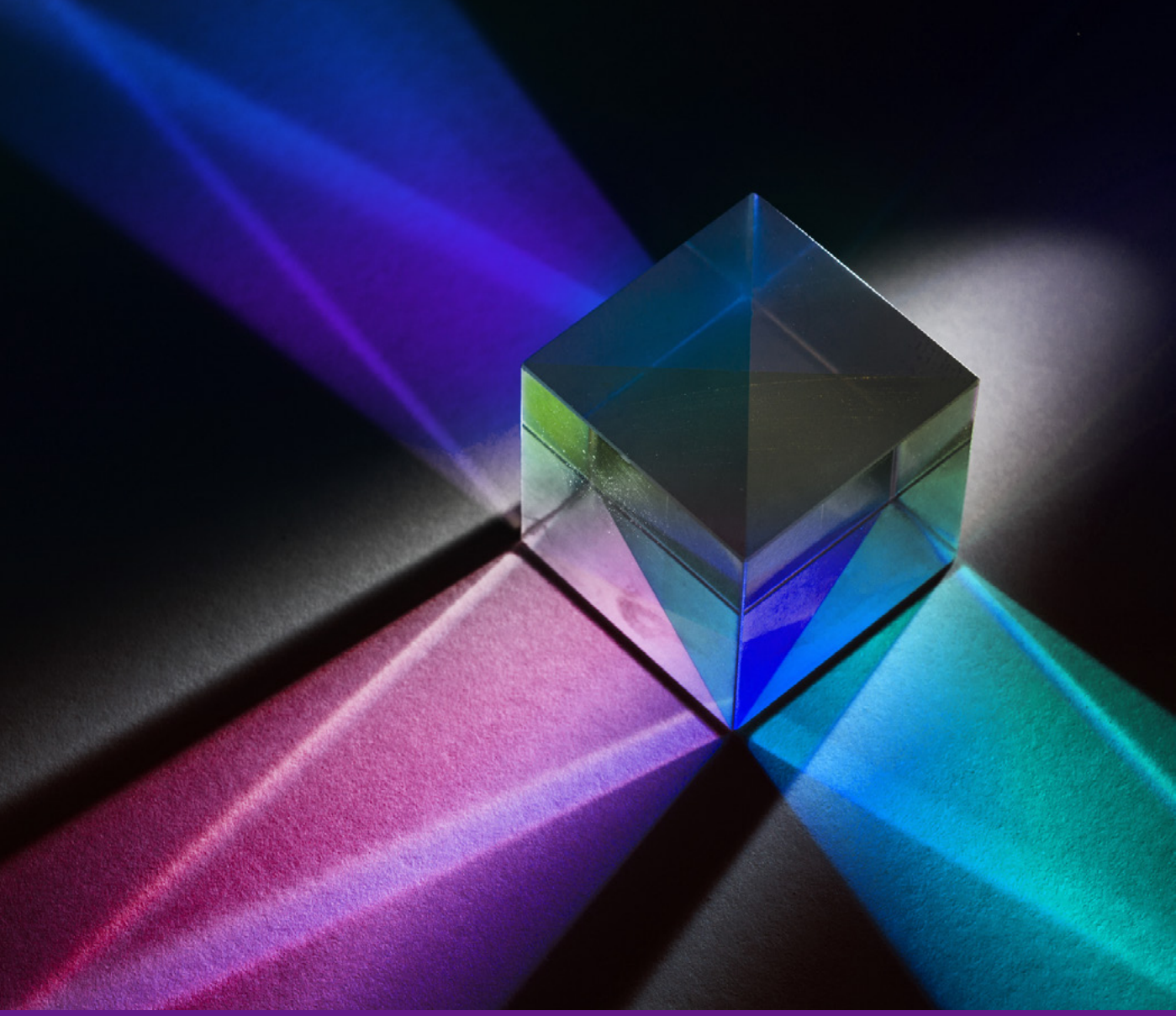


# 词元经济学：AI时代的新价值方程

文 马尼什·夏尔马、关岚、苏瑞亚·慕克吉

## 提要

随着人工智能（AI）投入不断攀升，其所带来的增长、体验优化和资金使用效率提升却往往没有被充分计算。当AI成本进入损益表，企业要看的不止成本本身，更要聚焦这些投入最终创造了什么价值。


