



求智启新人工智能研究院

QIUZHI QIXIN AI RESEARCH INSTITUTE

— 企业AI转型落地 · 培训 · AGENT定制 —

GEO

GENERATIVE ENGINE OPTIMIZATION

生成式引擎优化白皮书

(2 0 2 6)

—— 平台机制 · 九种验证过的玩法 · 五层优化体系 · 度量与案例 ——

求智启新人工智能研究院 编著

二〇二六年八月

版权声明与免责声明

版权声明

本白皮书版权归求智启新人工智能研究院所有。任何单位或个人转载、摘编、引用本白皮书全部或部分内容，须注明来源为“求智启新人工智能研究院《GEO 生成式引擎优化白皮书（2026）》”，且不得对内容进行歪曲、篡改。

资料来源与编制立场

本白皮书的事实性素材来源于同行评议的学术论文、第三方数据机构的公开报告、平台官方说明与权威媒体报道。本研究院对上述公开信息进行了独立整理、交叉校验与再分析；文中的五层优化体系、度量框架、判断与建议为本研究院独立研究成果，不代表任何被引用机构或平台的立场。

本白皮书未获得任何被引用机构、平台或服务商的授权或背书，亦未与其形成合作关系。文中提及的产品与服务名称仅为说明行业现象，其商标权归各权利人所有。

关于 GEO 数据的四点重要提示

GEO 是一个成型不足三年的领域，读者在使用本白皮书数据时请务必注意：

1. **平台机制在持续变化。** 生成式引擎的检索与引用逻辑由各平台自行决定且不对外公开完整规则，可能随版本更新而显著改变。本白皮书中的平台特性描述反映的是 2026 年年中的观测结果。
2. **第三方监测数据存在方法论差异。** 引用率、市场份额、引荐流量等指标高度依赖抽样方式、查询集设计与统计口径，不同机构的数字不可直接比较，本白皮书均标注了来源。
3. **服务商披露的案例数据缺乏独立核验。** GEO 服务市场处于早期，营销性质的效果宣称较多。本白皮书对此类数据统一标注为“服务商披露”，并对无法交叉验证的具体数字采取保守呈现。
4. **“有效”具有强时效性与领域依赖性。** 学术研究已证明同一优化方法在不同领域的效果差异显著（详见第三章），照搬他人做法而不做本领域验证，是这一领域最常见的浪费。

免责声明

本白皮书所载信息截至 2026 年 8 月，仅供研究参考与决策辅助之用，**不构成任何营销投放建议、服务商选型建议或效果承诺**。GEO 效果受平台规则、行业竞争、内容基础等多重因素影响，任何单位或个人依据本白皮书内容作出的决策及由此产生的后果，本研究院不承担责任。

本白皮书不鼓励任何试图操纵生成式引擎输出的行为。文中所述方法均以“让真实、准确、结构清晰的信息更容易被正确理解和引用”为原则，任何以虚假内容、批量低质内容或误导性表述获取曝光的做法，均不在本白皮书的讨论与推荐范围内（见第七章）。

目 录

CONTENTS

版权声明与免责声明	2
版权声明	2
资料来源与编制立场	2
关于 GEO 数据的四点重要提示	2
免责声明	2
前言：问题从“排第几”变成了“被不被提到”	6
这本白皮书想回答四个问题	6
本白皮书的两条立场	6
怎么读这本白皮书	7
摘要	8
一、核心判断	8
二、关键数据速览	8
三、本白皮书的原创工具	9
四、一句话结论	9
编制说明与研究方法	10
研究范围	10
术语约定	10
资料来源分层	10
数据处理原则	11
本白皮书的局限	11
第一章 GEO 是什么	12
1.1 入口迁移：三组数据	12
1.2 定义与边界	12
1.3 根本差异：从“抢位置”到“被引用”	13
1.4 一个反直觉的学术发现	13
1.5 GEO 不适合谁	14
第二章 生成式引擎怎么选内容	15
2.1 三种取源架构	15
2.2 三大国际平台的偏好差异	15
2.3 跨平台通用的四类信号	16
2.4 国内平台的情况	17
2.5 引擎正在变成“有广告位的引擎”	17
第三章 九种被验证过的玩法	19
3.1 方法一：添加引文（效果最高）	20
3.2 方法二：补充统计数据	20
3.3 方法三：流畅度优化（性价比最高）	21
3.4 方法四：引述来源	21
3.5 方法五：技术术语	22
3.6 方法六：易理解性优化	22
3.7 方法七：权威性表达	22
3.8 方法八：独特词汇	23
3.9 方法九：关键词堆砌（负效果，应停止）	23

3.10 最重要的一句话：效果因领域而异	23
第四章 五层优化体系 25	
4.1 L1 事实层：让 AI 先说对	25
4.2 L2 内容层：让 AI 能读懂并转述	26
4.3 L3 信号层：让 AI 认为你可信	27
4.4 L4 分发层：出现在 AI 取信息的地方	28
4.5 L5 监测层：知道有没有用	29
4.6 五层的推进节奏	29
第五章 度量体系：怎么证明有用 31	
5.1 为什么 SEO 的指标在这里失效	31
5.2 五个核心指标	31
5.3 怎么建基线：一套零成本的手工方法	32
5.4 归因难题：必须向管理层说清楚的一件事	33
5.5 什么时候该停	34
第六章 落地案例 36	
6.1 一级证据：学术实验	36
6.2 二级证据：平台观测	36
6.3 三级证据：服务商与行业披露案例	37
6.4 横向观察：五条结论	38
第七章 六个陷阱与风险边界 40	
7.1 陷阱一：把 GEO 做成 SEO 换皮	40
7.2 陷阱二：批量灌注低质内容	40
7.3 陷阱三：只做自有渠道	40
7.4 风险一：错误信息的放大与固化	41
7.5 风险二：平台规则变化与广告位挤压	41
7.6 风险三：服务商选型与效果承诺	42
第八章 分角色行动建议 43	
8.1 给市场与品牌负责人	43
8.2 给内容与运营团队	43
8.3 给技术负责人	43
8.4 给管理层	43
第九章 趋势展望（2026—2030） 45	
9.1 自然可见度的成本会持续上升	45
9.2 度量会从“手工”走向“标准化”，但标准会滞后	45
9.3 “被引用源”本身会成为一个竞争市场	45
9.4 多模态与实时性权重上升	45
9.5 国内会形成与国际不同的生态路径	45
附录 A 参考信息源清单 47	
A.1 A 级：学术研究与国家统计局	47
A.2 B 级：第三方数据机构与平台观测	47
A.3 C 级：行业分析与服务商披露	48
附录 B GEO 自查清单（32 项） 49	
B.1 L1 事实层（8 项）	49
B.2 L2 内容层（8 项）	49
B.3 L3 信号层（6 项）	49

B.4 L4 分发层（5 项）	50
B.5 L5 监测层（5 项）	50
附录 C 基线测试记录表模板	51
C.1 测试执行规范	51
C.2 记录字段	51
C.3 月度汇总表	51
附录 D 平台适配速查表	53
D.1 跨平台通用的四类信号（资源有限时优先做这四项）	53
附录 E 术语表	54
关于求智启新人工智能研究院	55
我们是谁	55
GEO 在研究院服务体系中的位置	55
研究院的白皮书序列	55
联系我们	56
关于本白皮书	56

前言：问题从“排第几”变成了“被不被提到”

先看三个不需要解释的数字。

第一个：6.02 亿。这是国家互联网信息中心统计口径下我国生成式人工智能的用户规模，普及率 42.8%。

第二个：10.6%。这是艾瑞咨询调研中，仍以传统搜索引擎为首选的用户比例——与之相对，超过 41% 的用户已几乎完全转向 AI 搜索，47.8% 大幅增加了 AI 搜索的使用频率。

第三个：超过 80%。这是购买前会通过 AI 搜索辅助决策的用户比例，在 3C 数码、家电、旅游等决策周期长的品类中更高。

这三个数字合起来说明一件事：**用户提问的地方变了。**过去他打开搜索框输入几个关键词，然后在十条蓝色链接里挑；现在他直接问“预算五千左右适合摄影的相机推荐”，然后看 AI 给出的那一段回答。

对企业而言，这个变化的含义相当直接：

过去的问题是“我的网站排在第几”。现在的问题是“AI 在回答这个问题时，会不会提到我，以及会怎么描述我”。

这两个问题看起来相近，实际上要求的能力完全不同。排名是一场关于位置的竞争，你可以通过技术手段争取更靠前；而被引用是一场关于**可理解性与可信度**的竞争——AI 必须先能准确理解你是谁、你提供什么，并且认为这个信息足够可靠，才会在生成答案时把你写进去。

这就是 GEO（Generative Engine Optimization，生成式引擎优化）要解决的问题。

这本白皮书想回答四个问题

1. **生成式引擎到底怎么挑内容？** ChatGPT、Google AI Overviews、Perplexity 以及国内的豆包、DeepSeek、通义、夸克，它们的取源机制差在哪里，为什么同一份内容在不同平台的命运可能完全不同。
2. **有哪些玩法是真的有效的？** 不是靠经验和感觉，而是有学术实验数据支撑的——包括哪些做法实际上是负效果。
3. **具体怎么落地？** 从品牌事实梳理到被引用源建设，一套可以照着做的动作清单。
4. **怎么证明有效？** GEO 最大的实际困难不是做，而是说不清做了有没有用。第五章专门讨论度量。

本白皮书的两条立场

- **第一，GEO 不是 SEO 换个名字。** 第三章会给出一个来自学术研究的关键证据：在生成式引擎中，**关键词堆砌是负效果**——它会降低而不是提升可见性。沿用 SEO 思维做 GEO，可能做的是无用功，也可能是反效果。
- **第二，GEO 的正当边界是“让真实信息被正确理解”，不是“让引擎说我想让它说的话”。** 本白皮书不讨论任何形式的操纵性做法。除了伦理原因，还有一个现实原因：生成式引擎的核心竞争力就是答案可信度，任何被识别出的操纵手段都会被快速反制——投入即沉没。

怎么读这本白皮书

表 0-1 不同读者的建议阅读路径

读者身份	最相关的章节	建议读法
市场与品牌负责人	第一章、第五章、第七章、第八章	第五章度量体系最重要 ——不解决度量，预算申请与效果复盘都无从谈起
内容与运营团队	第三章、第四章、附录 B	第四章五层体系与附录 B 自查清单可直接作为工作手册
SEO 转型者	第一章 1.4 节、第三章、第七章 7.1 节	先看哪些 SEO 习惯要改掉， 再看新增什么
技术负责人	第二章、第四章 L3 层、第五章 5.3 节	关注 schema、可抓取性、结构化数据与监测实现
服务商与代理商	全书，重点第五章、第七章	第七章列了六类需要向客户说清楚的边界

