# Fuel Burned, Distance Traveled Unknown

*Rethinking Performance Management in an AI-Native World*



*-Amit Gupta, a1009us@gmail.com*

*LinkedIn profile: https://www.linkedin.com/in/amit-gupta-85300774/*

**23rd March 2026**

In March 2026, NVIDIA CEO Jensen Huang dropped a rhetorical bomb into the laps of engineering leaders everywhere. Speaking on the All-In Podcast and at his company's GTC conference, Huang proposed a new metric for evaluating top talent: token consumption.

> *"If that $500,000 engineer did not consume at least $250,000 worth of tokens, I am going to be deeply alarmed."*
>
> — Jensen Huang, All-In Podcast & GTC, March 2026

He added that he would essentially "go ape" if a highly paid engineer only burned $5,000 in AI compute. He likened a developer underutilizing AI to a modern chip designer stubbornly insisting on using paper and pencil. The provocation was clear: in an age where AI acts as a multiplier of human capability, the sheer volume of AI horsepower a worker deploys should become a primary indicator of their value.

This statement sparked a fierce debate about the nature of work, the measurement of productivity, and the future of human performance evaluation. Are tokens — the fundamental unit of compute consumed by large language models — the new key performance indicator (KPI)? Or is this just the latest, most sophisticated way for organizations to fall victim to Goodhart's Law, which famously states that when a measure becomes a target, it ceases to be a good measure?

As we move deeper into an AI-augmented workplace, the question is no longer just what tools we use, but how we measure the humans using them. The paradigm of performance management — built over decades through Management by Objectives (MBOs) and

Objectives and Key Results (OKRs) — is fracturing under the weight of generative AI. We are standing at an inflection point, forced to ask a profound question: **What does "performance" even mean when AI is doing "all the work"?**

---

## The Measurement Trap: Adoption vs. Consumption

To understand the current crisis in performance management, we must first look at how organizations are attempting to measure AI integration. Currently, there are two distinct philosophies emerging: the **Adoption Model** and the **Consumption Model**.

The Adoption Model, championed by companies like Amazon and GitHub, focuses on enablement. Dashboards for tools like Amazon Q and GitHub Copilot track metrics such as active users, suggestion acceptance rates, and lines of code generated. This approach asks a simple, binary question: *Did the team actually adopt the capability, and by how much?*

Adoption metrics are excellent leading indicators during a rollout phase. They tell leaders who needs training, which teams are resisting change, and whether a new tool is actually entering the business workflow. They are simple to explain, comparable across users, and relatively resistant to manipulation. However, adoption metrics are inherently shallow. A worker might "use AI" once a day to draft a trivial email and register as an active user, while another might orchestrate complex, multi-agent workflows that fundamentally transform their output. Binary adoption misses the depth of integration.

Enter the Consumption Model, as articulated by Jensen Huang. By measuring token usage, leaders attempt to capture the intensity and depth of AI interaction. Token burn reflects agentic workflows, autonomous retries, and the sheer scale of infrastructure deployed by an individual. The logic is seductive: if AI is the engine of modern work, the amount of fuel burned is a proxy for how much leverage the worker is achieving.

Yet, using token consumption as a standalone performance metric is a dangerous proposition. It fundamentally confuses input with output, measuring the compute consumed rather than the value delivered. More troublingly, it actively rewards waste. Verbose prompting, poorly tuned agents, and redundant API calls all burn more tokens. If a worker knows their performance is tied to token consumption, the incentive is to be as inefficient as possible — a textbook manifestation of Goodhart's Law. Furthermore, token usage is highly dependent on the specific model used, the vendor's pricing structure, and the nature of the employee's role, making cross-company or even cross-team comparisons virtually impossible.

| Metric Philosophy | Core Question | What It Measures | Key Vulnerability |
|---|---|---|---|
| **Adoption Model** | Are they using AI at all? | Active users, acceptance rates | Shallow; fails to capture depth of integration or value created |
| **Consumption Model** | How much AI horsepower are they deploying? | Token burn, compute cost | Confuses input with output; highly gameable; rewards inefficiency |
| **Outcome Model** | Is the work actually getting better? | Cycle time, quality, business impact | Harder to attribute; requires longer measurement horizons |

The uncomfortable truth is that none of these models, in isolation, is sufficient. And yet organizations are being pressured — by the pace of AI adoption, by competitive anxiety, and by leaders like Huang making bold public proclamations — to pick one and run with it.

---

## The Workload Paradox

Complicating the measurement debate is a growing body of evidence suggesting that our fundamental assumptions about AI and productivity may be flawed. For years, Silicon Valley sold a utopian vision: AI would automate routine tasks, freeing up human workers to focus on high-level strategy, creative problem-solving, or simply enjoying a four-day workweek.

The reality on the ground looks very different. A 2026 study by ActivTrak, which analyzed 443 million hours of work activity across more than 160,000 employees, found that AI adoption is actually *increasing* the pace and density of work. (For more on this, see my recent post: Is AI freeing me up or keeping me busier than ever?)

**+104%**
Increase in time spent on email

**+145%**
Increase in chat & messaging time

**+94%**
Increase in business management tool use

**−9%**
Decrease in deep, focused work time

*Source: ActivTrak, 2026 State of the Workplace — 443M hours analyzed across 160,000+ employees*

Similarly, an eight-month ethnography conducted by UC Berkeley Haas researchers at a technology company found that generative AI did not free up time; instead, it redistributed work into evenings, weekends, and lunch breaks. A Harvard Business Review report

corroborated this sentiment, noting that **83% of workers feel AI is increasing their overall workload.**

This creates a profound paradox for performance management. AI is being utilized as an additional productivity layer, accelerating throughput but failing to act as a substitute for existing work. If AI allows a worker to generate ten times as many reports, the organization simply demands ten times as many new reports, while the human worker spends their "saved" time managing the increased communication overhead required to coordinate this massive output.

