WHITE PAPER WORKFORCE PLANNING FOR ENTERPRISES

Measuring AI Spend by Role: A Method

A token count shows how much AI a team used. This paper sets out how to join it to what each role's work is and what the role produced, which usage data each AI vendor gives an administrator, and how to read the result without overclaiming.

By Mahesh Tatiparthy · November 11, 2026 · 24 min read

Most large companies can now say what they spend on AI. Deloitte surveyed 3,500 leaders in 25 countries in May and June 2026 and found that 71% of their companies track total AI inference and API spend and 51% track token volume (Deloitte, October 29, 2026). A token is the unit AI vendors bill by, roughly a piece of a word. A token count tells a finance team how much AI a group used. With the vendor's prices or cost report it also gives the cost. It does not tell anyone whether the work in that group suits AI, or whether the group produced more because of it.

This paper sets out a method for closing that gap. It joins three feeds for each role: token use, the share of the role's work AI can do, and the role's own output measure. It then places every role on a simple map that a CFO and a CHRO can both read. The paper also documents which usage data each major AI vendor gives a company administrator, because that data sets the limit on what can be measured.

KEY TAKEAWAYS

- A token count measures input. Amazon's guidance for its software engineers said of input measures such as token use that they are "disconnected from what matters" (Amazon, October 27, 2026). The method in this paper sets the count beside the role's work and the role's output.

- Vendors report different things. On official documentation read on November 11, 2026, Anthropic, the OpenAI API and Cursor document token counts that can be tied to a user. The Microsoft 365 Copilot usage report documents activity dates per user and no tokens. The GitHub Copilot and Google Gemini pages do not mention tokens.

- The map has two axes: token use, and the share of the role's work AI can do. High use where few tasks suit AI is spend to question. Low use where many tasks suit AI is the place to train and roll out first.

- Output decides whether the right place is paying off. Each role needs its own measure, read from the system where the work happens and paired with a quality measure.

- The result is an association. GitHub says its own cohort comparisons "show associations, not proof" (GitHub Docs). Report by role and team. In a performance review, a token count on its own is not evidence about a person. Read beside the role's task score and its output, it is a question for the manager to ask.

What does a token count tell a leader, and what does it leave out?

A token count answers one question well: which teams are using the AI tools the company pays for, and how much. Turning the count into a cost takes the vendor's prices or its cost report.

The count stops there. In guidance for software engineers published on October 27, 2026, an Amazon vice president and distinguished engineer wrote that input measures such as token use are "deeply flawed", mainly because they are "disconnected from what matters" (Amazon). The same guidance adds that if a team's work does not benefit from current AI tools, "that is a finding, not a failing". That sentence describes the first gap in a token count: it carries no information about whether the work suits AI.

The second gap is output. Uber's president and chief operating officer, Andrew Macdonald, said of the connection between the company's AI spending and its results: "That link is not there yet" (Fortune, August 7, 2026). Fortune's reporter also wrote that Uber had ranked engineers by AI use on leaderboards, which is the reporter's account and was not a company statement. People ranked by use have a reason to run AI on work that does not need it. A leader therefore needs two more facts beside every token count: what the role's work is, and what the role produced.

What are the three feeds?

Feed one is token use. It comes from the usage reports of the AI tools the company already pays for. The fields vary by product. Some reports give tokens, the model used and cost. Others give activity records only, as the next section shows.

Feed two is the role's tasks. Every role is a bundle of tasks, and AI can do some of them and cannot do others. JobRoute scores each task in a role

from 0 to 100 for AI exposure. It uses O*NET, the US Department of Labor's public list of occupations and their tasks, and the Anthropic Economic Index, which records the tasks people use Claude for. The role's task exposure is the average of its task scores, with more important tasks counting for more. The same data records which class of model each task needs, from the smallest and cheapest class to the largest. The formula and sources are published on the JobRoute methodology page. This feed answers the question a token count cannot: how much of this role's work can AI do at all?

Feed three is the role's output measure. It is the number the team already uses to judge its own work, read from the system where the work happens. The company chooses the measure for each role, because the company knows which number its managers trust.

The records are matched by role and team for the same reporting period, usually a month. A company roster with job titles connects a user in a usage report to a role. From that point on the method works with groups. JobRoute joins the three feeds for each role and sets them side by side, as described on the AI spend by role page.

Which usage data does each AI vendor expose to an administrator?

