九章格物 NineGrids

GENERATIVE ENGINE OPTIMIZATION · INDUSTRY PANORAMA 2026

星阙实验室GEO研究中心 · 第 10 册

2026 年生成式引擎优化行业全景与趋势研判

从排名位置到答案引用。需求、供给、市场、玩家、渗透、红利、度量与治理的完整拆解，以及 2026 至 2028 的分年推演与十条可证伪判断。

出品：九章格物 · 星阙实验室

作者：李现龙

版本：V2.0

时间：2026 年 11 月

权威信源：基于 CNNIC 统计报告、国家互联网信息办公室规章与通报、新华网与经济日报报道、arXiv 已发表研究及火山引擎官方文档编制

微信扫我加好友

扫码添加微信 · 获取 GEO 方案

© 2026 九章格物 · 星阙实验室 · ninegrids.cn

目录

核心摘要 3

01 全景　定义、边界与三次入口迁移 3

02 需求　谁在问、问什么、问到哪一步 5

03 供给　一次回答的机制解剖 7

04 市场　三个口径与四块细分 11

05 玩家　能力矩阵与单位经济 13

06 渗透　七行业的条件与失败模式 15

07 红利　六条优势的机制与算账 17

08 度量　从被提到到可结算 20

09 治理　标识、投毒与合规动作 22

10 趋势　2026 至 2028 的分年推演 23

11 结语　三类主体的行动清单 27

术语、复算与研究分级 28

编制与修订记录 29

参考来源 29

00 核心摘要

生成式引擎优化在 2026 年完成了从概念到生意的转身：需求侧的入口已经迁移，供给侧的平台已经闭环，市场侧的规模已经可测，制度侧的规则已经强制。这四件事同时发生，才是这门生意值得认真做的理由。

第一，入口已经换位。截至 2025 年 12 月，国内生成式人工智能用户 6.02 亿，普及率 42.8%，一年净增 3.53 亿、增幅 141.7%，占网民规模的 53.5%（本白皮书复算）。[4]用户不是在“多用一个工具”，而是在把决策前的信息获取整体挪到问答框里。

第二，市场已经成型。经济日报援引信通院口径：2025 年国内生成式引擎优化市场 42 亿元，年复合增速 38%，超过 68% 的大中型企业已把它列进营销预算；同文援引 Gartner 口径的全球支出从 2024 年底的 12 亿美元升至 2026 年预计的 35 亿美元，年复合增速超过 140%。[5]两组数字均为该篇报道的转述，本白皮书不另作折算；两个口径不可相加，只能并列阅读。

第三，优势是结构性的。答案通常只给一个结论，谁被引用谁独占；疑问句的引用率远高于其他句式；跨引擎的引用结果高度一致，一套内容可以多引擎复用；引用份额可以按句位加权精确计算。这四条决定了它的投入产出比优于点击计费型投放。

第四，动作有分层。45 项研究的横向比较给出的排序很清楚：语义相关性与上下文位置支持度最强，可抽取证据次之，时效中等，文档结构方向不一，流畅度与简化偏弱，权威语气与关键词堆砌排在最后。[2]把钱花在最后一档，是最常见的浪费。

第五，度量必须先于交付。没有指标口径就没有验收，没有抽样设计就没有结论。第八章把可见性、过程与资产三层指标、最小样本量与报表字段一次写清，这是这个行业最缺的一块，也是甲方最容易吃亏的一块。

第六，窗口不长，这是判断而非事实。市场处于成型早期，服务商四层尚未固化，可信评测刚刚起步（首批 9 家通过专项评测），标识制度自 2025 年 9 月 1 日施行[6][7]。制度一旦收紧，先入场者拿的是资质与语料位置，后来者只能拼执行。

表 0-1 一页看懂这门生意的六个数

指标	数值	时点	含义
生成式 AI 用户	6.02 亿	2025 年 12 月	入口规模的基数[4]
普及率	42.8%	2025 年 12 月	已跨过早期采用者阶段[4]
生成式用户占网民	53.5%	2025 年 12 月	本白皮书复算的派生值
国内市场支出	42 亿元	2025 年	信通院口径，经济日报援引[5]
市场年复合增速	38%	2025 至 2028	成型期而非饱和期[5]
已列预算的企业占比	68%	2026 年	大中型企业口径[5]

说明：前六行均为官方或机构公开口径；派生值由公开数值复算，不与机构口径合并成同一趋势判断。

01 全景　定义、边界与三次入口迁移

理解生成式引擎优化，先要理解它替换的是什么。过去二十年，内容方争夺的是搜索结果里的一个位置；现在争夺的是一段答案里的一句引用。位置可以并列十条，答案通常只给一个结论。这个变化没有提高门槛，它改变了分配方式。

一个定义 三个限定

生成式引擎优化指通过内容与信源的系统性建设，提高品牌或机构被生成式引擎在回答中采用并署名的相对份额。定义里有三个限定词：**系统性**意味着不是改一次标题，而是持续的内容与信源工程；**采用并署名**意味着目标不是被收录，而是被写进答案且能追溯；**相对份额**意味着它天然是与竞品比较的指标，而不是绝对曝光。

这三个限定词决定了它不能按投放逻辑管理。投放买的是展示位，可以按千次曝光结算；引用买的是被选中，只能按份额、位置与可溯源性验收。把它当作“更便宜的曝光”来做，投入一定会失焦。

和三个相邻概念的边界

行业里最常见的混淆，是把生成式引擎优化与搜索引擎优化、应用商店优化、品牌公关当成同一件事。它们的争夺对象、成功形态与计价方式完全不同，混为一谈会导致预算科目与验收标准一起错位。

表 1-1 四个相邻概念的对照

概念	争夺对象	成功形态	计价方式	可否度量
生成式引擎优化	答案中的一句引用	被采用、可溯源、份额提升	按项目或订阅	可按份额复算
搜索引擎优化	结果页位次	排名上升、点击增加	人力与内容成本	排名可查
应用商店优化	商店内搜索位	下载与转化提升	人力与素材成本	榜单可查
品牌公关	媒体口径与声誉	报道数量与倾向	按篇或按年框	需人工评估

说明：四者的边界在于“争夺对象”而非“交付形态”；同一份内容可以同时服务多个目标，但指标不能互相替代。

三次入口迁移与规则变化

信息入口的迁移在二十多年里发生过三次，每次都在重排注意力的分配规则。第一次把内容搬上网页，规则是目录与链接；第二次用关键词把网页排序，规则是相关性与权重；第三次把答案直接写在提问下面，规则变成语料与可信度。2023 年之后的生成式引擎，是第三次迁移的落地形态。

图 1-1 三次入口迁移与规则变化



说明：阶段划分按公开产品形态归纳，时间边界为示意，不代表单一产品的上线时间。

三次迁移的共同点是：每次迁移都让“内容做得更多”失效，让“内容做得对”生效。第一次拼的是收录量，第二次拼的是关键词覆盖，第三次拼的是能不能被抽出来、被信任。这也解释了为什么在 2026 年，内容数量已经不再是壁垒——语料池的规模由平台决定，不由内容方决定。

四类引擎 四种引用行为

“生成式引擎”不是一类产品。按检索与生成的关系，可以分成四类，它们的引用行为差异很大，优化抓手也不同。把四类当成一类做，是许多方案落空的原因。

表 1-2 四类引擎的特征与优化抓手

类型	检索与生成关系	引用行为	优化抓手
检索增强对话	先检索再生成，上下文里放候选片段	常带逐句或段落引用	片段可抽取性与信源权威
搜索增强摘要	在原搜索结果上加一段生成摘要	多标站点，少标具体句	站点整体权威与页面结构
任务型代理	多轮检索并调用工具完成动作	引用不稳定，偏重结构化数据	接口、结构化字段与时效
端侧与私有库	只在本地或企业语料内检索	几乎不对外署名	语料入库与权限配置

说明：四类按公开产品行为归纳，同一平台可能同时具备多种形态；引用行为的判定以实际输出为准，不以产品宣传为准。

全球一条线 国内一条线

这个行业有两条并行的技术曲线。海外起点是 2023 年的学术研究，把“如何让内容被生成式引擎选中”变成可测量的问题，随后在 2024 至 2025 年进入产品化与工具化阶段[1]。国内的顺序不同：先是平台自建检索与内容生态，2025 年再把标识与评测做成制度，学术与方法论的沉淀反而发生在产业之后[6][7]。

顺序不同，导致两个市场的竞争形态不同。海外更强调工具、监测与可复现实验，服务方以 SaaS 与咨询混合为主；国内更强调平台适配、内容供给与合规资质，服务方以项目制与年框为主。读者若把海外的工具清单直接搬回国内，会遇到两个问题：平台接口不同，引用口径也不同。

2026 年成为分水岭的三件事

第一件是制度落地。《人工智能生成合成内容标识办法》自 2025 年 9 月 1 日施行，配套强制性国家标准同步执行，未落实标识的应用在 2025 年 11 月被集中查处通报。[7]标识从“倡议”变成“入场条件”，行业的合规成本第一次被写进价格。

第二件是规模可测。2025 年国内支出 42 亿元、增速 38%[5]，意味着这个市场已经大到可以被单列预算，而不是挂在“内容营销”科目下报销。预算科目独立，管理动作才会独立。

第三件是评测出现。首批 9 家机构通过可信专项评测[6]，行业第一次有了“谁可信”的可核定口径。评测不解决效果问题，但解决采购时的资格问题，这对甲方比任何方法论都实用。