随着生成式AI从试点逐渐走向规模化部署，模型调用、推理、基础设施以及软件服务等支出开始持续进入企业的损益表，并像其他技术投入一样受到财务部门的关注。虽然成本端日益清晰，AI所创造的价值，却难以在企业账面上被同样清晰地衡量和呈现。

这一现状也使词元经济学（Tokenomics），即围绕AI词元（Token）消耗建立起来的经济管理体系，开始成为企业管理的重要议题。其意义并不是简单地控制词元数量，而是建立一种新的投入产出关系：让企业在AI上的每笔投入，都尽可能创造更高的业务价值。

### 什么是词元经济学

词元是AI处理信息的基本单位。每一次提问、回答、信息检索和智能体交互，都会消耗词元。

过去，词元只是账单上的一个计费项目。如今，随着AI进入越来越多的业务流程，词元正在成为一种新的生产要素。和资金、人才、设备一样，它也需要被追踪、衡量和管理。

词元经济学，就是把AI资源消耗和业务产出放在一起管理的方法。它关心的不只是单次使用能不能更便宜，更重要的是判断哪些工作值得使用更高水平的AI，以及每一笔AI投入是否带来了可计算的结果。




## 一笔越来越难算的账

传统软件的成本通常相对容易预测：通过企业购买多少许可证、覆盖多少用户、签订何种合同，大体就能判断未来支出。而AI的成本结构更加动态——它不仅取决于有多少人在使用AI，还取决于用户提出什么问题、输入多少上下文、模型需要进行多复杂的推理、生成多少内容，以及工作流中调用了多少次模型。随着智能体开始代替人持续执行任务，企业面对的已不再是“有多少用户”，而是“有多少工作正在被AI持续执行”。

一个员工偶尔向模型提出简单问题，与一个持续运行的智能体工作流，其资源消耗可能存在很大差异。在企业内部，少数重度用户、复杂工作流或智能体，往往会占用绝大部分AI支出。如果企业仍然按照传统软件的许可证思维来管理AI，就很难解释这种差异，也很难找到真正应该优化的地方。

与此同时，AI成本正在经历一个看似矛盾的变化：单位词元越来越便宜，但总体AI支出未必下降。企业可能部署更多工作流、运行更多智能体、调用更多模型，最终令整体使用量大幅增加。埃森哲研究[1]显示，当词元价格下降25%，仅有15%的组织会选择保留由此节省下来的成本，相当一部分企业倾向于将这些资源重新投入新的AI应用、更大的工作负载或更高性能的模型中。

而智能体的快速发展，正在进一步放大这一趋势。智能体能够自主规划任务、调用工具、检索资料、进行多轮推理，再根据结果继续执行下一步操作。一个看似简单的工作，在后台可能已经变成由多个模型调用构成的长链条工作流。未来，当单个智能体开始与其他智能体协作时，企业需要管理的是一整套持续运行的数字劳动力。

这一趋势正在快速到来。预计到2031年，中国企业活跃智能体将达3.5亿个，较2026年增长72倍[2]。智能体数量的指数级增长，意味着企业对AI算力、数据、治理和成本管理的需求同步激增。

届时，AI成本不再只是模型调用费用，而是覆盖算力基础设施、数据治理、RAG工程化、智能体运维、工具链整合，以及安全合规等环节的长周期投入。许多企业会发现，隐性成本增长的速度甚至快于显性成本。

因此，企业不能简单地将“少用AI”作为成本控制方案，而是要思考：如何才能用对AI，在控制成本和创造价值之间找到最佳平衡。

1. 埃森哲通过模拟研究分析了随着AI逐步成为企业日常运营成本组成部分，词元消耗持续增长将对组织产生何种影响。研究结果基于对来自美洲、欧洲中东和非洲地区（EMEA）以及亚太地区（APAC）的6000名高级管理者画像开展的AI行为建模。这些管理者所在企业的年收入均超过10亿美元。该模拟研究与Aaru联合开展。Aaru结合公开、私有及专有数据，构建能够反映目标高管群体人口特征、心理特征及行为模式的合成画像，从而支持模拟分析。