求智启新人工智能研究院

二〇二六年八月

摘要

一、核心判断

- 1. **入口迁移已经发生，且不可逆。** 我国生成式 AI 用户达 6.02 亿、普及率 42.8%；调研显示仅 10.6%的用户仍以传统搜索引擎为首选，超 80%的用户在购买前用 AI 搜索辅助决策。全球生成式 AI 平台的月均访问量已达约 95 亿次、同比增长约 70%。**"要不要做 GEO"已不是一个有讨论价值的问题。**
- 2. **GEO 与 SEO 的差异是机制层面的，不是技巧层面的。** KDD 2024 收录的 GEO 研究给出了一个关键证据：在九种被测试的优化方法中，**关键词堆砌是唯一负效果的方法（约-7.8%）**，而添加引文（约+40.9%）、补充统计数据（约+34.2%）、优化行文流畅度（约+28.0%）效果最好。**SEO 时代的核心动作，在生成式引擎中失效甚至反向。**
- 3. **"平台通吃"的方案不存在。** 观测显示 ChatGPT 与 Perplexity 的引用域名重叠率仅约 11%，两者的引用率相差可达数十倍。ChatGPT 偏好百科式权威（Wikipedia 约占其引用的 7.8%）与第三方评测站；Perplexity 高度偏好新鲜内容（约 82%的引用来自 30 天内发布）与社区来源；Google AI Overviews 更看重语义完整性与多模态。**必须按平台分别设计。**
- 4. **GEO 最大的实际困难是度量，不是执行。** 引用发生在 AI 的回答里，不产生传统意义上的点击与排名；且 AI 回答具有随机性，同一问题两次结果可能不同。本白皮书第五章提出五个核心指标与一套无需付费工具即可建立基线的手工方法。
- 5. **这个领域正在快速商业化，也因此鱼龙混杂。** 国内 GEO 服务市场 2026 年已达数十亿元量级、年增速超 100%，同时出现大量效果宣称缺乏验证的服务。**本白皮书第七章列出了六类需要警惕的做法与承诺。**

二、关键数据速览

表 0-2 GEO 相关关键数据

维度	数据	来源与口径
我国生成式 AI 用户	6.02 亿人 ，普及率 42.8%	中国互联网络信息中心 统计报告
用户搜索行为迁移	超 41% 几乎完全转向 AI 搜索； 47.8% 大幅增加使用频率； 仅 10.6% 仍以传统搜索为首选	艾瑞咨询《2026 年 GEO 生成式引擎优化行业研究报告》
购买决策影响	超 80% 用户购买前 通过 AI 搜索辅助决策	同上；3C 数码、家电、旅游等 长决策周期品类渗透更高
国内 GEO 市场规模	2025 年约 6 亿元 ， 2030 年预计突破 500 亿元	艾瑞咨询
全球 AI 平台访问量	月均约 95 亿次 ， 同比增长约 70%	Similarweb（2025.06—2026.05）
平台份额变化	ChatGPT 从约 76% 降至约 53% ； Gemini 从不足 9%升至约 27—28% ； Claude 从 2%升至近 9%	Similarweb，全球网页流量份额， 2025 年 6 月对比 2026 年 5 月

维度	数据	来源与口径
引用率变化	ChatGPT 引用率从 1.6% 升至 6.8%	Similarweb；品类差异极大：旅游约 23%、汽车约 20%、专业服务不足 4%
平台间引用重叠	仅约 11% 的域名同时被 ChatGPT 与 Perplexity 引用	第三方审计；两平台引用率相差可达数十倍
学术验证的最佳方法	添加引文约 +40.9% 、统计数据约 +34.2% 、流畅度优化约 +28.0% ；关键词堆砌约 -7.8%	KDD 2024 《GEO: Generative Engine Optimization》，相对无优化基线

资料来源：求智启新人工智能研究院根据公开资料整理。不同机构统计口径不一，不宜跨来源加总或直接比较。

三、本白皮书的原创工具

- **GEO 五层优化体系**——事实层、内容层、信号层、分发层、监测层，每层配动作清单与优先级。见第四章。
- **五指标度量框架**——提及率、首位推荐率、引用份额、答案正确率、AI 引荐流量，并给出无需付费工具的手工基线建立方法。见第五章。
- **平台适配速查表**——按主流生成式引擎分列取源机制、偏好来源与优化重点。见第二章与附录 D。

四、一句话结论

GEO 不是让引擎多说你的名字，而是让引擎在需要说你的时候，能说对。前者是操纵，会被反制；后者是基础建设，会持续复利。

编制说明与研究方法

研究范围

本白皮书研究的对象是**"提升品牌与内容在生成式 AI 回答中被正确理解、引用与推荐的方法体系"**。讨论范围包括：生成式引擎的取源与引用机制、可操作的优化方法、效果度量方式、组织落地路径与风险边界。

以下不在讨论范围内：传统搜索引擎排名优化的具体技术（仅在与 GEO 对比时提及）、付费广告投放（尽管 AI 回答中的广告位正在出现，见第七章）、以及任何以操纵引擎输出为目的的手段。

术语约定

表 0-3 相关术语的关系

术语	含义	与 GEO 的关系
GEO 生成式引擎优化	让内容在生成式 AI 的回答中被正确理解、引用与推荐	本白皮书的主题
SEO 搜索引擎优化	提升网页在传统搜索结果中的排名	部分基础重叠 (可抓取性、权威性)， 但核心动作不同 ，见 1.4 节
AEO 答案引擎优化	面向"直接给答案"的场景优化， 如精选摘要、语音助手	GEO 的前身概念之一； 多数场景下与 GEO 动作重合
AI 可见度	品牌在 AI 回答中出现的 广度与质量的统称	GEO 的目标变量
引用 (Citation)	AI 在回答中标注的信息来源	GEO 的核心可观测结果
提及 (Mention)	AI 在回答正文中提到品牌， 不一定附带链接	比引用更重要 —— 零点击场景下， 提及本身就是曝光

资料来源分层

层级	来源类型	用途
A 级	同行评议的学术论文、 国家统计局口径数据	用于机制性结论与基础事实， 直接引用并标注出处
B 级	第三方数据机构的公开报告、 平台官方说明	用于市场规模、行为迁移与平台机制， 引用时标注机构与统计窗口
C 级	行业分析文章、服务商披露案例	用于玩法参考与案例说明， 标注为"服务商披露"， 不作为效果承诺依据

数据处理原则

1. **学术数据与商业数据分开呈现。** 第三章的九种方法效果来自可复现的学术实验，第六章的案例数据来自商业披露，两者的可靠性不在同一量级，本白皮书不将其混用。
2. **平台观测数据标注统计窗口。** 生成式引擎的行为变化很快，一个没有时间戳的引用率数字没有意义。
3. **对服务商效果宣称采取保守呈现。** 无法交叉验证的具体百分比，本白皮书以定性表述呈现并注明来源性质。
4. **明确区分"已验证"与"业界共识"与"本研究院判断"。** 三者在正文中分别标注，读者可自行取舍。

本白皮书的局限

- **平台机制不透明。** 所有关于生成式引擎取源偏好的结论，都是外部观测的统计规律，而非平台公布的规则。它们可能随时改变。
- **国内平台的公开研究极少。** 关于豆包、DeepSeek、通义、夸克等平台引用机制的系统性第三方研究目前罕见，本白皮书中国内部分的确定性低于国际部分。
- **缺少长周期效果数据。** GEO 作为实践不足三年，尚无跨越完整商业周期的效果追踪研究。
- **发布即过时的风险。** 这是本领域所有研究的共同问题。建议读者把本白皮书的**方法框架**作为长期参考，把**具体数字**作为时点快照。

第一章 GEO 是什么

这一章要把定义说清楚，尤其是把 GEO 与 SEO 的差异说清楚——因为绝大多数关于 GEO 的误解和无效投入，都来自"以为它是 SEO 的新版本"。

1.1 入口迁移：三组数据

先确认这件事的规模。

表 1-1 搜索行为迁移的三组证据

维度	数据	含义
用户基数	我国生成式 AI 用户 6.02 亿人 ，普及率 42.8%	不再是早期采用者的小众行为，已进入大众渗透阶段
行为替代	超 41% 用户几乎完全转向 AI 搜索； 47.8% 大幅增加使用频率； 仅 10.6% 仍以传统搜索为首选	替代而非补充。 传统搜索的首选地位已经失守
决策影响	超 80% 用户购买前通过 AI 搜索辅助决策； 3C 数码、家电、旅游等长决策周期品类渗透更高	影响的不是浏览， 是 购买决策的前置环节

资料来源：中国互联网络信息中心统计报告；艾瑞咨询《2026 年 GEO 生成式引擎优化行业研究报告》。

全球侧的数据方向一致：生成式 AI 平台的月均访问量已达约 95 亿次、同比增长约 70%；同时平台格局本身也在剧烈变动——ChatGPT 的全球网页流量份额从 2025 年 6 月的约 76% 降至 2026 年 5 月的约 53%，Gemini 从不足 9% 升至约 27—28%，Claude 从约 2% 升至近 9%。

这组份额变化对 GEO 有一个直接含义

不要把 GEO 做成"ChatGPT 优化"。

一年之内头部平台份额从 76% 掉到 53%，说明这个市场的格局远未稳定。把全部资源押在单一平台上，既承担平台规则变化的风险，也承担平台份额变化的风险。

第二章会说明：**不同平台的取源机制差异极大，"押一个平台"和"覆盖多平台"在动作上根本不是同一套方案。**

1.2 定义与边界

GEO (Generative Engine Optimization, 生成式引擎优化) 是指：通过优化信息的可理解性、结构、可信度与分布，使品牌与内容在生成式 AI 的回答中被正确理解、引用与推荐的方法体系。

这个定义里有三个词需要展开：

1. **"正确理解"排在第一位。** GEO 的第一道门槛不是曝光，而是准确——如果 AI 对你的产品参数、服务范围、公司信息的理解是错的，那么越多的曝光越是伤害。这也是第四章把"事实层"放在五层体系最底层的原因。

- 2. **"引用与推荐"是两件事。**引用（Citation）指 AI 在回答中标注了你的来源链接；提及（Mention）指 AI 在正文里说到了你但未必给链接。在零点击场景下，**提及往往比引用更有商业价值**——用户看到"推荐考虑 A 品牌"这句话就已经完成了心智占位。
- 3. **"分布"与"内容"同等重要。**生成式引擎从哪里取信息，决定了你的内容写得再好、如果不在它取信息的地方，也不会被看到。第四章的"分发层"专门讨论这件事。