02 需求　谁在问、问什么、问到哪一步

需求侧的判断不能只看用户规模。六亿人用不用，和他们怎么问、问到哪一步、信不信答案，是三件不同的事，对内容的含义也完全不同。这一章把需求拆成三层：人、问句、阶段。

规模已经跨过临界点

截至 2025 年 12 月，国内生成式人工智能用户 6.02 亿，普及率 42.8%，一年净增 3.53 亿。[4]按 11.25 亿网民基数复算，占网民 53.5%。这三个数放在一起看，结论是：它已经从“早期尝鲜”进入“主流人群的日常工具”。

表 2-1 需求侧的五个关键数值

指标	数值	时点	对内容的含义
生成式 AI 用户	6.02 亿	2025 年 12 月	覆盖主流消费与从业人群[4]
普及率	42.8%	2025 年 12 月	已跨过 40% 的采用拐点[4]
一年净增	3.53 亿	2025 年	增量来自非技术人群[4]
同比增幅	141.7%	2025 年	入口扩张速度接近翻倍[4]
占网民比例	53.5%	2025 年 12 月	本白皮书复算，6.02 ÷ 11.25

说明：前四行为官方公开数值，第五行为复算派生值；复算保留一位小数，不进入任何预测模型。

不是多用一个工具 而是换了一个入口

判断“换入口”而不是“加工具”，看三件事：增量来自谁、问的是什么、问完还去不去别处。增量主要来自非技术人群，说明它已进入生活决策；问句从“查资料”转向“做判断”，说明它承担了比较与筛选；问完不再点开十个链接，说明注意力已经被截留在答案层。

三件事里最关键的是第三件。过去搜索的商业模式建立在“导流”上，内容方靠点击获益；答案层把导流环节去掉了，内容方只能靠“被引用”获益。这就是为什么这门生意必须谈份额，而不能谈流量。

四类提问者 四种问法

同一个品类，四类人的问法完全不同，能匹配的内容形态也不同。把四类人的问题写成一种语气，等于放弃其中三类。

表 2-2 四类提问者的问法与内容含义

提问者	典型问法	真正想要的	内容含义
初次了解者	“这是什么、和那个有什么区别”	一个能听懂的定义与边界	定义段要短、要能独立成段被抽出
比较决策者	“A 和 B 哪个更适合我这种情况”	条件化的对比与适用边界	逐项对比表加一句适用结论
价格敏感者	“大概多少钱、有没有更便宜的”	价格区间与影响因素	区间要给口径，不给单点数字
风险规避者	“会不会有问题、出问题谁负责”	风险、资质与责任边界	风险提示与资质信息必须同段出现

说明：四类为归纳分类，同一用户在不同阶段会切换角色；分类用于组织内容，不用于人群画像。

问到哪一步 决定写什么

决策链有四段：认知、比较、验证、成交。每一段对应不同的问句形态与不同的引用敏感度。答案层在比较与验证两段的介入最深，因为这两段最需要第三方口径。

表 2-3 决策链四阶段的问句与内容要求

阶段	问句形态	引用敏感度	必须提供的内容
认知	是什么、有没有用、怎么开始	中	定义、适用边界、最常见的三个误解
比较	选哪个、差在哪、哪个更划算	高	逐项对比、条件化结论、不适用的情形
验证	真的吗、谁说的、有没有案例	高	出处、时间、样本、可复核的细节
成交	多少钱、多久、怎么退	中	价格区间、交付周期、退出与责任条款

说明：引用敏感度为归纳判断，依据是问句类型与答案构成的对应关系，非统计值。

图 2-1 决策链四阶段的推进与流失



说明：宽度仅示意各阶段可被提问的相对规模，非统计比例；阶段之间可回退，不是单向流失。

四类场景 对不同句式的偏好

按问题性质分四类场景：信息型（要事实）、比较型（要差异）、应急型（要立刻能做的动作）、交易型（要价格与流程）。信息型吃定义与数据，比较型吃表格，应急型吃短句，交易型吃区间与条款。

这四类场景解释了前文的一个现象：为什么长文不一定要写长，但要写得“能拆”。同一篇内容里四类场景并存时，要用小节把它们分开，让引擎可以只抽走其中一段，而不是把整篇压缩成一句模糊的概括。

三处空位 现在还在

第一处是新兴品类与细分工艺。问题已经被提出，但没有可靠来源，引擎只能在零散信息上拼答案。第二处是地方性服务与区域差异，通用内容覆盖不到具体区域的价格、时效与资质。第三处是长尾比较类问题，用户在两方案之间摇摆时提出的问题，往往最接近成交，但公开内容里缺少逐项对比与适用条件。

三处空位有一个共同点：都不靠文采，靠事实的组织方式。这也解释了为什么第七章所说的优势是结构性的，而不是投放技巧的延伸。空位不会永远存在，但它关闭的速度取决于竞争者何时到场，而不是取决于平台政策。

需求侧需要防的三个偏差

第一是问句漂移。同一品类的问题表述会随时间变形，去年有效的问法今年可能不再被触发。第二是口径变化，平台对引用与署名的展示规则并不稳定，同一份内容在不同版本的界面上表现不同。第三是样本偏差，用少量高频问题代替全部提问，会系统性高估头部问题的价值。

三个偏差都指向同一个动作：需求侧要有自己的问题集，并且按季度维护。没有问题集，所有“需求旺盛”的判断都是别人的口径。

03 供给　一次回答的机制解剖

需求侧决定了有多少问题被提出，供给侧决定了谁能被写进答案。这一章拆的是后一半：一次回答在几百毫秒里发生了什么，内容在哪一层被筛掉，什么样的写法能穿过全部筛选。理解这套机制，才能判断哪些投入是有效的，哪些只是自我安慰。

一次回答要走的五步

生成式引擎不是“把网页读一遍再改写”。它更接近一条流水线：先把问题拆成检索意图，再从索引里召回候选，接着按权威度与相关性重排，然后把选中的片段交给模型生成答案，最后给答案挂上来源。五步之中，任何一步失手，内容都不会出现在答案里。

图 3-1 一次生成式回答的五步流水线

1. **拆解意图** 把问句拆成实体、条件与时效要求，决定要检索什么。
2. **召回候选** 从索引按语义与关键词并行取回，形成候选集。
3. **重排** 按权威度、相关性与时效重排，决定谁进入上下文。
4. **生成** 抽取片段、改写、合并，形成一段带结论的答案。
5. **署名** 标注来源，部分场景只标站点或完全不标。

说明：五步为按公开文档与已发表研究归纳的通用流程，不同引擎的模块划分与顺序略有差异，不代表任一平台的实际架构。

五步里的五个可干预点

把五步拆开看，会发现内容方能干预的只有特定的几处，别处投入再多也不会改变结果。这是“无效动作”产生的根本原因，不是执行不力。

表 3-1 五步中的可干预点与不可干预点

环节	内容方能干预的	不能干预的
拆解意图	用词与问句的贴合度，决定被抓到的概率	意图分类模型、同义扩展规则
召回候选	可抓取性、站点结构、站点整体权重	索引规模、召回阈值
重排	信源权威、事实一致性、时效标注	重排模型的权重与训练数据
生成	片段的自足性、结论前置、单位与条件	生成模型的改写偏好
署名	站点可识别性、作者与资质可核	平台的署名展示策略

说明：可干预点的判定依据是公开文档对检索与重排环节的描述[9][10]，不涉及平台未公开的参数。

引擎看的五类信号

在召回与重排阶段，引擎实际依赖的不是“文笔”，而是可计算的信号。2026 年的十引擎受控检索实验给出了一个关键结论：不同引擎对同一批内容的取舍高度一致，原因是它们看重的是同一组信号。[3]这既解释了为什么优化动作可以跨引擎复用，也解释了为什么“只讨好一个引擎”的做法回报有限。

表 3-2 引擎侧五类可观察信号

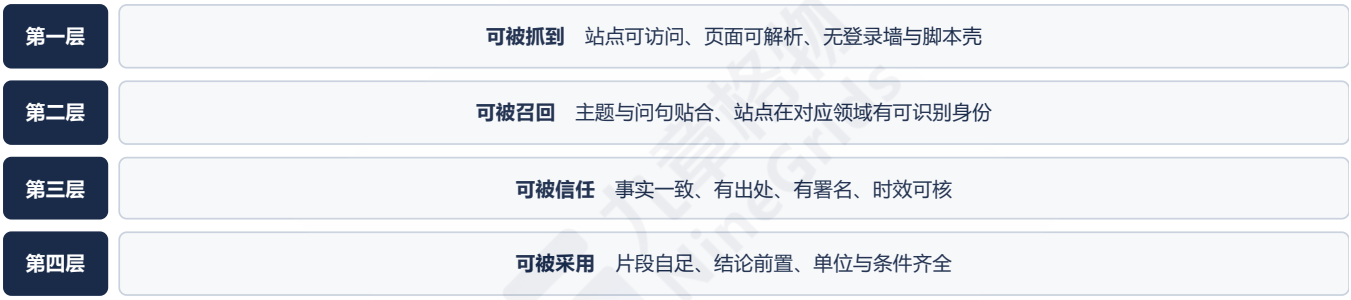
#	引擎侧信号	对内容供给的含义	可否外部观察
1	**来源站点与权威分级**	站点身份先于单篇质量被评估，机构站与垂类权威站占优	部分可观察
2	**相关度评分**	片段与问句的语义贴合度决定能否进入上下文	不可直接观察
3	**正文可解析格式**	纯文本与 markdown 更易被抽取，脚本壳页面易被截断	可观察
4	**时效与地域范围**	时间敏感问题优先新鲜内容，跨地域问法结果不同	可观察
5	**引用来源标记**	被引用可被记录，成为可回溯的效果证据	可观察

