

AI Dark Output: The Visible Cost of Invisible Output

AI 暗产出：不可见产出的可见成本

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

为什么 AI 日益增长的产出将成为历史上最难的经济衡量问题之一。AI“暗产出”最终可能占据经济活动的绝大部分，但其衡量却是一项挑战。

MALCOLM SPITTLER AND DYLAN PATEL

MALCOLM SPITTLER 和 DYLAN PATEL

MAY 30, 2026 2026 年 5 月 30 日



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.](#)

在 20 世纪 80 和 90 年代，宏观经济数据无法检测到新兴计算机革命所带来的贡献。罗伯特·索洛（Robert Solow）曾留下那句名言：“除了在生产率统计数据中，你可以随处看到计算机时代。”然而，尽管经历了互联网泡沫的兴衰，“美股七巨头”

（Magnificent 7）现在的市值已达欧洲市场的 1.8 倍。人工智能领域也出现了类似的问题：宏观经济数据尚不具备捕捉 AI 所创造价值的能力，而世界各地的头条新闻、公众情绪和政府却能迅速捕捉到以美元、瓦特、加仑和就业岗位为代价的成本。Matt Drach 对此也有不同于我们的有趣见解。

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.

2013 年一项枯燥的方法论修订将研发和知识产权投资纳入了 GDP 核算，使 20 世纪 90 年代的总产值增加了 ~\$3.6 万亿。在官方账目中，这部分增量被均匀分摊，因此增长率仅略有上升，但其数额已接近 2000 年全年 GDP 的 30%。AI 带来的衡量问题的严重程度令以往的衡量问题相形见绌，我们将国民经济核算目前无法看到的 AI 工作成果称为“暗产出”。由于 AI 的新产出大多集中在服务业，而国家统计数据在捕捉服务业生产率增长方面长期存在困难，因此更多的 AI 产出可能会处于“隐形”状态。

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.

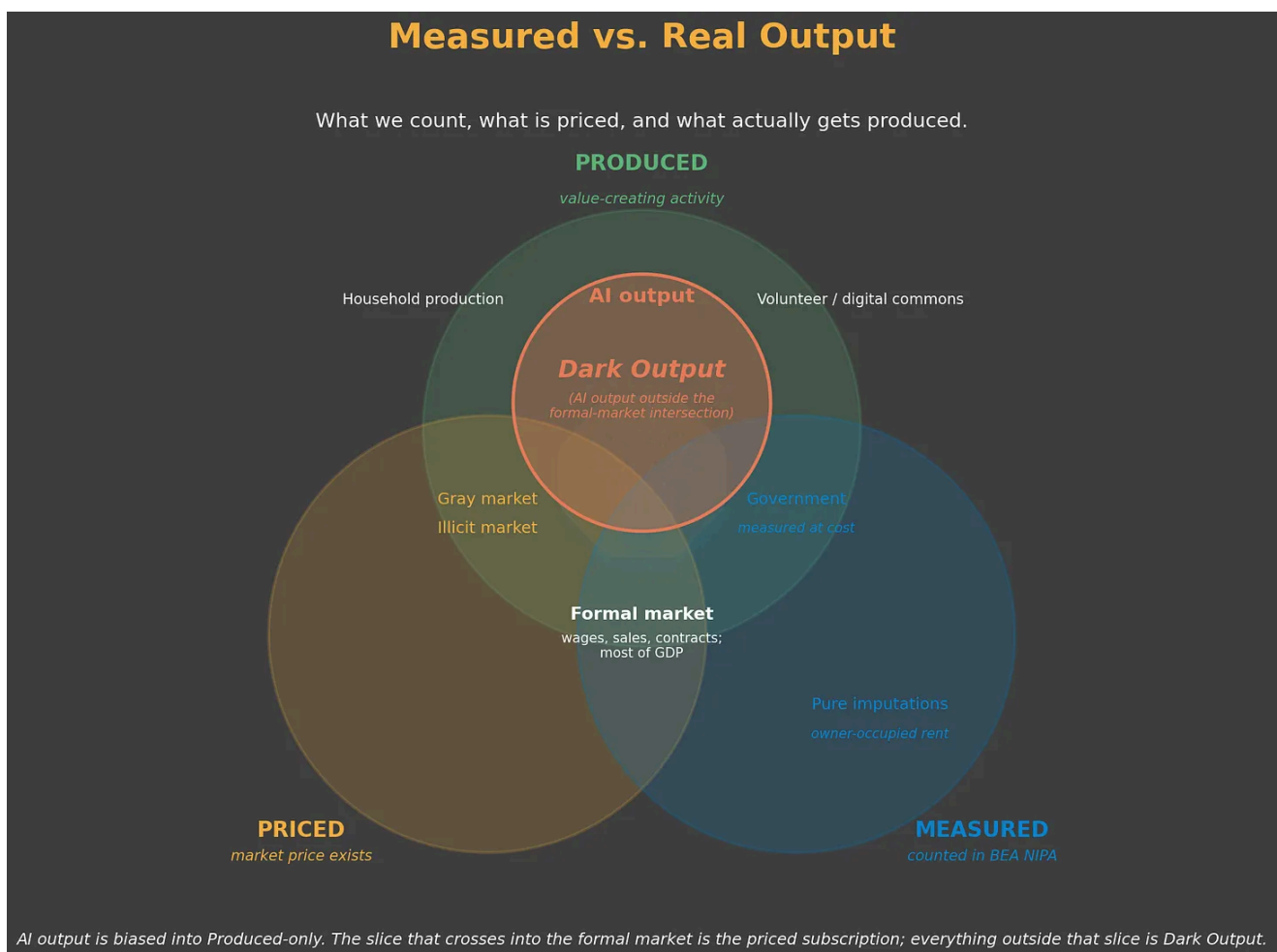
候任美联储主席凯文·沃什（Kevin Warsh）在 2025 年 12 月也承认了这一点：“如果你只盯着数据看，我认为你是在向后看。你会迟到的。你不会意识到这个国家能够实现更快的非通胀性增长。所以你必须做出抉择。”随着 AI 增长转向更活跃的资本市场融资，任何未能显示 AI 成果的衡量指标都将受到审视，以寻找泡沫的迹象。

Dark Output 暗产出 (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.

AI 的产出在变得可衡量之前，就已经真实存在了。我们可以捕捉到 Token 的支出，也可以捕捉到岗位的流失。但除非 AI 的产出以可见的价格出售，否则 GDP 只能记录下 Token 的支出。通常情况下，当某种事物的价格崩塌时，我们可以观察到这种通缩并将其结果称为生产力。然而，由于服务业中众所周知的难题（见附录 1），GDP 会将这些记录为下降，价格甚至可能显示为通胀。就像构成我们宇宙的暗能量一样，“暗产出”(Dark Output) 可能只会通过其对经济其他要素的影响而显现，而非通过直接观察。其中最明显的影响之一就是岗位流失，我们目前正在通过“暗产出监测器”对其进行追踪。



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

来源: 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.

我们正面临一场规模堪比工业革命的变革风险，即尽管企业在 AI 服务上的支出日益增加，但大部分新增产出却是不可见的。

Types of Dark Output 暗产出的类型

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:

暗产出（Dark output）是指由 AI 赋能、客观存在但在 GDP、价格、劳动力统计或行业核算中不可见或被严重扭曲的经济价值。我们将其分为两大类：

- **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.

