Accel

The Four AI ratios

There is no right level of AI spend — measure consumption and hold it against business performance

Every technology company today is wondering if they are optimising the use of AI for their product and internal productivity, but benchmarks have been scarce and hard to tie to business metrics. In our opinion, it is very hard to define a target for AI cost or token consumption given that:

- technology is evolving faster than anything we have seen in tech before
- cost of token is decreasing quickly
- use cases are multiplying with the gain of model performance
- adoption is growing exponentially

How can you set targets when the foundation keeps changing at such a fast pace?

However, this does not mean that AI consumption cannot be measured. We just need to focus on measuring the consumption and assess how an increase in consumption ties to business performance. Instead of trying to define a metric to optimise towards a number, we are defining AI usage ratios that we can then tie to business metrics. At the end of the day, a business is a business, and the leverage of AI does not change the business metrics driving value creation: revenue growth, churn, customer acquisition costs, EBITDA and cash flows. What matters is how the consumption of AI impacts the value creation.

In summary, piloting the AI adoption of your business comes down to two fundamental questions:

1. Is the company maximising its AI adoption?
2. Is the use of AI driving business benefits?

The first is the measurement question and the second is the performance question, and they have to be answered in that order — you cannot tie AI to business outcomes until you can measure the AI investment.

We have developed four ratios to answer the first question and will then discuss how they can be tied to business performance. All four ratios are built primarily from two inputs, the cost of token[1] and the cost of labour, which means they can be computed easily from the ledger rather than assembled from a survey.

[1] Cost of token includes the cost of model providers, the cost of AI subscriptions as well as the cost of AI infrastructure a company uses to train or deliver open source or proprietary models.

RATIO	DEFINITION	SUB-QUESTION IT ANSWERS	HELD AGAINST
AI intensity	cost of token used in the delivery of the product ÷ revenue	Is the company leveraging AI to deliver intelligent products?	Growth rate and churn rate
AI burn	total AI costs ÷ revenue	How many points of EBITDA is the company investing in AI?	Business performance (EBITDA margin or growth)
AI labour	cost of efficiency token ÷ (cost of efficiency token + labour)	What share of production is done by AI?	Each function's own output metric
AI maxing	cost of efficiency token ÷ headcount	How much is each individual or team leveraging AI?	Peer teams/individuals in the same function

PART 1

Is the company maximising its AI adoption?

Let's look at each ratio and what insights they provide on the AI usage for a given company.

1) AI intensity ratio

The AI intensity ratio is focusing on the top line of the business and is measuring how much the company is leveraging AI to deliver intelligent products. A tech company's product or service used to be lines of code. Now these lines of code are complemented by intelligence, quantified in the form of token. This ratio calculates how much token is used by customers in the delivery of the service and is defined simply by the following formula:

$$\text{AI intensity ratio} = \frac{\text{cost of token used in the delivery of the service}}{\text{revenue}}$$

WHAT ARE WE SEEING TODAY

If the company is a SaaS company, their ratio started at zero and is moving up as their product includes new intelligent functionality using AI models. If the company is AI native, the ratio will be typically in the 50–80% range at the beginning. With technology maturing, the cost of token is expected to continue to decrease by an order of magnitude every two years, so it will be interesting to see where the optimal ends up. SaaS companies will see their ratio increase as they develop more AI functionalities and AI native companies will benefit from the use of cheaper models and more software functionalities.

2) AI burn ratio

The second ratio focuses on the total AI investment from the company. To make things simple to understand, we have defined it as the number of EBITDA points the company is investing in AI. It is calculated with the following formula:

$$\text{AI burn ratio} = \frac{\text{total AI costs}}{\text{revenue}}$$

The total AI costs in the formula include the token delivered through the product, the efficiency token used internally, the token used to train or fine-tune proprietary models, and the tooling and inference infrastructure around them all. Because these costs are expensed as they are incurred, the ratio reads directly as the EBITDA impact of the AI investment — a company running an AI burn ratio of 6% is spending six points of EBITDA margin on AI. That is what makes this ratio an easy representation of the relative investment in AI by the company.

The AI intensity ratio sits inside this one: the token delivered through the product is part of total AI costs. The two are therefore not independent, and the relationship between them is itself informative. Intensity divided by burn is the share of the AI investment that reaches the customer, and a company with a low figure there is spending most of its AI budget on internal optimisation.

WHAT ARE WE SEEING TODAY

For a legacy SaaS company, the ratio is a few percent or less. For an early AI-native company it can exceed 100% — if the AI intensity ratio alone is 50–70% and there is internal token consumption and inference infrastructure on top, total AI costs can be larger than revenue. It is the clearest arithmetic statement of what it costs to build an AI-native business today. We would expect it to compress meaningfully as these companies scale and as token prices fall.

3) AI labour ratio

The third ratio has been designed assuming that token is a form of labour provided by a machine. If we consider the total work going into a company, the natural question becomes: how much of the work is actually being done by machines rather than by people? The AI labour ratio answers this question and is defined by the following formula:

$$\text{AI labour ratio} = \frac{\text{cost of efficiency token}}{\text{cost of efficiency token} + \text{labour cost}}$$

This AI labour ratio tells you what percentage of your total production comes from AI. Unlike the AI intensity ratio, which can vary a lot by company based on the service they deliver, this ratio is directly comparable across functions and companies.