1.3 根本差异：从"抢位置"到"被引用"

SEO 和 GEO 的差异不在技巧层面，在**结果形态**层面。

表 1-2 SEO 与 GEO 的六个根本差异

维度	SEO（传统搜索优化）	GEO（生成式引擎优化）
竞争对象	排名位置—— 十条链接里的第几条	是否被写进那一段回答—— 不存在"第二页"这个概念
用户行为终点	点击进入你的网站	多数情况下不点击（零点击）； 在 AI 回答里完成认知
核心资产	域名权重、外链、关键词覆盖	语义完整性、事实一致性、 可被引用的权威来源分布
内容单位	页面（Page）	可被独立引用的信息块—— 一段能自成答案的内容
成功信号	排名、点击率、自然流量	提及率、首位推荐率、 引用份额、答案正确率
结果稳定性	排名相对稳定，可持续观测	同一问题两次提问结果可能不同—— 必须用统计方法而非单次观察

"零点击"这件事需要被正视

很多企业 GEO 的第一反应是："AI 回答里提到我，但用户不来我网站，那有什么用？"

这个问题问反了。正确的问法是：**在用户已经不点链接的情况下，你还有没有别的方式进入他的决策？**

答案只有一个——**进入那段回答本身**。品牌在 AI 回答中被提及并被正面描述，其作用等同于过去搜索结果首位加上一段权威推荐语。它不带来流量，但它带来的是更上游的东西：**候选集资格**。用户后续可能直接搜索你的品牌名、直接去电商平台找你——这些流量在归因上不会算作"AI 来源"，但它们的起点在那段回答里。

第五章 5.4 节专门讨论这个归因难题。

1.4 一个反直觉的学术发现

如果这一章只能记住一件事，本研究院建议记住下面这个。

KDD 2024 收录的论文《GEO: Generative Engine Optimization》（Aggarwal 等）构建了 GEO-bench 基准，系统测试了九种优化方法对生成式引擎回答可见性的影响。结果显示，相对于无优化基线，九种方法中有八种为正效果，唯一的负效果来自“关键词堆砌”（约 -7.8%）。

而效果最好的三种是：添加引文（约 +40.9%）、补充统计数据（约 +34.2%）、优化行文流畅度（约 +28.0%）。

这个结果的含义是：**SEO 时代最核心的动作（围绕关键词组织内容）在生成式引擎中不仅失效，而且有害；而过去被认为“锦上添花”的动作（引用权威来源、给出具体数据、把话写通顺）反而成了主要杠杆。**

本研究院认为，这个发现解释了当前市场上一个普遍现象：**很多企业投入了资源做“GEO”，但做的其实是 SEO——把关键词换成长尾问句，把标题堆得更密，然后困惑于为什么没有效果。**第三章会逐一拆解这九种方法。

1.5 GEO 不适合谁

为避免误导，本研究院明确列出 GEO 优先级较低的情形：

表 1-3 GEO 优先级判断

情形	优先级	理由
决策周期长、 用户会主动做功课的品类 (B 端服务、专业服务、 高客单消费品、 软件与工业设备、 教育、医疗健康)	高	用户会向 AI 提出复杂的比较型问题， AI 的回答直接进入候选集筛选
信息不对称大、 用户不知道该问什么的品类	高	AI 承担了“帮用户定义需求”的角色， 品牌被写进定义里价值极大
品牌信息在网上 存在错误或不一致	极高，且 紧急	错误会被 AI 反复复制传播， 且纠正成本随时间上升
冲动型、低客单、 依赖货架流量的品类	中低	用户不会先问 AI 再买； 资源应优先投向货架与内容平台
纯本地生活服务 (如社区餐饮)	低	决策依赖地图与点评， AI 搜索介入有限
内容基础极差 (官网信息不全、 全网无第三方提及)	先补基础	GEO 优化的是“被理解和被引用”， 没有可被理解的东西时， 优化无从谈起

第二章 生成式引擎怎么选内容

这一章拆解机制。理解了引擎从哪里取信息、偏好什么样的来源，第三、四章的具体做法才不是记口诀。

必须先声明：**以下所有结论都是外部观测的统计规律，不是平台公布的规则。** 平台可能随时改变，读者应把它们当作"当前的经验模型"而非"确定的机制"。

2.1 三种取源架构

生成式引擎回答一个问题时，信息可能来自三个地方，不同平台的配比差异很大。

表 2-1 三种取源架构

架构	机制	对优化的含义
训练内嵌	信息在模型训练时已被吸收，回答时直接从参数中生成	无法短期影响。 依赖长期的、大范围的公开信息沉淀； 且存在明显的时间滞后
实时索引调用	调用已有的搜索索引，再由模型综合成答案	与传统搜索基础重叠最多。 可被抓取、可被索引 仍是前提条件
实时检索 (RAG)	每次提问都发起网络检索，读取当前页面内容后生成	响应最快、可影响性最强。 内容新鲜度与 页面可读性直接决定命中

实践中三种架构混合存在。观测显示：**ChatGPT** 以训练数据为主、辅以选择性检索；**Google AI Overviews** 主要调用其实时索引；**Perplexity** 则对每次查询都执行实时检索。这个差异直接导致了下一节的偏好分化。

2.2 三大国际平台的偏好差异

表 2-2 主要国际生成式引擎的取源偏好（2026 年观测）

平台	取源机制	偏好的来源类型	优化重点
ChatGPT	训练数据为主 + 选择性检索	百科式权威 —— Wikipedia 约占其引用的 7.8% ； 依赖第三方评测与 评论平台信号 (如 G2、Capterra 类)	强化 Bing 侧可见性； 获取第三方评测收录； 完善主题集群与 百科类条目的准确性
Google AI Overviews	实时索引调用，整合自然搜索结果	社区来源 (Reddit 约 2.2%) ； B 端场景中 LinkedIn 是 关键来源； 多模态内容 (图文视频结合 可显著提升被选中概率)	投资多模态内容； 维护视频平台存在； 保持页面定期更新； 语义完整性最关键

平台	取源机制	偏好的来源类型	优化重点
Perplexity	每次查询 实时检索	社区来源占比极高 (Reddit 约占其顶级来源的 46.7%) ; 强烈偏好新鲜内容—— 约 82% 的引用来自 30 天内发布; 单条回答平均引用约 21.87 条	提高发布频率; 采用问题式标题结构; 建立社区与评测平台影响力

资料来源：第三方审计与行业研究（2026 年）。上述比例为观测统计，非平台公布数据，且随版本更新可能变化。

2.2.1 一个必须知道的数字：11%

观测显示，仅约 11% 的域名同时被 ChatGPT 与 Perplexity 引用；两个平台的引用率相差可达数十倍（有审计给出的数字为 ChatGPT 约 0.59% 对 Perplexity 约 13.05%）。

这个数字的含义非常直接：“做好内容自然会被各家引用”这个假设是不成立的。两个都在做生成式回答的平台，取源结果几乎不重叠——说明它们看重的信号根本不同。

研究院观点：不要买“平台通吃”的方案

市场上常见的 GEO 服务承诺是“帮你在所有 AI 平台提升曝光”。基于上述观测，本研究院建议对这类承诺保持警惕。

更现实的做法是：先选定 1—2 个对你的业务最重要的平台，把动作做深；再考虑扩展。

怎么选？两个判断依据：**其一，你的目标客户实际在用哪个**（B 端与 C 端、国内与出海差异极大）；**其二，你的内容禀赋更接近哪个平台的偏好**——如果你有大量原创数据与研究，ChatGPT 路径更有利；如果你有高频内容生产能力，Perplexity 路径更有利；如果你有视频与多模态资产，AI Overviews 路径更有利。

2.3 跨平台通用的四类信号

虽然平台偏好差异极大，但有四类信号在各平台观测中都呈现正相关。**资源有限时，应优先做这四件事。**

表 2-3 四类通用引用信号

信号	含义	观测到的效果	怎么做
① 语义完整性	一段内容能自成答案，不需要读者再去别处补充	与被引用概率高度相关 (有研究给出相关系数约 0.87)	每个内容块回答一个完整问题； 不要写“详见下文”
② 结构化标记	FAQ、HowTo、Product 等 schema 标记	FAQ schema 与 ChatGPT 引用权重正相关（有观测给出约 +40%）	在关键页面部署对应的结构化数据

信号	含义	观测到的效果	怎么做
③ E-E-A-T 信号	可验证的作者身份、 更新日期、外部引用	各平台共同看重	署名到人并附资历； 标注更新时间； 引用权威外部来源
④ 原创数据与直接引述	自有统计数据、 可被引用的原话	统计数据可提升可见性 约 22% ； 直接引述约 37%	产出自有调研数据； 提供可直接引用的 观点陈述句

这四类信号与第三章的学术实验结果高度一致——引文、统计数据、流畅表达在两套独立的研究中都排在前列。这种交叉印证，是本白皮书对这四点给出较高置信度的原因。

2.4 国内平台的情况

必须坦率说明：关于国内生成式引擎（豆包、DeepSeek、通义千问、夸克、百度 AI 搜索、Kimi 等）取源机制的系统性第三方研究目前罕见，本节的确信性显著低于前两节。

基于行业实践的公开分享与本研究院的观察，可以给出以下几点判断：

- 1. **共性大于差异的部分：**结构化、语义完整、事实一致、有权威来源支撑的内容，在国内平台同样受益。第三章与第四章的基础动作跨平台通用。
- 2. **内容生态来源不同。**国际平台高度依赖 Wikipedia、Reddit、专业评测站；国内的对应生态是百科类条目、知乎、小红书、垂直社区、行业媒体与政务/协会类信息源。分发层的动作必须按此调整（详见第四章 L4 层）。
- 3. **部分平台与自有生态强耦合。**有搜索或内容平台背景的产品，其取源与自有索引、自有内容生态的关联度更高——这意味着在对应生态内的存在感会被放大。
- 4. **中文语义与实体识别是额外挑战。**品牌名的简称、别称、英文名、错别字变体，会显著影响 AI 能否把不同来源的信息关联到同一实体上。这一点在国内场景比国际场景更突出，也是第四章事实层的重点。

给国内企业的一条务实建议

在缺少可靠第三方研究的情况下，不要依赖任何人对国内平台机制的断言（包括本白皮书），而应自己做基线测试。

方法很简单，成本接近于零：列出 20—30 个你的目标客户真实会问的问题，在每个主流平台各问三遍，记录回答里提到了谁、怎么描述的、引用了哪些来源。一个下午就能完成，得到的却是针对你自己行业的一手数据。