替代性暗产出（Substitution dark output）是指以前由人类完成、现在由 AI 完成的工作。在我们的“暗产出监测器”中，我们已经识别出价值约 1.5 万亿美元的任务，这些任务可以被当前一代 AI 大幅增强或自动化。

- **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.

新增暗产出（New dark output）是指由 AI 完成的、以前并非由人类完成的新工作（可能是因为在 AI 使其变得廉价之前，这些工作的成本太高）。从长远来看，这一部分的规模可能远超替代性产出。

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

在这两种情况下，尽管统计系统未能察觉，但价值确实存在。这并非一个孤立的问题（参见附录 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 <i>real labor</i>	Priced <i>market price</i>	Measured <i>captured by GDP</i>	Verdict
AI work inside a flat subscription	✓	—	—	<i>Inside the gap</i>
Household / unpaid care	✓	—	—	<i>Inside the gap</i>
Informal & criminal economy	✓	✓	—	<i>Inside the gap</i>
Government services (cost-based)	✓	—	✓	<i>Measured at cost</i>
Owner-occupied housing (imputed rent)	—	—	✓	<i>Measured by imputation</i>
Capital gains, transfers, asset sales	—	✓	—	<i>Priced but not output</i>
Ordinary market output	✓	✓	✓	<i>Captured by GDP</i>

Rows in orange are Dark Output candidates: real activity that doesn't clear the three-test gauntlet.

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

来源：SemiAnalysis Tokenomics: Dark Output

Substitution Dark Output 替代型暗产出

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.

替代型暗产出 (Substitution Dark Output) 的一个例子是简单的法律文件。从理论 GDP 的角度来看, 无论是由律师起草还是由 AI 起草, 该文件对用户而言应具有相同的经通胀调整后的价值。然而, 服务业的 GDP 和通胀极难估算 (见附录 2), 法律服务并没有统一的“单位”, 只有律师的收款单据以及针对企业服务成本的调查。当 AI 接管这项任务时, 由于成本被吸收到 Token (代币/令牌) 费用中, 原有的收款单据便消失了。而当政府官员调查律师的服务成本时, 他们可能会发现平均价格反而上升了, 因为最简单的文件现在由 AI 而非律师完成。从 GDP 的视角来看, 这笔交易实际上已经消失了, 取而代之的只是经济体系中另一个无关部门产生的几美元 Token 支出。

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).

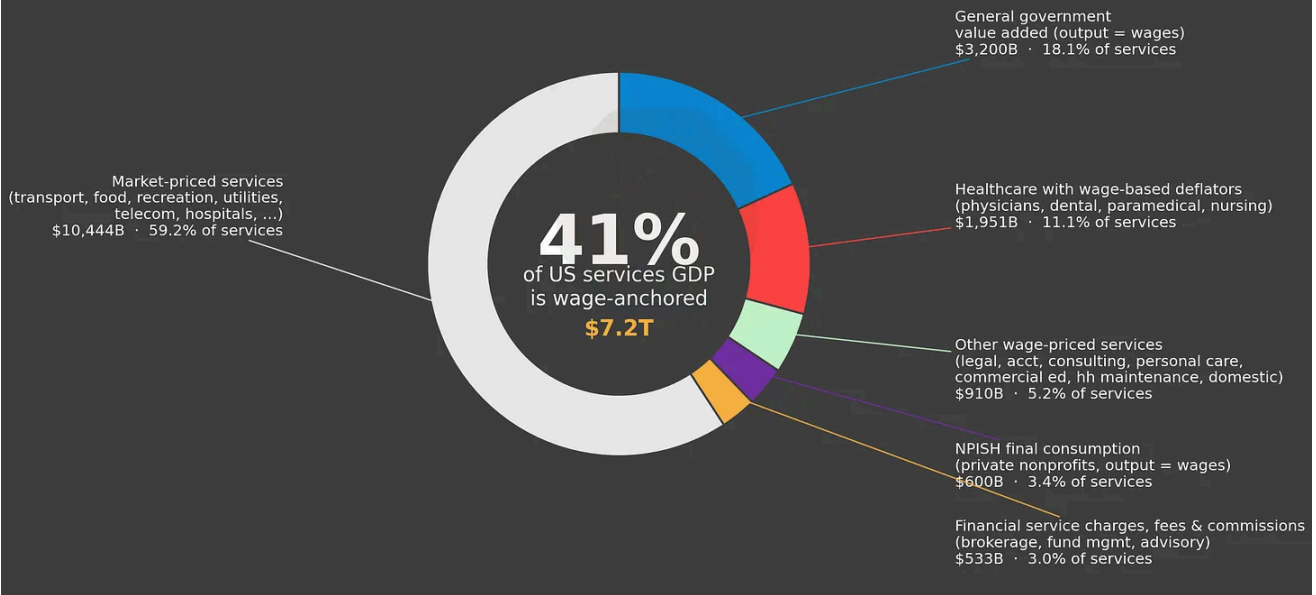
对于 Tokenomics 的订阅者, 我们持续追踪市场信号所显示的、当前 AI 有潜力取代的前沿任务。根据 AI 执行这些任务的方式, 它们可能会从国民经济核算中完全消失 (请参阅下文的“暗产出监测”部分)。

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.

除了住房以外, 国民账户中的大多数服务都是通过这种收据和标价系统来衡量的, 该系统通过将支出除以价格来反推所完成任务的“数量”。这种核算方式无法体现生产率的提高。当账户记录到较低的收入时, 会被解读为产出的下降。

Where US service-sector output is anchored to wages, 2025

Output measured as 'inputs' (gov, NPISH, domestic) PLUS market-priced services where the deflator is a wage index



