STRATOS
Research

The AI Automation Arms Race

An Analysis of the Costs and Benefits of the AI Race

Samuel Trif

July 20, 2026

---

1. Introduction

Between 2024 and the first half of 2026, the share of United States layoffs explicitly attributed to artificial intelligence rose from a marginal 0.6% to more than one in four cuts announced in April and May 2026. Over the same time span, the leading technology companies committed on the order of 700 billion dollars to AI infrastructure. Whether these two figures describe the same underlying phenomenon, a genuine substitution of capital for labour, or two loosely connected narratives that happen to move together, is the question this article sets out to disentangle.

The adoption of artificial intelligence has become a central driver in corporate strategy, and is increasingly cited as the driving force behind workforce reductions. Yet the evidence is divided; alongside the many cases where automation has delivered measurable efficiency gains, there are documented situations in which the expected results failed to materialise, and in which the cost of reversing course exceeded the savings that motivated the decision in the first place. Understanding what distinguishes one outcome from the other is essential for assessing the phenomenon accurately, without assuming either that automation always pays off, or that it is merely a reputational exercise.

The analysis draws on a number of documented cases, considering both positive and problematic outcomes, before turning to a broader dynamic, the competitive pressure to adopt AI, and its consequences for the economy as a whole.

2. The Productivity Evidence

Any credible assessment must begin by acknowledging that a substantial body of evidence documents concrete and measurable productivity gains. Research by Brynjolfsson et al. found that the introduction of AI assistants in customer service increased average operator efficiency by 14%, with gains of up to 34% among less experienced workers, who were able to close the gap with their senior colleagues more quickly. A broader review of task-level productivity studies places typical gains between 15% and 30% in ordinary working environments, with higher peaks under controlled experimental conditions. At the sector level, 95% of retail and consumer goods companies report annual cost reductions attributable to AI, and the Federal Reserve estimates that generative tools save workers an average of roughly 5.4% of their working hours, the equivalent of more than two hours per week.

The benefits, however, are unevenly distributed across task types. Gains concentrate in codifiable and output-verifiable tasks, drafting text, data analysis, code generation, document summarisation, while remaining modest or negligible in roles defined by judgement, relationship management, and coordination. The 2025 METR study further observed that some experienced developers took 19% longer to complete tasks when using AI tools, despite believing themselves to be faster; a finding that highlights how perceived efficiency does not always align with actual performance.

The distinction is decision-relevant as a substantial share of business decisions relies on self-reported or throughput-based speed metrics, which measure *perceived* rather than *actual* efficiency, and therefore tend to overestimate the value genuinely produced.

3. The Adoption Race

Adoption decisions respond to competitive positioning as well as to internal returns. Investors and analysts read AI adoption as a marker of innovation, and a firm that moves more slowly than its peers risks being priced as structurally less efficient, independently of whether the technology has yet paid off inside the firm. This penalty is reflected in valuation, and it accrues whether or not the automation succeeds. Two documented company cases from the same period show how far this incentive can carry a decision beyond the available evidence. They measure different quantities. The Uber figures track the internal adoption of AI tools among engineers, an input to production; the Klarna record concerns the substitution of AI for customer service staff, an outcome in the labour market. In both cases the decision preceded the evidence needed to evaluate it, and in both cases the corrective information arrived through channels that the original decision metrics did not track.

Uber illustrates the cost side. Tokens have never been cheaper, and AI bills have never climbed faster. The perception that artificial intelligence is a cheap technology rests on the unit price of a single token. That price has indeed fallen sharply, compressed by intensifying competition among model providers, by the diffusion of open-weight alternatives, and by the entry of international competitors, of which the Chinese developer DeepSeek is the most prominent example. The unit price, however, is the wrong object of analysis. What determines the cost borne by an organisation is the product of that price and the volume consumed, and the volume is neither fixed nor exogenous; it responds to availability, to workflow design, and to the elasticity of internal demand once the marginal cost to the individual user falls to near zero.