If we evaluate employees purely on output volume in this environment, we risk driving unprecedented burnout. Leaders must recognize that an AI-augmented worker is not just a faster human; they are a human managing a localized industrial revolution. Performance management must therefore account for the cognitive load of orchestrating AI, rather than simply measuring the sheer volume of widgets produced by the human-machine hybrid.

---

## Amazon's Approach: Measuring the System, Not the Individual

If both adoption metrics and consumption metrics are flawed, and if AI is fundamentally altering the nature of workload, how should leaders evaluate performance? Interestingly, Amazon — despite its reputation for rigorous, individual-level metrics — offers a compelling alternative framework.

In July 2025, Amazon's internal developer experience team introduced a new metric called "Cost-to-Serve-Software" (CTS-SW). This metric abandons the attempt to measure individual developer productivity through lines of code or AI tool usage. Instead, CTS-SW tracks the total dollars spent per unit of software that actually reaches the customer.

Amazon recognized that the software development lifecycle is too complex, and changing too rapidly due to AI, to rely on activity-based costing. By working backward from the final output (the deployed software) and modeling all input costs, they discovered that the single largest predictor of efficiency was not individual AI usage, but **team velocity**.

> *"Our science validates that software is a team sport, and framing this as a team-level outcome instead of an individual one prevents using flow as a performance metric for individual engineers."*
>
> — Amazon Science, July 2025

This represents a vital paradigm shift. As AI commoditizes individual task execution, the true differentiator becomes system-level efficiency. An individual engineer burning $250,000 in tokens is irrelevant if the broader organizational system is bogged down by poor CI/CD practices, half-baked security, or broken deployment pipelines. Performance management in the AI era must shift its gaze from the individual node to the broader network.

---

## A New Contract for the AI Era

The integration of AI into the workforce requires a new psychological contract between employers and employees. When performance management is reduced to surveillance — tracking keystrokes, monitoring active time, or counting tokens — it erodes trust. As Jasmine Escalera, a career expert, noted regarding Amazon's recent push for employees to rigorously prove their impact:

> *"When it's talked about in terms of monitoring, surveillance... That means you don't trust your employees. You're not trying to help your employees."*
>
> — Jasmine Escalera, Career Expert, via HR Brew, January 2026

To build a sustainable performance management paradigm for the AI era, leaders must embrace a more nuanced, multi-layered approach:

1. **Decouple Telemetry from Evaluation:** Organizations should track AI adoption and token consumption strictly for operational and financial telemetry. These metrics answer questions about infrastructure costs, training needs, and budget allocation. They must never be used as direct proxies for human performance.
2. **Shift Focus from Output Volume to Outcome Quality:** Because AI makes the generation of volume trivial, measuring the sheer quantity of output (emails sent, code written, reports generated) is no longer a valid indicator of performance. Leaders must evaluate the impact of the work. Does it move the business forward? Does it solve a complex customer problem?
3. **Evaluate the Orchestrator, Not the Executor:** Employees should be evaluated on their ability to design, manage, and refine AI workflows. Performance discussions should focus on how well an employee delegates to AI, how effectively they verify AI outputs and do evals, and how they apply critical thinking to the results.
4. **Measure at the Team Level:** As individual tasks become automated, value is increasingly created in the spaces between individuals. Evaluating team-level outcomes,

as Amazon has done with CTS-SW, prevents the toxic gaming of individual metrics and encourages collaboration.

Jensen Huang is right about one thing: the engineer who refuses to utilize AI is akin to the designer clinging to paper and pencil. They will inevitably be outpaced. However, measuring the fuel they burn rather than the distance traveled — or the tokens they consume — is the wrong way to evaluate their genius.

The new paradigm of performance management must recognize that AI is not merely a tool; it is a collaborator. Our metrics must evolve to measure not how hard the human works the machine, but how effectively the human-machine partnership solves the problems that matter the most.

---

## Conclusion

> *The transition to an AI-driven workplace demands a fundamental rethinking of how we measure value and performance. Clinging to outdated metrics or simply swapping them for raw consumption data like token burn will only lead to misalignment and burnout. Leaders must instead focus on systemic efficiency, outcome quality, and the effectiveness of the human-AI partnership. Ultimately, performance management in this new era must be rooted in trust, empowering employees to navigate and orchestrate this localized industrial revolution rather than merely surviving it.*

---

### References


1. Business Insider — "Jensen Huang says he would be 'deeply alarmed' if his $500,000 engineer did not consume at least $250,000 of tokens." March 19, 2026. businessinsider.com
2. MIT Sloan Management Review — "What AI Can Teach Us About Designing Better KPIs." Spring 2026. sloanreview.mit.edu
3. ChatGPT Conversation — "Tokens as Performance Metric." chatgpt.com/s/t_69c04f7585ac8191853b1f27b47f1550
4. ActivTrak — "2026 State of the Workplace: AI is Accelerating Work, Not Replacing It." March 12, 2026. activtrak.com
5. Fortune / UC Berkeley Haas — "In the workforce, AI is having the opposite effect it was supposed to." February 10, 2026. fortune.com
6. Harvard Business Review — "AI Doesn't Reduce Work — It Intensifies It." February 9, 2026. hbr.org
7. Amazon Science — "Measuring the effectiveness of software development tools and practices." July 29, 2025. amazon.science

8. HR Brew — "Amazon wants employees to show their impact in performance reviews." January 21, 2026. hr-brew.com