

AI Dark Output: The Visible Cost of Invisible Output

Why AI's increasing output is going to be one of the hardest economic measurement problems in history. AI "Dark Output" could end up being the majority of economic activity, but a challenge to measure

MALCOLM SPITTLER AND DYLAN PATEL

MAY 30, 2026



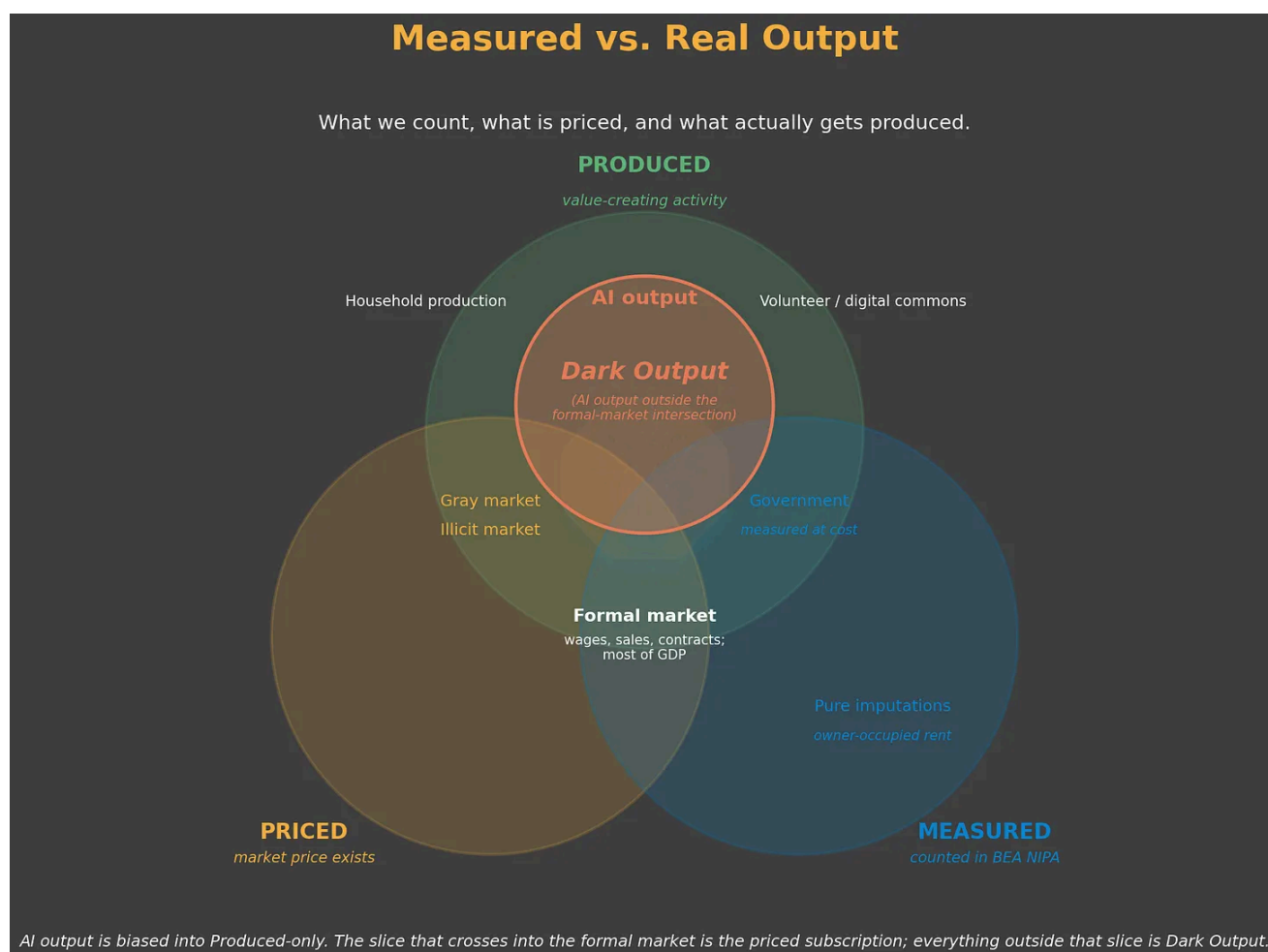
During the 1980s and 90s, macroeconomic data could not detect the contribution of the emerging computer revolution. Famously, Robert Solow quipped “You can see the computer age everywhere, but in the productivity statistics.” And yet, despite the dot com boom and bust the Magnificent 7 now have a market cap 1.8x that of Europe. A similar issue is arising with AI where the macroeconomic data is not yet equipped to capture the value produced by AI while the headlines, public sentiment, and governments around the world are quick to capture the costs incurred in dollars, watts, gallons and jobs. [Matt Drach had an interesting take separately from us on this.](#)

A boring 2013 methodology revision added R&D and investment in intellectual property to GDP accounting boosting total production for the 1990s by ~\$3.6T. In the official accounts it was spread evenly, so the growth rate only rose marginally, but it amounted to nearly 30% of full year 2000 GDP. The magnitude of the measurement problem from AI dwarfs prior measurement issues, we call the work AI does that national accounts can't currently see Dark Output. Even more of the new output from AI is likely to be invisible as it is clustered in the service sector where national statistics have longstanding issues with capturing productivity growth.

Incoming Fed Chairman Kevin Warsh acknowledged as much in December 2025 “If you're looking at the data, my view is you're backward looking. You're going to be late. You're not going to realize the country is able to have non-inflationary growth faster. So you're going to have to make a bet.” With the transition of AI growth to more active capital market funding, any measures that fail to show results from AI will be scrutinized for signs of a bubble.

Dark Output

AI output will be real before it is measurable. We can capture token spend, and we can capture jobs lost. But unless AI's output is sold at a visible price, only token spend is captured in GDP. Normally when the price of something collapses, we can see this deflation and call the results productivity. Due to well-known difficulties in the service sector (see Appendix 1), GDP will record those as declines, and prices may even show inflation. Like the dark energy that makes up our universe, Dark Output will likely only be visible in its effects on other elements of the economy and not through direct observation. One of the most visible effects is job displacement which we are now tracking on our Dark Output Monitor.



Source: SemiAnalysis [Tokenomics: Dark Output](#)

We are at risk of having an event on the scale of the Industrial Revolution where most of the new output is invisible even as businesses spend increasingly large amounts on AI services.

Types of Dark Output

Dark output is AI-enabled economic value that exists but is not visible, or is badly distorted, in GDP, prices, labor statistics, or industry accounts. We categorize this into two buckets:

- **Substitution dark output** is work that was previously done by humans and is now done by AI. In our Dark Output Monitor we have identified roughly \$1.5T in tasks that current generation AI could substantially augment or automate.
- **New dark output** is new work done by AI that wasn't previously being done by humans (probably because it was too expensive to do until AI made it cheap). In the long run this is likely to be much larger than the substitution side.

In both cases, value exists despite the statistical system failing to see it. This is not a unique problem (see Appendix 1).

Three tests, one ledger — what GDP sees and what it doesn't				
Each row is an activity in the economy. Three columns ask whether real labor produced it, whether a market price exists, and whether GDP captures it. An activity has to clear all three to land in the headline.				
Activity	Produced real labor	Priced market price	Measured captured by GDP	Verdict
AI work inside a flat subscription	✓	—	—	Inside the gap
Household / unpaid care	✓	—	—	Inside the gap
Informal & criminal economy	✓	✓	—	Inside the gap
Government services (cost-based)	✓	—	✓	Measured at cost
Owner-occupied housing (imputed rent)	—	—	✓	Measured by imputation
Capital gains, transfers, asset sales	—	✓	—	Priced but not output
Ordinary market output	✓	✓	✓	Captured by GDP
Rows in orange are Dark Output candidates: real activity that doesn't clear the three-test gauntlet.				