Uber granted its roughly 5,000 engineers access to Claude Code, an AI coding tool, in December 2025. The share of engineers using it in agentic mode rose from 32% in February 2026 to 84% in March, and by spring some 95% of the engineering workforce was using AI tools monthly, with around 70% of committed code originating from them. An internal leaderboard that ranked teams by the intensity of their AI usage accelerated the climb; usage was rewarded for its own sake, decoupled from any downstream verification of value. Imprecise prompting added a further inflationary factor, since poorly specified requests force the model to rework the same task repeatedly. By mid-April the entire annual AI budget had been exhausted, twelve months of planned spending compressed into four, at a run-rate of between 500 and 2,000 dollars per engineer per month in API calls alone. The company's Chief Technology Officer, Praveen Neppalli Naga, confirmed the overrun and stated that the firm was revising its budgeting assumptions. Total research and development spending stood at 3.4 billion dollars in 2025, so the strain did not come from the absolute size of the outlay. Consumption-based pricing does not fit annual budget cycles built on fixed per-licence assumptions, and internal demand for a tool whose marginal cost to the user is zero expands far faster than any forecast calibrated on the prior period. The effect is a mismatch between how the tool is priced and how cost centre budgets are typically structured.

Rapid uptake does reflect genuine perceived utility among engineers, and that signal should not be dismissed. Yet Uber's Chief Operating Officer, Andrew Macdonald, acknowledged that the company could not yet correlate the increase in token consumption with a measurable improvement for end users.

Klarna illustrates the labour side. The Swedish payments and financial services company froze hiring in December 2023 and reduced its overall headcount from roughly 5,500 at the end of 2022 to about 3,400 at the end of 2024, largely through attrition, while an AI assistant built on OpenAI technology absorbed customer service work the company described as equivalent to 700 full-time agents. The early data were encouraging; the chatbot handled two thirds of customer requests and was estimated to generate annual savings of 40 million dollars.

Had the trajectory held, the case would read as a textbook demonstration of labour substitution. Yet, over time, customer complaint volumes rose, responses were frequently generic, and the system struggled with the non-routine cases that require more contextual judgement, disputes, fraud claims, account escalations. In May 2025 the chief executive, Sebastian Siemiatkowski, publicly conceded to Bloomberg that the company had "gone too far", and a renewed hiring phase began under a flexible remote staffing model. The reversal preceded Klarna's United States listing of September 2025 by a few months, and by July 2026 the stock traded at roughly half its offering price, an indication that the market's reward for an AI-first narrative can be short-lived when service quality moves the other way. The cost of reversing course, rehiring, retraining, and repairing service quality, exceeded the savings originally booked. This asymmetry recurs across failed automation cases and has a structural explanation. The savings from a headcount reduction are recognised immediately and are highly visible; the costs of degraded service quality accrue with a lag, are diffuse, and surface as attrition, reputational erosion, and remediation expense that rarely gets attributed back to the original decision. An evaluation conducted at the moment of the cut therefore sees a favourable figure that a fuller accounting, taken over a longer horizon, would revise downward.

Klarna is not an isolated instance. A parallel dynamic appeared at IBM, which deployed an automated HR platform, AskHR, to absorb routine back-office activity, only to rehire staff for more sensitive functions after delays and errors accumulated. A recurring proximate cause is the phenomenon of hallucination, the generation of confident but factually baseless output, a non-existent internal policy, or a discount that was never authorised yet becomes binding once communicated to a customer. In such cases, the labour-cost savings are transformed into dispute resolution costs, legal exposure, and reputational damage. It is also telling that, on BCG estimates, roughly 74% of generative-AI pilot projects never reach full-scale production, most often stalling on data quality and governance issues, which places much of the failure in the organisational conditions surrounding deployment.

Taken together, these cases document the same governance failure at different points of the production process. Competitive pressure drove Uber's tool adoption faster than its finance function could track, and drove Klarna's disinvestment from human labour faster than its service quality could bear. In each case the decision outran the evidence available at the time it was taken.

4. Structural Shift or AI-Washing?

Aggregate data confirm that the race is accelerating. According to Challenger, Gray & Christmas, AI was cited as a cause in 0.6% of United States layoffs in 2024 and roughly 5% in 2025. In 2026 the monthly share climbed from 7% in January to 25% in March, when AI became the single most cited reason for the first time, a position it held through June, reaching 31% of announced cuts that month. Over the first half of 2026 the reason was cited in roughly 102,000 job cuts, close to a quarter of the total. In the first five months of the year the United States technology sector alone recorded roughly 142,000 job eliminations, a 33% increase on the previous year. These figures record the stated cause of layoffs, which is not the same as their true cause, and the distinction carries the weight of this section.