The first feed depends on what each vendor reports. The table below records what the vendors' own documentation said on November 11, 2026. An API, in the product names below, is the interface a company's administrator uses to pull a report. Where a page gives activity counts and no tokens, the table says so. Products change often, so check the linked page before building on any row.

PRODUCT AND REPORT	TIED TO A USER?	TOKEN COUNTS?	WHAT ELSE IT REPORTS
Claude Enterprise Analytics API	Yes, daily	Yes, with cost, on usage-based Enterprise plans	Conversations, messages, projects, Claude Code and Cowork activity
Claude Code Analytics API	Yes, daily, by email or API key name	Yes, by model, with estimated cost	Sessions, lines of code, commits, pull requests
Anthropic Usage and Cost API (legacy)	No	Yes, with cost	API key, workspace and model
OpenAI API usage	Yes, when grouped by user	Yes: input, output and cached	Request counts; project, API key and model
ChatGPT Enterprise	Not confirmed	Not confirmed	A Compliance API with raw records for audit
Microsoft 365 Copilot usage report	Yes	No	Last activity date for each Copilot app
GitHub Copilot usage metrics	Yes, daily	Not mentioned; an AI credits figure per user	Adoption, acceptance, lines of code, pull request activity
Cursor Admin API, daily usage	Yes, daily	No	Lines added and accepted, request counts, most used model
Cursor Admin API, usage events	Yes, per event	Yes, with amount charged	Model, repository
Google Workspace Gemini audit logs	Yes, per event	Not mentioned	App, feature and action

Usage data documented for company administrators, by AI product, as read on November 11, 2026

Source: Official documentation from Anthropic, OpenAI, Microsoft, GitHub, Cursor and Google, linked in the paragraphs below

Anthropic. The Claude Enterprise Analytics API provides "per-user daily metrics" for chat, Claude Code and Cowork, with data from January 1, 2026. Its cost and usage reports give "per-user and organization-level token usage and cost over time" on usage-based Enterprise plans. On seat-based plans the documentation says the same endpoints "reflect usage credits only". The Claude Code Analytics API returns one day at a time for each user or API key, with input, output and cache tokens by model and an estimated cost. It does not include people who sign in to Claude Code with a Claude subscription plan. The older Usage and Cost API, now marked legacy, reports tokens and cost by API key, workspace and model and has no user dimension. Its documentation points to a successor, the Usage and Cost Analytics API, for cost per user in Claude Code.

OpenAI. The organization usage endpoints of the OpenAI API can group usage by user, beside project, API key and model, and return input, output and cached token counts by the minute, hour or day. That covers use of the API. For ChatGPT Enterprise, the governance documentation describes a Compliance API that supplies raw records for audit and is "not an adoption or productivity dashboard". The help pages that would list the fields in the per-user analytics export could not be opened for this paper, so the table marks ChatGPT Enterprise as not confirmed.

Microsoft 365 Copilot. The Graph usage report, on a page updated August 26, 2026, returns one row per user with the last activity date in each Copilot app: Teams, Word, Excel, PowerPoint, Outlook, OneNote, Loop and Copilot Chat. It returns no token count and no prompt count. Microsoft also documents a Consumption Dashboard that shows how "Copilot Credits and GitHub AI Credits" are consumed across agents, services and users. Credits are a Microsoft unit, and the pages read for this paper give no conversion to tokens.

GitHub Copilot. The usage metrics documentation describes user-level daily reports, available within two full UTC days. The REST endpoints return download links to daily and 28-day user reports, with history from October 10, 2025. Neither page mentions tokens. On June 19, 2026 GitHub added `ai_credits_used` to each user record and described it as "a metrics signal for analyzing consumption, not a billed total" (GitHub changelog).

Cursor. The Admin API has two relevant reports. Daily usage data gives one record per user per day with lines added and accepted, request counts and the most used model, and no tokens. Filtered usage events give one record per event, with input, output and cache token counts for token-based calls and the amount charged in cents.

Google Workspace with Gemini. Google announced on July 1, 2025 that Gemini audit logs are available through the Reporting API, showing the app where Gemini was used, the feature and the action for each user (Google Workspace Updates). The post does not mention tokens.

Read together, the documentation sorts the products into three kinds of record: tokens with cost, credits in a vendor's own unit, and activity only. These cannot be added together. A role that works mostly in an activity-only tool has no token figure at all, and the method has to say so instead of showing a zero.