Source: SemiAnalysis [Tokenomics: Dark Output](#)

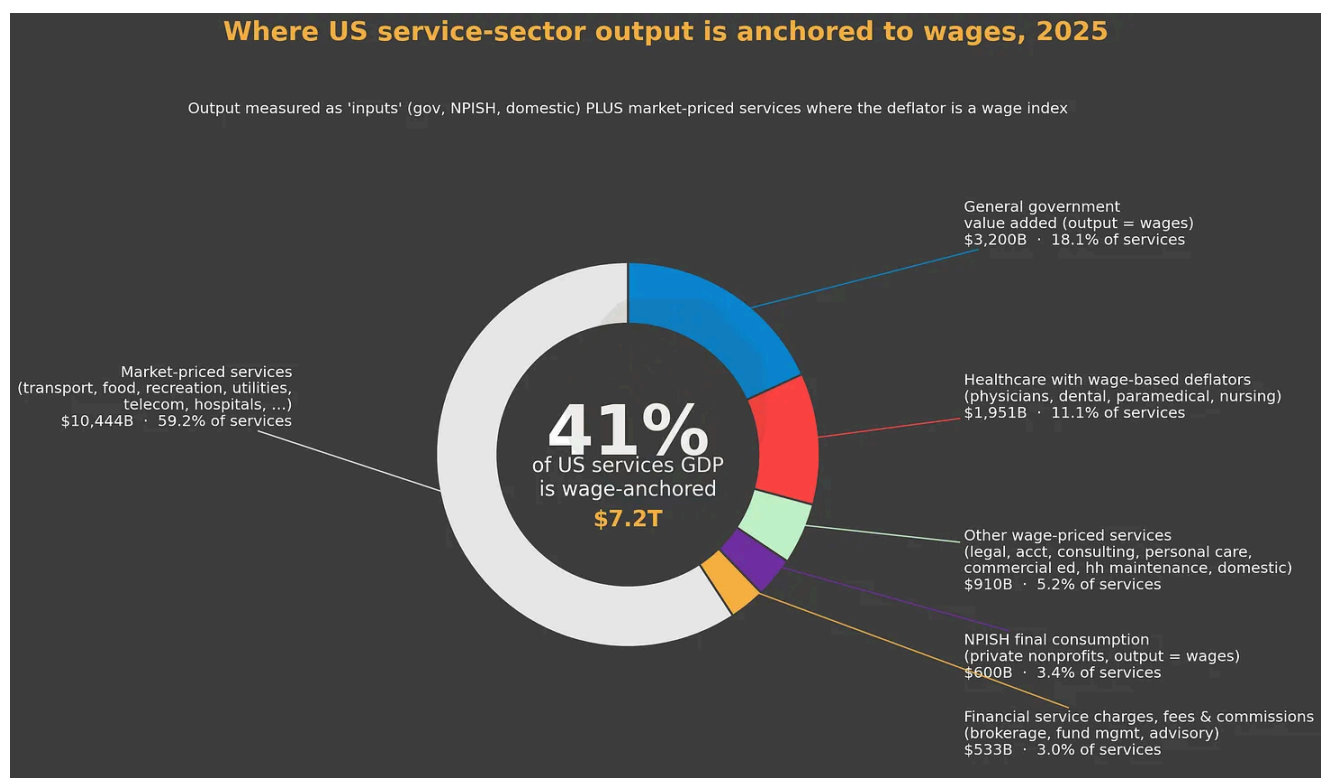
Substitution Dark Output

An example of substitution Dark Output is a simple legal document which in theoretical GDP should have the same inflation adjusted value to a user whether a lawyer drafts it or AI drafts it. But service sector GDP and inflation is hard to estimate (see Appendix 2), there is no 'unit' of legal services, just lawyers' receipts and surveys

of firms for the cost of services rendered. When AI takes over the task, the receipts vanish as the cost is absorbed in tokens, and when government officials survey lawyers on the cost of services they may find that the average price has gone up, as the simplest documents are now completed by AI and not lawyers. From the perspective of GDP, the transaction has effectively vanished except for a few dollars of tokens sitting in an unrelated sector of the economy.

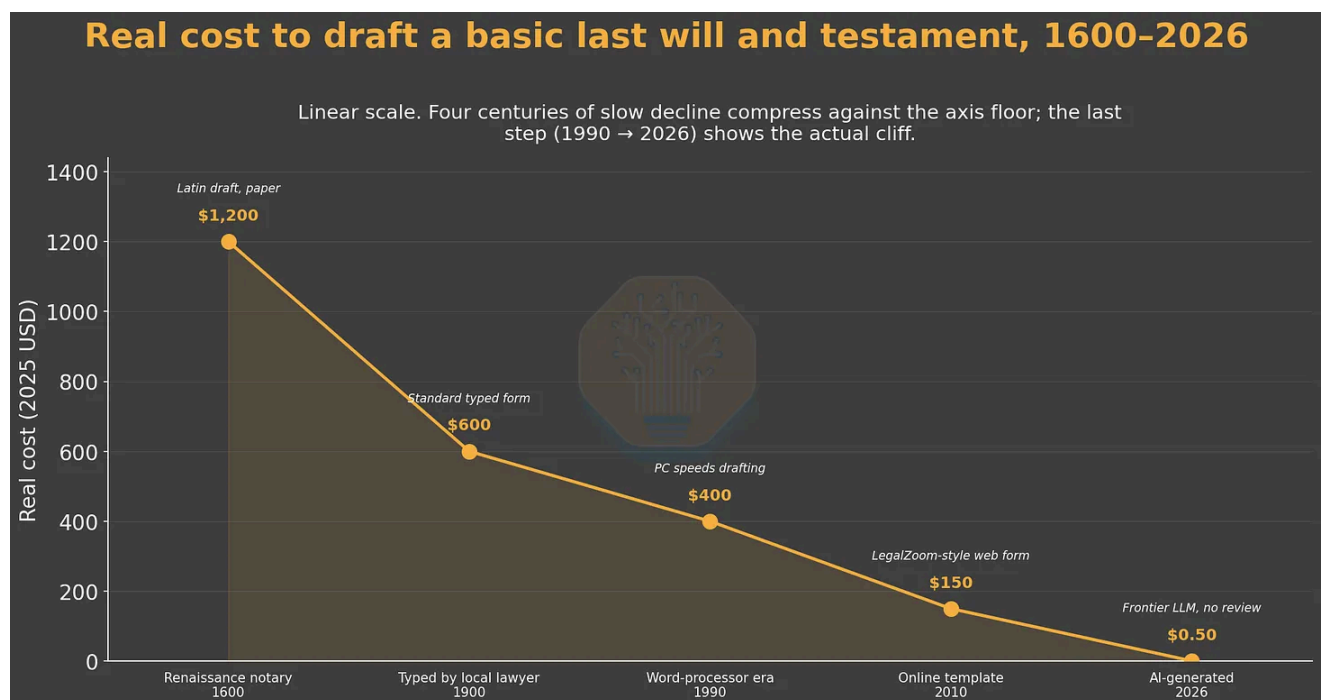
For [Tokenomics subscribers](#), we track the frontier of tasks that market signals show current AI has the potential to replace. These tasks, depending on how they are performed by AI may vanish from the national accounts all together (see the Dark Output Monitor section below).

Other than housing, most services are measured in the national accounts through this sort of receipts and list prices system which backs into ‘quantity’ of tasks being done by dividing spend by price. This sort of accounting doesn’t allow for productivity gains. When the accounts record lower receipts, they will read this out as an output decline.