说明：第 1、2 项为平台内部评分，外部只能通过结果差异反推；第 3 至 5 项可在接口返回内容中直接确认[9][10]。

内容进入答案的四层结构

从“被收录”到“被署名”，中间有四道门。四道门是串联的，任何一道关闭，后面的努力都归零。这也是为什么很多方案在“收录良好”的情况下依然没有引用。

图 3-2 内容进入答案的四层结构



说明：四层的顺序为串联关系，非并列评分；某层失守通常表现为“收录正常但零引用”。

表 3-3 四层失守的表现与修复方向

失守层	典型表现	可能原因	修复方向
第一层	站点从未出现在候选集	抓取被阻、页面结构断裂	先修可访问性
第二层	收录有、相关问句无候选	主题分散、站点身份不清	收窄主题与栏目
第三层	进入候选但被重排挤掉	事实冲突、无出处、时效过期	补出处与复核时间
第四层	被引用但只截半句或截错	段落不自足、结论埋在中段	改首句与拆段

说明：四类表现来自项目复盘的归纳，不是统计分布；同一页面可能同时失守多层。

四种引用形态 决定优化重点

“被引用”不是一个统一的形态。按粒度可以分成四种：逐句引用、段落引用、站点署名、完全不署名。四种形态对内容的要求不同，用一套写法通吃，通常只能拿到最弱的一种。

表 3-4 四种引用形态与对应的写法

形态	答案里的表现	对内容的要求	可控程度
逐句引用	答案中出现具体句子并挂来源	单句自足、含数字与条件	最高，句子可设计
段落引用	整段被改写后挂来源	段落首句即结论、段内无跳跃	较高，段落可组织
站点署名	只标站点名，不标具体内容	站点身份清晰、栏目稳定	中等，靠长期积累
完全不署名	答案正确但不注明来源	无法直接优化	低，只能靠间接占位

说明：形态分布随平台与问题类型变化；本白皮书把“零引用份额”作为完全不署名情况的代理指标，见第七章。

平台接口差异 决定能做到什么

不同平台的开放程度差别很大，直接决定优化能做到多细。开放检索接口的平台，能把“标题、站点、摘要、正文、权威度、时效”这些字段拿回来做对照；不开放的平台，只能靠结果观察反推。

表 3-5 三类平台的开放程度与可用能力

平台类型	可获得的字段	能做的分析	限制
开放检索接口	标题、站点、URL、摘要、正文、权威度[9][10]	候选集分析、引用归因、份额计算	单次返回条数有上限
摘要型搜索	站点与摘要，少标具体句	站点层份额、主题覆盖	无法做到句子层归因
封闭问答	仅答案文本与少量来源标记	结果观察与抽样比对	口径不稳定，需多次采样

说明：字段能力以 2026 年 10 月的官方文档为准[9][10]；平台调整接口时，本表需要重新核对。

能被抽走的七个格式要素

内容是否被采用，很大程度上取决于“能不能被干净地拿起来”。以下七个要素不是写作风格问题，而是抽取友好度问题，缺一项就会在某一层被削弱。

表 3-6 抽取友好度的七个要素

#	要素	具体做法	缺失后果
1	**首句即结论**	每段第一句给出可独立成立的判断	只能被整段改写，无法逐句引用
2	**数字带口径**	数值紧跟时间、范围与统计对象	数字被弃用或误用
3	**条件写全**	结论后紧跟适用与不适用条件	被当作绝对断言，可信度下降
4	**段落自足**	一段只回答一个问题，不依赖上下文	抽取后语义断裂
5	**对比成表**	选项、差异、适用场景三列齐全	比较类问题无内容可引
6	**出处可核**	官方文件名、发布机构与时间齐全	被重排阶段判为低可信
7	**纯文本正文**	正文不依赖脚本渲染，图片信息另附文字	内容不可解析，直接出局

说明：七个要素按抽取流程的顺序排列，与图 3-2 的四层门对应；它们是必要条件，不是充分条件。

为什么改写常常无效

最常见的做法是把已有文章“改得更通顺、更简短、更专业”，然后期待引用上升。这个动作落在第七个要素之外的维度上，属于阅读体验，而抽取阶段看的是结构、事实与来源。改动不触及评价维度，自然看不到变化。

真正改变结果的动作通常很朴素：把结论挪到首句，把数字补上时间和范围，把“因人而异”换成具体的适用条件，把散落的对比整理成表。这四件事做完，同一篇内容的可抽取性会明显不同，而这四件事都不需要额外预算。

04 市场　三个口径与四块细分

一个行业能不能成为独立赛道，看三件事：规模是否可被单列，产业链是否出现分工，外部资本是否给出定价。2025 至 2026 年，这三件事同时给出了正面答案。这一章把数字与结构放在一起，避免只报规模不讲来源。

三个口径 不能相加

不同机构的口径差异很大，混用会得出错误结论。国内口径统计的是服务与内容支出，全球口径统计的是更宽的技术与营销支出，资本市场口径反映的是预期而非收入。三者不能相加，只能并列阅读。

表 4-1 规模的三组公开口径

口径	公开数值	统计范围	读法
国内服务与内容	2025 年 42 亿元，CAGR 38%	国内 GEO 服务、内容与工具支出[5]	看本土市场真实付费
全球技术与营销	2024 年底 12 亿美元，2026 年预计 35 亿美元，CAGR 超 140%	全球范围内的技术、工具与营销支出[5]	看增速与方向，不折算国内
资本市场预期	A 股相关板块涨幅超 50%，部分个股翻番	市场对赛道前景的定价[5]	看情绪与资金，不看收入

说明：三个口径的统计对象不同，不得相加或互相折算；汇率变动会进一步放大口径差。

三个口径里最有信息量的是增速差。国内 38% 说明本土还在“铺开阶段”，全球 140% 说明技术与工具层正在被资本快速推高。前者对应服务机会，后者对应工具机会。读者若要做的是内容与咨询，应盯国内口径；若要做的是产品与平台，全球口径更相关。

四块细分 四种生意

把 42 亿元拆开，会看到四种性质完全不同的生意：内容生产、监测工具、策略咨询、培训与评测。它们的客单、毛利与竞争密度差异很大，把四块当成一块谈，会让定价与交付一起失真。

表 4-2 四块细分的生意特征

细分	交付形态	客单与周期	毛利形态	竞争密度
内容生产	按篇或按品类交付	单篇至年框，按月结算	人力密集，毛利随规模下降	最密，价格透明
监测工具	账号制或接口调用	年费制，按查询量分档	前期投入高，后期边际成本低	较稀，技术门槛在前
策略咨询	诊断、方案、陪跑	项目制或季度制	依赖人，毛利随资历上升	中等，看案例
培训与评测	课程、认证、专项评测	按人头或按机构	标准化程度高，毛利稳定	最稀，资质门槛

说明：客单与毛利形态为本白皮书基于公开报价与项目实践的归纳判断，非统计值；竞争密度按 2026 年公开可见的供给数量估算。

图 4-1 产业链的三层与六个主体



说明：分层按“谁决定入口、谁交付服务、谁提供验证”划分，示意结构；同一机构可能同时出现在两层。

四块细分的关系是串联而不是并列：咨询定方向，内容做交付，工具做验证，培训与评测解决组织能力与采购资格。一个成熟的买方通常会同时采购其中两块以上，但很少在同一家买齐四块。这解释了为什么“全包型”服务商在市场上并不占优。

预算从哪里来

预算科目的迁移，是行业成型的另一个证据。钱不会凭空出现，它一定是从别的科目里挪过来的。从 2025 至 2026 年的项目实践看，主要的迁移来自三处：品牌内容生产、搜索与信息流投放、口碑与公关，另有一小部分从市场监测科目新增。

表 4-3 预算科目的迁移路径

来源科目	迁移强度	迁移理由
品牌内容生产	主来源	同属内容支出，只是换了分发目标，审批阻力最小
搜索与信息流投放	次要	投放停付即归零，预算方开始寻求可累积的替代
口碑与公关	次要	答案层的口径直接影响声誉，可并入声誉管理
市场监测	新增小额	原科目无对应能力，通常单列小额试点预算

说明：本表为项目实践的观察归纳，非行业统计；迁移强度指出现频次，不代表金额占比。

迁移路径决定了谈判方式。从内容科目挪钱，谈的是“同样的钱换个写法”；从投放科目挪钱，必须先证明“可累积”，否则预算方会拿 CPA 来对比；从公关科目挪钱，考核的是口径与风险，而不是份额。

采购怎么走 五步与五问

大中型企业的采购流程通常有五步：需求提出、方案比选、试点验证、合同签订、季度复盘。五步里最容易卡住的是第三步，因为试点缺少可验收的口径。这也是第八章要单独讲度量的原因。

表 4-4 采购五步的关键决策与卡点

环节	谁决定	看什么	常见卡点
需求提出	市场或品牌负责人	竞品是否已被引用	说不清与 SEO 的区别，预算立不住
方案比选	采购与业务双方	案例、口径、交付边界	报价口径不一，无法横比
试点验证	业务负责人	三个月内的可观察变化	无基线数据，无法判定效果
合同签订	法务与采购	验收条款与退出条件	承诺写得太满，验收无法执行
季度复盘	预算方	份额、零引用率、成本	报表口径与合同口径不一致