The data admit two interpretations, and the interesting question is under what conditions each dominates. The first reads the layoffs as structural transformation, a genuine reallocation of resources toward capital, evidenced by the roughly 700 billion dollars in infrastructure investment announced by leading technology firms in 2026 and by the nearly 20% contraction in employment among software developers under 26 recorded by Stanford HAI. On this reading, the rehiring at Klarna and IBM is an adjustment cost within an early phase, consistent with the productivity gains documented in Section 2. The concentration of the developer contraction among the youngest cohort is itself consistent with a substitution account, since entry-level tasks are the most codifiable and hence the most exposed.

The second reading is more sceptical, and rests on a growing body of evidence that the announced cuts and the underlying economics do not align. A 2026 Gartner study of 350 companies found no improvement in financial returns among the firms that cut most aggressively. If the layoffs reflected genuine automation of newly redundant labour, the deepest cutters should show the clearest margin gains, and they do not. Survey evidence points in the same direction. Yotzov, Barrero, Bloom et al. (2026), covering nearly 6,000 senior executives in the United States, the United Kingdom, Germany, and Australia, find that more than 90% of executives report no impact of AI on employment at their own firm over the past three years, despite the frequency of public announcements to the contrary. The reversal data reinforce the anomaly. According to a July 2026 Orgvue survey reported by CNBC, 55% of business leaders who made staff redundant on account of AI now consider the decision wrong, and Robert Half finds that 32% of hiring managers who eliminated a role because of AI have already rehired for the same or a similar position.

This pattern is commonly described as AI-washing. The term denotes the practice of presenting business decisions as AI-driven because the label is rewarded by investors and by the press, regardless of the actual role the technology played; cuts caused by over-hiring during the preceding cycle, by interest-rate normalisation, or by ordinary cost discipline are announced under an AI heading. The gap between what firms declare publicly and what their executives report in anonymous surveys is consistent with this practice. At the aggregate level, the Yale Budget Lab has so far detected no AI-driven acceleration in labour-market turnover, which is what one would expect if a meaningful fraction of the announced AI cuts were relabelled versions of ordinary ones.

The two readings are not mutually exclusive. The structural reading fits best in codifiable, entry-level roles, where the developer data and the productivity evidence point the same way. The AI-washing reading fits best in the aggregate numbers, where the promised effects do not show up. Firms have a reason to call a cut AI-driven whether or not AI caused it, because the market rewards that label. The reported layoff figures should therefore be treated as a ceiling on genuine substitution, not a measurement of it.

5. The Layoff Trap

The transition from firm-level to system-level analysis introduces effects that are invisible at the level of the individual decision and become dominant only in aggregate. Once the technology is sufficiently diffused, made progressively more accessible through cheaper inference, maturing usage practice, and improving models, automation ceases to be an isolated operating variable and acquires macroeconomic weight, at which point its private and social valuations can come apart.

This divergence is the subject of a study by Brett Hemenway Falk of the University of Pennsylvania and Gerry Tsoukalas of Boston University, first circulated in March 2026, which formalises what the authors term the “layoff trap”. The mechanism is a demand externality. A firm that automates captures the full cost saving, the wages no longer paid, while the corresponding reduction in aggregate demand, workers who, deprived of income, curtail their spending, falls only fractionally on the automating firm. In a market with N competitors, each firm bears roughly one N-th of the demand destruction it causes; the rest falls on its rivals. Each firm therefore weighs the full private saving against a fraction of the demand cost it creates, and the privately optimal level of automation exceeds the collectively optimal level. The wedge between the two is the externality, and it widens as the practice spreads.

The consequential result is that the outcome need not be a transfer of wealth from workers to owners. Under the conditions the model isolates, it is a net contraction for both. If reductions in labour income are simultaneous and economy-wide, the fall in disposable income propagates through to revenues across the entire system, and the erosion of aggregate demand ultimately compresses the very profits the automation was meant to protect. The worker enters the accounting twice, as a cost on the firm's balance sheet and as a source of demand on its revenue line, and a strategy that optimises against the first while ignoring the second can be individually rational and collectively self-defeating at the same time.