第五章 5.3 节给出了这套手工基线方法的完整步骤与记录表。本研究院认为这是任何 GEO 项目都应当做的第一件事——它同时解决了"现状如何"和"以后怎么衡量"两个问题。

2.5 引擎正在变成"有广告位的引擎"

还有一个正在发生、但讨论较少的变化需要提示：AI 回答中的广告位已经出现。

Similarweb 的观测显示，约 **26%** 的 ChatGPT 回答中已包含广告，且广告密度集中在对话早期——约三分之一的广告曝光出现在首轮回答中，其后显著下降但在长会话中持续存在。

这对 GEO 的含义有两层：

- **短期是竞争加剧。** 自然被引用的位置会被付费位置挤压，尤其是在商业意图明确的查询上。
- **长期是格局重演。** 传统搜索的历史是“自然结果逐步被广告位挤压”，生成式引擎大概率会走类似路径。**这意味着当下是自然可见度成本最低的窗口期**——本研究院认为这是当前投入 GEO 的一个重要理由，其价值随时间递减。

第三章 九种被验证过的玩法

这一章的内容有一个别处少见的特点：**它来自可复现的学术实验，而不是从业者的经验总结。**

KDD 2024 收录的《GEO: Generative Engine Optimization》（Aggarwal 等）构建了 GEO-bench 基准——覆盖多领域用户查询与相应的网页来源——并系统测试了九种优化方法对生成式引擎回答可见性的影响。整体结论是 GEO 可将可见性提升最高约 40%，且**不同方法在不同领域的效果差异显著。**

本章逐一拆解这九种方法，每种给出：论文中的效果、本研究院的实操转译、适用领域与常见坑。

表 3-1 九种方法的效果排序（相对无优化基线）

排序	方法	位置权重提升	主观印象提升	本研究院的实操定位
1	添加引文 (Cite Sources)	约 +40.9%	约 +28.0%	首选动作。 投入低、效果最高
2	补充统计数据 (Statistics Addition)	约 +34.2%	约 +22.8%	首选动作。 需要有自有数据或 可靠外部数据
3	流畅度优化 (Fluency Optimization)	约 +28.0%	约 +13.5%	基础动作。 几乎零成本，全领域有效
4	引述来源 (Quotation Addition)	约 +27.5%	约 +13.5%	与①配合使用
5	技术术语 (Technical Terms)	约 +17.6%	约 +10.9%	专业领域适用， 大众领域慎用
6	易理解性优化 (Easy-to-Understand)	约 +14.0%	约 +6.2%	面向大众领域
7	权威性表达 (Authoritative)	约 +10.4%	约 +18.7%	主观印象提升最高。 适合辩论、历史、科学类
8	独特词汇 (Unique Words)	约 +6.2%	约 +5.7%	边际动作
9	关键词堆砌 (Keyword Stuffing)	约 -7.8%	约 +4.7%	负效果。停止这么做

资料来源：Aggarwal 等，《GEO: Generative Engine Optimization》，KDD 2024。数值为相对无优化基线（基线可见性约 19.3%）的相对改进幅度，两列分别对应位置加权与主观印象两类可见性指标。实际效果因领域而异，详见 3.10 节。

3.1 方法一：添加引文（效果最高）

做法：在你的内容中引用可信的外部来源，并明确标注出处。| **效果：**位置权重约 +40.9%，主观印象约 +28.0%——九种方法中最高。

为什么有效

生成式引擎在组织答案时需要判断“这段内容可不可信”。一段带有明确外部出处的陈述，比一段没有出处的陈述更容易被采信并转述——**因为引擎可以把可信度部分转嫁给你引用的来源。**

怎么做

- 关键结论后附来源：“据【机构】【年份】的【报告名】，……”这个句式本身就是可被引擎识别与转述的结构。
- 优先引用引擎本身就信任的来源类型：官方统计、行业协会、学术论文、权威媒体、上市公司披露。
- 给出可点击的链接，但更重要的是把来源名称写进正文——很多引擎读取正文语义而非只看链接。

适用领域

论文观测显示该方法在**人物与社会、解释类、历史类**查询上效果最突出。本研究院补充：**任何“用户会追问依据”的领域都适用**——B 端服务、医疗健康、金融、法律尤其如此。

常见坑

- 引用了**不存在或已失效的来源**。这会在引擎侧形成负面信号，且一旦被用户发现，信任损失远大于收益。
- **只堆链接不写来源名**。裸链接的语义价值远低于“据 XX 机构报告”这样的完整表述。

3.2 方法二：补充统计数据

做法：在内容中加入具体的、可核实的数字。| **效果：**位置权重约 +34.2%，主观印象约 +22.8%。

为什么有效

生成式引擎生成的是“答案”，而具体数字是答案质量的直接体现。当用户问“这个方案能省多少成本”时，一段含有具体百分比与金额的内容，比一段说“能显著降低成本”的内容有用得多——引擎的选择反映了这一点。

怎么做

- **优先使用自有数据**。行业里没有的数字，是最难被替代的引用理由——这也是本研究院建议企业做行业调研的商业理由之一。
- 数字要带口径：“在**30家客户**的样本中，平均缩短约**42%**”比“效率提升**42%**”更容易被采信。
- 把**关键数字**放在段落靠前位置，而不是埋在长段落中间。

适用领域

论文显示在**法律政务、辩论、观点类**查询上最有效。本研究院补充：**所有涉及对比与选型的商业查询**（"A 和 B 哪个好""预算 X 能买到什么"）高度依赖数据。

常见坑

- **编造或夸大数据。**除了合规风险，交叉验证不一致的数字会削弱整体可信度信号。
- **数据没有时间戳。**过时数据被引用后会持续传播，纠正成本很高。

3.3 方法三：流畅度优化（性价比最高）

做法：把内容写通顺——句子结构清晰、逻辑连贯、无冗余表达。 | **效果：**位置权重约 +28.0%。

为什么有效

这是九种方法中最"朴素"也最容易被忽略的一种。生成式引擎需要理解并转述你的内容——**表达越清晰，被正确理解与转述的概率越高**。晦涩、绕、充满营销黑话的文本，引擎理解成本高，转述时容易失真甚至被跳过。

怎么做

- 一句话说一件事；避免超长复合句。
- **删掉营销形容词，换成具体描述。**"行业领先的智能化解决方案"对引擎几乎无信息量；"支持在离线环境部署的视觉质检系统，单条产线部署周期约 2 周"有信息量。
- 段落开头直接给结论，细节放在后面（金字塔结构）。

适用领域

论文显示在**商业、科学、健康**领域最有效。本研究院认为：**这是唯一一个全领域、零成本、无风险的动作，应当无条件执行。**

本研究院对这条的额外强调

在九种方法里，"把话写通顺"排名第三，而"堆关键词"是唯一的负效果。**这两个事实放在一起，基本上就概括了 GEO 与 SEO 的全部差异。**

SEO 时代内容是写给算法看的，所以充满了为算法服务的痕迹；GEO 时代内容是给一个"会读懂并转述"的模型看的，**写给人看的内容，恰好就是写给模型看的内容。**

这是一个罕见的、优化目标与用户利益完全一致的局面。

3.4 方法四：引述来源

做法：直接引用专家、机构或用户的原话（带引号的完整陈述）。 | **效果：**位置权重约 +27.5%。

与方法一（添加引文）相近但不同：引文是标注出处，引述是**给出可被整段搬运的原话**。有观测显示直接引述可提升可见性约 37%。

怎么做

- 在内容中提供**完整、独立、可直接引用的陈述句**——一句话就能说清一个观点，不依赖上下文。

- 引述来源要具名到人并附身份：“XX 研究院院长 XXX 表示：……”
- 为你自己的核心观点也准备“可引述句”——一段可以被原样搬进 AI 回答的表述。这是本研究院认为最被低估的内容技巧。

3.5 方法五：技术术语

做法：在内容中使用领域内准确的专业术语。 | **效果：**位置权重约 +17.6%。

专业术语帮助引擎判断内容的专业度与领域归属。但这是一把双刃剑：论文显示其在专业领域有效，而在大众领域会与方法六（易理解性）冲突。

判断标准

- 用户在提问时会不会用这个词？会——用；不会——先给通俗说法再附术语。
- 术语首次出现时给出一句解释。这样既保留专业信号，又不牺牲可理解性。

3.6 方法六：易理解性优化

做法：降低阅读门槛，用大众能懂的语言表达。 | **效果：**位置权重约 +14.0%。

与方法五方向相反，适用于面向大众用户的领域。两者不是矛盾，而是分场景选择——同一品牌的 B 端页面与 C 端页面应采用不同策略。

3.7 方法七：权威性表达

做法：以明确、有立场、有依据的方式陈述，而非模棱两可。 | **效果：**位置权重约 +10.4%，但主观印象约 +18.7%——九种方法中主观印象提升最高。

这个方法的特点是：它对“位置”的提升一般，但对“引擎如何描述你”的提升最大。也就是说，它影响的不是“有没有提到你”，而是“提到你说得好不好”。

怎么做

- 给出明确判断而非罗列可能性：“在 X 条件下应当选择 A”优于“A 和 B 各有优劣，需要综合考虑”。
- 判断后紧跟依据——权威性来自“有立场且给出理由”，不是来自语气强硬。
- 署名到具体的人与机构，附资历。

适用领域

论文显示在辩论、历史、科学类查询上最有效。本研究院补充：在选型建议类的商业查询中同样关键——用户问“该选哪个”时，模棱两可的内容不会被引用。

3.8 方法八：独特词汇

做法：使用区别于同类内容的独特表述。 | **效果：**位置权重约 +6.2%。

效果有限，属于边际动作。本研究院的实操建议是：**不要为独特而独特，但要确保你的核心概念有一个稳定的、专属的命名。**例如本研究院在系列研究中使用的"SVA-3D 模型""七步现场交付法""LEAP 四步法"——稳定的专有名称会成为可被引擎识别与归属的实体标记。