说明：流程为一般性归纳，不同企业的审批层级与周期差异较大；试点期通常为 2 至 4 个月。

资本信号说明了什么

2026 年 A 股相关板块的涨幅与部分个股的翻番[5]，反映的是预期而不是收入。预期的来源有两处：一是入口迁移带来的替代想象，二是平台开放检索接口带来的可测量性。两处都是真实的，但都不等于当下的现金流。

对从业者的实际含义是：资本热度会带来两类竞争者，一类是跨界的技术公司，一类是被热度吸引的投机型服务商。前者会抬高工具门槛，后者会压低内容价格。夹在中间的中小服务商，最需要的是明确的交付边界与可验收的口径，而不是更大的承诺。

市场还有三个结构性约束

第一，付费方分散。预算分散在市场、品牌、电商、公关多个部门，单个决策者能批的额度有限。第二，效果滞后。内容进入语料后需要时间被检索与重排接受，通常以月为单位体现，与季度考核节奏错位。第三，口径不一。同一份内容在不同平台的引用表现不同，缺少跨平台的统一口径。

三个约束都不会因为技术进步而消失，只会因为制度与工具成熟而缓解。判断一个市场是否成熟，看的不是规模数字，而是这三个约束是否被标准化产品解决。

05 玩家　能力矩阵与单位经济

行业里最容易被误读的是“谁在做 GEO”。真实情况是四层完全不同的供给同时存在，能力、客单与交付周期差异极大，把它们混为一谈，报价就会失真，验收也无法对齐。这一章先分层，再算账，最后给选商方法。

供给的四个层

第一层是平台与语料方，不直接接单，但决定规则；第二层是策略与内容服务商，做诊断、内容改造与信源建设；第三层是工具与监测方，卖的是份额与归因能力；第四层是个人与小团队，做单点执行。四层的客单价相差一个数量级，交付物也完全不同。

表 5-1 四层供给的能力与客单

层级	核心能力	客单区间	交付周期
平台与语料方	入口、规则、语料准入	不直接接单	规则按季度调整
策略与内容服务商	诊断、内容改造、信源建设、报表	年框 30 万至 200 万元	3 至 6 个月见效
工具与监测方	份额监测、竞品对比、归因	年费 3 万至 30 万元	开通即用，按月出数
个人与小团队	单点改写、批量发布	单篇 800 至 8,000 元	1 至 4 周

说明：客单区间为本白皮书基于公开报价与项目实践的归纳，非统计均值；不同行业与合规要求会造成显著差异。

能力矩阵 谁真正具备什么

把四层供给与五项能力交叉，会看到明显的错位：能算的不一定能写，能写的不一定能合规，能合规的不一定能规模化。买方最常见的错误，是按“报价最低”选层，然后要求对方具备全部五项能力。

表 5-2 四层供给的能力矩阵

层级	诊断	内容	技术	合规	度量
策略与内容服务商	●	●	○	●	○
工具与监测方	○	○	●	○	●
个人与小团队	○	●	○	○	○
自建团队	●	●	●	●	●

说明：●为通常具备，○为通常不具备或需外采；自建团队为理想配置，实际能力取决于编制与投入，不代表普遍水平。

这张矩阵最有用的一行是最后一行。它说明自建在理论上可以覆盖全部能力，但代价是编制与时间；而四层供给的组合，实际上是把自建拆成可采购的模块。所以“自建还是外包”不是二选一，而是“哪几个模块值得自己养”。

三档价格 对应三种交付

价格差异不是因为“谁更贵”，而是因为交付边界不同。低价档交付的是动作，中价档交付的是结果，高价档交付的是能力转移。买方若用低价档的预算要求高价档的交付，谈判一定破裂，验收一定扯皮。

表 5-3 三档价格的交付边界

档位	交付物	验收方式	适用对象
入门档	问题集、内容改造清单、结构化配置	按交付物验收，不承诺份额	试水型预算、单品类
标准档	季度份额目标、监测报表、迭代计划	按引用份额与零引用率验收	有明确品类与竞品的中型企业
共建档	团队培训、内容中台、度量体系	按能力转移与自运营程度验收	多事业部、强合规行业

说明：档位划分为便于采购决策的归纳，市场上存在混合形态；签约时应把交付物、验收口径与退出条件写清。

单位经济 三种生意的账

把四层供给的账摊开算，会看到一个反直觉的结论：人力密集的内容生意，毛利并不低于工具生意，但现金流的稳定性差得多。原因在交付节奏：内容按月结算，工具按年预收。

表 5-4 三种供给形态的单位经济

形态	成本结构	现金流特征	规模化的瓶颈
内容型	人力为主，边际成本随交付量上升	按月结算，回款周期长	人效，质量随规模下滑
工具型	前期研发重，后期边际成本低	年费预收，现金流好	数据覆盖与平台接口变动
咨询型	资深人力为主，人力成本高	项目制，节点收款	可复制的案例与交付模板

说明：本表为结构归纳，不含具体毛利率数值；实际毛利取决于团队构成、地域成本与交付深度。

这三种形态对从业者的含义不同。内容型要解决的是“如何在扩张时保持质量”，工具型要解决的是“平台接口一变如何自保”，咨询型要解决的是“案例能否脱离个人”。三者共同的解法都是同一件事：把写法固化成模板，把模板固化成验收标准。

自建 外包 还是混合

判断标准不是预算多少，而是三件事：品类是否核心、更新是否高频、数据是否敏感。三件事都成立，自建；都不成立，外包；只有一两件成立，混合。混合的典型形态是“外包内容、自建度量”。

表 5-5 三种路径的适配条件

路径	适配条件	风险
自建	核心品类、日更、数据敏感	编制与专业度不足，交付易空转
外包	非核心品类、低频、无敏感数据	能力不外溢，续约依赖同一家
混合	核心内容自建，度量与监测外采	口径需统一，否则双方各说各话

说明：条件判定以业务属性为准，不以预算规模为准；混合路径的口径统一是成败关键。

选供应商要问五个问题

第一，问题集是否由你提供，还是由供应商自定。自定问题集等于自定成绩。第二，基线数据是否在合作前采集。没有基线，后续任何变化都无法归因。第三，份额口径是否可复算，公式与抽样方法能否写进合同。第四，内容交付是否包含署名与资质信息，还是只交文本。第五，退出时语料、模板与报表是否留给甲方。

五个问题里，最容易被忽略的是第五个。它决定了合作结束后资产是否还在。行业里常见的续约依赖，本质上就是这个问题的答案写错了。

供应商的三种失败模式

第一种是承诺型，把 40% 这类受控实验数字写进合同，交付时靠口径调参兑现。第二种是铺量型，用批量内容占满长尾，短期份额上升，随后被平台清理。第三种是缄默型，只交内容不交报表，效果无法核对，续约靠关系。

三种失败模式的共同根源是同一件事：缺少可复算的口径。这也解释了为什么本书把度量单列一章——在这个行业里，口径不是技术细节，而是合同的基础设施。

06 渗透　七行业的条件与失败模式

行业全景不能只讲总量，要讲分布。同一个市场在不同行业的渗透速度差三到五倍，原因不是预算多寡，而是问题结构不同。这一章给出判断条件、七个行业的对照、三行业的细描，以及四种最常见的失败模式。

先判断一个行业会不会被先改写

四个条件同时成立，这个行业就会被先改写。**决策链长**：用户在成交前要比较多次，问句自然多。**信息不对称强**：用户无法自行判断，必须依赖第三方口径。**合规要求高**：答案里必须出现资质、边界与免责，泛泛内容无法过关。**单价高**：一次被引用可能带来可观收益，投入产出比成立。四个条件里成立三个，就值得提前布局；只成立一个，等别人试错更划算。

反过来，短决策链、低单价、纯审美判断的品类（如快消新品试饮、服饰搭配），答案对成交的影响有限，优化回报会被摊薄。这类行业适合做品牌层的语料占位，不适合按效果付费。

七个行业的条件与节奏

下表按四个条件给出行业排序与切入点。排序不是市场规模的排序，而是“同样的投入能换回多少引用份额”的排序。医疗与本地服务靠合规与地域，餐饮靠场景与门店事实，房产家居靠对比与预算区间，汽车靠参数与配置，教育靠政策与资质，B2B 软件靠集成与案例。

表 6-1 七个行业的渗透条件与切入点

行业	决定性条件	切入的可抽取单元	见效节奏
医疗健康	合规与资质	适应证、流程、风险提示、机构资质	6 至 12 个月
本地服务	地域与时效	区域、价格区间、上门时效、口碑	3 至 6 个月
餐饮	场景与门店事实	菜品、人均、营业时段、包间与停车	2 至 4 个月
房产家居	对比与预算	面积、材料、工期、报价区间、验收标准	4 至 9 个月
汽车	参数与配置	配置差异、保养成本、落地价、交付周期	4 至 8 个月
教育培训	政策与资质	课程结构、师资资质、收费与退费规则	5 至 10 个月
B2B 软件	集成与案例	接口能力、部署方式、行业案例、报价模型	9 至 18 个月

说明：见效节奏为本白皮书在医疗、餐饮、房产家居、汽车四个行业成册研究中的观察区间，非行业统计值；实际节奏受平台规则与竞争强度影响。

同样的条件 不同的内容形态

七个行业的条件不同，落到内容形态上差异更大。医疗要长文与资质，餐饮要短事实与高频更新，房产家居要对比与报价区间，B2B 软件要文档与接口说明。用错形态，等于在错误的入口上投入。