Source: BEA NIPA annual via Haver USNA, 2025. Healthcare wage-anchored = physicians + dental + paramedical + nursing homes (excludes hospitals, transaction-priced via DRG). Market-priced rest includes housing imputed rent and FISIM, which are imputed but not AI-relevant

A basic will as seen in the figure below has fallen in price for generations as technology changed the process of creation, but because it was gradual, the induced error was less extreme. A drop from \$400 to \$150 in 30 years is less than 5% a year. A drop from \$150 to \$0.50 in a year is more than a 99% cost decrease. One introduces bias, the other vanishes from the dataset. Legal services prices were only added to the CPI in 1987, and since then the price index is up 4.6x (as of September 2024). The price index is effectively an employment cost index because there is no accounting for the increased productivity.



Source: SemiAnalysis. Illustrative. Anchors are representative costs deflated to 2025 USD: medieval parchment scribe (Pirenne, Economic and Social History of Medieval Europe), Renaissance notary (Mokyr, The Lever of Riches), 1900 attorney from BLS historical occupational wages, 1990 solo-practitioner Martindale-Hubbell billing rate × 3 hours, 2010 LegalZoom basic product, and 2026 frontier-model API cost for ~5,000 words. The 2026 figure assumes no lawyer review — adding one hour of review (~\$300) would put it near the LegalZoom band; the chart shows the ceiling on unattended drafting cost, not the all-in legal-services cost.

New Dark Output

In contrast, new Dark Output is work that did not happen before AI made it cheap enough to do. No wage bill disappears because no firm or household would have paid a human to do that work at prevailing prices. For example, when literature reviews fall from \$2,000 to \$2, we do not do the same number and pocket the savings, we do them before every project! Summarizing the last six months of emails on a theme in your inbox is useful. Running an academic literature review before an interview is useful.

Both can create real value, but neither leaves a clean economic trace beyond the tokens, API calls, cloud spend, or subscription that made the task cheap enough to run.

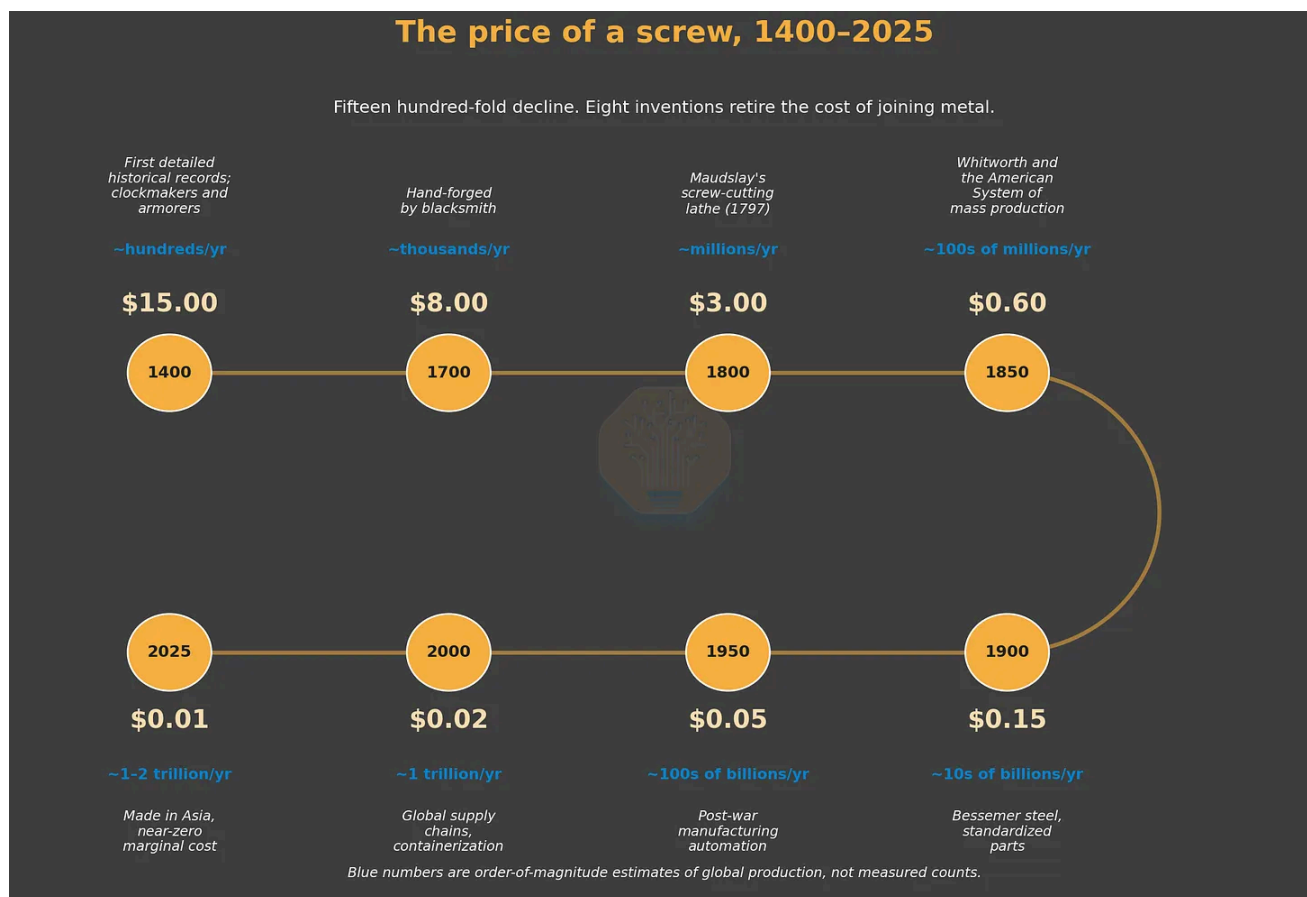
There are anecdotal signs that a large fraction of current token spend is for new work that wasn't previously paid for rather than replacing existing work. But the exact magnitude is opaque as it sits behind the anonymizing curtain of tokens. Identifying if a specific AI task is creating value and how much would likely be difficult even if you had the full conversation trace, as it is the national accounts will at best see AI revenue.

Captured AI Output

A final category of AI output is work that was previously done by humans and now is done by AI, but that can still charge the same amount as before. This captured AI will only exist where companies have genuine market power, and can protect prices in the face of declining costs of production. Consider two scenarios, first a firm that used to buy a \$10,000 HR service from an outside provider now buys that HR service for \$10,000 from an AI HR provider. In that case the output still is captured in national accounts and all that disappeared was the wages and workers. In the second version that \$10,000 service is now done internally for \$10 of tokens. In that scenario GDP has declined by \$9,990 despite the same work being done.