3.9 方法九：关键词堆砌（负效果，应停止）

做法：围绕目标关键词高密度重复。 | **效果：**位置权重约 -7.8%——九种方法中唯一负效果。

这是本章最重要的一条，因为它是大量企业当前正在做的事。

为什么在生成式引擎中它会变成负效果

传统搜索引擎的排序依赖词频与匹配度等信号，因此关键词密度有正向作用。

生成式引擎则不同：它要读懂内容并转述。**高密度重复的文本在语义上是冗余的、不自然的**，既降低了单位篇幅的信息量，也可能被识别为低质内容特征。

换句话说：**你在为一个"数词频的系统"优化，而现在读你内容的是一个"要复述给用户听的系统"。**

行动建议：对存量内容做一次审查，把为关键词而写的段落改写为为回答问题而写的段落。这件事的收益是双向的——既去掉负效果，又同时提升了流畅度（方法三，+28%）。

3.10 最重要的一句话：效果因领域而异

论文明确指出，不同方法在不同领域的最佳适用性差异显著：

表 3-2 方法与领域的适配关系

方法	论文中效果最突出的领域
权威性表达	辩论、历史、科学
流畅度优化	商业、科学、健康
引述来源	陈述类、事实类、法律政务
添加引文	人物社会、解释类、历史
补充统计数据	法律政务、辩论、观点类

资料来源：Aggarwal 等，KDD 2024。

本研究院的使用建议

不要把这九种方法当成一张待办清单全部执行。论文本身强调"内容创作者应根据目标领域进行有针对性的调整"。

本研究院建议的优先级是：

第一批（无条件做）：方法三流畅度优化、方法九停止关键词堆砌。零成本、全领域有效。

第二批（有资源就做）：方法一添加引文、方法二补充统计数据。效果最高，但需要内容生产投入。

第三批（按领域选）：方法五技术术语与方法六易理解性**二选一**；方法七权威性表达用于选型建议类内容。

并且：无论选哪几种，都必须按第五章的方法建立基线并实测，因为你的领域未必与论文的分类一致。

第四章 五层优化体系

第三章给的是"单个动作有多大效果"，这一章给的是"整件事该怎么组织"。

本研究院把 GEO 的工作拆成五层，从下到上依次是：**事实层、内容层、信号层、分发层、监测层**。这个顺序不是随意的——**下层不成立时，上层的投入会被浪费甚至产生反效果**。

表 4-1 GEO 五层优化体系总览

层级	要解决的问题	核心产出	典型周期	不做的后果
L5 监测层	有没有用、 下一步改什么	基线数据与 迭代记录	持续	无法证明效果， 预算难以为继
L4 分发层	AI 从哪里取信息， 我在不在那里	被引用源的 覆盖清单	3—12 个月	内容再好也 进不了取源范围
L3 信号层	引擎凭什么 认为我可信	E-E-A-T 信号、 结构化数据、 更新机制	1—3 个月	内容被读到 但不被采信
L2 内容层	内容能不能被 读懂和转述	可独立引用的 信息块	1—3 个月	被读到但 转述失真
L1 事实层	AI 对我的理解 准不准确	统一的 品牌事实库	2—6 周 (最先做)	曝光越多 伤害越大

为什么事实层必须最先做

设想一个场景：AI 在回答里推荐了你的产品，但把价格说错了、把不支持的功能说成支持了、把已停产的型号说成在售。

这种情况下，曝光量越大，损失越大——客户带着错误预期而来，转化率下降、客诉上升、口碑受损。更麻烦的是纠正成本：**错误信息一旦被多个来源复制，会在生成式引擎中形成自我强化的循环，纠正难度随时间显著上升。**

因此本研究院的建议是明确的：**在做任何曝光提升动作之前，先花 2—6 周把事实层做干净。**这不是流程洁癖，是风险控制。

4.1 L1 事实层：让 AI 先说对

目标：关于你的所有关键事实，在全网范围内一致、准确、可被机器读取。

要梳理的六类事实

表 4-2 品牌事实清单

类别	具体内容	常见问题
主体信息	公司全称、简称、英文名、 成立时间、总部地址、 所属集团关系	不同渠道的公司名不一致； 简称与全称无法被关联
业务定义	一句话业务描述、 服务范围、覆盖行业、 不做什么	"不做什么"几乎无人写， 导致 AI 把不相关需求 也推荐给你
产品事实	产品线、型号、 关键参数、价格区间、 在售状态	停产产品仍在网上流传； 参数在不同页面不一致
资质与背书	资质证书、认证、 获奖、合作方、 代表客户	散落各处未集中呈现
人物信息	创始人与核心专家的 姓名、职务、资历	人物是重要的实体锚点， 但多数企业只写职务不写资历
联系与渠道	官网、官方账号矩阵、 客服渠道	存在大量废弃或 非官方的同名账号

具体动作

1. **建立一份"品牌事实基准表"**，由市场与业务共同确认，作为全渠道口径的唯一来源。
2. **做一次全网一致性排查**：官网、各内容平台账号、百科类条目、行业目录、媒体报道、电商店铺、招聘页面——找出与基准表不一致的地方并逐一修正。
3. **统一实体命名**。明确品牌名的标准写法，以及所有需要被关联的变体（简称、英文名、常见错写）。中文场景下这一步尤其重要。
4. **处理"过时但仍在流传"的信息**。停产产品、旧价格、已离职的对外人物、旧地址——这些是错误答案最主要的来源。

常见坑

- **只改官网不改第三方**。引擎的取源范围远大于你的官网，第三方渠道的旧信息会继续被引用。
- **没有指定"事实负责人"**。事实库需要持续维护，无人负责则三个月后重新失效。

4.2 L2 内容层：让 AI 能读懂并转述

目标：内容以"可被独立引用的信息块"形式存在，而不是以"需要通读的长文"形式存在。

核心概念转变：GEO 时代的内容单位不是"页面"，而是"一段能自成答案的内容块"。

信息块的四个标准

1. **自成答案**。单独把这一段拿出来，能完整回答一个问题，不需要前后文。这对应第二章的"语义完整性"信号。
2. **问题即标题**。用用户真实会问的问句作为小标题，而不是营销式短语。

- 3. **结论前置。** 第一句就是答案，后面才是展开与依据。
- 4. **带出处与数据。** 对应第三章的方法一与方法二——效果最高的两个动作。

具体动作

- **从真实问题反推内容结构。** 收集销售、客服、社区里客户真实提问的原话，每个高频问题对应一个信息块。**注意是原话，不是关键词。**
- **为核心观点准备"可引述句"**——一句话说清、可被原样搬运的完整陈述（对应方法四）。
- **建立对比型内容。** 用户在 AI 里大量提出"A 和 B 哪个好""预算 X 该选什么"，而多数品牌官网从不提竞品——这导致在对比类查询中集体缺席。**在合规与客观的前提下提供横向对比，是当前性价比很高的空白位。**
- **改造存量内容：**删除关键词堆砌段落（方法九）、把长句拆短（方法三）、给结论补数据与出处。

常见坑

- **把长文切碎当成结构化。** 结构化的关键是每块自成答案，不是物理切分。
- **用"详见下文""如前所述"。** 这类表述直接破坏语义完整性，是引用的天敌。
- **只做官网内容不做问答型内容。** 用户的提问方式与官网的组织方式相差很远。

4.3 L3 信号层：让 AI 认为你可信

目标：提供引擎用于判断可信度的可机读信号。

表 4-3 信号层的四组动作

信号组	具体动作	优先级
结构化数据	部署 FAQ、HowTo、Product、Organization、Person 等 schema； FAQ schema 与 ChatGPT 引用权重正相关（有观测给出约+40%）	高 （一次性投入、长期生效）
E-E-A-T 信号	内容署名到具体的人并附资历； 标注发布与更新时间； 引用外部权威来源； 建立作者页面	高
新鲜度机制	建立定期更新节奏并保留更新记录； Perplexity 约 82%的引用来自 30 天内发布的内容	高 （对实时检索型平台尤其关键）
可抓取性	确认页面可被各 AI 爬虫访问； 检查 robots 规则； 避免关键内容依赖前端渲染导致无法读取	基础项 （不做则以上全部失效）
关于"要不要屏蔽 AI 爬虫" 这是一个需要企业明确决策的问题：允许 AI 爬虫抓取，内容可能被用于训练与生成而不带来点击；屏		

信号组	具体动作	优先级
蔽 AI 爬虫，则基本退出生成式引擎的可见范围。 本研究院的判断是：对绝大多数以获客为目的的企业，应当允许抓取。 理由是第一章 1.3 节所说的——在零点击成为常态的情况下，进入回答本身就是目标，而不是手段。 例外情形：以内容付费为主要商业模式的媒体与知识产品方，需要单独权衡。这类主体的合理做法通常是“开放摘要层、保护正文层”。		

4.4 L4 分发层：出现在 AI 取信息的地方

目标：在引擎实际取源的渠道中建立存在。

这一层是 GEO 与 SEO 差异最大的地方，也是最容易被忽略的一层。第二章的数据已经说明：**ChatGPT 的引用中 Wikipedia 约占 7.8%，Perplexity 的顶级来源中 Reddit 约占 46.7%**——这些都不是你的官网。

表 4-4 分发层的渠道地图

渠道类型	国际对应	国内对应	优先级判断
百科类	Wikipedia	各类百科条目	高。 ChatGPT 路径的关键； 但需严格遵守 各平台的收录规则
社区问答	Reddit、Quora、 Stack Exchange	知乎、小红书、 垂直论坛、 各行业社区	高。 Perplexity 与 AI Overviews 的 重要来源
评测与目录	G2、Capterra 等 第三方评测平台	行业目录、 垂直评测媒体、 协会名录	B 端场景高。 ChatGPT 依赖 第三方信号
专业社交	LinkedIn	行业媒体专栏、 专业社群	B 端场景中 AI Overviews 的 关键来源
多模态	YouTube	视频平台、 图文平台	AI Overviews 侧 显著加分 (有观测显示 图文视频结合 可大幅提升选中率)
权威背书	行业报告、 学术引用、 主流媒体	政务与协会信息源、 行业白皮书、 主流媒体报道	长期最有价值， 见效最慢

这一层的三条原则

- 1. **先做"别人说你"，再做"你说自己"。** 第三方来源的可信度权重高于自有渠道。一篇被行业媒体引用的数据，价值高于十篇官网软文。
- 2. **遵守各平台规则，不做批量灌注。** 百科类条目、社区问答都有明确的内容规范与反营销机制，违规内容不仅会被删除，还可能给品牌带来负面记录。**这一点见第七章。**
- 3. **产出"值得被引用的东西"，而不是"想被引用的东西"。** 最有效的分发层资产是：自有调研数据、行业方法论、可验证的案例——它们会被别人主动引用，这比自己四处发布有效得多。

本研究院的一个观察

在服务企业的过程中我们反复看到一个现象：**投入产出比最高的分发层动作，往往是把企业内部已有的东西公开出来。**

很多企业手里有真实的项目数据、有行业经验总结、有客户问题清单——这些东西只存在于内部 PPT 里。把它们整理成一份公开的、有数据有出处的行业材料，其被引用的概率远高于任何一篇为了曝光而写的稿子。