WHAT USAGE REPORTS CANNOT SEE

The reports above cover tools the company pays for and administers. They may omit work done through personal AI accounts. A role that looks like a low user may be using tools the company cannot see, so treat low use as a reason to talk to the team before acting.

How do you turn usage records into a figure for a role?

Start with the roster. Where a usage record identifies a user or an account, the roster connects that identity to a job title and a team, and the job title maps to an occupation with a task list. Records that belong to service accounts and API keys go to the team that owns them or to a separate line for automated work, so that an overnight batch job does not appear as one very busy person.

Then aggregate. Sum the use for each role and team over the period and divide by the number of people in the group, so that a team of 40 and a team of 8 can be compared. Keep each vendor's unit separate. Tokens from one model are also not equal to tokens from another, because larger models usually cost more per token, so cost is the better common measure wherever the vendor reports it.

"High" and "low" are relative. A practical rule is to rank each role against the company's own median for the same tool and period. AI use is spread unevenly across occupations. In Anthropic's June 2026 Economic Index, computer and mathematical occupations were "roughly 30% of survey respondents" among the Claude users surveyed, against a 4% share of US employment (Anthropic, June 26, 2026). The map therefore does not ask every role to match engineering. It asks whether a role's use is in line with how much of its work AI can do.

For roles whose main tool reports activity only, record the share of the group that was active in the period and label it as activity. That is a weaker signal than tokens, and the readout should say so.

Finally, set a minimum group size and do not report below it. Small groups make the numbers unstable and make individuals identifiable. Microsoft's Copilot Business Impact report handles both concerns: it has a "minimum group size" setting that follows company policy, and for the analysis it asks

organizations to include at least 65 Copilot users and 65 other users (Microsoft Learn). The right floor for a role-level report is a decision for the company's HR and privacy leads.

How do you choose an output measure for each role?

Engineering has an obvious candidate in the pull request, a proposed change to software that a colleague reviews before it is merged. Other roles need their own measure. A good one meets five tests.

1. The team already tracks it. A measure invented for the AI project has no history to compare against.
2. It counts finished work. Tickets resolved is finished work. Messages sent is activity.
3. The system where the work happens records it. A figure typed into a spreadsheet at month end is easier to bend than one the help desk or the finance system logs by itself.
4. The role can influence it. Revenue is a poor measure for a support agent because too much else moves it.
5. It has a quality partner. Volume paired with reopen rate, error rate or review findings shows whether more work also means worse work.

ROLE	OUTPUT MEASURE	QUALITY PARTNER	READ FROM
Software developer	Pull requests merged, review time	Defects, rework	The code hosting system
Support agent	Tickets resolved, time to resolve	Reopened tickets, customer rating	The help desk
Finance analyst	Days to close, invoices processed	Corrections after close	The finance system
Paralegal	Documents drafted, matters closed	Review findings	The document and matter systems
Sales representative	Meetings booked, pipeline created	Conversion to the next stage	The customer relationship system (CRM)

Example output measures a company can start with. The company chooses the measure for each role.

Source: JobRoute analysis

Research on single occupations supports measuring each role in its own unit. A study of 5,179 customer support agents measured "issues resolved per hour" and found a 14% average rise with an AI assistant, and 34% for novice and low-skilled workers (Brynjolfsson, Li and Raymond, NBER Working Paper 31161). A Stanford working paper on 79 accounting firms reported a "7.5-day reduction in monthly close time" among firms that adopted AI, and describes itself as early evidence (Choi and Xie, May 2025). A study with upper-level law students found productivity gains of 50% to 130% in five of six legal tasks (Schwarcz and co-authors, 2026). None of these figures should be read as a forecast for another company.

Measured output matters more than reported output. In a 2025 trial, 16 experienced open-source developers took 19% longer on their tasks when AI

was allowed, and afterwards “still believed AI had sped them up by 20%” (METR, July 10, 2025). METR’s February 2026 update found different speeds with newer tools. The gap between belief and measurement is what matters for method: a survey that asks people whether AI helps cannot stand in for a number from the system of record.

How does the map place a role?

The map uses two of the three feeds. The vertical axis is token use, high or low against the company’s own median. The horizontal axis is the share of the role’s work AI can do, taken from the role’s task exposure. Each role and team lands in one of four cells.