2. 《【重磅】2026 IDC中国“人工智能创新奖”揭晓 | 人工智能与数据峰会圆满里程》，IDC咨询，2026年8月18日，https://mp.weixin.qq.com/s/6doMV7KwEQ9agxm3Oa20JQ。

## 成本优化：避免“高配低用”

并不是所有任务都需要最先进、最昂贵的模型。对于规则明确、步骤清晰、能够被拆解的工作，轻量级模型通常已能胜任；而对于需要复杂推理、处理大量上下文，或者一旦出错就可能产生严重后果的工作，才需要更高水平的模型能力。如果企业不加区分地为所有任务配置最高性能的模型，就会在大量低价值工作上产生不必要的支出；但如果一味地选择低成本模型，则可能在真正重要的任务上牺牲结果质量。

例如，一项常规的数据整理任务可能只需要一个轻量模型；但如果首席战略官需要综合竞争格局、宏观趋势、市场信息以及潜在收购目标的财务数据，并从相互矛盾的证据中形成战略判断，那么更强的推理能力就具有明显价值。又或者，一项智能体工作流中任何一个环节的错误都可能沿流程逐级放大，那么企业就不能只依据模型价格做决定，而应该把错误成本、业务风险以及最终结果的价值一起纳入判断。

埃森哲在实践中观察到：以一个内部平台为例，该平台每周处理约8.7万亿个词元。通过在自有基础设施上运行，并根据任务复杂度自动匹配最合适的模型，我们能够将成本控制为全部使用前沿大模型方案成本的六分之一。

不过，仅仅提供不同模型还不够。有时即使系统默认使用中等能力的模型，仍有不少员工习惯选择最强模型。一个人的选择看起来微不足道，放大到整个组织，却足以让成本悄然攀升。

对中国企业而言，模型选择还存在额外复杂性。调研显示，在评估AI智能体嵌入式产品定价时，中国企业面临“不同供应商定价模型不一致”与“无法与自身财务及采购系统整合”两方面的困扰，这一水平显著高于全球及美国[3]。这意味着，管理层在定价对比管理、系统对接协调，以及涉及多部门、多供应商的沟通、排期与推进管理等工作中，通常需要投入更多的人力与时间。

因此，模型选择不仅是技术问题，还是管理问题。企业需要建立明确规则：根据任务复杂度匹配模型能力，按照风险等级设置权限与审批机制，从组织层面引导资源向最能产生价值的场景倾斜。

## 兑现价值：实现物尽其用

成本优化并不是终点，真正的挑战在于将节省下来的资源转化为业务价值。

AI价值最容易衡量的成果，往往是效率提升。以制作一份季末总结演示文稿为例，过去需要三名财务分析师工作15个小时；使用AI后，一名员工两小时就能完成，词元成本为几百美元，同时释放出价值数千美元的工作时间。

但效率提升并不等同于价值创造，真正重要的是，释放出来的人力流向何处。只有当这些时间被投入客户服务、经营分析、产品创新等更高价值的工作，AI投入才真正转化为业务回报。



3. 《IDC调查：CIO预算制定指南，2025——企业智能投资与安全影响》，IDC，2025年9月，https://my.idc.com/getdoc.jsp?containerId=CHC53139625。

这也解释了为什么单纯观察AI账单往往不足以指导管理决策。财务负责人看到的是成本，技术团队看到的是使用量，而业务负责人看到的是效率、收入或客户体验。三者如果彼此割裂，就很难判断一笔AI支出究竟是浪费，还是对未来增长的必要投资。

因此，企业在AI时代需要建立一种新的管理语言——领先企业正在搭建一套横跨财务、技术和业务的统一看板，以一致的指标体系衡量成本、使用和产出，明确决策权、责任边界、治理机制，从而清晰识别AI价值创造的来源、成本节约的实现时间，以及如何将其重新投入。

## 词元经济学决定企业长期AI竞争力

埃森哲调研[4]显示，仅23%的高管认为，AI已在其组织范围内持续、广泛地创造业务价值。面对“为什么企业难以把AI支出与业务结果相连”这个问题，57%的受访者提到业务与技术协同不足，51%指出数据和分析能力不够，另有50%表示组织很难将AI贡献与其他影响因素区分开来。