表 6-2 七行业的内容形态与关键指标

行业	主力内容形态	关键指标	最常见的坑
医疗健康	长文加资质块，风险提示同段	权威来源引用率	只讲效果不讲边界
本地服务	区域页加价格与时效	地域问句覆盖率	一套内容投全国
餐饮	门店事实卡，按月更新	场景问句命中率	菜单信息过期
房产家居	对比表加工期与验收	比较类引用份额	报价只给单点数字
汽车	参数表加配置差异	配置问句引用率	参数抄错、版本混淆
教育培训	课程结构加资质与退费	政策类问句引用率	资质表述违规
B2B 软件	文档加接口说明与案例	集成类问句引用率	文档滞后于产品

说明：内容形态与指标为本白皮书归纳，不同平台对形态的偏好不同；指标命名与第八章的口径保持一致。

三个行业的细描

医疗健康。关键不是写得动人，而是写得可核：谁能出这个方案、依据是什么、风险在哪、出了问题谁负责。这类内容天然带资质与免责，写得越具体，越容易被引用，也越经得起监管检查。医疗的另一个特点是“宁可少说，不可错说”，一条过期的适应证足以让整站可信度归零。因此医疗的核心工作量不在写作，而在复核——每篇内容都要有明确的复核人与复核时间。

餐饮。用户问的是“这家店今天中午适不适合带孩子去”，答案要的是营业时间、人均、有没有包间、停车方不方便、招牌菜是什么。这些是随时变化的事实，更新的价值远大于文采。餐饮的难点是一致性：同一家店在点评站、地图站、自有站上的信息必须一致，任何一处过期都会让引擎在冲突中选择更保守的那条。因此餐饮的核心工作量在“对齐”，不在“创作”。

房产家居。用户要的是“我这套房、这个预算、这个工期，能不能做、多少钱、怎么做”。这是一组典型的比较与验证类问题，最吃逐项对比与报价区间。房产家居的误区是把报价写成单点数字，一旦被引用就变成纠纷源头；正确写法是给区间并说明影响因素，让答案既能被引用，也不至于把交付风险写死。

三个行业放在一起看，能说明“打法不能通用”：把医疗的内容标准套到餐饮，会写出没人看的合规长文；把餐饮的更新节奏套到医疗，会踩监管红线；把房产的单点报价套到家居，会在验收阶段引发争议。

渗透的三个阶段

第一阶段是试点，选一个问题集、一个品类，用三个月验证能否被引用，成本最低。第二阶段是规模化，把验证过的写法复制到全部品类，同时建监测与迭代机制。第三阶段是资产化，把语料、署名、度量沉淀为组织能力，交付从“买服务”转为“养资产”。

表 6-3 三个阶段的交付重点与退出条件

阶段	交付重点	进入下一阶段的条件	典型周期
试点	一个问题集、一个品类的可抽取改造	基线可比、出现可观察的引用变化	2 至 4 个月
规模化	模板化写法、监测体系、迭代节奏	模板可被非原作者复用且质量不降	6 至 12 个月
资产化	语料库、署名体系、内训与度量	自运营能力可独立跑一个季度	12 个月以上

说明：周期为本白皮书观察区间，实际取决于品类数量与合规强度；未达进入条件而强行扩张，通常导致质量下滑。

多数企业卡在第二阶段：试点有效，但复制时内容质量下滑，署名不统一，监测跟不上。原因往往不是能力不足，而是没有把第一阶段的写法固化成模板与验收标准。这个问题在第七章的有效动作分层里会有对应解法。

四种失败模式 与它们的早期信号

失败很少是突然发生的，通常有可识别的早期信号。识别信号比事后复盘有用，因为它能止损。

表 6-4 四种失败模式与早期信号

模式	早期信号	成因	止损动作
只改不测	三个月内没有任何基线对比	无问题集与监测	先补基线再继续
只写不改	内容量上升，引用份额不动	写法未击中被抽取要素	回到首句与拆段
只扩不固	品类翻倍，单品类质量下滑	模板未固化即扩张	暂停扩张，补模板
只承诺不界定	合同承诺具体百分比	用受控实验数字当交付	改按份额与零引用率

说明：早期信号按项目复盘的共性归纳，非统计分布；四条信号可同时出现。

07 红利　六条优势的机制与算账

这一章回答两个问题：这门生意的优势到底在哪，这份优势能不能算成钱。答案不靠判断，靠可核数据；凡是算不出来的部分，都会标明是估算而不是结论。

六条结构性优势

优势之所以“结构性”，是因为它们来自分配方式本身，而不是来自执行水平。执行水平会被追上，结构不会。六条优势里，前三条来自引擎的取舍机制，后三条来自成本与制度。

表 7-1 六条优势与它们的边界

优势	机制	可核依据	边界
单一位次	答案通常只给一个结论，赢者独占该问	十引擎实验零引用份额 20% 至 28%[3]	同一问句的赢家会被轮换
问句红利	疑问句触发检索与引用的概率显著更高	疑问句 64.7%，整体 13.7%[2]	陈述句与品牌词不吃这套
跨引擎复用	不同引擎看重同一组信号，取舍高度一致	十引擎一致性 Spearman 0.704[3]	私有语料与闭源引擎仍有差异
成本结构	内容为一次性投入，引用可重复发生	投放按点击计费，停止即归零	需持续维护时效与事实
可度量	引用份额可按句位加权、按词数归一化	指标定义与实验方法公开[3]	样本量不足时波动大
制度门槛	标识与可信评测构成准入，先入者占位	办法 2025 年 9 月施行，首批 9 家通过[6][7]	门槛会随时间抬高，也会筛掉自己

说明：第三列为本白皮书引用的公开实验与制度事实；第六行的制度优势依赖合规投入，不构成对所有机构的普遍结论。

空位比竞争更值得关注

十引擎受控检索实验里，有一个数字经常被忽略：零引用份额在 20% 至 28% 之间，个别配置高达 67%。[3]意思是，相当比例的提问，引擎给出了答案却没有引用任何外部来源。这不是“没有竞争”，而是“还没有人占据”。

空位有两类。一类是问题已被提出但没有可靠来源，比如新兴品类、地方性服务、细分工艺。另一类是来源存在但不可抽取，比如藏在图片、视频、表单里的信息。前者要补内容，后者要补结构。两类空位的共同点是：先占的人不用打败谁，只需要先到场。

空位的时间价值可以用一句话概括：它只对“先到者”有用。等到零引用份额降到 10% 以下，同样的内容投入只能买到更小的增量，这时候拼的是执行与预算，而不是时间。

问句是被引用的高发区

2026 年的公开研究给出了一个明确的结构差异：在 AI 问答场景中，整体引用率为 13.7%，而疑问句的引用率达到 64.7%，样本覆盖 55,393 条查询、连续 40 天。[2]差距不是几个百分点，而是接近五倍。

这个数字对内容策略的含义很直接：把标题写成名词短语，等于放弃了引用率最高的入口。把每一节写成“用户会怎么问”，再在首句给出结论，才是顺着机制做。要注意的是，这个比例属于特定引擎与特定时间窗的观测值，不能当作所有平台的通用常数。

五倍差不只影响写法，也影响组织方式。以问句为单位组织内容，意味着编辑的工作单位从“篇”变成“问”，排期表也要跟着变。这一步不完成，写法改进会在两周内退回原样。

一次投入 多引擎复用

同一批查询送到十个不同引擎，结果之间的一致性达到 Spearman 0.704，样本为 229 个查询、1,220 组结果、1,145 个被引段落。[3]这个强度说明：内容被选中与否，主要由内容本身与站点信源决定，而不是由某个引擎的偏好决定。

对投入决策的意义在于：不必为每个引擎单独建一套内容体系。一套可抽取的事实底座，加上按平台规则做的分发适配，就能覆盖大部分场景。真正需要单独处理的，是私有语料与闭源模型带来的差异，以及各平台对时效与地域的不同要求。

成本结构因此成立：内容是一次性投入，引用是反复发生的收益。这与按点击计费的投放有本质区别——投放停止，曝光立即归零；内容留在语料里，仍可能被引用。前提是内容的事实与时效被持续维护，否则一次过期数据就会让整段内容失效。

优势要落到可验收的指标上

可度量是这门生意能进入预算表的前提。引用份额的计算方式已经公开：按被引用句子在答案中的位置加权，再按词数归一化，避免长答案天然占优。这个口径让“我们被提到了”变成“我们占了多少”。

> **引用份额 = Σ（被引句权重 × 该句词数）/ Σ（全部候选来源的加权词数）**
>
> 位置权重按句序递减，句首高于句末；同一答案内多句引用按句累加；跨答案汇总时按查询分层，避免高频问句淹没长尾。

口径清楚了，才算得清账。下面用一个示例说明算法怎么落地——示例中的数字仅为演示，不代表任何真实客户。

表 7-2 引用份额算例

来源	句位	句词数	加权词数
品牌自有站	第 1 句	42	42 × 1.0 = 42.0
垂类媒体	第 3 句	38	38 × 0.6 = 22.8
竞品官网	第 2 句	35	35 × 0.8 = 28.0
其余候选	—	150	150 × 0.3 = 45.0

说明：句位权重取 1.0、0.8、0.6 三档为示例设定，实际权重以合作的监测口径为准；本表数字为演示数据，非真实项目。

按上表口径，品牌自有站的份额为 42.0 ÷（42.0 + 22.8 + 28.0 + 45.0）= 30.5%。同一组数据若只数“出现次数”，三家各一次，看不出差距；按句位加权，第一位与第三位的价值差就出来了。这就是为什么口径必须前置到合同。