	FEW OF THE ROLE’S TASKS SUIT AI	MANY OF THE ROLE’S TASKS SUIT AI
High token use	Question the spend	Right place. Check output and model size
Low token use	Leave alone	Missed. Train and roll out first

The map: token use against the share of the role's work AI can do

Source: JobRoute method

The third feed, output, is then read for every cell. It does the most work in the top right cell, where it separates use that is paying off from use that is not. The readout for a leader is one line per role, ending in a plain reading of what the three feeds say together.

Where the company draws the line between “few” and “many” is a choice. State the cut used and keep it fixed between periods.

How do you read each part of the map?

High use, few tasks fit: question the spend. A role is using a lot of AI and little of its work suits it. The team may have found a use the task data does not capture: the methodology page notes that the Anthropic Economic Index reflects Claude use and that many O*NET tasks have no direct observation in it. The spend may also come from one automated job. Or the spend may simply not be helping. Ask the team what the tool is used for and look at the output measure before deciding.

Low use, many tasks fit: missed. Much of this role's work is work AI can do, and the role is doing it without AI. A usage dashboard on its own does not surface this cell, because it has no view of the work. Possible explanations include no licence, no training, a tool that does not fit the systems the role works in, or use on personal accounts. This is where training and rollout go first.

High use, many tasks fit: right place. Two checks remain. The first is output: did the measure move in the intended direction, and did quality hold? If so, the spend has a case. If the measure is flat, people are using AI on suitable work and the result has not shown up. One possible explanation is the workflow around the tool, where review or a later step absorbs the saved time. The second check is model size. JobRoute records which class of model each task needs, and a team that sends routine tasks to the largest model may be paying more per token than the work requires. Model choice varies widely by product: the June 2026 Economic Index reports that 54% of Claude Code use is "served by Opus, against 10% of chat and Cowork conversations" (Anthropic).

Low use, few tasks fit: leave alone. An adoption target for this role would only produce use for its own sake. Revisit the cell when the task data is updated, because what AI can do changes.

Amazon's guidance points leaders at the same two tails. It asks them to watch the 10th percentile of AI adoption, to find teams that are not using AI effectively, and the 90th percentile, to check whether heavy users see delivery increases to match (Amazon). The map adds a view of whether the work suits AI.

What statistical cautions apply?

The join produces a comparison, and seven cautions keep the reading honest.

- **Association is not cause.** GitHub says of its own cohort comparisons: "These comparisons show associations, not proof that adoption caused a change in productivity" (GitHub Docs). Microsoft says of the differences in its Business Impact report that they "have not been evaluated for statistical significance" and that Copilot use "is unlikely to be the only variable responsible" (Microsoft Learn, updated October 21, 2026). The same limits apply to this method.
- **Heavy users differ from light users.** Teams that take up AI early may already be the stronger teams, or may have easier work. METR found selection strong enough to weaken its own study: in its February 2026 update it called the new data "very weak evidence" because developers and tasks dropped out in ways related to AI use (METR, February 24, 2026).
- **Compare a team with itself first.** The most useful comparison is the same team before and after rollout, beside a similar team that has not yet rolled out. A single snapshot across teams mixes AI with everything else that makes teams different.
- **Small groups mislead.** One long project or one absent person can move a small team's numbers more than any tool. In the METR update, the estimate for 10 returning developers was an 18% speedup, and the 95%

confidence interval, the range the true figure plausibly lies in, ran from a 38% speedup to a 9% slowdown. State the group size next to every figure.

- **Output lags use.** People need time to learn a tool, and some measures, such as days to close, are produced once a month. Choose the window to fit how long the work takes to complete.
- **Other things change at the same time.** A new system or a reorganization can move an output measure. Microsoft's documentation names "role responsibilities and seasonality" as other influences (Microsoft Learn). Record known changes beside the readout.
- **A measure that becomes a target gets gamed.** This applies to output measures as much as to token counts. If tickets resolved is rewarded, tickets get split. The quality partner is the defence, along with the rule that the readout is for deciding where to spend and train, and is never a score for a person.

> WHY THIS MATTERS
>
> The readout supports decisions about where to spend and where to train. It does not support a statement that AI caused a specific gain. A company that needs the stronger claim needs a study design that deals with the other explanations, such as a staged rollout in which comparable teams start at different times.

A worked example with a fictional company

The company below is fictional. Larkmoor Mutual is an invented insurer, created for this paper. The rows are illustrative and use qualitative values only. They are not customer data, and the "work AI can do" labels are placeholders for the example and are not JobRoute scores for these occupations.