这也说明，AI不能只由技术部门管理，企业需要建立一套能够让AI规模化，同时让价值规模化的经营体系。我们认为，有四项行动需要尽早着手：

**第一，AI经济账需要进入最高管理层的议程。**首席财务官（CFO）和首席信息官（CIO）不能再分别从财务和技术角度管理AI，而需要共同建立企业级视图，将AI支出、词元消耗、模型使用、工作流以及业务结果连接起来。AI不能只是IT部门预算中的一个新科目，而应该成为企业整体经营管理的一部分。

这意味着企业需要从“账单管理”转向“价值管理”。传统的IT成本分析通常能够告诉管理者某项服务花费了多少资金，但在AI时代还需要进一步回答：这些成本具体由哪些工作流产生？哪些团队或智能体正在消耗最多资源？这些消耗是否创造了相应的业务价值？哪些工作可以通过更合适的模型或架构降低成本，同时不影响最终结果？

**第二，企业需要在AI规模化之前就建立治理机制，而不是事后补救。**企业应当从AI应用上线之初，就明确谁可以使用哪些模型、什么类型的任务应该使用什么模型、哪些工作流需要设置预算或限制，以及安全、合规和风险控制应该如何嵌入系统。



4. 埃森哲变革脉动调研，埃森哲，2026年7月，https://www.accenture.com/us-en/insights/pulse-of-change。

其中最容易被忽略的问题恰恰是最简单的问题：为什么要使用AI？

如果一个AI工具没有明确的业务目标，没有对应的工作流程，也没有人对最终结果负责，那么即使它单次调用成本很低，规模化之后也可能变成一项没有责任归属的持续性支出。AI治理不能只关注“能不能用”，还必须关注“为什么用”“在哪里用”以及“用了以后产生了什么成果”。

**第三，建立更加精细的智能资源匹配机制。**不能因为某个模型性能最好，就默认将其应用于所有场景。企业可以把工作按照复杂度、风险、业务价值和错误成本进行分类，再决定采用什么模型。随着模型能力和价格持续变化，这套匹配机制也应该不断调整。

这意味着企业需要把AI管理从一次性的采购决策变成持续的运营。今天最具性价比的模型，可能几个月后就不再是最佳选择；一个原本只能由前沿模型完成的任务，也可能随着模型能力提升而转向成本更低的模型。因此，企业需要持续比较不同模型的成本与性能，并根据实际业务结果不断调整。

**第四，把词元经济学变成持续运营管理能力，而不是一次性的成本削减项目。**企业可以建立持续的“诊断—优化—监测”循环：首先弄清楚AI支出集中在哪里；其次找到造成不必要成本的模型选择、提示词、上下文、工作流或架构问题；最后实施优化，并持续观察优化是否真正带来了业务价值。埃森哲提出的词元经济学框架特别强调这一点，即从“看见”问题，到“修复”问题，再到“持续保持”优化成果，最终沉淀为组织能力。

这种持续管理尤其重要，因为AI经济账不会保持静态。模型不断升级，词元价格持续变化，智能体不断承担新的工作，企业的AI使用模式也会随之改变。一次性的成本优化很快就可能失效，真正能够产生长期效果的，是把经济账管理嵌入企业日常运营。

AI时代的企业竞争或将逐渐分化为两种截然不同的路径。一类企业将AI视为新的技术成本中心，不断应对日益复杂且难以预测的支出；另一类企业则将AI视为新的生产要素，通过持续优化模型、工作流和资源配置，让AI创造价值的速度持续超越成本增长速度。两者的差异，并不在于谁拥有更多模型或是更早部署AI。真正的差异在于，谁能够回答一个越来越重要的问题：每一单位AI智能究竟创造了多少业务价值。

如果企业能够把这个问题真正纳入经营体系，那么词元就不再只是模型服务商账单上的计费单位，而会成为衡量企业AI生产效率的重要指标。企业也无需通过简单地限制AI使用来控制成本，而是可以更加有意识地扩大那些真正能够创造价值的使用场景，同时减少低价值、低效率的消耗。

**马尼什·夏尔马（Manish Sharma）**
埃森哲首席战略与服务官

**关岚（Lan Guan）**
埃森哲首席AI与数据官

**苏瑞亚·慕克吉（Surya Mukherjee）**
埃森哲商业研究院高级研究总监

于雅对本文亦有贡献。

业务垂询：contact-us@accenture.cn