有效动作与无效动作

把 45 项研究的结论落到动作层面，会得到一张相当不客气的清单。[2]有些被行业当作标准动作的做法，实证支持度接近于零；有些看起来“不高级”的动作，反而是最稳的。

表 7-3 动作的有效性分层

动作	判定	条件说明
按真实问句重写段落首句	高回报	语义相关性与上下文位置双强，成本最低
把数字补上口径与时间	高回报	可抽取证据为中等偏强，越具体越稳
做逐项对比与适用条件表	高回报	比较类问句最缺这类内容
价格与库存按周更新	中回报	时效信号有效，但只对时间敏感问题成立
结构化数据与 llms.txt[11]	中回报	决定能否被抓到，不保证被引用
追求更流畅更简短的改写	低回报	影响阅读体验，不直接影响采纳
堆“专家指出”等权威语气	无效	权威来自署名与资质，不来自措辞
关键词密度、批量低质铺量	负向	稀释语义，且可能触发平台降权

说明：判定为条件化结论，成立范围以综述所列样本与任务类型为限；跨行业使用需先做小样本验证。

那个 40% 只能这样用

行业里流传最广的数字是“优化后可见性提升 40%”。它出自 2023 年的原始研究，条件是模拟检索环境下、五篇候选文档中只优化其中一篇、对比相对可见性。[1]换成真实竞争、换成多篇同时优化、换成问答任务，这个数字都不成立；2026 年的综述已明确把它列为应当被拒绝的通用主张。[2]

> **可以这样用。**把它当作“优化动作在受控条件下确实能改变结果”的存在性证据，用来支撑立项；不要把它当作报价承诺、KPI 基数或对外宣传口径。给客户承诺 40% 的服务商，交付时只能靠口径游戏兑现。

与之配套的反向证据更需要记住：一项覆盖 6 个领域、约 1,900 查询、16,360 文档的研究中，54 个优化组合只有 3 个显著为正，问答任务为 0；另一项覆盖 171,003 文档、2,700 查询的实验里，仅优化正文使 top-20 下降 9%，重排后 top-10 下降 16%，引用下降 6%。[2]这些结果不是否定价值，而是划出边界：动作要对，场景要对，单点改写不够。

三步把优势算成钱

优势要变成预算语言，需要三步：先算可触达的问题量，再算单位问题的获取成本，最后算一次引用值多少钱。三步都是估算，但估算过程可复核，比拍一个百分比可靠得多。

表 7-4 投入产出的三步估算

步骤	算法	需要注意的偏差
可触达问题量	问题集内可被自有内容覆盖的问句数	问题集本身可能偏头部，需按长尾分层抽样
单位获取成本	内容与人力总投入 ÷ 覆盖问句数	越到长尾，单问成本越高，不宜线性外推
单次引用价值	该品类的成交转化 × 客单 × 引用影响系数	影响系数无公开基准，应按小样本 A/B 校准

说明：第三步的“引用影响系数”目前没有可引用的公开基准，本白皮书不建议直接套用外部数字；应以自有小样本实验逐步校准。

三步之外还要加一条约束：估算结果必须与“零引用份额”联动。空位多时，同样的投入能覆盖更多问句；空位减少后，边际成本上升。把所有年份按同一成本外推，是最常见的测算错误。

08 度量　从被提到到可结算

这个行业最缺的不是方法，是口径。没有口径，甲方无法验收，乙方无法复用经验，双方只能靠感觉续约。这一章把指标、抽样、报表与合同条款一次写清，目的是让“效果”变成可以签字的东西。

三层指标 各管一件事

指标不是越多越好，而是三层各管一件事：可见性指标回答“赢了多少”，过程指标回答“为什么赢或没赢”，资产指标回答“能不能持续”。只报第一层，报表好看但无法改进；只报第二层，动作勤奋但无法结算。

表 8-1 三层指标体系

层级	指标	定义	采样频率
可见性	引用份额	按句位加权、按词数归一化的占比[3]	按月，按品类分层
可见性	零引用率	该问题集内无任何外部引用的问句占比	按月
可见性	首位引用率	品牌出现在答案第一句的问句占比	按月
过程	覆盖问句数	问题集中被自有内容覆盖的问句数量	按季度
过程	可抽取率	抽样页面中满足七个格式要素的比例	按季度抽检
资产	署名可追溯率	被引用句可回溯到具名主体与资质的比例	按季度
资产	模板复用率	新增内容中由既有模板产出的比例	按季度

说明：指标定义为操作性界定，具体权重与抽样规则应在合同中固定；首位引用率为本白皮书建议指标，非行业通行标准。

抽样决定结论的可信度

同一批问题，抽样方式不同，结论可能完全相反。三条规则必须前置：**分层**，按品类、场景、决策阶段分层抽样，不能只抽头部问题；**重复**，同一问句在不同时段重复询问，取多次结果的平均，单次结果不足以判定；**固定**，问题集在一个考核周期内不随结果调整，否则等于每期换考卷。

本白皮书建议的起步规模是每品类每月不少于 200 个问句、每问句至少 3 次采样。这个数字不是行业标准，而是让波动可控的最低量级：问句太少时，一次偶发引用就能让份额翻倍，报表会失真。

波动还有两个来源需要交代：平台侧的模型与检索策略会更新，同一份内容的表现会跳变；季节与事件会让问句结构变化，节前的问句与节后的问句不是同一批。因此月报必须附“口径未变”的声明，否则跨月对比没有意义。

报表要有的十一个字段

报表不是越详细越好，而是要让读者在三分钟内判断“该不该调整”。十一个字段足以支撑一次季度复盘，多出来的字段通常是装饰。

表 8-2 月度报表的字段设计

#	字段	内容	用途
1	考核周期	起止日期与采样总次数	判断可比性
2	问题集版本	本期使用的问句清单版本号	防止换考卷
3	引用份额	按品类分层的加权份额	看结果
4	零引用率	无引用的问句占比	看空位与风险
5	竞品对照	同问题集内前三名主体	看位置
6	覆盖问句数	本期新增与累计	看产能
7	可抽取率	抽检样本的合格比例	看质量
8	署名可追溯率	可回溯的引用占比	看资产
9	异常说明	平台变更、口径调整、数据缺口	看可信度
10	本期动作	完成与延期的改造项	看执行
11	下期计划	目标品类与预期变化区间	看方向

说明：字段设计为操作性建议，可按企业现有报表体系合并；第 9 项异常说明是判断报表可信度的关键字段。

合同要写清的五条

第一条，问题集归属。问题集由甲方提供或共同确认，版本变更需双方签字，避免“成绩单由考生出”。第二条，基线。合作开始前完成一次基线采集，基线数据作为合同附件。第三条，口径。份额公式、句位权重、分层方式与采样次数写进合同，不留“以实际报告为准”的模糊表述。

第四条，验收区间。约定“提升区间”而不是“具体百分比”，并明确区间之外的处理方式：不达标如何补做，超额如何结算。第五条，资产归属。语料、模板、报表与问题集在合作结束后归甲方，或明确授权继续使用。

五条里最容易漏的是第四条。合同写成“提升 30%”，等于把受控实验的数字当成交付承诺；写成“份额相对基线提升 8 至 15 个百分点”，才既有约束力又可执行。

三类验收争议 与处理方式

争议不是靠讲道理解决的，靠的是合同里有没有预留口径。

表 8-3 三类验收争议与处理

争议类型	典型说法	处理方式
口径争议	“你的份额算法和我的不一样”	**以合同附件的公式为准，双方各出一份报表交叉核对**
基线争议	“合作前没有数据，无法判断”	**用回溯采样重建基线，并在合同注明重建方法**
归因争议	“这次上升是投放带来的，不是内容”	**按品类分设对照，未投入品类作为参照组**

说明：三类争议覆盖项目复盘中出现的多数分歧；对照组的设置需要在项目启动时确定，事后补设通常不被认可。

三类争议有一个共同解法：把口径写在前面。这也是为什么度量章放在红利章之后、治理章之前——它承接优势的可结算性，也为合规留痕提供数据结构。

09 治理　标识、投毒与合规动作

行业的另一条主线是治理。2025 至 2026 年，规则从倡议变成强制，乱象从个案变成公开查处，平台从被动清理转向机制防御。对合规经营者而言，这不是负担，而是护城河。

标识制度从倡议变成强制

《人工智能生成合成内容标识办法》于 2025 年 3 月 14 日发布，2025 年 9 月 1 日施行，配套强制性国家标准 GB 45438-2025 同步执行，第十四条明确了服务提供者的标识义务。[7]2025 年 11 月 25 日，监管部门对未落实标识要求的应用进行了集中查处通报。[8]

这条制度对内容方的直接含义是：机器生成的内容必须可识别。过去那种用批量生成内容占位的做法，从“效果不好”变成“合规风险”。反过来，人工核实、署真名、带资质的内容，获得了制度性的相对优势——这也是第七章所说的制度门槛。

还有一个常被忽略的细节：标识义务落在“服务提供者”身上，而多数品牌方是“使用者”。这意味着品牌方即使不直接开发工具，也要对合作方产出的内容负责。合同里因此需要一条：合作方须保证生成内容已完成标识，并留存记录。