这也是"做白皮书"这件事在 GEO 时代价值上升的原因——它同时满足了方法一（引文）、方法二（数据）、方法七（权威性）与分发层的权威背书要求。

4.5 L5 监测层：知道有没有用

目标： 建立可持续观测的基线与迭代机制。

这一层的内容较多，且是 GEO 落地中最普遍的短板，因此本白皮书用**第五章整章**展开。这里只给出一条原则：

在做任何优化动作之前，先建立基线。 没有基线，三个月后你无法回答"到底有没有变好"这个问题——而这个问题回答不了，预算就续不上。

4.6 五层的推进节奏

表 4-5 90 天启动节奏建议

阶段	时间	重点层级	关键动作	完成标志
第一阶段 摸底	第 1—2 周	L5 监测层	手工建立基线 (见第五章 5.3 节)： 30 个真实问题×3 个平台×3 次	一张有数字的 基线表
第二阶段 清账	第 3—6 周	L1 事实层	梳理品牌事实基准表； 全网一致性排查与修正	关键事实 全渠道一致
第三阶段 改造	第 7—12 周	L2 内容层 + L3 信号层	存量内容改造 (去关键词堆砌、 补数据与出处)； 部署 schema； 建立更新机制	核心页面完成改造； 结构化数据上线

阶段	时间	重点层级	关键动作	完成标志
第四阶段 扩散	第 13 周 起	L4 分发层	产出可被引用的 原创材料； 按平台建设第三方存在	出现 非自有渠道 的引用
持续	每月	L5 监测层	重测基线问题； 记录变化与动作对应关系	一条可画出来的 趋势线
关于节奏的两条提醒 第一，不要跳过第一阶段。 摸底只需要一到两周、几乎零成本，但它决定了后面所有工作能不能被证明有效。跳过它是本领域最常见也最昂贵的错误。 第二，L4 分发层见效最慢但价值最高。 前三个阶段的动作大多在 1—3 个月内可见变化，而权威来源的建设通常需要 6—12 个月。 如果只做见效快的部分，天花板会很低。 建议在第三阶段就并行启动 L4 的内容生产。				

第五章 度量体系：怎么证明有用

这一章解决 GEO 落地中最大的实际障碍。

本研究院在与企业交流中反复听到同一句话：“我们做了一些，但说不清有没有用。”这不是执行问题，是度量问题——而度量问题不解决，GEO 项目通常活不过两个季度。

5.1 为什么 SEO 的指标在这里失效

表 5-1 三个失效的旧指标

旧指标	为什么失效	替代方向
排名	生成式回答里不存在“第几位”； 同一问题两次提问的答案可能不同	改用提及率与 首位推荐率 (统计意义上的位置)
点击率 / 自然流量	零点击是常态； 大量影响发生在用户不点击的情况下	改用引用份额 + 分离出的 AI 引荐流量
关键词覆盖	用户向 AI 提的是完整问题， 不是关键词； 且关键词堆砌为负效果	改用真实问题集覆盖率

还有一个技术性差异必须强调：**生成式回答具有随机性**。同一个问题问三次，可能得到三个不完全相同的答案。这意味着**任何基于单次观察的结论都不可靠**——GEO 的度量必须是统计性的。

5.2 五个核心指标

表 5-2 GEO 五指标框架

指标	定义	怎么算	为什么重要
① 提及率 Mention Rate	在问题集的回答中， 品牌被提到的比例	被提及的回答数 ÷ 总回答数 (每问题问 3 次取平均)	最基础的可见度指标。 零点击场景下， 被提及即已产生价值
② 首位推荐率 Top-Pick Rate	在推荐类问题中， 品牌被列在第一位 或被单独推荐的比例	首位提及数 ÷ 推荐类问题回答数	商业价值最高的指标。 用户对第一个 推荐的注意力显著更高
③ 引用份额 Citation Share	在 AI 给出的引用来源 中， 你的域名或你被引用的 第三方来源所占比例	你的来源出现次数 ÷ 全部引用来源次数	反映 分发层 的成效； 可拆分为自有渠道 与第三方渠道

指标	定义	怎么算	为什么重要
④ 答案正确率 Answer Accuracy	回答中关于你的事实性描述正确的比例	正确描述数 ÷ 全部涉及描述数 (按事实基准表核对)	风险指标，也是事实层的验收标准。 应设为不低于某个下限
⑤ AI 引荐流量 AI Referral	可归因于 AI 平台的站点访问与转化	流量分析工具中按来源筛选	唯一能进财务口径的指标， 但会显著低估真实影响 (见 5.4 节)
五个指标的权重建议 本研究院建议的考核权重是： 提及率 25%、首位推荐率 30%、引用份额 20%、答案正确率 15%、AI 引荐流量 10%。 两点说明： 其一，首位推荐率权重最高。 它最接近商业结果——被提到和被首推的转化价值差距很大。 其二，AI 引荐流量权重最低，但不能不设。 它是唯一能被财务认可的口径，必须跟踪；但把它当主指标会严重低估 GEO 的价值，因为大量影响以"用户看完 AI 回答后直接搜品牌名"的方式发生，归因上会记在别的渠道（见 5.4 节）。			

5.3 怎么建基线：一套零成本的手工方法

市面上已有若干 GEO 监测工具，但本研究院建议**第一次基线必须手工做一遍**——因为这个过程本身会让团队理解"AI 是怎么谈论我们的"，这个认知比数字更重要。

第一步：构建问题集（半天）

1. **收集 30 个真实问题。**来源优先级：客服与销售记录里客户的原话 > 社区里的真实提问 > 团队推测。**必须是完整问句，不是关键词。**
2. **按四类均衡分布**（见下表），避免全是品牌词问题——那会得到虚高的结果。

表 5-3 问题集的四种类型

类型	示例句式	建议占比	考察什么
品类型	"XX 类产品有哪些推荐"	30%	有没有进入候选集
场景型	"预算 X、用于 Y 场景，选什么"	30%	商业价值最高—— 最接近真实购买决策
对比型	"A 和 B 哪个更适合 Z"	25%	在竞争性问题中的表现
品牌型	"XX 这家公司怎么样"	15%	答案正确率的主要考察场

第二步：执行测试（一天）

1. **每个问题在每个目标平台各问 3 次**，分别记录。3 次是为了应对随机性——这是最低要求。
2. **用干净环境提问：**退出登录或使用无痕，避免个性化历史影响结果。

3. **逐条记录五项内容：**是否提及你、位置（第几个被提到）、是否首推、对你的描述是否准确、回答引用了哪些来源。

第三步：整理成基线表（半天）

表 5-4 基线记录表模板（每行一个问题×平台）

字段	记录内容
问题编号 / 问题原文	完整问句
问题类型	品类型 / 场景型 / 对比型 / 品牌型
平台	ChatGPT / 豆包 / DeepSeek / 夸克 / ...
提及次数（3 次中）	0—3
首位提及次数（3 次中）	0—3
被提及的竞品	列出全部， 这是竞争基线
描述准确性	准确 / 部分错误 / 严重错误（附错误内容）
引用来源	列出全部引用域名
测试日期	YYYY-MM-DD

这张表一做完，你会立刻得到四个此前不知道的答案

其一，你的真实提及率是多少——多数企业第一次测完的结果远低于预期。

其二，谁是 AI 眼中的竞品——这个名单往往与你自己认定的竞品清单不一致，这是极有价值的信息。

其三，AI 在哪些地方说错了你——直接构成事实层（L1）的待办清单。

其四，AI 在从哪里取信息——直接构成分发层（L4）的目标渠道清单。

换句话说：一次基线测试，同时产出了后续三层工作的任务清单。这是本研究院坚持“先测后做”的根本原因。

第四步：建立复测节奏

- **每月复测同一问题集**，保持问题、次数、环境一致——变量只能有一个，就是你做的优化动作。
- **每季度扩充问题集**（新增不超过 20%），并单独标记新增问题，避免影响趋势可比性。
- **记录每次优化动作的时间点**，与指标变化对齐——这是唯一能建立因果判断的方式。

5.4 归因难题：必须向管理层说清楚的一件事

这是 GEO 汇报中最容易出问题的地方，本研究院建议在项目立项时就说明。

GEO 的大部分商业影响，在流量归因中是看不见的。

机制是这样的：用户向 AI 提问“XX 类产品推荐”，AI 的回答中提到了你的品牌；用户**没有点击**，而是记住了名字，随后在电商平台搜索你的品牌名、或者直接访问你的官网、或者在下次采购时把你列入询价名单。

这条链路上，最后一步会被归因为“品牌搜索”或“直接访问”，而它的真实起点在那段 AI 回答里。

表 5-5 归因缺口的三种表现与应对

表现	为什么发生	应对方式
AI 引荐流量占比很低 (常见为个位数)	零点击是常态	不要把它当主指标。 它是“下限证据”， 不是“全部效果”
品牌搜索量上升， 但说不清来源	AI 回答带来的 心智占位在别处兑现	跟踪品牌词搜索量 与提及率的相关性； 两者同步上升是有力旁证
销售端反馈 “客户说是 AI 推荐的”	真实链路的 唯一直接证据	在销售与客服的 来源问卷中 增加“AI 助手推荐”选项—— 这是成本最低、 说服力最强的一招
给汇报者的一句实话 不要试图用单一数字证明 GEO 的 ROI。在当前的归因技术条件下做不到，勉强算出来的数字经不起追问。 更稳妥的汇报结构是三层：“提及率与首位推荐率的变化（可精确测量）+ 品牌搜索量的同步变化（相关性证据）+ 销售端的来源反馈（直接证据）”。 三者同向变化时，结论是可靠的；只有其中一项变化时，应当承认不确定性——这比给一个精确但站不住的 ROI 数字更能建立长期信任。		

5.5 什么时候该停

与本研究院其他白皮书一致，这里也给出停止条件。

表 5-6 三个月后的判断

情况	判断	建议动作
提及率与首位推荐率 均有可观测上升	继续并加码	把资源向 L4 分发层倾斜， 争取更高的天花板
答案正确率上升， 但提及率无变化	继续	事实层的价值已经兑现 (风险下降)； 曝光提升需要 L4 发力， 周期更长
各项指标均无变化， 且已执行满三个月	先查执行，再查 方向	九成情况是执行没到位： 检查是否只做了 L2 而没做 L1 和 L4； 确认后仍无变化再调整方向