Larkmoor pays for a coding assistant, a general AI assistant on a usage-based plan and an office suite assistant. The first two report tokens and cost per user. The office suite assistant reports activity dates only. The company gives JobRoute read access to the three usage reports, its roster with job titles, and one output system for each of the roles it wants to study first.

ROLE	TOKEN USE	WORK AI CAN DO	OUTPUT MEASURE	OUTPUT
Software developer	High	High	Pull requests merged	Up
Accounts payable analyst	High	High	Invoices processed	Flat
Claims support specialist	Activity only: few of the team active	High	Claim files closed	Flat
Recruiter	High	Low	Positions filled	Flat
Field inspector	Low	Low	Inspections completed	Flat
Customer service representative	Activity only: most of the team active	High	Tickets resolved	Up

Illustrative readout for a fictional company, one quarter. All values are made up for the example.

Source: Illustrative; no real company or JobRoute data

In the developer row, use is high, the work suits AI and merged pull requests rose while the defect measure held. The one open question is cost: most of the team's tokens went to the largest model, including for routine tasks that a smaller model handles.

The accounts payable row is the most informative. The analysts use the assistant heavily and on suitable work, yet invoices processed did not move. Larkmoor's finance lead talks to the team and learns that every AI-prepared invoice still goes through the same two manual approvals as before. The time saved in preparation is waiting in an approval queue. The fix is in the workflow.

The claims support row is the likely missed opportunity. Much of the role's work is drafting and summarizing, the team has licences for the office suite assistant only, and few people have opened it. That assistant reports activity and no tokens, so the row sits outside the token comparison and is labelled as activity. Larkmoor puts this team first for training.

In the recruiter row, the team explains that one person built an automated job that screens every inbound application with the general assistant, which accounts for most of the tokens. Larkmoor's HR and legal leads review that job before anything else, because rules on automated hiring decisions may apply to it. What began as a spend question turns out to be a governance question.

The field inspector row needs no action. The customer service row has the same data limit as claims support: the readout can say that most of the team is active and tickets resolved rose, and it cannot give a cost per ticket.

How does this differ from other AI spend and usage tools?

Several products measure AI spend or usage, and a company choosing an approach should look at them. As read on November 11, 2026: Larridin describes "token usage, seat licenses and cloud model costs in one spend view" with adoption by team and role. Rippling's AI Spend Console, announced August 6, 2026, breaks costs down "by teams, roles, or departments". Microsoft Copilot Analytics relates Copilot use to outcome

data that the customer uploads. For engineering organizations, Jellyfish reports token use and spend by tool and model, and the DX AI Measurement Framework covers utilization, impact and cost. This method differs from those five in its starting point. It begins with what each role's work is, task by task, so that it can say whether use sits where the work suits AI. The five pages linked here do not describe that starting point, and they may not describe everything the products do. Other vendors assess which tasks in a role AI can do, so the task view alone is not new either. What this method contributes is the join of the three feeds.

What are the limits of this method?

It measures roles and teams. A token count is not a performance measure, and the method does not produce a score or a ranking for a person.

Rules on worker monitoring may apply. Laws on automated decisions about workers and on employee monitoring differ by country and state. The JobRoute AI employment compliance calendar for 2027 lists the main ones with their primary sources. This is general information, not legal advice.

Task exposure is an estimate. JobRoute's task scores draw on observed AI use where the Anthropic Economic Index has it and on a documented heuristic where it does not. The index reflects use of one vendor's products. A company's own version of a role can also differ from the standard task list. The methodology page lists these limits.

The result is an association. A causal claim needs a study design that deals with other explanations.



ABOUT THE AUTHOR

Mahesh Tatiparthy

Mahesh Tatiparthy is the co-founder and CEO of JobRoute and the author of The Seven Frictions (2026). He founded QueryNow in 2014 and has led its enterprise AI work since.