Why Services aren't like Goods

Manufacturing automation gave statisticians something to count. If machinists got better at making screws, the factory would report they made more screws, at lower costs, or better margins. Real GDP could rise because it was based on the quantity of output. So as the price of screws fell by 99+% over the past 6 centuries, we can count that the quantity of screws also went up on the order of 10 billion times. Real GDP correctly captures this as growth and productivity



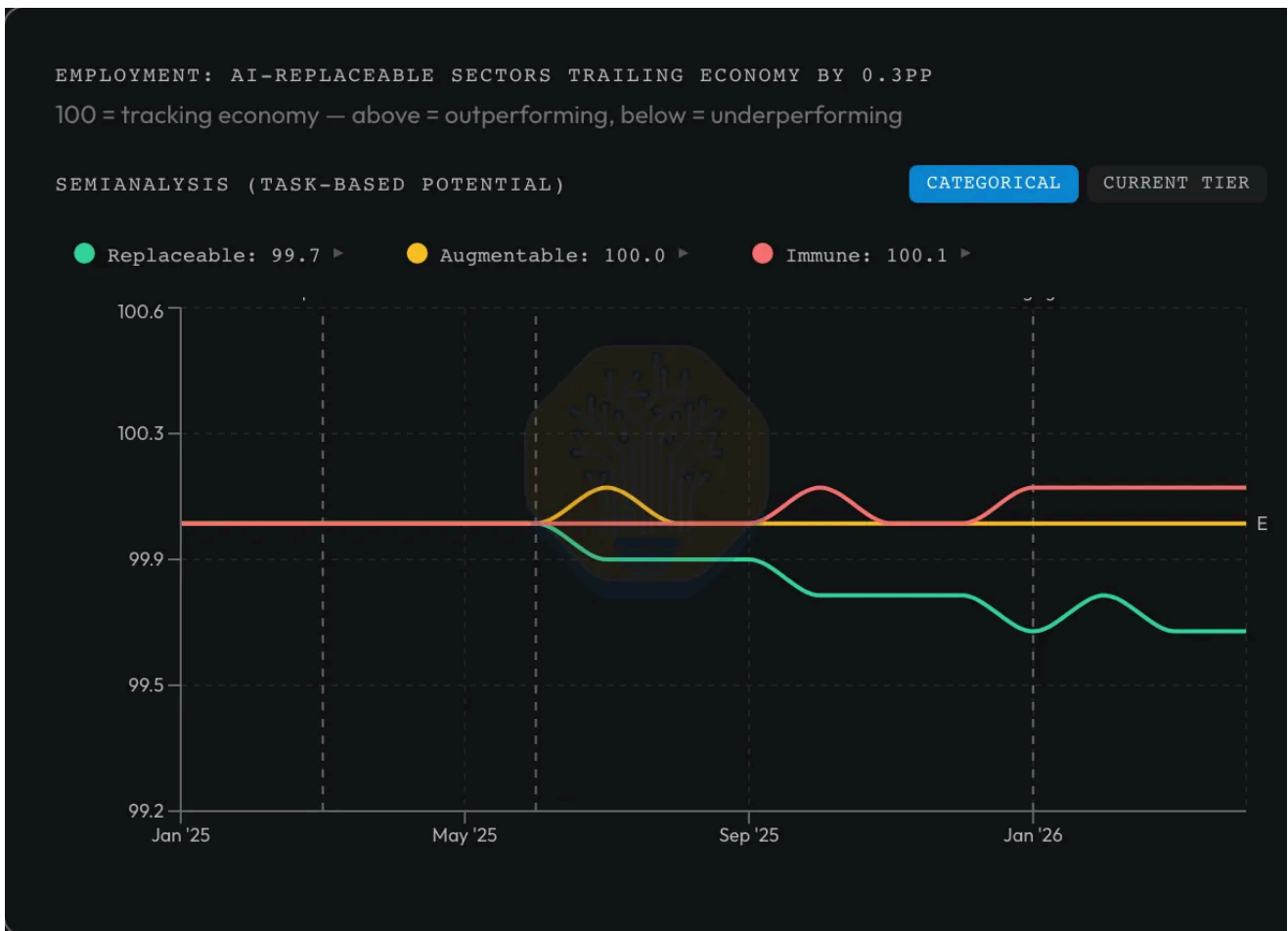
Source: Real price of a single common iron or steel screw in 2025 USD with order-of-magnitude estimates of global production. Pre-1900 figures are reconstructed from craft-shop prices, journeyman wage rates, and industrial-revolution case histories (Adam Smith, *Wealth of Nations*; Robert Allen, *The British Industrial Revolution*; Mitchell, *British Historical Statistics*); modern endpoint via Home Depot / McMaster-Carr retail. CPI deflation via [MeasuringWorth.com](https://www.measuringworth.com).

We lack a functional vocabulary for units of services, and mental work. As useful as it would be, there is no measure of ‘mind power’ that does for AI what horsepower did for the Industrial Revolution. Horsepower gave people a way to compare machine output with animal and human labor. Tokens do not do that. A million tokens can produce junk, a useful email summary, a legal document, or a decision that changes a company’s strategy. The economic value depends on the output, not the token count.

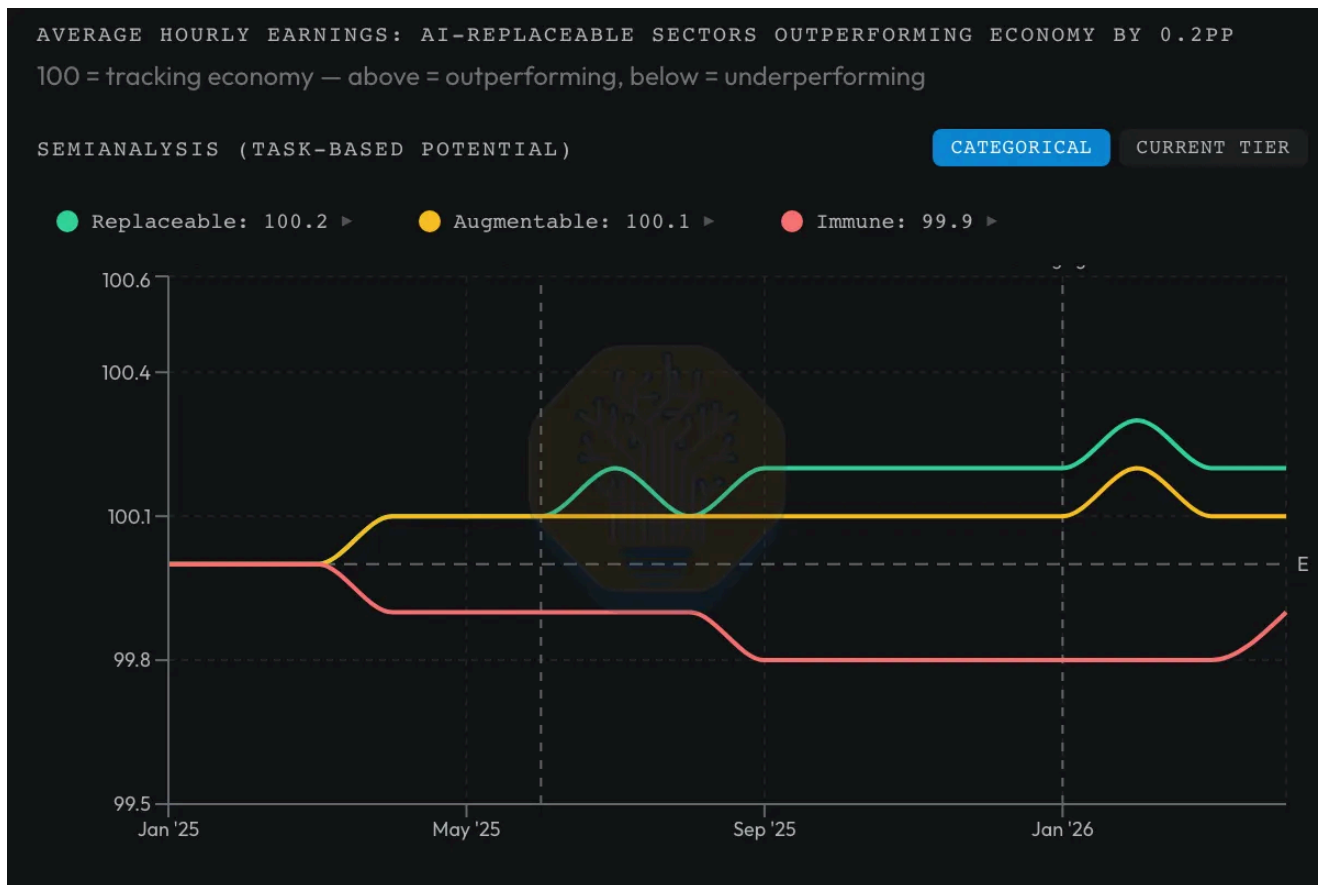
Finger Prints of Dark Output

A common observation in AI commentary is that junior staff are being displaced from routine work first. The corollary is that average wages in exposed occupations can rise because the lower-paid workers leave the sample. The cheapest workers disappear from the data. No one got a raise, and yet wages rose.