情况	判断	建议动作
提及率上升但 答案正确率下降	立即暂停曝光动作	这是最危险的组合—— 错误信息正在被放大。 回到 L1 事实层

第六章 落地案例

这一章需要一个前置说明：**GEO 是一个新领域，公开的、可核验的完整案例目前很少。**大量流传的“效果案例”来自服务商的营销材料，缺乏第三方验证。

因此本章采取如下处理：**分层呈现，标明证据等级**——学术实验证据、平台观测数据、服务商披露案例分开列出，读者可自行判断可信度。

阅读提示

证据等级从高到低：①学术实验（可复现）→②第三方平台观测（有统计窗口）→③服务商披露（未经独立验证）。

本研究院建议：**用①和②做判断，用③做参考。**在向管理层汇报时，引用③类数据应明确标注其性质。

6.1 一级证据：学术实验

案例 6.1.1 GEO-bench：九种方法的效果验证

证据等级：学术实验（KDD 2024） | 可复现性：高 | 核心结论：GEO 可将可见性提升最高约 40%，且关键词堆砌为负效果

研究设计

Aggarwal 等构建了 GEO-bench 基准——覆盖多领域用户查询与相应网页来源的大规模数据集——并以黑盒优化框架系统测试九种方法对生成式引擎回答可见性的影响，采用位置加权与主观印象两类可见性指标。

关键结果

- 整体：GEO 可将可见性提升**最高约 40%**
- 效果最好：添加引文约 **+40.9%**、补充统计数据约 **+34.2%**、流畅度优化约 **+28.0%**
- 唯一负效果：关键词堆砌约 **-7.8%**
- **效果高度依赖领域**——同一方法在不同领域差异显著（详见第三章表 3-2）

可复用要点

- **这是目前唯一可复现的 GEO 效果证据，应作为方法选择的第一依据。**
- “效果因领域而异”这一结论，直接否定了任何“通用最优方案”的说法。

6.2 二级证据：平台观测

案例 6.2.1 平台份额与引用率的一年变化

证据等级：第三方流量观测（Similarweb，2025.06—2026.05） | 用途：判断资源如何在平台间分配

关键观测

- **平台份额剧变**：ChatGPT 全球网页流量份额从约 **76%**降至约 **53%**；Gemini 从不足 9%升至约 **27—28%**；Claude 从约 2%升至近 **9%**（比例增速最快）。
- **引用率整体上升**：ChatGPT 引用率从 **1.6%**（2025 年 6 月）升至 **6.8%**（2026 年 5 月）。
- **品类差异极大**：旅游与酒店约 **23%**、汽车约 **20%**，而专业服务不足 **4%**。观测指出，"**选项可验证、可比较**"的品类引用率更高。
- **行为变化**：ChatGPT 界面调整后，主页型引荐占全部引荐流量的比例从约 26—29%升至约 **62—63%**，反映用户越来越把回答中的品牌名当作"研究终点"。
- **广告出现**：约 **26%**的 ChatGPT 回答中已包含广告。

对企业的三条含义

1. **不要只优化一个平台**。一年内头部份额从 76%到 53%，单平台策略的风险敞口很大。
2. **先判断你所在品类的引用率水平**。专业服务类不足 4%意味着：即使做到品类前列，绝对曝光量也有限——**投入规模应与之匹配，不对标旅游品类的预期**。
3. **"用户把品牌名当研究终点"是关键信号**。它印证了第一章的判断：被提及本身就是价值，而不只是引流手段。

案例 6.2.2 跨平台引用重叠仅 11%

证据等级：第三方引用审计 | 核心发现：**ChatGPT 与 Perplexity 的引用域名重叠率仅约 11%**，引用率相差可达数十倍

观测内容

- **重叠率约 11%**——两个平台几乎在引用不同的内容源。
- **引用率差异**：有审计给出 ChatGPT 约 0.59% 对 Perplexity 约 13.05%。
- **来源偏好差异**：ChatGPT 引用中 Wikipedia 约占 7.8%；Perplexity 顶级来源中 Reddit 约占 46.7%，且约 82%的引用来自 30 天内发布的内容。

可复用要点

- **"做好内容自然会被各家引用"是错误假设**。必须按平台设计动作。
- **内容新鲜度对不同平台的价值天差地别**。对实时检索型平台，更新频率本身就是竞争力；对训练内嵌型平台，长期沉淀更重要。

6.3 三级证据：服务商与行业披露案例

以下案例来自行业分析文章与服务商披露，**未经独立验证**。本白皮书保留其定性描述，对无法交叉验证的具体百分比采取保守呈现。

案例 6.3.1 白酒品牌的首位推荐率优化

证据等级：服务商披露 | 行业：消费品（白酒） | 披露成效：优化前 **AI 首位推荐率不足 50%**，优化后核心产品线首位提及率显著提升，**AI 问答总提及次数增长多倍**

披露的动作路径

1. **品牌语义校准**——构建标准化知识库，统一产品参数与工艺标准表述
2. **关键词矩阵搭建**——覆盖选购、场景需求、品质认知等维度（注：此处的“关键词”指问题维度覆盖，非关键词堆砌）
3. **结构化内容输出**——生产问答化、场景化的 AI 友好型内容
4. **数据监测迭代**——监测提及率、排名与舆情并持续优化

本研究院的解读

- **这四步与本白皮书的五层体系高度对应**：①对应 L1 事实层、②③对应 L2 内容层、④对应 L5 监测层。**值得注意的是缺少 L4 分发层**——这可能解释了为什么效果被描述为“提及次数增长”而非“权威来源引用”。
- **“优化前首位推荐率不足 50%”这个表述值得推敲**。对一个头部品牌而言，50%的首位推荐率已经不低；这提示读者注意服务商案例中基线选取的问题。
- 行业侧披露的**见效周期为“2—4 周观察到提升、3 个月排名稳定”**——本研究院认为这个节奏偏乐观，且未区分不同层级的动作（L1 快、L4 慢）。

可复用要点

- **四步路径本身是合理的**，尤其是把“语义校准”放在第一步，与本白皮书“事实层优先”的判断一致。
- **但缺少分发层意味着天花板有限**。参见第四章 4.4 节。

案例 6.3.2 国内 GEO 服务市场的整体情况

证据等级：行业分析与咨询机构报告 | 用途：判断市场成熟度与选型风险

公开数据

- 市场规模：**2025 年约 6 亿元，2030 年预计突破 500 亿元**（艾瑞咨询）；另有行业分析称 2026 年已达“数十亿元人民币”量级、年增速超 100%。
- 企业采纳：有行业分析称**超 70% 企业已将 GEO 纳入年度营销预算**。
- 增长来源：**存量端来自传统 SEO 与 PR/舆情管理预算的转移；增量端来自知识库构建、网页优化、API 对接等技术服务支出**（艾瑞咨询）。

本研究院的提示

- **不同来源的市场规模数字差异较大**（6 亿 vs 数十亿），说明统计口径尚未统一——这本身就是市场早期的特征。
- **“超 70% 企业已纳入预算”这类数字应谨慎对待**，其调研样本与定义均不明确。
- **增长来源的判断很有价值**：如果增量主要来自“知识库构建、网页优化、API 对接”，说明 **GEO 的实质工作偏向技术与内容基建，而非营销投放**——这与本白皮书五层体系的构成一致。

6.4 横向观察：五条结论

表 6-1 从三级证据中得到的共性结论

结论	证据来源	可信度
引用权威来源与提供具体数据是效果最高的两个动作	学术实验（+40.9%/+34.2%） + 平台观测（+37%/+22%）	高 （两套独立研究交叉印证）
关键词堆砌应当停止	学术实验（-7.8%）	高
必须按平台分别设计	平台观测 （重叠率 11%、引用率差数十倍）	高
事实准确性应优先于曝光提升	本研究院判断 + 服务商案例中 "语义校准第一步"的一致做法	中高
品类决定天花板	平台观测 （旅游 23% vs 专业服务<4%）	高

第七章 六个陷阱与风险边界

这一章列出本研究院认为需要明确警示的六类问题。其中前三类是**做法上的陷阱**，后三类是**风险与边界**。

7.1 陷阱一：把 GEO 做成 SEO 换皮

症状：把关键词换成长尾问句，把标题堆得更密，其余照旧。

这是当前最普遍的问题。第三章的学术数据已经说明：**关键词堆砌在生成式引擎中是唯一的负效果方法（约-7.8%）**，而 SEO 时代不被重视的引文、数据、流畅表达才是主要杠杆。

怎么自检

- 打开你最近三个月产出的"GEO 内容"，问一个问题：**这段话里有几个具体数字？有几处标注了外部来源？**如果答案为零，那么做的还是 SEO。
- 检查内容单位：**是"页面"还是"可独立引用的信息块"？**如果每段都需要上下文才能读懂，语义完整性不成立。

7.2 陷阱二：批量灌注低质内容

症状：用 AI 批量生成大量内容，铺满各个平台，期待"总有一篇被引用"。

这个做法有三个问题：

1. **在语义层面无效。**生成式引擎要的是"能自成答案的高质量信息块"，不是数量。大量同质化内容不提升被引用概率。
2. **在平台层面违规。**百科类条目、社区问答平台都有明确的反营销与反垃圾机制，批量灌注会被识别、删除，并可能给品牌留下负面记录。
3. **在长期层面反噬。**如果这些内容确实进入了引擎的取源范围，它们承载的低质表述会成为 AI 描述你的素材——**你会得到大量关于自己的、平庸甚至错误的描述。**

本研究院的判断

GEO 是一个"少而精"占优的领域，这与 SEO 时代的"多而广"逻辑相反。

原因在机制上：搜索引擎返回十条链接，覆盖面有价值；生成式引擎只写一段话，它需要的是**最值得引用的那一个来源**，而不是一百个备选。

一份有真实数据、有明确出处、结构清晰的行业材料，价值高于一百篇批量生成的软文。

7.3 陷阱三：只做自有渠道

症状：把所有精力放在官网与自有账号上，不做第三方来源建设。

第二章的数据已经说明问题：**ChatGPT 引用中 Wikipedia 约占 7.8%，Perplexity 顶级来源中 Reddit 约占 46.7%**——这些都不是企业的自有渠道。

本研究院的表述是：“你说自己好”在生成式引擎中的权重，远低于“别人说你好”。这与 SEO 时代外链的逻辑相似，但要求更高——引擎看的不只是链接，还有来源本身的可信度类型。

优先顺序建议

1. 先产出值得被引用的东西（自有数据、方法论、可验证案例）
2. 再争取权威来源的收录与引用（行业媒体、协会、研究机构）
3. 同时维护社区与评测平台的真实存在（遵守平台规则，不做灌注）
4. 最后才是自有渠道的内容优化

