there, with no structured onboarding or proficiency measurement. Within the same team, some engineers reach over 50% AI-assisted output while others hover near zero. Leaders cannot distinguish between them.

*“We’re coding faster, and we’re getting work done faster. But this then becomes a change management problem — how do we actually reap the benefits of that as an organization?”*

— VP of Engineering, LeadDev 2025 Survey

### The Perception-Reality Gap

Perhaps the most striking finding comes from METR’s randomized controlled trial, which revealed a dangerous disconnect between how productive developers feel with AI and how productive they actually are.



This finding directly validates the need for objective proficiency measurement. Self-reported productivity and actual productivity are inversely correlated for many AI users — creating a halo effect from the dopamine hit of instant AI suggestions that masks time spent debugging and rewriting AI-generated output.

### Quality Degradation and Technical Debt

Even where AI tools accelerate code production, quality concerns are mounting. The Cortex 2026 State of Engineering Report documents increasing change failure rates alongside velocity gains — a signal that AI is accelerating output without improving quality. Engineering leaders describe increasing code review burden, decreased codebase understanding among engineers who rely heavily on AI output, and accumulated technical debt when generated code is not properly evaluated before acceptance.

### Governance and Shadow AI

AI tool usage is outpacing organizational governance. Only 32% of companies have formal AI governance policies despite 90% active AI usage. Engineers routinely use personal AI accounts with production code, creating security and IP risks that most organizations have not yet quantified.

## 4. The Industry Standard Gap

A comprehensive review of ISO, IEEE, and major standards body outputs reveals that no established standard exists for measuring human-AI collaborative proficiency. This is not an oversight — it is a gap that every major institution has identified and none has filled.

Category	What Exists	Status
International Standards	AI management & lifecycle	None address individual AI

(ISO/IEEE)	standards	collaboration proficiency
<b>Academic Frameworks</b>	3+ proposed theoretical models	All conceptual; none operational or validated
<b>Industry Certifications</b>	CompTIA, AWS, Google Cloud certs	Focus on traditional SE or cloud ops; not collaboration
<b>Global Institutions (WEF)</b>	Public call to action for AI skill measurement	Declaring the need — not building the solution
<b>Operational Assessment Platforms</b>	NONE	<b>This is the white space</b>

*“As AI becomes increasingly embedded in professional workflows, there is a pressing need for standardized frameworks to measure and develop AI literacy across the workforce. Even soft skills — long considered immeasurable — will become quantifiable.”*

— World Economic Forum, January 2025

The WEF is calling for exactly what the research confirms is needed. Academic researchers have proposed theoretical frameworks — AIQ, HAIC — but none have been operationalized. Existing certifications from CompTIA, AWS, and cloud providers measure tool-specific knowledge, not the cognitive skill of human-AI collaboration.

The result: every organization deploying AI tools is doing so without a shared vocabulary for what ‘good’ looks like.

## 5. The Hiring Assessment Crisis

73% of engineering leaders report that AI has fundamentally changed the skills they look for when hiring. Yet the interview and assessment industry has not kept pace. The dominant formats — LeetCode-style algorithm tests, take-home assignments, live coding sessions — were designed to measure skills that are now largely delegable to AI.

The new problem is not whether a candidate can write a binary search tree. The new problem is whether they can decompose a complex problem effectively for an AI model, critically evaluate AI-generated output for correctness and security vulnerabilities, iterate intelligently when AI output is wrong or incomplete, and verify and validate AI contributions before they reach production.

None of these behaviors are measured by current assessment formats. Worse, take-home assessments are now trivially completable with AI assistance, making it nearly impossible to distinguish AI-dependent from AI-augmented engineers.

*“Prompt engineering is a learnable and trainable skill — but we have no way to assess it in candidates or measure it in our existing team.”*

— Director of Engineering, LeadDev 2025 Survey

## 6. Conclusion: The Measurement Gap Is the Opportunity

The data converges on a single conclusion: organizations are flying blind on their most significant technology investment of the decade. They have adoption metrics without proficiency metrics. They have velocity data without quality signals. They have hiring processes that cannot distinguish the candidates they actually need.

The path forward requires a new measurement category: AI collaboration proficiency — the skill of working effectively with AI tools across the full spectrum of knowledge work. This requires behavioral measurement, not self-reporting. It requires multi-dimensional frameworks, not single-metric dashboards. And it requires assessment that is embedded in real work, not artificial test environments.

The research is unambiguous. Every institution from the World Economic Forum to leading academic labs has declared the need. No one has built the solution.

### Pairbility is building the D1–D8 Framework:

*the first operational standard for measuring human-AI collaboration proficiency*

Read the framework — [pairbility.com/research](https://pairbility.com/research)

## . Sources & Methodology

All findings in this report are derived from publicly available sources. Research methodology included primary and secondary analysis across industry reports, academic publications, standards body documentation, developer forums, and enterprise case studies.

#### LeadDev 2025 AI Impact Report

[leaddev.com/the-ai-impact-report-2025](https://leaddev.com/the-ai-impact-report-2025)

Survey of 880 engineering leaders on AI adoption challenges, hiring changes, and measurement gaps.

#### METR Randomized Controlled

Controlled study with 16 experienced open-source developers showing 19%

<b>Trial (July 2025)</b> <i>metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study</i>	slowdown with AI tools despite 20% perceived speedup.
<b>Jellyfish AI Impact Analysis</b> <i>jellyfish.co</i>	Analysis across 20 million pull requests, 200,000 developers, and 1,000 companies on AI tool adoption patterns.
<b>Cortex 2026 State of Engineering Report</b> <i>cortex.io/report/engineering-in-the-age-of-ai-2026-benchmark-report</i>	Engineering health benchmarks showing increased change failure rates alongside velocity gains with AI tool adoption.
<b>IBM Institute for Business Value</b> <i>ibm.com/think/insights/ai-roi</i>	Enterprise AI ROI data: only 1 in 8 companies achieving both revenue gains and cost savings from AI investment.
<b>DX AI Measurement Framework</b> <i>getdx.com/research/measuring-ai-code-assistants-and-agents</i>	Framework developed in partnership with GitHub, Dropbox, Atlassian, and Booking.com for measuring AI developer tool impact.
<b>World Economic Forum (January 2025)</b> <i>weforum.org</i>	Global call to action for standardized AI skill and proficiency measurement frameworks.
<b>ISO / IEEE Standards Review (January 2026)</b> <i>iso.org, ieee.org</i>	Landscape review of existing AI-related standards bodies confirming zero standards for human-AI collaborative proficiency.