To make sure the numbers are meaningful, the token costs should include only token used for internal consumption. Typically, model providers are combining all the token costs in the invoice, so it is important to have tagged the token separately at source to differentiate between delivery token and efficiency token. The tagging taxonomy should be defined based on the granularity of the analysis you would like to conduct (function, team etc...). Without proper tagging, it will be hard to compute properly the ratios, and the split cannot be reconstructed afterwards. The labour cost has to be fully loaded.

WHAT ARE WE SEEING TODAY

Most companies today sit in the low single digits. For example, a team spending 600 dollars a month per engineer on AI tooling against a fully loaded cost of $15,000 a month is running an AI labour ratio of 4%. The most aggressive engineering organisations are pushing into double digits. We expect this ratio to increase significantly over the next few years, and to be the one where the spread between leaders and laggards becomes most visible and comparable.

4) AI maxing ratio

The first three ratios take a wide lens across the organisation. The fourth ratio zooms in and provides more granularity into how each team or individual is leveraging AI to perform their work. It shows how much AI is used by a person or a team and is calculated by the following formula:

$$\text{AI maxing ratio} = \frac{\text{cost of efficiency token}}{\text{headcount}}$$

This ratio comes out as a dollar figure per employee per year, which is immediately communicable, and is not distorted by where your people happen to sit. It shows the true usage by each person or team.

The AI maxing ratio and the AI labour ratio are two views of the same underlying numbers but achieve different goals. The AI labour ratio is anchored around hard currency and will be compared against business output. The AI maxing ratio should be used for comparing teams and individuals inside a single function and can be tied to specific team KPIs or individual MBOs.

WHAT ARE WE SEEING TODAY

Most companies are currently spending a few hundred dollars per employee per year on AI. Engineering-heavy organisations are in the thousands, and the frontier is an order of magnitude above that. This number will vary internally by team and by function, and the benchmarking should be based on high versus low performing teams and how they use AI to enhance their performance. This will help level up the organisation and spread best practices.

If the ratio is applied to larger teams, the mean is interesting, but it should also be read alongside the distribution.

PART 2

Is the use of AI driving business benefits?

The four ratios provide strong insights into how an organisation is using AI and how it compares to other organisations in the same industry. As we discussed in the introduction, no optimal level has emerged yet for any of them, but they can be extremely helpful as a company hold them against a business performance metric, over time, on the company's own trailing series. The correlation – or not – of the increase in AI usage and business performance is the finding.

> *For fast-growing start-ups, many things are changing over a short period of time, so it can be hard to isolate the benefits of AI adoption. The benchmarking exercise we are conducting provides some insights as companies can be compared by sector and performance vs. AI adoption vs. peer group is quite instructive. The benchmark is available* ***www.xxx.com***

RATIO	HELD AGAINST
AI intensity	Growth rate and churn rate
AI burn	Business performance (EBITDA margin or revenue growth)
AI labour	Each function's own output metric
AI maxing	Peer teams/individuals in the same function

Is AI improving the top line?

The AI intensity ratio needs to be compared over time against the revenue growth rate (is delivering more token through your product or service accelerating adoption?) and against the churn rate (is delivering more token through your product making it stickier and more differentiated?). There is no optimal target today as everything is changing fast, but over time there should be an optimal enabling tech companies to have a good gross margin while delivering a very intelligent product or service.

Is the total AI investment driving business performance?

The AI burn ratio calculates the total investment a company is making in AI, and it should be held against the main business objective the company is driving towards.

For a SaaS company trying to reaccelerate its top line, the AI burn ratio gives you the EBITDA investment to be held against the additional growth rate achieved (e.g., EBITDA deteriorated by 10 points but growth accelerated by 20 points).

For a company aiming to maintain its growth rate but trying to improve profitability, the AI burn ratio should be compared to the improvement in EBITDA margin. If AI costs rise by three points of revenue and the EBITDA margin is flat, the rest of the business improved by three points and the AI investment has paid for itself exactly. If AI costs rise is correlating to an increase in EBITDA margin, then the AI investment is making the company more productive.

This is a simple way to measure if AI is driving any positive impact in the business overall. The failure case is equally legible: an AI burn ratio rising while the EBITDA margin falls or growth stays flat says the investment is not being recovered anywhere in the P&L.

Two cautions. The identity assumes AI costs are expensed rather than capitalised, so any capitalised development needs adding back before the comparison works. And it attributes everything else that changed in the period to "underlying improvement", which is why it should be read over several quarters rather than one.

Is AI impacting the productivity of each function?