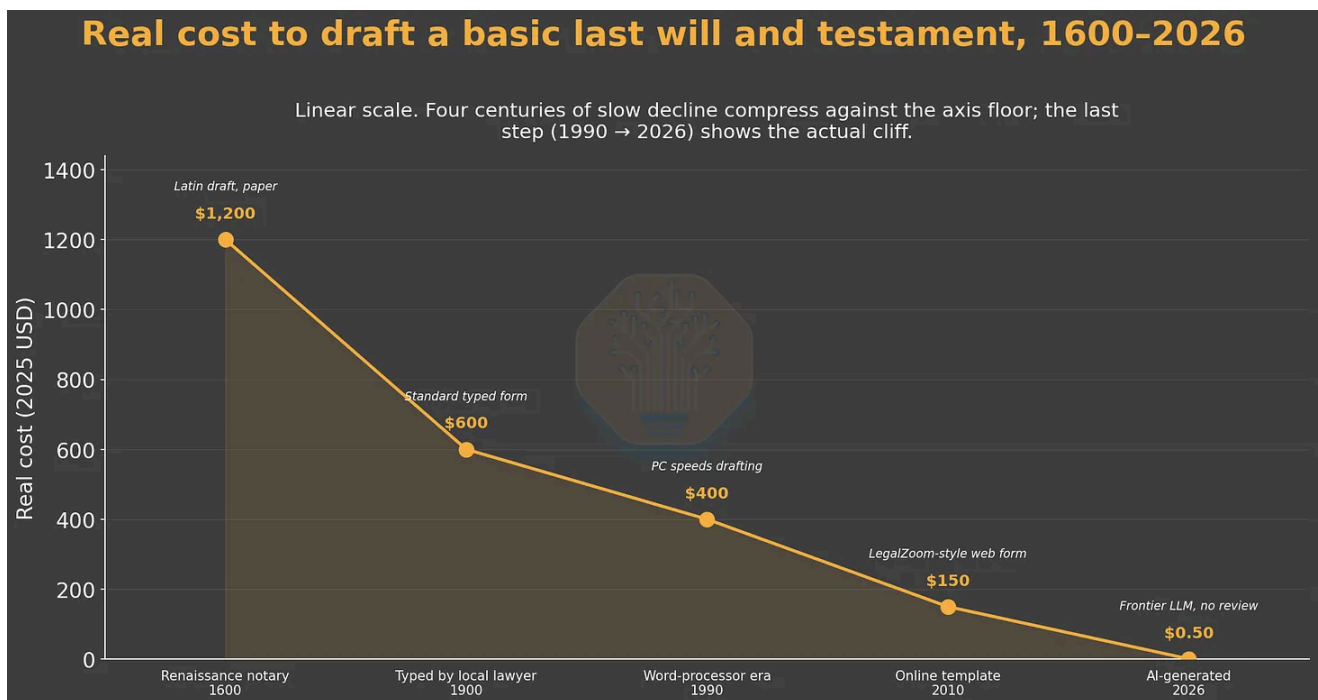
Source: BEA NIPA annual via Haver USNA, 2025. Healthcare wage-anchored = physicians + dental + paramedical + nursing homes (excludes hospitals, transaction-priced via DRG). Market-priced rest includes housing imputed rent and FISIM, which are imputed but not AI-relevant

来源：BEA NIPA 年度数据，经由 Haver USNA，2025 年。医疗保健工资锚定部分 = 医生 + 牙科 + 辅助医疗 + 护理院（不包括医院，医院通过 DRG 进行交易定价）。市场定价的其余部分包括住房虚拟租金和 FISIM，这些虽为虚拟计算但与 AI 无关。

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.

如下图所示，一份基础遗嘱的价格随着技术改变创作流程而在几代人的时间里持续下降，但由于这种下降是渐进的，所引发的误差并不极端。30 年间价格从 400 美元降至 150 美元，年均降幅不足 5%。而一年内从 150 美元降至 0.50 美元，成本降幅则超过了 99%。前者引入了偏差，而后者则直接从数据集中消失了。法律服务价格直到 1987 年才被纳入 CPI（消费者价格指数），自那时起，该价格指数已上涨了 4.6 倍（截至 2024 年 9 月）。由于没有将生产力的提升计算在内，该价格指数实际上已变成了一项雇佣成本指数。



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.

来源：SemiAnalysis。图示说明。锚点为折算至 2025 年美元的代表性成本：中世纪羊皮纸抄写员（Pirenne, 《中世纪欧洲经济社会史》）、文艺复兴时期公证人（Mokyr, 《富饶之杠杆》）、根据美国劳工统计局（BLS）历史职业工资计算的 1900 年律师、1990 年独立执业律师 Martindale-Hubbell 计费费率 × 3 小时、2010 年 LegalZoom 基础产品，以及 2026 年前沿模型 API 生成 5,000 字的成本。2026 年的数据假设无需律师审核——若增加一小时审核时间（约 300 美元），其成本将接近 LegalZoom 的区间；本图表显示的是无人值守起草成本的上限，而非法律服务的总成本。

New Dark Output 新的暗产出

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.

相比之下，新的“暗产出”(Dark Output)是指在 AI 使其成本变得足够低廉之前，根本不会发生的工作。由于没有任何公司或家庭愿意按现行价格雇人完成这些工作，因此并不会导致工资支出的消失。例如，当文献综述的成本从 2,000 美元降至 2 美元时，我们并不会维持原有的工作量并把省下的钱装进兜里，而是在开展每个项目之前都会做一遍！总结收件箱中过去六个月关于某个主题的邮件是有用的；在面试前进行学术文献综述也是有用的。这两者都能创造真正的价值，但除了产生这些任务所需的 Token、API 调用、云端支出或订阅费用外，它们几乎没有留下任何清晰的经济痕迹。

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.

有传闻迹象表明，目前很大一部分代币 (token) 支出是用于以往未曾付费的新工作，而非替代现有工作。但由于这些支出隐藏在代币这一匿名幕帘之后，其确切规模尚不透明。即便拥有完整的对话追踪记录，要识别特定的 AI 任务是否创造了价值以及创造了多少价值可能都很困难；就现状而言，国民经济核算体系充其量只能看到 AI 带来的收入。

Captured AI Output 捕获的 AI 输出

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.

最后一类 AI 产出是以前由人类完成、现在由 AI 完成，但仍能收取与以前相同费用的工作。这种被捕获的 AI 产出只存在于那些拥有真实市场势力、且在生产成本下降时仍能维持价格的公司中。考虑两种情形：第一种，一家公司以前从外部供应商购买价值 10,000 美元的人力资源服务，现在以 10,000 美元的价格从 AI 人力资源供应商处购买该服务。在这种情况下，产出仍被记录在国民账户中，消失的只是工资和工人。在第二种情形中，这项价值 10,000 美元的服务现在由内部通过价值 10 美元的代币 (tokens) 完成。在这种情况下，尽管完成了同样的工作，GDP 却减少了 9,990 美元。

Why Services aren't like Goods

为什么服务不同于商品

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

制造业自动化为统计学家提供了可量化的指标。如果机械师制造螺丝的水平提高了，工厂就会报告他们制造了更多螺丝，且成本更低或利润更高。实际 GDP 能够增长，是因为它是基于产出数量的。因此，随着螺丝价格在过去 6 个世纪里下降了 99% 以上，我们可以计算出螺丝的数量也增加了约 100 亿倍。实际 GDP 将其准确地捕捉为增长和生产力。

The price of a screw, 1400-2025

Fifteen hundred-fold decline. Eight inventions retire the cost of joining metal.



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.

2025 年美元计价的单枚普通钢铁螺丝的真实价格，以及全球产量的数量级估算。1900 年以前的数据是根据手工作坊价格、熟练工人工资水平和工业革命案例史（亚当·斯密《国富论》；罗伯特·艾伦《英国工业革命》；米切尔《英国历史统计》）重建的；现代终点数据通过 Home Depot / McMaster-Carr 零售价获取。消费者物价指数（CPI）通胀调整参考 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.

我们缺乏一套针对服务单元和脑力劳动的实用词汇。尽管“脑力”的衡量标准会非常有用，但目前还没有任何指标能像马力之于工业革命那样，为 AI 提供同样的衡量基准。马力让人们能够将机器产出与动物及人力进行比较。而 Token（标记）却做不到这一点。一百万个 Token 可能产生一堆垃圾，也可能生成一份有用的邮件摘要、一份法律文件，或是一个改变公司战略的决策。其经济价值取决于产出结果，而非 Token 的数量。

Finger Prints of Dark Output

暗输出的指纹

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.

AI 评论中一个普遍的观察结果是，初级员工正首先从常规工作中被取代。其推论是，受影响职业的平均工资可能会上升，因为低薪工人离开了样本。数据中最廉价的劳动力消失了。没有人涨薪，但工资水平却上升了。

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.