Employment in the most AI exposed sectors of the economy is falling relative to the broader economy. Yet those same underperforming segments are showing relative wage increases.



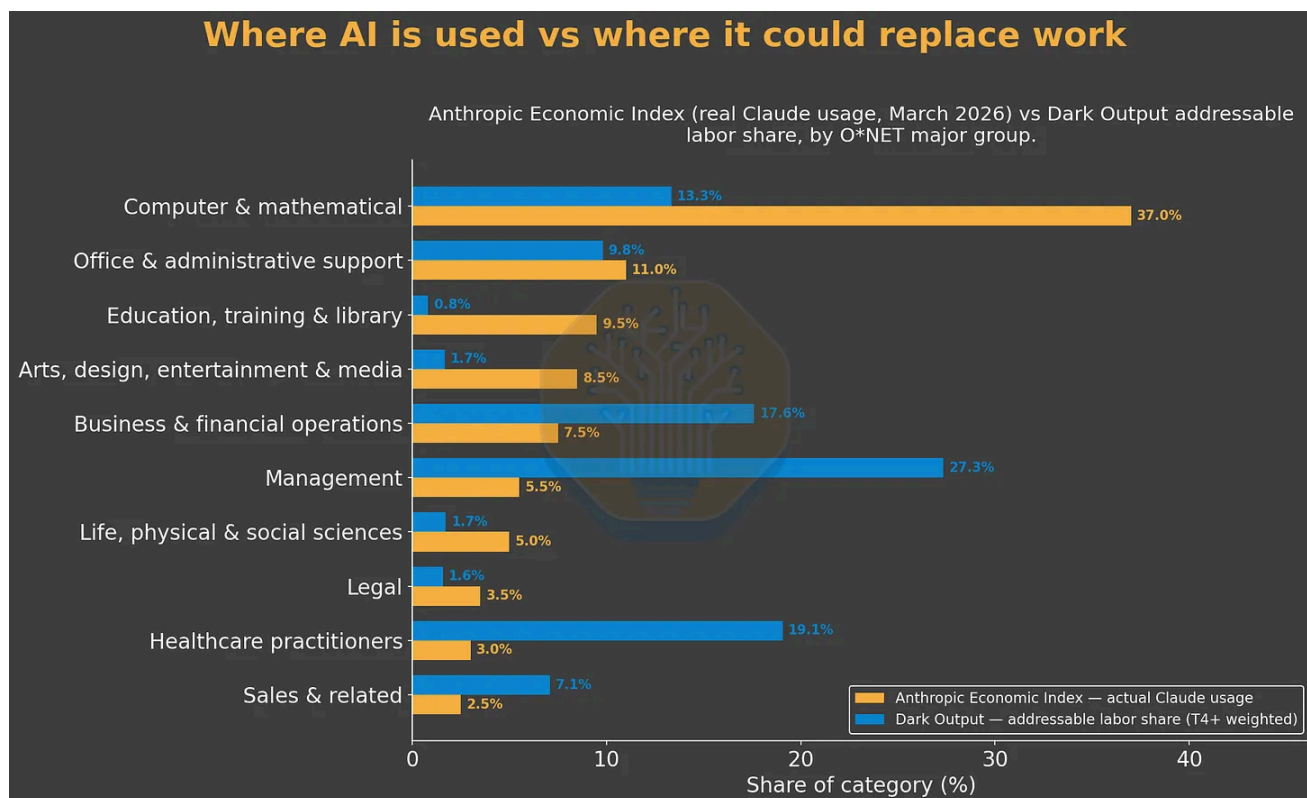
Source: SemiAnalysis Tokenomics: Dark Output



Source: SemiAnalysis [Tokenomics: Dark Output](#)

This mismatch between employment and wages is one fingerprint of AI displacement we track in our Dark Output dashboard. This is not a direct measurement of dark output, but the sort of odd measurement artifact that dark output would create as it became more prevalent.

An initial sign of new dark output is the heavy prevalence of Token usage in segments of the economy that are not showing signs of rapid labor deterioration. In Anthropic's Economic Index from March 2026 they show 37% of tokens are being used in computers and mathematics and yet the contribution to GDP from investment in software has not broken from its pre-AI trend and wasn't even at an all time high.



Source: SemiAnalysis Task Benchmark (T4+ verification, April 2026) overlaid on the Anthropic Economic Index (September 2025). DWA classification via O*NET; usage shares from Anthropic AEI public data.

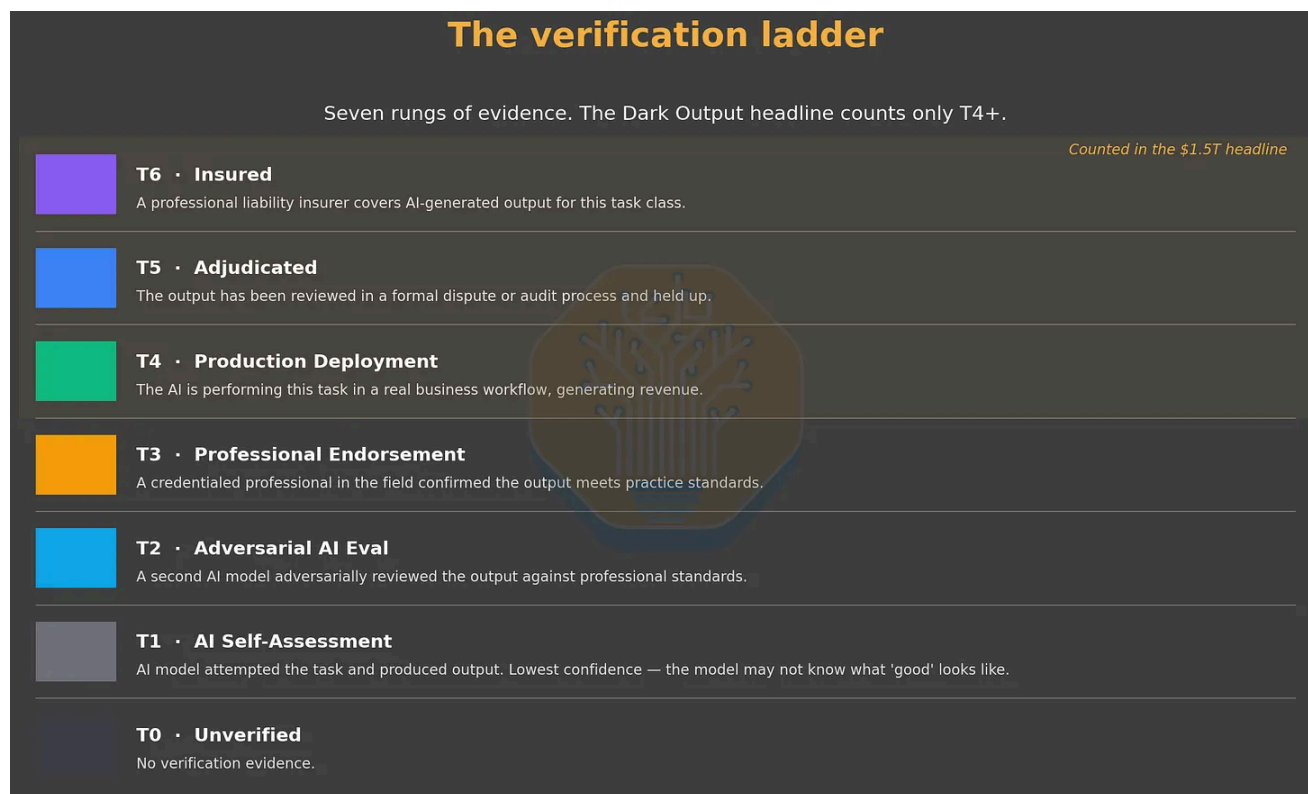
Why We Use Market Signals, Not Benchmarks