Competition does not correct the distortion; in the model it enlarges it. Each firm reasons that automating faster than its rivals will win market share. The calculation holds for any single firm considered in isolation, and fails when every firm makes it simultaneously, because the relative advantages cancel, no firm gains durable share, and the cumulative destruction of demand remains. The authors describe this as a "Red Queen" dynamic, borrowing the figure from Lewis Carroll who must keep running simply to remain in place. The meaning in this setting is precise. Firms sustain continuous automation expenditure without any improvement in relative competitive position, while the resource they collectively depend on, aggregate consumer demand, deteriorates. Market structure shapes the outcome. A monopolist, facing the entire economy's demand curve, would internalise the full contraction its own automation generates and would have a private incentive to stop short of the point of net demand destruction. In a fragmented market this restraint disappears, since each firm internalises only its own fraction of the aggregate demand loss while capturing the whole of its own saving, and the process continues past the point that serves any participant. In the model, additional competition and more capable AI increase the size of the excess.

The model has already attracted scrutiny, and the debate is instructive in its own right. A replication study by McEntire verifies the internal logic of all ten propositions and shows that the catastrophic conclusion depends on parameter choices. Under the authors' baseline, low demand elasticity, no spending out of profits, homogeneous goods, no re-employment feedback, the model produces collapse; under equally defensible parameters reflecting standard features of real economies, product differentiation, modest owner spending, endogenous re-employment, it produces stability or even under-automation. Khalifeh accepts the core externality and questions the static assumptions instead, a fixed number of firms, a stable task structure, and displaced workers treated as consumers only, with no possibility of entrepreneurship or entry into new activities. Kimura decomposes the welfare loss into two margins, the excess automation itself and a shift of real purchasing power from workers to owners, and finds that the parameter which shrinks the over-automation wedge is the same one that raises the owners' consumption share, so that correcting the inefficiency and improving the distribution pull in opposite directions. None of these critiques dissolves the mechanism; together they establish that the trap is a contingent risk whose severity depends on empirical parameters that remain unsettled.

The empirical record is consistent with that reading. The Yale Budget Lab has so far observed no macroeconomic evidence of a demand spiral, in line with the absence of any measured turnover acceleration noted in the previous section. The contribution of the framework is to identify a failure mode before it becomes visible in the data, when it is still cheaper to avert than to reverse, and to specify the conditions, scale, simultaneity, and slow labour-market reabsorption, under which it would materialise.

6. Conclusions

The evidence assembled here supports a differentiated assessment of AI-based automation. The productivity gains are real, measurable, and concentrated in codifiable, verifiable tasks. Three qualifications, however, complicate the balance-sheet case at successive levels of analysis. At the level of measurement, the metrics on which adoption decisions rest capture perceived efficiency, and the error runs in the technology's favour. At the level of cost, the true expense of the technology depends on the skill with which it is used and on the elasticity of internal demand, and is systematically underestimated at the point of decision, as Uber's budget cycle demonstrated. At the level of the firm, the reversals at Klarna and IBM show that the savings from a cut are recognised early and visibly while its costs accrue late and diffusely, biasing any contemporaneous evaluation upward.

Above these sits the systemic dimension, which no individual firm has an incentive to price. When automation occurs at scale, its savings remain private while the damage to aggregate demand is socialised across the whole market, and competitive pressure widens the imbalance. Whether the resulting trap binds in practice is an open empirical question; the parameters that govern its severity are contested, and the aggregate data show no spiral to date. The direction of the incentives, however, is not contested, and it is the same direction observed at every level of this analysis, in metrics that overstate the benefits of adoption, in narratives that inflate its role in layoffs, and in a competitive logic that rewards moving faster than the evidence.

The question is therefore how, where, and to what extent to automate, and it can be answered honestly only when both sides of the ledger are measured with equal rigour. At present the cost side is instrumented far more precisely than the benefit side, and the systemic side is not instrumented at all. Until that changes, the rational decision of the individual firm and the desirable outcome for the system cannot be assumed to coincide.

Sources: Brynjolfsson et al.; International AI Safety Report 2026; Federal Reserve 2026; METR 2025; BCG 2025; Gartner 2026; Challenger, Gray & Christmas monthly reports 2024–2026; Yotzov, Barrero, Bloom et al., *Firm Data on AI*, NBER WP 34836, 2026; Stanford HAI AI Index 2026; Yale Budget Lab 2026; Orgvue via CNBC 2026; Robert Half 2026; The Information / Forbes 2026 (Uber); Bloomberg 2025 and company filings (Klarna); Hemenway Falk & Tsoukalas, *The AI Layoff Trap*, SSRN 6448898 / arXiv 2603.20617, 2026; McEntire, SSRN 6592220, 2026; Khalifeh, SSRN 6576198, 2026; Kimura, SSRN 6936479, 2026.