在经济体中，受 AI 影响最严重的部门，其就业人数相对于整体经济正在下降。然而，正是这些表现不佳的细分领域，却显示出相对的工资增长。

EMPLOYMENT: AI-REPLACEABLE SECTORS TRAILING ECONOMY BY 0.3PP

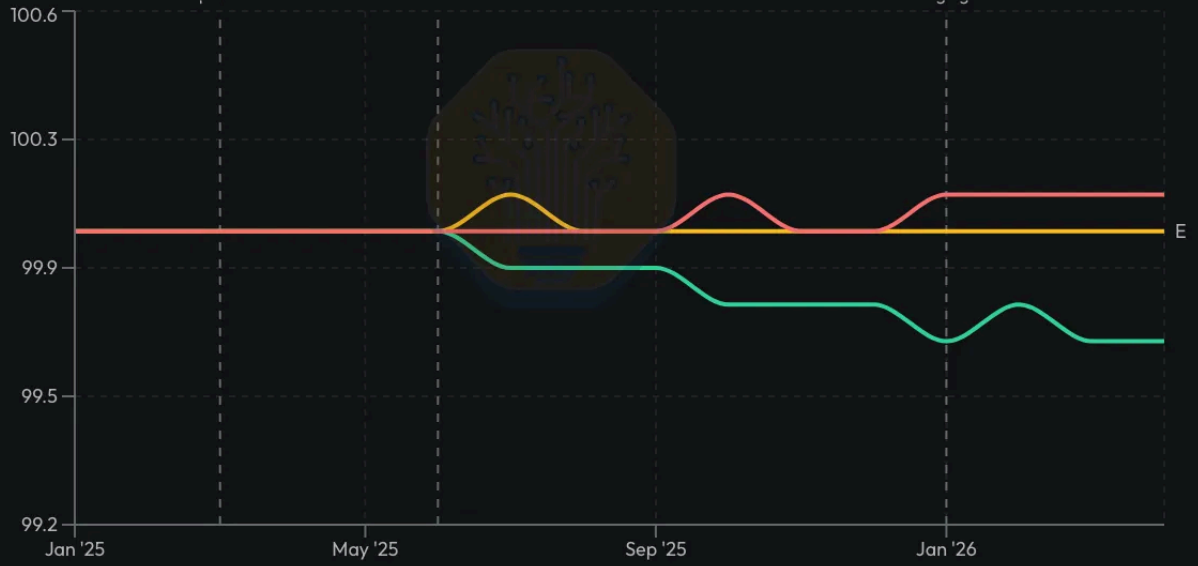
100 = tracking economy — above = outperforming, below = underperforming

SEMIANALYSIS (TASK-BASED POTENTIAL)

CATEGORICAL

CURRENT TIER

● Replaceable: 99.7 ▶ ● Augmentable: 100.0 ▶ ● Immune: 100.1 ▶



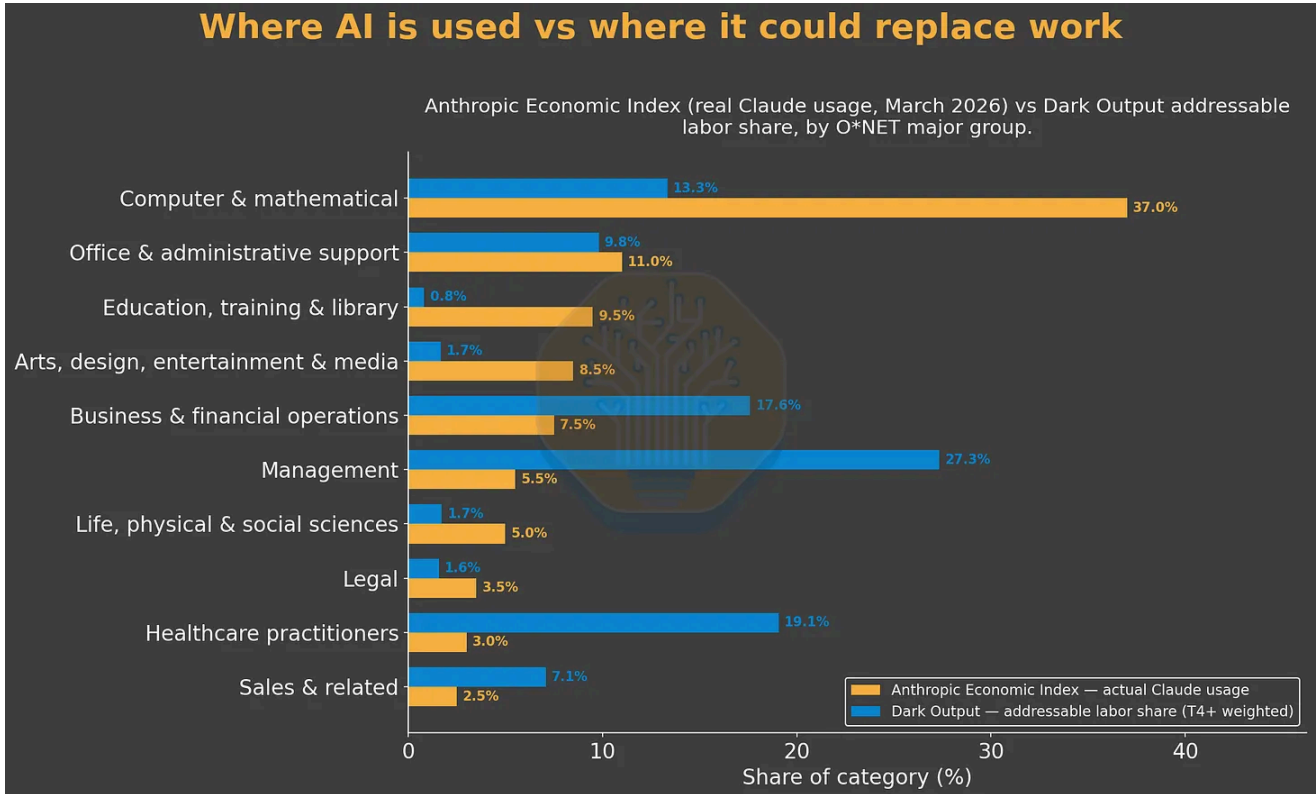
Source: SemiAnalysis [Tokenomics: Dark Output](#)

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

Source: SemiAnalysis Tokenomics: Dark Output

software has not broken from its pre-AI trend and wasn't even at an all time high.

新型暗产出的一个初步迹象是，在那些尚未表现出劳动力快速恶化迹象的经济部门中，Token 的使用率极高。在 Anthropic 2026 年 3 月发布的经济指数中显示，37% 的 Token 被用于计算机和数学领域，然而软件投资对 GDP 的贡献并未突破 AI 出现前的趋势，甚至没有达到历史最高水平。



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.

SemiAnalysis 任务基准 (T4+ 验证, 2026 年 4 月) 叠加 Anthropic 经济指数 (2025 年 9 月)。DWA 分类通过 O*NET 进行; 使用份额来自 Anthropic AEI 公开数据。

Why We Use Market Signals, Not Benchmarks

为什么我们使用市场信号，而非基准测试

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.