Benchmarks answer the wrong question, and they answer it late. Expert evaluations ask whether AI can satisfy an evaluator under test conditions, often an evaluator who expects expert work. They are expensive, slow, subjective, and backward-looking because expert time is scarce. Labor augmentation and displacement does not require AI to beat the best lawyer, analyst, or engineer. It requires AI to be good enough, cheap enough, and reliable enough to aid or replace the worker who would have done the task at prevailing wages. That is why we monitor public claims companies make about their own business practices, rather than abstract claims about what another firm could do.

The Evidence Ladder

Market signals vary in strength. Tiers 1 and 2 are benchmark-driven. They suggest a model can complete a task under test conditions, we only use them to estimate the cost of AI completing the task. Tier 3 is the hype layer: a public claim that a product or company can do the work. Unfortunately it is also the human verified benchmark layer. In our view a business saying the tools are in use in production is a stronger level

of evidence. A court fight where a firm successfully defends AI work is stronger than that. An insurer underwriting the risk is the strongest signal because a third party has priced the failure mode and taken on that risk. Our analysis shown in the dark output monitor treat these as an evidence ladder, not a binary yes or no.

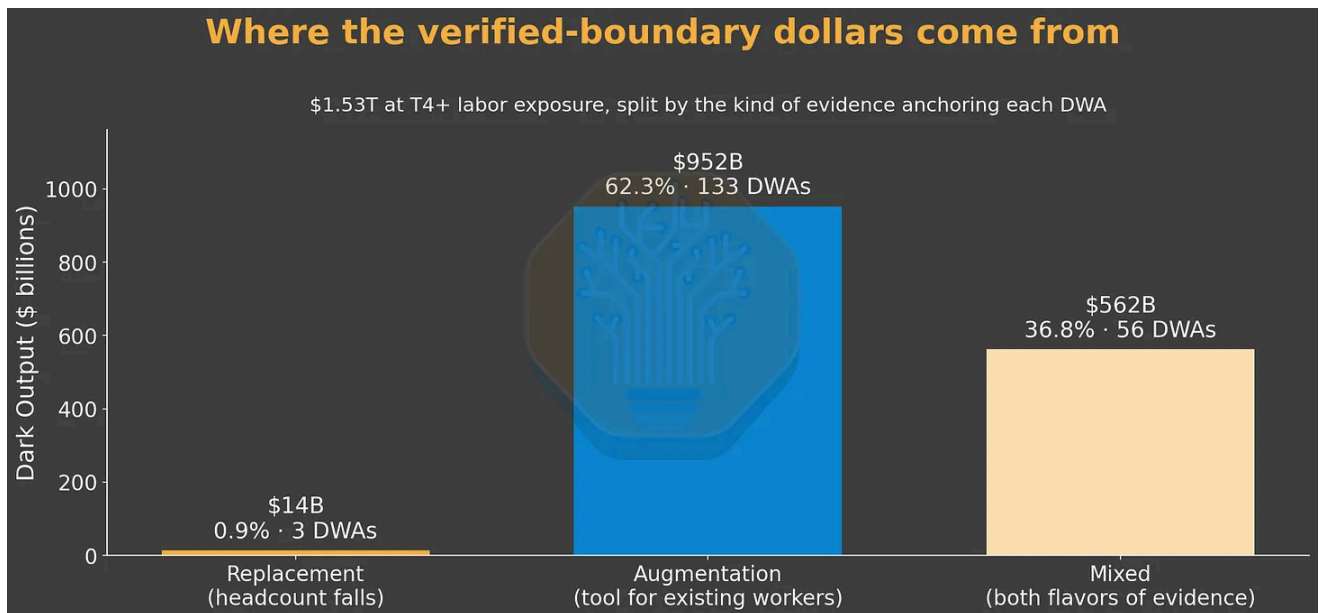


Source: SemiAnalysis [Tokenomics: Dark Output](#)

From Exposed Labor to Dark Output

The headline \$1.5T estimate is based on Tier 4 and above evidence from this ladder. It is not a claim that \$1.5T of labor has disappeared. It is a claim that tasks tied to roughly \$1.5T of labor cost sit inside categories where current AI has credible displacement potential. The number should be read as exposed labor, not missing output. We have not yet seen evidence of Tier 5 or Tier 6 activities, and that should read as a cautionary note on AI boosterism.

Most of the evidence we have collected to date points to AI augmentation and not AI replacement.



Source: SemiAnalysis and BLS OEWS wages and CES employment counts, mapped via O*NET DWAs. April 2026

When AI aids or takes over a task, the output does not automatically disappear. It only vanishes from the national accounts if the prices fall or worse the task gets moved inside the purchasing firm from the outside. If the market is uncompetitive a firm could still charge the same price as when a human did the work and the value would be captured as an explosion in margins and correctly show up in the national accounts.

What the Dark Output Monitor Can and Cannot Say

Our Dark Output monitor currently shows a map of pressure, not a forecast of layoffs or dark output. It identifies where firms have incentives to move from human labor to AI labor, where importantly the cost gap is largest.

It tracks tasks, occupations, wages, evidence tiers, token costs, and possible FTE displacement. Those are labor-side and input-side measures. They are enough to show where the transition can start, and in the areas where large numbers of tokens are being used without displacing labor we get a hint of where new Dark Output is being created.

A high-exposure sector should not be read as a sector where jobs have already disappeared. It should be read as a sector where the economics of substitution are visible: the task is identifiable, the wage pool is large, the market evidence is strong, and the token cost is low enough to matter. Only time will tell whether demand is

elastic enough for new output and labor augmentation to keep employment near historical levels. Or if civic, legal and government pressures are enough to prevent transition.

A useful corollary is the slow rollout of self-driving cars, where challenges in culture, insurance and market structure have been as important or more important than fundamental technological hurdles.

Where the Statistics Break

There are multiple measurement errors that could make output go dark. Different economic data sets will capture those errors in different ways. Treating all of those failures as 'GDP missing AI' makes the problem sound simpler than it is.

Boundary Shift

Boundary shift is work that used to be bought in the market that moves inside the firm or household. A paid research brief becomes an internal AI workflow. A contractor task becomes an employee prompt. The value may remain, but the transaction that made it visible disappears.