maheshtatiparthy.com LinkedIn

Sources and further reading

1. Enterprise AI Enters an ROI Reckoning: press release for The State of AI in the Enterprise (71% track total AI inference and API spend, 51% track token volume, 44% track per-unit inference cost; 3,500 leaders in 25 countries, surveyed May and June 2026) Deloitte, 2026
2. Amazon software engineer role guideline by Terk Terkowitz, VP and Distinguished Engineer (input metrics described as deeply flawed; 10th and 90th percentile guidance; published October 27, 2026) Amazon, 2026
3. Uber's CTO says the tokenmaxxing era is ending (COO Andrew Macdonald on the link between AI spend and results; reporter's account of usage leaderboards; August 7, 2026) Fortune, 2026
4. Analytics APIs (Claude Enterprise Analytics API: per-user daily activity; per-user token usage and cost on usage-based Enterprise plans; data from January 1, 2026) Anthropic, 2026
5. Claude Code Analytics API (daily records per user or API key; tokens and estimated cost by model; sessions, commits and pull requests) Anthropic, 2026
6. Usage and Cost API, legacy (grouping by API key, workspace and model; no user dimension; points to the Usage and Cost Analytics API for per-user Claude Code cost) Anthropic, 2026

7. Admin API reference: organization usage (completions usage grouped by user_id, project_id, api_key_id or model; input, output and cached tokens) OpenAI, 2026
8. ChatGPT Enterprise governance (Compliance API described as not an adoption or productivity dashboard) OpenAI, 2026
9. reportRoot: getMicrosoft365CopilotUsageUserDetail, Microsoft Graph beta (last activity date per user and per app; page updated August 26, 2026) Microsoft, 2026
10. Copilot Analytics introduction (Consumption Dashboard for Copilot Credits and GitHub AI Credits; business impact reporting from uploaded outcome data; page updated October 9, 2026) Microsoft, 2026
11. Copilot Business Impact report (differences not evaluated for statistical significance; minimum group size setting; at least 65 Copilot and 65 non-Copilot users; page updated October 21, 2026) Microsoft, 2026
12. GitHub Copilot usage metrics (user-level detail; data within two full UTC days; comparisons show associations, not proof) GitHub, 2026
13. REST API endpoints for Copilot usage metrics (user-level daily and 28-day reports for enterprises and organizations; reports from October 10, 2025) GitHub, 2026
14. AI credits consumed per user now in the Copilot usage metrics API (ai_credits_used; a metrics signal, not a billed total; June 19, 2026) GitHub, 2026
15. Admin API (daily usage data per user without tokens; filtered usage events with token counts and charged cents) Cursor, 2026
16. Access Gemini audit logs using the Reporting API and the security and audit investigation tools (app, feature and action per user; July 1, 2025) Google Workspace Updates, 2025
17. Generative AI at Work, NBER Working Paper 31161 (5,179 support agents; issues resolved per hour up 14% on average, 34% for novice and low-skilled workers) Erik Brynjolfsson, Danielle Li and Lindsey R. Raymond, National Bureau of Economic Research, 2023

18. Human + AI in Accounting: Early Evidence from the Field, Working Paper 4261 (79 firms, 277 accountants; monthly close 7.5 days shorter among AI adopters) Jung Ho Choi and Chloe Xie, Stanford Graduate School of Business, 2025

19. AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice (gains of 50% to 130% in five of six legal tasks) Daniel Schwarcz and co-authors, Journal of Law and Empirical Analysis, 2026

20. Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity (16 developers, 246 issues; 19% longer with AI while believing it was 20% faster) METR, 2025

21. Developer productivity experiment update (estimates with confidence intervals that include zero; selection effects; February 24, 2026) METR, 2026

22. Anthropic Economic Index report: Cadences (about 9,700 survey respondents; computer and mathematical occupations about 30% of respondents against 4% of US employment; 54% of Claude Code served by Opus against 10% of chat and Cowork; June 26, 2026) Anthropic, 2026

23. Larridin AI Spend and ROI product page (token usage, seat licenses and cloud model costs in one spend view; adoption by team and role) Larridin, 2026

24. Introducing AI Spend Console (spend by teams, roles or departments; mapped to performance ratings or pull request volume; August 6, 2026) Rippling, 2026

25. Jellyfish Advances Developer and Agent Productivity Insights for the AI-Native SDLC (token usage and spend by tool and model for engineering organizations; October 7, 2026) Jellyfish, 2026

26. DX AI Measurement Framework (utilization, impact and cost for engineering organizations) DX, 2026

Frequently asked questions

What is a token, and why is a token count a poor measure of value?

What are the three feeds in this method?

Do all AI vendors report token use per user?

How do you choose an output measure for a role outside engineering?

What does the map show?

Does a link between AI use and output prove that AI caused the change?

Does this method score individual employees?