基准测试回答了错误的问题，而且回答得太晚。专家评估考察的是 AI 是否能在测试条件下满足评估者的要求，而这些评估者通常期待的是专家级的工作成果。由于专家时间稀缺，这类评估往往昂贵、缓慢、主观且具有滞后性。劳动力的增强与替代并不要求 AI 击败最顶尖的律师、分析师或工程师。它只要求 AI 足够好、足够便宜且足够可靠，能够辅助或取代那些以现行工资水平执行该任务的员工。这就是为什么我们监控公司对其自身业务实践的公开声明，而不是关注关于另一家公司能做什么的抽象陈述。

The Evidence Ladder **证据阶梯**

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.

市场信号的强度各不相同。第一层和第二层是由基准测试驱动的，它们表明模型可以在测试条件下完成任务，我们仅将其用于估算 AI 完成任务的成本。第三层是炒作层：即公开声称某个产品或公司可以胜任该工作。遗憾的是，这一层也是人工验证的基准层。在我们看来，一家企业表示这些工具已投入生产使用，是更有力的证据。而一场企业成功为 AI 作品辩护的法庭诉讼，其效力则更强。保险公司承保相关风险则是最强的信号，因为第三方已经对失败模式进行了定价并承担了该风险。我们在“暗产出监测器”(dark output monitor) 中展示的分析将这些视为一个证据阶梯，而非简单的“是”或“否”。

The verification ladder

Seven rungs of evidence. The Dark Output headline counts only T4+.

Counted in the \$1.5T headline



T6 · Insured

A professional liability insurer covers AI-generated output for this task class.



T5 · Adjudicated

The output has been reviewed in a formal dispute or audit process and held up.



T4 · Production Deployment

The AI is performing this task in a real business workflow, generating revenue.



T3 · Professional Endorsement

A credentialed professional in the field confirmed the output meets practice standards.



T2 · Adversarial AI Eval

A second AI model adversarially reviewed the output against professional standards.



T1 · AI Self-Assessment

AI model attempted the task and produced output. Lowest confidence — the model may not know what 'good' looks like.

T0 · Unverified

No verification evidence.

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

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

From Exposed Labor to Dark Output

从显性劳动力到暗产出

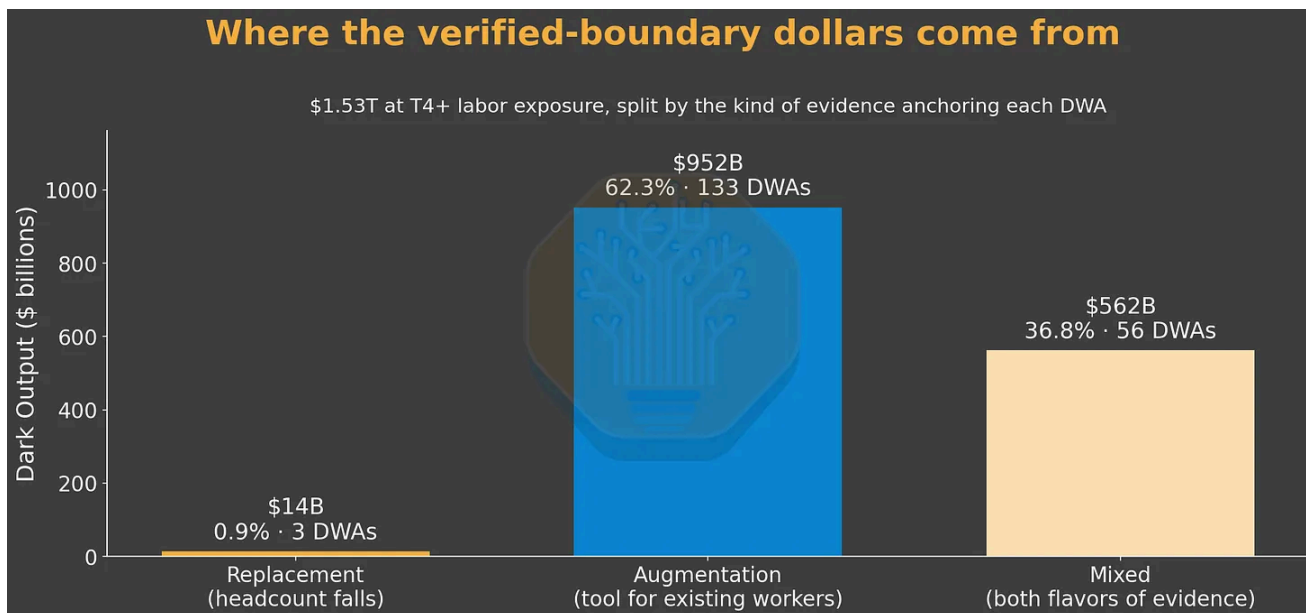
The headline \$1.5T estimate is based on Tier 4 and above evidence from this ladder. It is not a claim that \$1.5T of labor has disappeared. It is a claim that tasks tied to roughly \$1.5T of labor cost sit inside categories where current AI has credible displacement potential. The number should be read 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.

1.5 万亿美元的头条估算数据是基于该阶梯中第 4 级及以上的证据得出的。这并非声称已有 1.5 万亿美元的劳动力消失，而是指与约 1.5 万亿美元劳动力成本相关的任务，正处于当前 AI 具有可靠替代潜力的类别中。这一数字应被解读为“受影响的劳动力”，而非“缺失的产出”。目前我们尚未看到第 5 级或第 6 级活动的证据，这应当被视为对 AI 乐观主义的一种警示。



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

迄今为止，我们收集的大多数证据都指向 AI 增强，而非 AI 替代。



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

数据来源：SemiAnalysis、BLS OEWS 工资数据及 CES 就业人数，通过 O*NET DWA 进行映射。2026 年 4 月

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.

当人工智能辅助或接管某项任务时，其产出并不会自动消失。只有当价格下跌，或者更糟的情况——该任务从外部外包转为由采购公司内部承担时，它才会从国民经济核算中消失。如果市场缺乏竞争，公司仍可以收取与人工操作时相同的价格，而这部分价值将体现为利润率的爆发式增长，并正确地显示在国民经济核算中。

What the Dark Output Monitor Can and Cannot Say

暗输出监测器能说与不能说的事

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.

我们的暗产出（Dark Output）监测器目前显示的是一张压力图，而非对裁员或暗产出的预测。它识别了企业有动力将人力劳动转向 AI 劳动的领域，更重要的是，它指出了成本差距最大的地方。

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.

它追踪任务、职业、工资、证据层级、Token 成本以及可能的等效全职员工（FTE）替代情况。这些属于劳动力端和投入端的衡量指标。它们足以显示转型可能从何处开始；而在大量使用 Token 却未替代劳动力的领域，我们可以窥见新的暗产出正在何处产生。

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.

高暴露行业不应被解读为工作岗位已经消失的行业。而应被视为替代经济学效应显现的行业：任务可识别、工资池规模庞大、市场证据充分，且 Token 成本低到足以产生实质影响。只有时间才能证明，需求弹性是否足以支撑产出增加和劳动力增强，从而使就业维持在历史水平附近；或者公民、法律和政府的压力是否足以阻止这一转型。

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.

一个有用的推论是自动驾驶汽车的缓慢推广，其中文化、保险和市场结构方面的挑战与基础技术障碍同样重要，甚至更为重要。

Where the Statistics Break

统计数据失效之处

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.