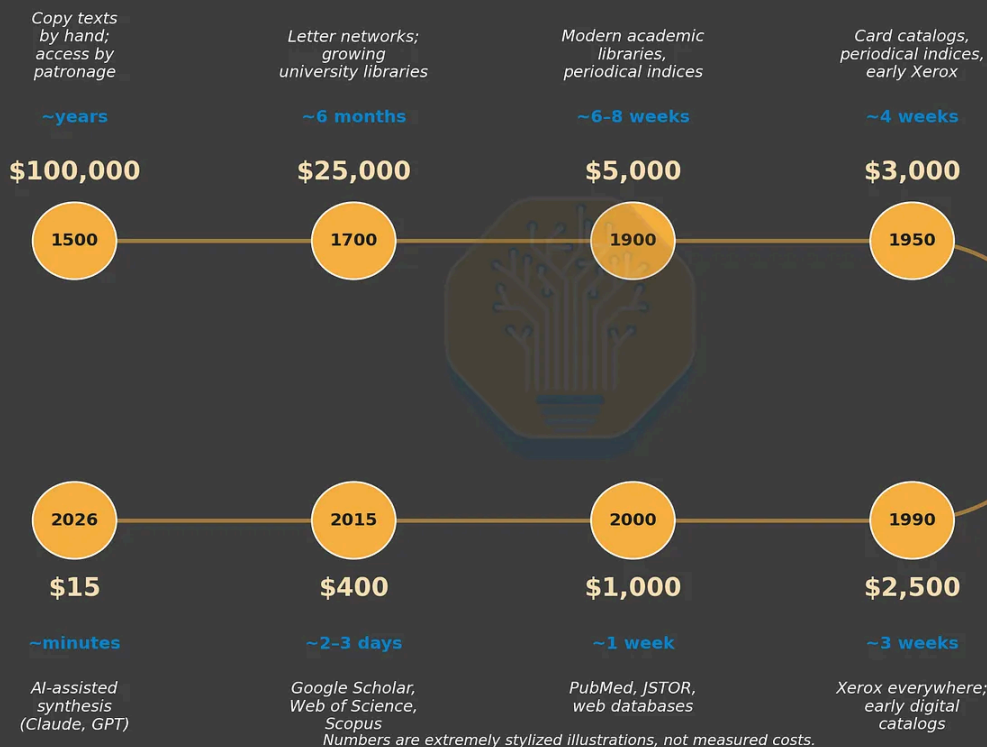
Price Collapse

There are no truly separate measures of quantity and quality in services. Receipts, wages, and hours worked, are captured but not quantity. As mentioned above there is no standard unit of legal services, no metric ton of literature reviews, and no barrel of consulting. If the accounts see lower receipts (because prices fell) and higher average wages (because junior staff are displaced), it will read as higher inflation and falling productivity and output.

These are genuinely hard questions, but there have unambiguously been productivity gains in the service sector from the march of technology. When those gains were relatively slow and constant it was a smaller issue than now that productivity has soared, and seems to be accelerating.

The cost of a literature review, 1500-2026

Seven thousand-fold decline. The AI-era parallel to the screw arc — compressed into 526 years.



Source: SemiAnalysis — stylized reconstruction. Pre-modern anchors from Mokyr (A Culture of Growth) and Goldgar (Impolite Learning); 20th-century points from NSF Survey of Earned Doctorates and Bowen (Higher Education in the Digital Age); 2026 figure from Dark Output Monitor. Costs deflated to 2025 USD via MeasuringWorth.com. Numbers are illustrative of the arc, not measured prices.

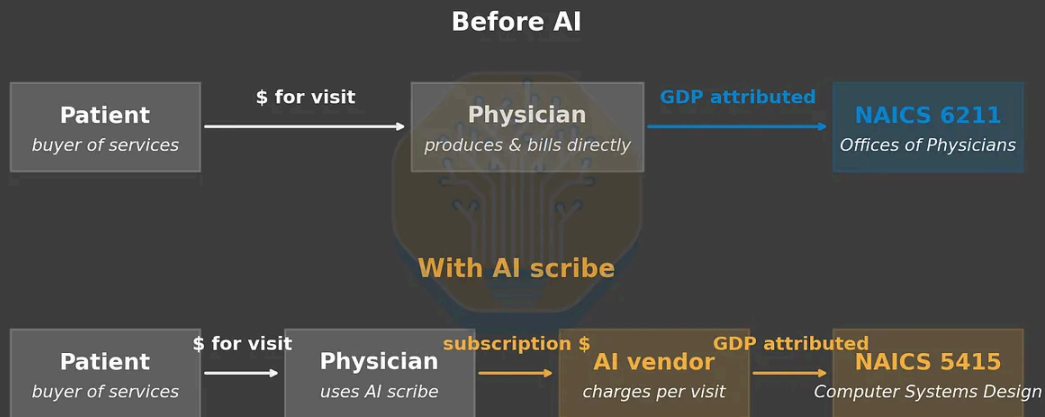
Sector Misrouting

Sector misrouting can occur when AI creates value in one sector while the transaction appears in another. The accounts end up counting the screws while missing the houses being built with them. A hospital may use AI to process paperwork faster, but if the only place AI shows up is in the revenue for an AI company or software provider, it will skew the national statistics. GDP-by-industry can make AI vendors look like the source of the value while the adopting sector looks stagnant.



AI-mediated transactions land in the wrong sector

A medical consultation note-taking task; same work, different national-accounts attribution. The bill ends up in the tech sector, not healthcare.



Same medical work. Healthcare-sector value-added falls or stagnates; Information / Computer-services value-added rises. The arithmetic of disinflation and output measurement breaks at the sectoral seam.

Source: SemiAnalysis. Conceptual flow diagram, no data series — illustrates how an AI-mediated service routes revenue from the producing NAICS sector (e.g., 6211 Offices of Physicians) into the AI vendor's NAICS sector (e.g., 5415 Computer Systems Design). Mechanism applies wherever a third-party AI tool replaces an internal labor input.