三类乱象与其定性

2026 年的公开报道把行业乱象归为三类：向语料池注入虚假或误导信息（业内称“AI 投毒”）、以测评与榜单为名的付费排序、批量生成低质内容抢占答案位。[5][6]三类乱象的共同点是短期有效、长期自毁：一旦被平台识别，站点与品牌会同时受损。

表 9-1 三类乱象的表现与处置

乱象	典型表现	定性依据	后果
AI 投毒	向公开语料注入虚假参数、虚构资质	司法层面已有相关定性[6]	**法律与平台双风险**
付费榜单	以测评、推荐为名收取排序费用	广告与测评规范约束[5]	**行政处罚与信任损失**
批量低质	大批量生成同质内容抢占长尾	标识办法与平台规则[7]	**降权与清理**

说明：定性依据为公开报道与规章条款的归纳，具体案件以司法机关与监管部门的正式结论为准。

平台自净的两道防线

第一道是来源侧的防线。百度上线答案溯源能力，360 建立信源优先级与内容安全机制[5]；豆包搜索的接口支持权威来源过滤与网页权威度重排，并可标注引用来源[9][10]。这些能力的共同方向是让“来源可查”，把不可查的内容挤出答案。

第二道是生态侧的防线。中国广告协会推动信源分级，知乎等平台建设统一语料层，行业联合发布可信生态报告，首批 9 家机构通过可信专项评测[6]。评测与分级的作用，是把“谁可信”从主观判断变成可核定的资质。

两道防线合起来，改变了竞争的性质：过去比谁能把内容塞进语料，现在比谁能被认定为可信来源。对愿意做长期投入的机构，这是明显有利的结构变化。

合规不是观念 是五个动作

合规在项目里通常被写成原则，落到执行却没人知道该做什么。把它拆成五个可检查的动作，才能进入排期表。

表 9-2 五个可检查的合规动作

#	动作	执行标准	留存证据
1	标识	机器生成内容按办法与国标完成标识	标识记录与截图
2	署名	内容具名到主体与作者，资质可核	作者与资质文件
3	复核	每篇内容有复核人与复核时间	复核台账
4	溯源	关键数字标注出处，可回溯到原始文件	来源清单
5	纠错	发现过期或错误内容后的下线时限	下线记录

说明：五个动作按执行顺序排列；第 5 项常被忽略，但它是平台清理时唯一能自证已尽责的证据。

合规成本已经进入价格

合规不是免费的。标识、留痕、资质核验、内容复核都要人力，这部分成本最终会体现在报价里。市场上同一份方案报价差三倍，很多时候差的就是这部分：一方只做内容，另一方承担了可追溯与合规责任。

买方在比价时需要问清一件事：报价里是否包含合规复核与留痕。如果不包含，后续的补做成本往往高于差价。这也是本书在第五章强调“交付边界”的原因——价格差异背后是责任差异。

10 趋势　2026 至 2028 的分年推演

趋势章最容易写成愿景。这一章换一种写法：把判断拆到年份上，每条判断都挂条件与观察点，读完之后能核对现实走到了哪一步，而不是只能点头或摇头。

研判的方法 三类信号

趋势判断之所以经常失准，是因为把三类不同成熟度的信号混在一起讲。本白皮书把信号分三级。**已发生**：有官方文件或公开数据，可以直接引用，不需要预测。**在发生**：有可观察的早期迹象，但规模与速度未定，需要给出区间。**待观察**：只有逻辑推演，没有可核证据，只能给触发条件，不给时间表。

分级的好处是可追责。读者可以问：这条判断属于哪一级，依据是什么，什么时候能验证。回答不了这三问的判断，不应当写进研判章节。

六个已经发生的变化

前三个变化有制度或数据支撑，后三个变化有公开可见的产业事实，属于“已发生”而非“将要发生”。把已发生的事情说成趋势，是这个行业最常见的冗余。

表 10-1 六个已发生的变化

变化	已发生的证据	对内容方的含义
入口迁移	6.02 亿用户，占网民 53.5%[4]	决策前的信息获取已部分移出搜索框
市场规模成型	42 亿元、增速 38%、68% 企业列预算[5]	可以单列预算与独立考核
制度强制	标识办法 2025 年 9 月施行，11 月集中查处[7]	合规成为入场条件，成本进入报价
平台开放检索	检索接口返回站点与权威度字段[9][10]	份额可归因，度量成为可能
评测体系出现	首批 9 家通过可信专项评测[6]	采购有了资格口径
投毒被定性	司法与监管公开处置[5][6]	短期套利路径被压缩

说明：六项均有公开文件或数据支撑，属事实陈述；其“含义”一列为本白皮书的判断，需与事实分开阅读。

分年推演 三年三种重心

把 2026 至 2028 拆成三年，每年有一个主导矛盾。2026 年的矛盾是“有需求没口径”，2027 年的矛盾是“有口径没产能”，2028 年的矛盾是“有产能没差异”。三年的重心因此不同：2026 拼度量，2027 拼组织，2028 拼资产。

表 10-2 三年的主导矛盾与重心

年份	主导矛盾	行业重心	可观察的验证点
2026	需求已起，口径未定	度量体系与合同口径成为竞争点	头部服务商是否开始提供可复算报表；甲方是否把份额写进合同
2027	口径有了，交付跟不上	组织化交付与模板复用成为分水岭	同一模板能否被非原作者复用且质量不降；人均覆盖问句数是否上升
2028	交付普及，差异消失	语料资产、署名体系与评测资质成为壁垒	语料与报表是否可留存在甲方；是否出现以资质为门槛的采购条款

说明：本表为条件化推演，成立前提是国内市场维持 30% 以上增速且平台继续开放检索接口；前提变化时，年份重心需重新判断。

三年的推演有一个共同底色：竞争从“会不会做”转向“能不能证明”再转向“有没有留下”。这个顺序不是猜测，而是由制度与工具成熟度决定的——先有度量，才谈组织；先有组织，才谈资产。

四个技术变量 会改变做法

技术变量不需要预测具体产品，只需要判断它改变哪个环节。四个变量分别冲击召回、生成、署名与语料边界，对内容方的影响差异很大。

表 10-3 四个技术变量与影响环节

变量	冲击环节	对内容方的含义	确定性
多模态检索	召回	图片、视频、表格中的事实也需要可检索	在发生
任务型代理	生成与调用	答案从“说”变成“做”，结构化数据更重要	在发生
端侧与私有库	语料边界	公开语料的边际价值下降，自有语料变重要	待观察
闭源与自建语料	署名	不对外署名的场景增多，份额口径需分层	待观察

说明：确定性一列按三类信号判定；待观察项不给时间表，只给触发条件，即出现大规模商用或平台规则明确。

四个变量里，对当下影响最直接的是多模态检索。大量门店与产品的关键事实仍以图片形式存在，一旦检索能读图，这批事实会重新进入竞争，而目前多数机构还没有为它们建立文本版本。这是可以提前做的事，不需要等产品成熟。

三情景 不是三种情绪

情景不是“乐观、中性、悲观”的情绪标签，而是三条可判定的路径。每一条都给出触发信号与对应策略，读者按信号对照，而不是按心情选择。

图 10-1 三情景的判定路径



说明：判定路径为简化示意，实际判断需同时观察接口开放度、制度强度与竞争密度三项；节点不代表概率分布。

表 10-4 三情景的触发信号与策略

情景	触发信号	策略含义
规模扩张	接口字段继续开放，评测机构扩容，预算科目独立	加大内容与度量投入，抢空位
封闭化	引用字段收窄，平台自建语料，第三方监测受限	转向平台内合作与自有语料
合规收紧	标识与资质要求扩围，查处频次上升	把合规能力做成进入门槛

说明：三个情景可以同时发生，不是互斥选项；策略应做组合，而非押注单一情景。

六个可证伪指标

研判是否准确，不能靠事后回看叙事，要靠指标。下面六个指标各自对应一条判断，任一越界，就意味着相应判断需要修正。

表 10-5 六个可证伪指标

指标	当前值	证伪阈值	被证伪的判断
生成式用户增速	同比 141.7%[4]	低于 20%	入口仍在扩张
市场规模增速	38%[5]	低于 15%	市场处于成型期
企业列预算比例	68%[5]	连续两季下降	预算科目已独立
跨引擎一致性	0.704[3]	低于 0.5	一套内容可多引擎复用
零引用份额	20% 至 28%[3]	低于 10%	空位仍然存在
可信评测机构数	首批 9 家[6]	超过 30 家	评测具备稀缺性

说明：阈值为本白皮书设定的判定线，非机构预测；指标应按季度复核，越界即修订对应判断，不修改阈值。

六个指标里，最值得盯的是第四个与第五个。前者决定“复用”这条优势是否还成立，后者决定“空位”这条优势还剩多少。两个都失效时，这个行业的打法会退回到纯粹的内容运营，红利期结束。

十条会与不会的判断

为了让研判可核对，下面把判断写成分条形式，每条给出条件与观察窗口。凡条件不成立，判断即不成立，不再解释。

表 10-6 十条判断与观察窗口

#	判断	成立条件	观察窗口
1	答案层不会回到十条并列的形态	用户留存不出现明显下降	12 个月
2	平台不会长期免费开放全量引用字段	商业化压力持续存在	18 个月
3	内容数量的边际回报会继续下降	平台继续加强语料治理	12 个月
4	合规成本会继续上升，不会回落	标识与评测要求不缩小	24 个月
5	服务商四层结构会收敛为三层	工具与咨询出现合并型产品	24 个月
6	按份额结算会替代按篇结算成为主流	甲方普遍建立基线数据	18 个月
7	不会出现统一的跨平台份额标准	平台之间不共享引用数据	24 个月
8	医疗与金融的合规门槛会先于其他行业抬高	监管延续当前强度	12 个月
9	端侧与私有库不会在两年内取代公开语料	企业侧部署成本不出现断崖下降	24 个月
10	“40% 提升”类承诺会继续被滥用，但会开始被拒签	甲方采购专业度上升	12 个月