存在多种测量误差可能导致产出被隐匿。不同的经济数据集会以不同的方式捕捉这些误差。将所有这些失效统称为“GDP 遗漏了 AI”，会使问题听起来比实际情况简单。

Boundary Shift 边界偏移

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.

边界转移是指过去在市场上购买的工作现在转移到了公司或家庭内部。一份付费的研究简报变成了一个内部 AI workflow；一项外包任务变成了一名员工的提示词。价值可能依然存在，但使其可见的交易却消失了。

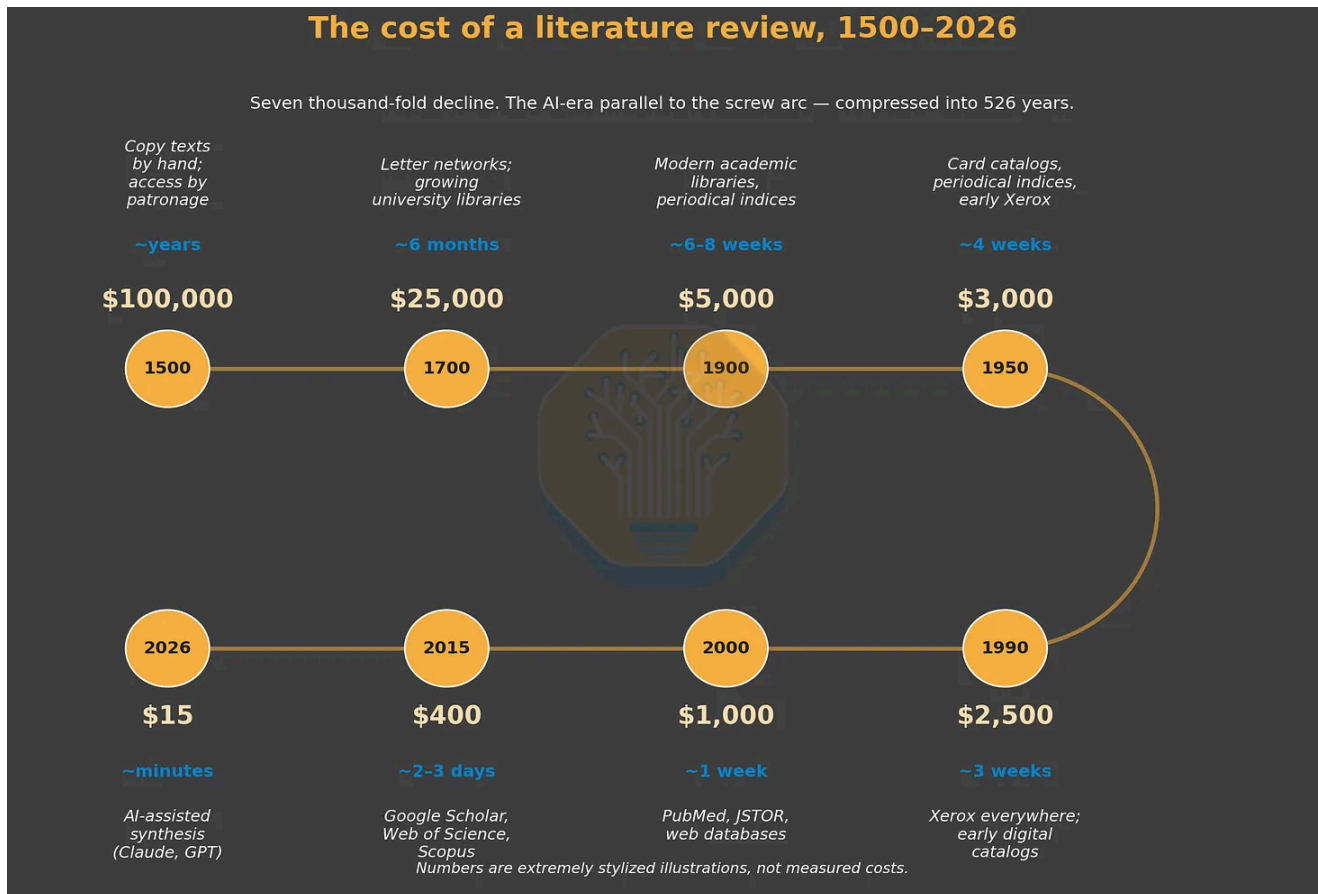
Price Collapse 价格崩塌

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.

在服务业中，并没有真正独立的数量和质量衡量标准。收据、工资和工作时长被记录了下来，但数量却没有。正如上文所述，法律服务没有标准单位，文献综述没有公吨，咨询服务也没有桶。如果会计账目显示收入降低（因为价格下跌）且平均工资升高（因为初级员工被取代），这将被解读为通胀上升以及生产率和产出的下降。

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.

这些确实是棘手的问题，但随着技术的进步，服务业无疑已经获得了生产率的提升。当这些增长相对缓慢且稳定时，其影响尚不如现在这般显著——如今生产率已经飙升，且似乎正在加速增长。



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

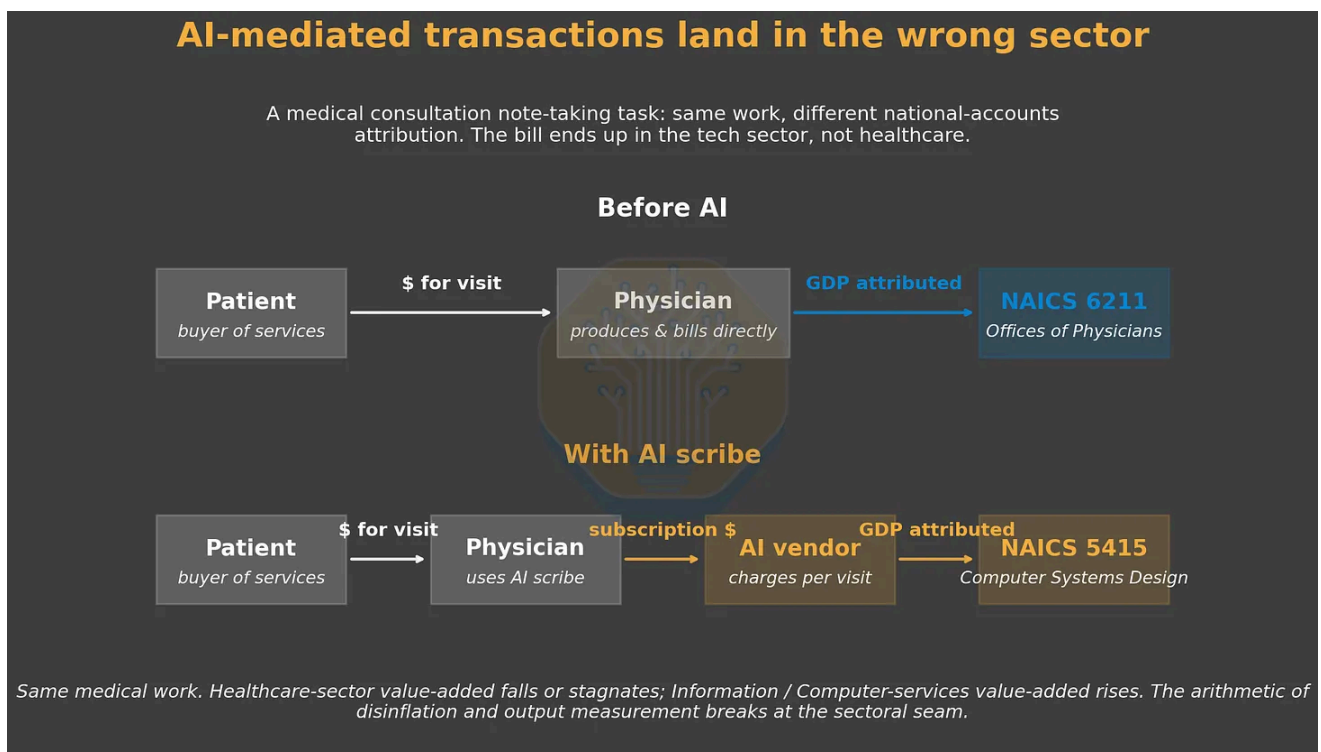
MeasuringWorth.com. Numbers are illustrative of the arc, not measured prices.