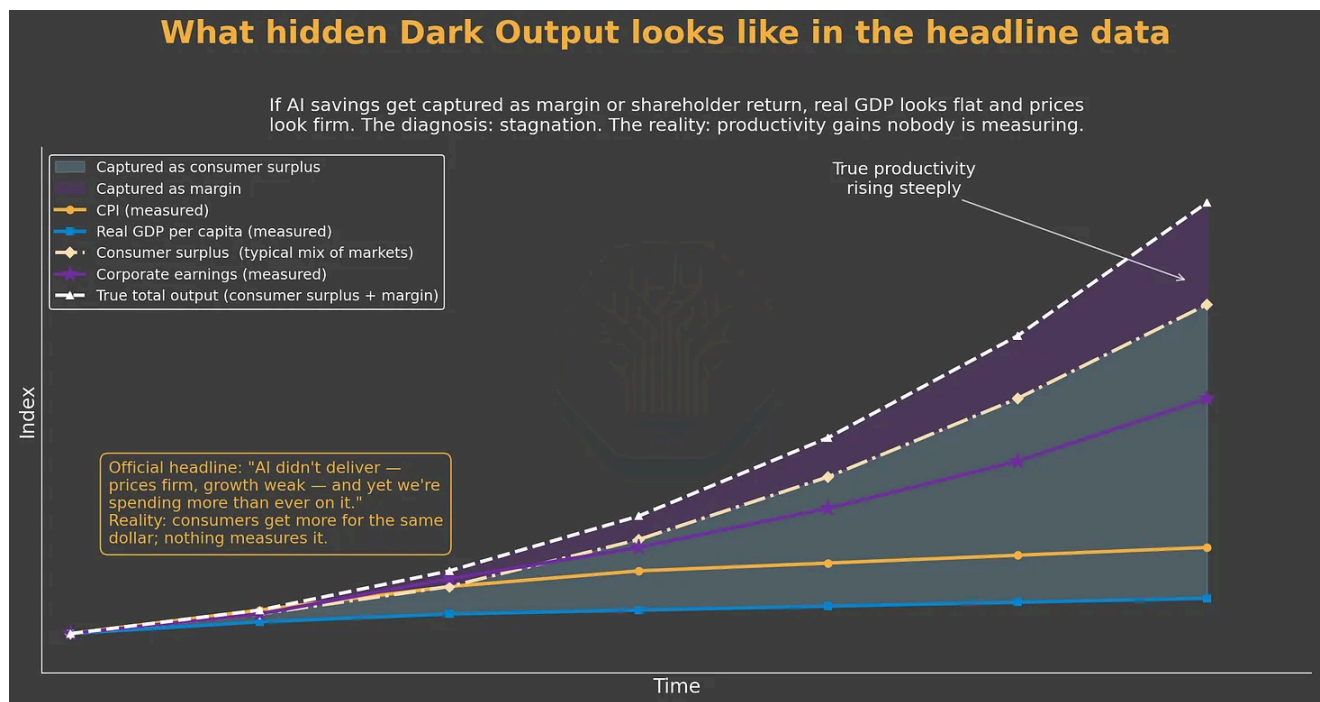
New Work Invisibility

We have already discussed new work invisibility, but if there is no receipt beyond tokens, the work is only visible at the cost of tokens. Real economic work is being done, we are better prepared for a meeting because AI wrote a dossier on the person across the table for a few tokens, but that value does not show up anywhere. It can seem ludicrous to estimate what it would have cost in dollars and hours even a few years ago to accomplish many of the tasks that are facilitated now for pennies and minutes. Any reasonable measure of the macro economy must in some way account for this or the AI boom may read to the data as an AI bust.

Macro Data Is Our Best View of the Economy. Breaking It Risks Errors

Macro data is our best view of the economy. It is always imperfect, and playing catch-up with changes in the economy. But it is part of how investors decide whether a boom is real, how policy makers decide to weigh the balance of unemployment and inflation, and how firms decide whether to hire, automate, or build. If AI breaks the data links

between labor, output, prices, and sectors, decisions get worse. The risk is that we keep using the same data even as they are becoming less accurate.



Source: SemiAnalysis

The economy can clearly see the costs of AI. Data centers, GPUs, electricity, water, and token spend are all visible. Knowing whether token spending represents real economic output that is durable or whether firms are playing with a shiny new tool is critical for good decisions.

Monitoring the Invisible: The Work Ahead

The Dark Output monitor shows us one measurable corner of the problem. It shines a light on labor that may be displaced, and token costs are low enough to matter. We will work to find other shadows of Dark Output and include them in the monitor.

These measurement errors cannot stay inside one dashboard. If AI creates real value that our statistics cannot see, everyone has a stake in learning how to measure it. With visible costs and revenue, but invisible output critics can dismiss AI as a bubble and not adapt. Politicians and investors need ways to classify, price, and tax new forms of output as the labor-based tax base weakens. AI has the potential to create surplus for communities to share but it also creates disruptions for workers and governments alike.

Dark output is not a reason to dismiss AI's costs. It is a call to work to measure the other side of the ledger. Labor displacement, power demand, water use, and land use are visible now. Tokens spent is visible. The output is harder to see. Cheap screws became countable output. Cheap AI work may not. If AI is creating an event on the scale of the Industrial Revolution, we need economic data that can see more than the displacement it causes.

Appendix 1. AI and Feminist Economics

There is an economic precedent worth naming explicitly: Care Economics, within the broader Feminist Economics tradition. It helps us think seriously about goods and services that have no line item in GDP. Even granting that GDP, as currently constructed, is the right metric for evaluating AI's economic impact, if AI generates enormous consumer surplus while displacing a great deal of formal output for a relatively small quantity of tokens, the score will still be misread. The Stiglitz Commission argued exactly this in 2009, and they were right that GDP is a poor proxy for well-being.

The Feminist Economics literature documented these issues at scale. Marilyn Waring showed in 1988 that the committees who drafted the original System of National Accounts were 91.7% male. One sentence in the founding document dismissed much of women's economic contribution, raising children, maintaining households, caring for the elderly and sick, as "of little or no importance" to the national accounts. Duncan Ironmonger calculated that Australia's household economy was 78% the size of its entire market economy. The UK's Office for National Statistics put household production at 63.1% of measured GDP. The International Labour Organization estimated 16.4 billion hours of unpaid care work performed daily, worth \$11 trillion a year, three times the global technology industry. By the conventions of national accounting, all of it has zero value.

This is not ancient history. It is not a resolved methodological debate. It is the same production boundary, updated but structurally unchanged, that is about to encounter AI-generated output at industrial scale. AI could push a large fraction of work out of the priced-and-produced region into just-produced, with the cost delinked from the production.

The economist Margaret Reid proposed a test in 1934 that remains the sharpest diagnostic: if work could be delegated to a paid third party, it is productive. When a family hires a housekeeper, the housekeeping enters GDP. When a family member does the same work, it does not. The act is identical. The accounting treatment depends entirely on whether money changes hands.

AI makes virtually every information task delegable. A large language model can draft a legal brief, analyze a financial statement, write a marketing plan, triage a patient complaint, generate code, or compose a research summary. In each case, the work was previously performed by a paid human and counted in GDP.

If an AI is asked to take notes on a medical consultation today, the only place that transaction can show up in the national accounts is buried in the bill from the AI company. Nowhere is the use itself reported in a way that would let disinflation or output be calculated correctly. We are using the same old GDP ruler we always have, while the production function pushes more of the economy into the no-man's-land of Dark Output.

A candid note on what our own framework inherits from this history. Displacement Dark Output measures only paid market labor: BLS wages, BLS employment counts, O*NET work activities. It does not measure AI's impact on unpaid care work, household production, or the informal economy. We invoke Waring and Ironmonger to establish that the production boundary is constructed and politically contested, then build a measurement system that operates entirely within that boundary. This is a deliberate choice, not an oversight. The data infrastructure for measuring market labor displacement (BLS, O*NET, employer surveys) exists and is auditable. The infrastructure for measuring household AI adoption does not, and inventing it would stack measurement uncertainty on measurement uncertainty. But the limitation is real. Dark Output reproduces a known exclusion. The 16.4 billion daily hours of unpaid care work that the International Labour Organization documented are no more visible in our framework than in the one we critique. We do not claim otherwise. Our key observation is that the problem documented by these alternative frameworks is about to get worse. Much of the displacement we measure also risks falling disproportionately on occupations with high female employment shares (administrative work is 72% female, BLS). We do not yet disaggregate by gender, but it is a logical extension.

All AI uses in the non-transactional production sphere are another form of Dark Output. When someone uses AI to do a domestic task faster, more easily, or better than before, the activity does not move from produced/priced into produced/unpriced; it just enlarges the produced/unpriced economy.

Appendix 2. AI and Feminist Economics

Since the 1990 Griliches conference, service accounting has improved, but in targeted ways. BLS expanded service producer price indexes, with PPI service coverage reaching more than 70% of the services sector by 2009 and the headline PPI system moving to Final Demand-Intermediate Demand in 2014 to include services, construction, government purchases, and exports. BEA moved to chain-type Fisher quantity and price indexes, integrated GDP-by-industry with input-output accounts, capitalized software and R&D, and improved treatment of difficult sectors like finance, insurance, and R&D.

But the core problem remains. BEA still says most detailed NIPA components are measured in dollars, not units, so real quantity is usually estimated by deflating current-dollar spending with a price index. That works tolerably when the transaction, product, and price index all still describe the same thing. It breaks down when AI moves service work into subscriptions, tokens, or internal production. The accounts can see receipts, wages, and sampled prices, but not necessarily the legal memo, literature review, HR task, or code review that still got done. It also has no unit of quality, if an AI augmented literature review is 10x more exhaustive there is no current method to capture that fact.

SemiAnalysis is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.



Recommend SemiAnalysis to your readers

Bridging the gap between the world's most important industry, semiconductors, and business.

Recommend



184 Likes · 5 Restacks

← Previous

Discussion about this post

Comments Restacks



Write a comment...