说明：十条判断均为条件化断言，条件不成立时判断自动失效；观察窗口自 2026 年 11 月起算。

怎样核对这一章是否准确

核对方法只有一种：按季度把表 10-5 的六个指标更新一遍，再看表 10-6 的条件是否仍成立。这不是“预测—验证”的学术流程，而是经营动作——判断错了要改预算，不是改文章。

本白皮书不承诺这三年的推演会全部命中。它承诺的是：每条判断都写清了条件与观察点，读者可以自己核对，不必相信作者的权威。这比给一个漂亮但无法验证的结论更有用。

11 结语　三类主体的行动清单

十一万字不如一张清单。这一章只做一件事：把前十章拆成三类主体各自该做的事，按顺序排好，能直接进排期表。

五句话的落点

第一句，入口已经换位，这不是趋势而是既成事实。第二句，市场已经成型，42 亿元与 68% 的预算渗透率足以支撑独立立项。第三句，优势是结构性的，单一位次、问句红利、跨引擎复用与可度量决定了投入产出比。第四句，动作要分层，首句、口径、条件、对比表是最稳的四件事。第五句，窗口不长，先入场者拿的是位置与资质，后来者只能拼执行。

五句话合起来是一个判断：这门生意值得做，但要按可验收的方式做。做得早不如做得对，做得对不如做得可证明。

甲方 先建度量再谈内容

顺序不能颠倒。先有问题集与基线，再谈内容；先有合同口径，再谈份额目标。多数项目失败不是因为内容差，而是因为没有基线，效果无法归因，最后只能靠关系续约。

具体动作有四个：一是自建问题集，按品类与场景分层，每季度维护；二是合作前完成基线采集，把数据写进合同附件；三是把份额、零引用率、首位引用率三项写进验收条款；四是在合同里约定资产归属，语料与模板要留在自己手里。

服务商 交付边界比承诺更重要

承诺越满，交付越难。可执行的做法是把交付拆成三档，按档位报价，把边界写在报价单上。入门档只交动作，标准档交结果，共建档交能力，三档不混用。

另外三件事值得投入：一是把写法固化成模板，模板要能让非原作者用；二是把报表做成可复算的，字段与公式提前约定；三是把合规留痕做成流程，标识、署名、复核、溯源、纠错五项都要有记录。这三件事做完，续约理由从“关系”变成“资产”。

内容团队 先改结构再改文笔

内容团队最常见的误区是先打磨文笔。文笔影响阅读体验，不影响抽取结果；结构影响抽取，进而影响引用。顺序应当是：先让每段首句能独立成立，再让数字带上口径，再补适用条件，最后才处理表达。

三个可直接执行的动作：一是把标题从名词短语改为用户会问的问句；二是一段只回答一个问题，长段落拆开；三是把散落在正文里的对比整理成表。三件事不增加预算，但会改变结果。

表 11-1 三类主体的行动清单

主体	前三个月	第一年
甲方	建问题集、采基线、定口径	把份额与零引用率写进合同，按季度复盘
服务商	拆交付档位、固化模板雏形	报表可复算、合规留痕成流程、案例可脱敏复用
内容团队	改首句、拆段落、补条件	以问句为单位排期，模板覆盖主要品类

说明：清单按执行顺序排列，前三个月完成度决定第一年的上限；未完成前置动作直接进入规模化，通常在第 6 至 9 个月出现质量下滑。

最后一句留给读者自己判断：这个行业最大的风险不是技术变化，而是用不可验收的方式投入，然后在一年后发现说不清效果。避开这一条，剩下的都是可解的问题。

附录 术语、复算与研究分级

表 A-1 术语表

术语	本白皮书的界定
生成式引擎优化	通过内容与信源建设，提高品牌被生成式引擎采用并署名的相对份额
引用份额	按句位加权、按词数归一化后，本方在候选来源加权词数中的占比
零引用	引擎给出答案但未引用任何外部来源的情形
可抽取性	内容能被检索、切分并作为独立片段使用的程度
四层门	被抓到、被召回、被信任、被采用四道串联筛选
可证伪指标	设定了明确阈值、越界即判定相关判断失效的观测指标

说明：术语界定为本书自用，与平台内部术语不必然一致；引用他处材料时以其原定义为准。

表 B-1 复算方式说明

派生值	计算方式	原始来源
生成式用户占网民 53.5%	6.02 亿 ÷ 11.25 亿，保留一位小数	CNNIC 第 57 次报告[4]
引用份额算例 30.5%	42.0 ÷（42.0+22.8+28.0+45.0）	本白皮书示例数据
年净增 3.53 亿	2025 年用户数减 2024 年用户数	CNNIC 第 57 次报告[4]

说明：复算仅做算术派生，不引入模型外推；任何含假设的推演均在正文标注为推演。

表 C-1 证据分级

级别	界定	本书中的用法
一级	官方文件、统计公报、强制性标准	作为事实直接引用，标注文件名称与时间
二级	同行评议或公开实验的研究成果	标注样本与条件，跨条件使用需重新验证
三级	公开报道、机构报告、项目复盘	标注为归纳判断，不承担统计含义

说明：三级证据不得用于支撑数值型结论；本书凡数值结论均来自一级或二级证据，或标注为复算。

引用与转述约定

本书引用的第三方数据均标注编号，编号对应文末参考来源。凡直接引用原文观点，以引号或明确归属标示；凡转述并做归纳，标注“本白皮书归纳”。涉及预测的部分一律标注条件，不将推演表述为事实。

本书不引用无法访问的链接，不引用二手转述中的数字。若原始文件更新，以更新后的版本为准，并在版本记录中注明变更。文末清单不印网址，按机构、标题与文献编号可检索；11 条来源均于 2026 年 11 月 14 日逐条实测返回 HTTP 200，逐条探测记录与网址见配套核验报告。

参考来源

[1] Aggarwal P. et al. *GEO: Generative Engine Optimization*. arXiv:2311.09735, 2023.
[2] *A Survey of Generative Engine Optimization*（含 45 项研究的横向比较）. arXiv:2607.14035, 2026.
[3] *Controlled Retrieval Experiments across Ten Generative Engines*. arXiv:2610.38090, 2026.
[4] 中国互联网络信息中心.《第 57 次中国互联网络发展状况统计报告》. 2026-03.
[5] 经济日报. 生成式引擎优化市场观察（平台不下场、不干预；42 亿元、CAGR 38%、68% 预算、A 股涨幅）. 人民网转载, 2026-03-20.
[6] 新华网. 中国人工智能产业发展联盟可信生态研讨会与首批评测发布. 2026-05-18.
[7] 国家互联网信息办公室.《人工智能生成合成内容标识办法》. 2025-03-14.
[8] 国家互联网信息办公室. 2025 年网络内容生态治理年度综述（含未标识应用集中查处）. 2025-12-31.
[9] 火山引擎. 豆包搜索接口文档（联网搜索 API，文档编号 87772）. 2026-11 版.
[10] 火山引擎. 搜索能力与权威度说明（联网搜索文档子页，文档编号 2272953）. 2026-11 版.
[11] llmstxt.org. *The llms.txt specification*（2024-09-03 首发，2026-08-10 修订）.

体例与阅读说明

数据时点同样需要交代。需求侧数值取 2025 年 12 月的官方统计，市场侧取 2025 年支出与 2026 年预算口径，研究侧取 2026 年 10 月前的已发表成果，制度侧截至 2026 年 11 月的公开文件。凡跨时点比较，正文都会标出两个时点，不做隐含的同比替换。

全书四类标记需要先说明。**带编号的角标**指该段数据或结论出自参考来源的对应条目，均为官方发布或已发表研究。**“本白皮书复算”**指出自公开数值派生的比率，未进入任何预测模型。**“条件化结论”**指结论只在标注的样本、语料与任务类型下成立，跨条件使用需要重新验证。**“归纳判断”**指本白皮书基于公开报价、公开报道与项目实践的归纳，不构成统计值，正文会逐处标注。

编制与修订记录

表 D-1 编制与修订记录

版本	日期	变更要点
V2.0	2026-11-15	目录页改为只放目录，取消页内导读与图目录；分割线改为蓝色，条目行距放大至铺满整页
V2.0	2026-11-15	册号由第 08 册改为第 10 册，与已发布的 01 至 09 册编号冲突一并消除，封面与配套产出同步换号
V2.0	2026-11-15	参考来源不再印制网址，11 条逐条发起 HTTP 探测，返回码与探测时点回填至来源探测记录
V1.9	2026-11-14	十章结构定稿；六幅图按图型分类核验，四十四张表按条文、判定、清单、数据四级分级核验
V1.0	2026-11-14	初版付印：定义与边界、需求与供给、市场与玩家、行业渗透、结构性红利、度量、治理、趋势十章

说明：版本号按付印时的口径记录，同日内多次改动合并计入当日版本；页数随内容增减调整，本白皮书为 29 页。

本白皮书所引数据均来自上述公开来源，标注编号与正文角标一一对应；文中涉及的判断与推演均已标注条件，不构成对任何机构的投资或采购建议。版本 v2.0，2026 年 11 月。