SemiAnalysis — 风格化重构。前现代数据参考自 Mokyr 的《增长的文化》(A Culture of Growth) 和 Goldgar 的《无礼的学习》(Impolite Learning); 20 世纪数据点来自 NSF 博士学位调查和 Bowen 的《数字时代的高等教育》(Higher Education in the Digital Age); 2026 年数据来自 Dark Output Monitor。成本已通过 MeasuringWorth.com 折算为 2025 年美元。数值旨在说明发展趋势, 而非实测价格。

Sector Misrouting 行业误导定向

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.

当 AI 在一个行业创造价值, 而交易却体现在另一个行业时, 就会发生行业误导。账目最终只统计了螺丝的数量, 却忽略了用这些螺丝建造的房屋。一家医院可能会使用 AI 来更快地处理文书工作, 但如果 AI 仅体现在 AI 公司或软件供应商的收入中, 就会扭曲国家统计数据。按行业划分的 GDP 可能会让 AI 供应商看起来是价值的源泉, 而采用该技术的行业则显得停滞不前。



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.

来源：SemiAnalysis。概念流程图，无数据序列——展示了 AI 介导的服务如何将收入从生产性 NAICS 部门（例如 6211 医生办公室）引导至 AI 供应商的 NAICS 部门（例如 5415 计算机系统设计）。该机制适用于任何第三方 AI 工具替代内部劳动力投入的场景。

New Work Invisibility **新工作的隐形性**

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.

我们已经讨论过新工作的隐形性，但如果没有代币（tokens）之外的凭据，这些工作只能以消耗代币为代价而存在。真实的经济活动正在发生，例如，因为 AI 只花了几个代币就为桌对面的人写了一份背景档案，让我们为会议做了更充分的准备，但这种价值在任何地方都没有体现出来。如果去估算在几年前完成这些任务需要花费多少美元和小时，而现在只需几分钱和几分钟就能搞定，这听起来似乎有些荒谬。任何合理的宏观经济衡量标准都必须以某种方式将此纳入考量，否则在数据层面，这场 AI 繁荣可能会被误读为 AI 萧条。

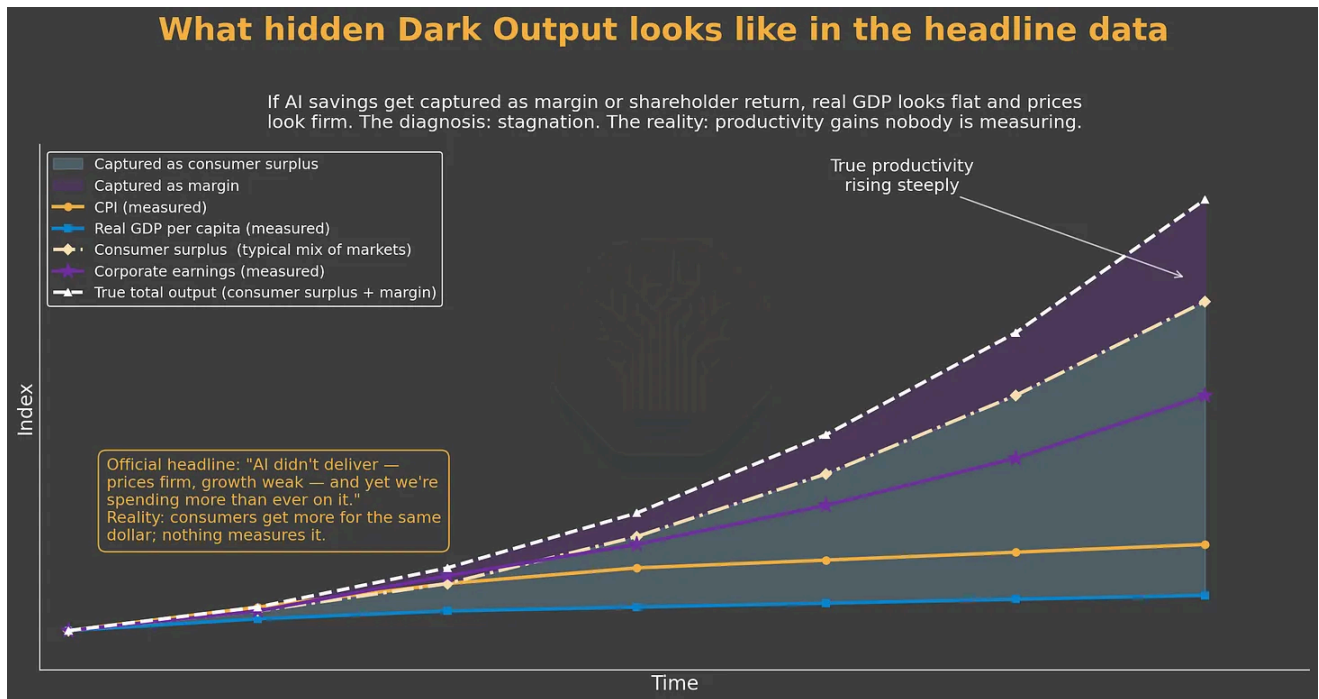
Macro Data Is Our Best View of the Economy. Breaking It Risks Errors

宏观数据是我们观察经济的最佳视角。破坏它将面临出错的风险

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.

宏观数据是我们观察经济的最佳视角。它总是存在缺陷，且往往滞后于经济的变化。但它是投资者判断繁荣是否真实、政策制定者权衡失业与通胀平衡、以及企业决定是招聘、自动化还是扩建的重要依据。如果 AI 破坏了劳动力、产出、价格和行业之间的数据关联，决策质量就会下降。风险在于，即便这些数据的准确性正在降低，我们仍在使用它们。



Source: SemiAnalysis [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.

经济领域可以清晰地看到 AI 的成本。数据中心、GPU、电力、水资源以及 Token 支出都是显而易见的。了解 Token 支出究竟代表了持久的真实经济产出，还是仅仅反映了企业在摆弄一个华而不实的新工具，对于做出明智决策至关重要。

Monitoring the Invisible: The Work Ahead

监测不可见产出：未来的任务

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.

“暗产出”(Dark Output) 监测器为我们展示了这一问题中一个可衡量的侧面。它揭示了那些可能被取代的劳动力，以及低到足以引起重视的 Token 成本。我们将致力于寻找“暗产出”的其他阴影区域，并将它们纳入监测范围。

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.