For this comparison the AI labour ratio should be calculated by function rather than blended, as it will provide more granular insights. The use and benefits of AI vary widely by function, and each function has its own output metric to hold it against:

FUNCTION	HELD AGAINST	THE QUESTION IT ANSWERS
Total	(cost of efficiency token + labour) ÷ gross margin	Is my combined machine and human labour making the company more productive?
S&M	CAC ratio — customer acquisition cost per $ of gross margin acquired	Are my customer acquisition costs for each dollar of gross margin going down?
R&D	Number of pull requests	Is each engineer shipping more (assuming equal code quality)?
G&A	Total G&A cost as % of gross margin	Is the cost of running the company falling as a share of what the company produces?
Support	Deflection rate (tickets never touched by a human ÷ total tickets), and support cost as % of gross margin	Is AI actually resolving tickets, and is that showing up in the cost line?

Support carries two comparisons deliberately. Deflection on its own can be improved by making the service worse — the ticket closes, the customer comes back through a channel

you measure less carefully, and the dashboard looks better than the business. The cost line is what confirms the deflection was real.

R&D's comparison is the one to treat with the most care. The number of pull requests is the most natural output measure and it is also the one AI inflates most directly, so it is worth pairing with the share of merged code reverted or rewritten within thirty days. Otherwise, the ratio and its comparison can both improve while the codebase quietly degrades.

Another point to note: business ratios should be computed with gross margin, not revenues, as the use of AI in the product is likely to have an important impact in the gross margin of the business and gross margin – not revenue – is paying the bills.

Are the individuals and teams using AI doing better than their peers?

The AI maxing ratio is the only one of the four that is read across the company at a point in time rather than against its own history. Take a single function, split the teams or individuals by their maxing ratio, and hold them against the same output metric. Engineering teams in the top decile of AI spend per head against engineering teams in the bottom decile, on pull requests per engineer per month. Sales pods in the top decile against the bottom on pipeline creation or quota attainment.

Because the function, the output metric and the period are all held constant, this is the closest thing to a controlled comparison a company can run. Every other comparison in this framework is a time series, which means it is exposed to everything else that changed over the period — a pricing change, a new product, a reorganisation. The cross-sectional version is not.

It is worth keeping an eye on selection bias. The teams that adopt AI most aggressively are frequently the strongest teams already, so a level comparison will overstate the effect of AI and understate the effect of talent. The correction is to compare trajectories rather than levels: not whether the high-maxing teams are performing better, but whether their output improved faster from the point at which their maxing ratio diverged. Where the ratio has diverged sharply and recently, that is a natural experiment worth reading carefully.

This comparison is quickly actionable, because it tells you what to do next. If the high-maxing teams are pulling away, the constraint is diffusion and the answer is enablement. If they are not, the constraint is the skillset and buying more token will not fix it.

One caveat: cost is falling

As noted above, the price of token falls by roughly an order of magnitude every two years at constant capability. All four ratios have the cost of token in the numerator, which means all four drift downwards over time even when usage is rising sharply. A flat AI labour ratio

through a year of aggressive adoption is not stagnation — it is what progress looks like once the price decline is netted out.

For companies who want to correct for it, you can fix the token price over the period. However, the price decrease will not affect comparison between companies at a point in time, nor the true magnitude of the business impact, as the ratios are compared against business metrics.

PART 3

Where to start?

If you would like to start using this framework for your company or start-up, our suggestion would be to focus on the following 11 metrics. You can then go into much more granularity by function, but this is a good starting point:

LEVEL	RATIO	FORMULA	PERIOD	COMPARE TO
Company wide	AI intensity ratio	delivery token ÷ revenue	Quarterly	Quarterly growth and churn
	AI burn ratio	total AI costs ÷ revenue	Quarterly	Key business performance metrics: growth, churn, EBITDA or free cash flow
	AI labour ratio	efficiency token ÷ (efficiency token + labour)	Quarterly	(efficiency token + labour) ÷ gross margin
Sales & Marketing	S&M AI labour ratio	S&M efficiency token ÷ (S&M efficiency token + S&M labour)	Quarterly	CAC ratio
	S&M AI maxing ratio	S&M efficiency token ÷ S&M headcount	Monthly	Quota attainment and/or pipeline generation for quota-carrying teams, BDRs/SDRs and lead-gen teams
R&D	R&D AI labour ratio	R&D efficiency token ÷ (R&D efficiency token + R&D labour)	Quarterly or monthly	Total pull requests per month or quarter
	R&D AI maxing ratio	R&D efficiency token ÷ R&D headcount	Monthly	Pull requests per engineer per month, split by backend and front end
G&A	G&A AI labour ratio	G&A efficiency token ÷ (G&A efficiency token + G&A labour)	Quarterly	Total G&A cost as % of gross margin
	G&A AI maxing ratio	G&A efficiency token ÷ G&A headcount	Quarterly	Individual performance within legal, finance and HR against MBOs

LEVEL	RATIO	FORMULA	PERIOD	COMPARE TO
Support	Support AI labour ratio	support efficiency token ÷ (support efficiency token + support labour)	Monthly or quarterly	Deflection rate (tickets never touched by a human ÷ total tickets), support cost as % of gross margin and cost per ticket solved
	Support AI maxing ratio	support efficiency token ÷ support headcount	Monthly	Tickets solved per agent

Accel