这些衡量误差不能仅停留在仪表盘内部。如果 AI 创造了统计数据无法察觉的真实价值，那么学会如何衡量它便与每个人的利益息息相关。如果成本和收入是可见的，而产出却是不可见的，批评者就会将 AI 斥为泡沫而拒绝适应。随着以劳动力为基础的税基削弱，政治家和投资者需要寻找新的方法来对这些新形式的产出进行分类、定价和征税。AI 有潜力为社区创造可共享的盈余，但同时也给劳动者和政府带来了冲击。

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.

暗产出 (Dark output) 并非忽视 AI 成本的理由。相反，它在提醒我们要努力衡量账本的另一面。劳动力流失、电力需求、用水量和土地占用在当下是显而易见的。消耗的 Token 也是可见的。但产出却更难被察觉。廉价的螺丝钉曾是可计量的产出，而廉价的 AI 工作可能并非如此。如果 AI 正在引发一场工业革命规模的变革，我们需要能够洞察其产出、而不仅仅是其造成的流失的经济数据。

Appendix 1. AI and Feminist Economics

附录 1. AI 与女性主义经济学

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.

有一个经济学先例值得明确提及：女性主义经济学传统中的关怀经济学（Care Economics）。它有助于我们认真思考那些在 GDP 中没有列支的商品和服务。即便承认目前构建的 GDP 是评估 AI 经济影响的正确指标，如果 AI 在产生巨大消费者剩余的同时，仅用相对少量的 Token 就取代了大量的正式产出，那么评估结果仍会被误读。斯蒂格利茨委员会（Stiglitz Commission）在 2009 年就提出了这一观点，他们正确地指出，GDP 并不是衡量福祉的良好指标。

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.

女性主义经济学文献大规模地记录了这些问题。玛丽莲·华林（Marilyn Waring）在 1988 年指出，起草最初国民经济核算体系的委员会成员中，男性比例高达 91.7%。该奠基性文件中的一句话将女性的大部分经济贡献——抚养孩子、维持家务、照顾老人和病人——斥为对国民账户“几乎没有或完全没有重要性”。邓肯·艾恩蒙格（Duncan Ironmonger）计算出，澳大利亚的家庭经济规模相当于其整个市场经济的 78%。英国国家统计局将家庭生产估算为衡量 GDP 的 63.1%。国际劳工组织估计，每天进行的无偿照护工作达 164 亿小时，每年价值 11 万亿美元，是全球科技产业的三倍。按照国民经济核算的惯例，所有这些的价值均为零。

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.

这不是遥远的历史，也不是一场已经解决的方法论争论。这正是那个即将以工业规模迎接 AI 生成内容的生产边界——虽然经过更新，但在结构上未曾改变。AI 可能会将大部分工作从“定价并生产”区域推向“仅生产”区域，从而使成本与生产脱钩。

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.

经济学家 Margaret Reid 在 1934 年提出了一项至今仍是最为犀利的诊断测试：如果一项工作可以委托给付费的第三方，那么它就是生产性的。当一个家庭雇佣管家时，家政服务会被计入 GDP。而当家庭成员从事同样的工作时，则不会。行为本身是完全相同的，会计处理方式则完全取决于是否发生了金钱交易。

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.

AI 几乎使每一项信息任务都变得可以委派。大语言模型可以起草法律摘要、分析财务报表、编写营销计划、对患者投诉进行分类评估、生成代码或撰写研究摘要。在上述每种情况下，这些工作此前都由受薪人员完成，并计入国内生产总值（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.

如果今天要求 AI 对一次医疗咨询进行记录，这项交易在国民账户中唯一能体现的地方，就是埋在 AI 公司的账单里。没有任何地方以能够正确计算通货紧缩或产出的方式报告这种用途。我们仍在使用那把陈旧的 GDP 标尺，而生产函数正将更多的经济活动推向“暗产出”这一无人地带。

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.

关于我们自己的框架从这段历史中继承了什么的坦诚说明。位移暗产出

(Displacement Dark Output) 仅衡量有偿市场劳动: BLS 工资、BLS 就业人数、O*NET 工作活动。它并不衡量 AI 对无偿护理工作、家庭生产或非正规经济的影响。我们援引 Waring 和 Ironmonger 的观点, 是为了确立生产边界是人为构建且在政治上存在争议的, 然后建立一个完全在该边界内运行的衡量体系。这是一个深思熟虑的选择, 而非疏忽。衡量市场劳动位移的数据基础设施 (BLS、O*NET、雇主调查) 是现成的且可审计的。而衡量家庭 AI 采用情况的基础设施并不存在, 发明这种基础设施会将衡量的不确定性层层叠加。但这种局限性是真实存在的。暗产出复制了一种已知的排斥。国际劳工组织记录的每日 164 亿小时无偿护理工作, 在我们的框架中与在我们所批判的框架中一样, 都是不可见的。我们并不否认这一点。我们的核心观察是, 这些替代框架所记录的问题即将恶化。我们衡量的许多位移风险也可能不成比例地落在女性就业比例较高的职业上 (根据 BLS, 行政工作 72% 为女性)。我们尚未按性别进行细分, 但这是逻辑上的延伸。

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.

非交易性生产领域中所有的 AI 使用都是“暗产出”的另一种形式。当有人利用 AI 比以前更快、更轻松或更好地完成家务劳动时, 这项活动并没有从“已生产/已定价”转变为“已生产/未定价”; 它只是扩大了“已生产/未定价”经济的规模。

Appendix 2. AI and Feminist Economics

附录 2. AI 与女性主义经济学

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.

自 1990 年 Griliches 会议以来，服务业核算有所改善，但具有针对性。美国劳工统计局（BLS）扩大了服务业生产者价格指数（PPI），到 2009 年，PPI 的服务业覆盖率已达到服务业总量的 70% 以上，且 PPI 总体体系在 2014 年转向“最终需求-中间需求”模式，以纳入服务、建筑、政府采购和出口。美国经济分析局（BEA）转向了链式费雪物量和价格指数，将分行业 GDP 与投入产出账户整合，将软件和研发（R&D）资本化，并改进了对金融、保险和研发等难测算部门的处理。

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.

但核心问题依然存在。美国经济分析局（BEA）仍表示，大多数国民收入和生产账户（NIPA）的细分组成部分是以美元而非单位来衡量的，因此实际数量通常是通过用价格指数对当前美元支出进行缩减来估算的。当交易、产品和价格指数描述的仍是同一事物时，这种方法尚可接受。但当 AI 将服务工作转化为订阅、Token 或内部生产时，这种模式就失效了。账户可以记录收据、工资和抽样价格，但未必能记录下那些实际完成的法律备忘录、文献综述、人力资源任务或代码审查。此外，它也没有质量单位，如果一份经 AI 增强的文献综述详尽程度提升了 10 倍，目前还没有任何方法能捕捉到这一事实。

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

SemiAnalysis 是一份由读者支持的出版物。如需接收新文章并支持我的工作，请考虑成为免费或付费订阅者。



Recommend SemiAnalysis to your readers

向您的读者推荐 SemiAnalysis

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

架起全球最重要的行业——半导体与商业之间的桥梁。

Recommend 推荐



184 Likes 184 次点赞 · 5 Restacks 5 次转推

← Previous 上一篇

Discussion about this post

关于此帖的讨论

Comments 评论 Restacks 转贴



Write a comment...