7.4 风险一：错误信息的放大与固化

这是 GEO 领域最被低估的风险。

机制是这样的：AI 在回答中给出了关于你的错误信息（旧价格、停产产品、错误参数、不实描述）；这个错误被用户看到、被转述、被其他内容引用；新的内容成为引擎的取源之一，**错误被自我强化**。

表 7-1 错误信息的三个特征

特征	说明	应对
自我强化	错误被复制到更多来源后，在引擎侧的“共识度”反而上升	发现即纠正，不要拖延——纠正成本随时间显著上升
难以定位	你不知道 AI 说错了什么，除非主动去问	把“答案正确率”列为常规监测指标（见第五章指标④）
曝光放大伤害	在事实层未清理前提升曝光，等于放大错误	事实层优先于曝光层（见第四章五层顺序）

本研究院的强制建议：把“答案正确率”设为不可妥协的底线指标。当该指标下降时，应暂停所有曝光提升动作，回到事实层。

7.5 风险二：平台规则变化与广告位挤压

两个正在发生的变化需要纳入规划：

- 其一，规则不透明且会变。生成式引擎的取源逻辑由平台自行决定、不完整公开，且随版本更新变化。任何基于当前观测建立的策略都应假设它会失效，并保持监测。
- 其二，广告位正在进入。观测显示约 **26%** 的 ChatGPT 回答中已包含广告，且集中在对话早期。传统搜索的历史表明，自然位置会被付费位置逐步挤压。

对投入时机的判断

本研究院认为这两个变化指向同一个结论：**当前是自然可见度成本最低的窗口期，且这个窗口正在收窄。**

这不是制造紧迫感——而是一个可以观察验证的判断：**跟踪你所在品类中 AI 回答里广告位的出现频率**，如果它在上升，说明自然位置的竞争成本正在上升。

7.6 风险三：服务商选型与效果承诺

国内 GEO 服务市场处于早期，本研究院建议在选型时警惕以下六类表述：

表 7-2 六类需要警惕的承诺

承诺	为什么有问题	应该怎么问
"保证在所有 AI 平台提升曝光"	跨平台引用重叠率仅约 11%， 机制差异极大	"你们针对哪几个平台？ 各自的动作有什么不同？"
"保证首位推荐"	生成式回答有随机性， 且平台规则不透明， 无人能保证	"你们如何测量首位推荐率？ 测多少次取平均？"
"X 周内见效"	不同层级动作周期差异极大 (事实层快、分发层慢)	"这个周期对应 哪一层的动作？"
"我们有独家渠道/技术"	生成式引擎不存在 可被购买的排名位	"具体是什么？ 能否说明其合规性？"
只报"提及次数增长 X 倍"	基数未知时倍数无意义； 且未区分正面与负面提及	"基线是多少？ 提及中正面描述占比多少？"
不提供基线测试	没有基线就无法证明效果—— 这是最大的红旗	"能否先做一次 基线测试再谈方案？"

给企业的一条选型原则

优先选择愿意先做基线、并把基线数据交给你的服务商。

理由很简单：基线数据是客观的，它同时也是对服务商自己的约束。愿意先测的服务商，说明它准备用数据说话；不愿意先测的，通常是因为**效果不可证伪对它更有利**。

而且，第五章 5.3 节已经说明——**基线测试你自己就能做，成本是一到两天的人力**。做完之后再谈方案，谈判位置完全不同。

第八章 分角色行动建议

8.1 给市场与品牌负责人

1. **第一件事是做基线，不是做内容。** 一到两天的人力投入，同时产出现状认知、竞品名单、错误清单与渠道清单（见第五章 5.3 节）。
2. **把"答案正确率"设为底线指标。** 在事实层清理干净之前，不要启动曝光提升——曝光会放大错误。
3. **汇报时用三层结构，不要给单一 ROI 数字。** 提及率与首位推荐率变化（可测）+ 品牌搜索量同步变化（相关性）+ 销售端来源反馈（直接证据）。
4. **先选 1—2 个平台做深。** 跨平台重叠率仅约 11%，"全平台覆盖"在动作上不成立。
5. **先判断你所在品类的引用率量级。** 旅游约 23%、专业服务不足 4%——投入规模应与品类天花板匹配。
6. **在销售与客服的来源问卷中加一个选项："AI 助手推荐"。** 成本最低、说服力最强的归因证据。

8.2 给内容与运营团队

1. **停止关键词堆砌。** 学术实验显示这是唯一负效果的方法（约-7.8%）。
2. **每个结论后面补两样东西：一个具体数字，一个外部出处。** 这两个动作对应效果最高的两种方法（+40.9%、+34.2%）。
3. **把内容单位从"页面"改成"能自成答案的信息块"。** 不写"详见下文""如前所述"。
4. **从客户原话反推内容，不从关键词反推。** 收集销售与客服记录里的真实提问。
5. **补上对比型内容。** 用户大量提出"A 和 B 哪个好"，而多数品牌从不提竞品——这是当前性价比很高的空白位。
6. **为核心观点准备"可引述句"——一句话说清、可被原样搬进 AI 回答的完整陈述。**

8.3 给技术负责人

1. **先确认可抓取性。** 检查 AI 爬虫的访问权限、robots 规则、关键内容是否依赖前端渲染——这一项不通，上面全部失效。
2. **部署结构化数据。** FAQ、HowTo、Product、Organization、Person schema。一次性投入、长期生效。
3. **建立更新时间戳机制。** 发布时间与更新时间应可被机器读取，对实时检索型平台尤其关键。
4. **支持监测的数据采集。** 在流量分析中单独标记 AI 平台来源；把品牌词搜索量纳入常规看板。
5. **建立品牌事实的单一数据源。** 让官网、小程序、各平台账号的关键事实来自同一处，避免人工同步失效。

8.4 给管理层

1. **把 GEO 理解为基础建设，不是投放。** 增长来源分析显示，其实质支出在知识库构建、网页优化、API 对接——这是技术与内容基建，不是媒介采买。
2. **接受归因不完整这个前提。** 要求精确 ROI 会逼出经不起追问的数字。用三层证据结构判断。

-
3. **给足周期。**事实层 2—6 周、内容与信号层 1—3 个月、分发层 6—12 个月。只做见效快的部分，天花板会很低。
 4. **明确一个决策：是否允许 AI 爬虫抓取。**对多数以获客为目的的企业应当允许；内容付费型主体需单独权衡。
 5. **把这件事的负责人放在能协调技术、内容、公关三方的位置上。**五层体系跨了三个职能，挂在任一单一部门下都会缺层。

第九章 趋势展望（2026—2030）

本章判断属于趋势推演而非预测，每条附观察指标。

9.1 自然可见度的成本会持续上升

约 26% 的 ChatGPT 回答中已出现广告，且集中在对话早期。传统搜索的历史表明自然位置会被付费位置逐步挤压。本研究院判断这一进程会在生成式引擎中重演，且速度可能更快——因为回答只有一段，位置更稀缺。

观察指标： 你所在品类的 AI 回答中，广告标识出现的频率变化。

9.2 度量会从“手工”走向“标准化”，但标准会滞后

当前 GEO 的度量高度依赖手工测试与各家自建口径。随着市场规模扩大，监测工具会成熟，但统一的行业指标定义预计仍将滞后 2—3 年——因为平台不提供官方数据，任何指标都是外部推断。

这对企业的含义： 短期内不要等待“行业标准”，先用第五章的五指标框架建立自己的口径并保持一致性。内部可比比外部可比更重要。

观察指标： 是否出现被多家头部企业共同采纳的 GEO 指标定义或第三方审计标准。

9.3 “被引用源”本身会成为一个竞争市场

既然引擎高度依赖少数几类来源（百科、社区、评测平台、权威媒体），这些来源的位置会成为稀缺资源。本研究院判断：围绕“如何进入被引用源”会形成一个新的服务市场，同时也会出现相应的操纵与反操纵博弈。

这对企业的含义： 早期通过正当方式建立的权威来源存在，其价值会随竞争加剧而上升；而通过灌注取得的位置，会随平台治理加强而归零。这是一个“早做且做对”收益最大的领域。

观察指标： 主要社区与百科平台针对营销内容的治理政策变化。

9.4 多模态与实时性权重上升

观测已显示图文视频结合的内容在 AI Overviews 侧被选中概率显著更高，而 Perplexity 约 82% 的引用来自 30 天内发布的内容。本研究院判断，随着引擎处理多模态与实时信息的能力增强，这两个维度的权重会继续上升。

这对企业的含义： 内容生产的形态要求会提高——只产出文字的组织会逐步失去优势。

观察指标： 你的品类中，AI 回答引用视频与图文来源的比例变化。

9.5 国内会形成与国际不同的生态路径

国际生成式引擎高度依赖 Wikipedia、Reddit 等开放生态；国内的对应生态是百科类条目、知乎、小红书、垂直社区与行业媒体，且部分平台与自有内容生态强耦合。

本研究院判断：**国内 GEO 的分发层动作会与国际显著分化**，直接照搬国际方法论的有效性有限。同时，**中文实体识别（品牌简称、别称、英文名、错写变体）会成为国内特有的技术难点**。

观察指标：国内是否出现针对本土平台取源机制的系统性第三方研究。

三个需要警惕的反向风险

一是过度投入于会失效的技巧。 平台规则不透明且会变，任何基于当前观测的具体技巧都可能失效。把资源投向“让真实信息更准确、更完整、更易被理解”这类不会失效的基础工作，风险最低。

二是操纵性做法的反噬。 生成式引擎的核心竞争力是答案可信度，平台有极强动机治理操纵行为。通过灌注与虚假内容取得的可见度，是负债不是资产。

三是把 GEO 当作可以替代产品与服务本身。 AI 回答会越来越多地整合真实用户评价与第三方评测——产品不行的情况下，更高的可见度只会更快地暴露问题。

附录 A 参考信息源清单

A.1 A 级：学术研究与国家统计

编号	来源	引用的要点
A1	Aggarwal, Murahari, Rajpurohit, Kalyan, Narasimhan, Deshpande: 《GEO: Generative Engine Optimization》, KDD 2024 (arXiv:2311.09735)	构建 GEO-bench 基准; GEO 可提升可见性最高约 40%; 九种方法的效果排序: 添加引文约+40.9%、统计数据约+34.2%、流畅度约+28.0%、引述来源约+27.5%、技术术语约+17.6%、易理解性约+14.0%、权威性约+10.4%、独特词汇约+6.2%、 关键词堆砌约-7.8% ; 方法效果因领域而异
A2	中国互联网络信息中心统计报告（2026）	我国生成式人工智能用户规模 6.02 亿人 , 普及率 42.8%

A.2 B 级：第三方数据机构与平台观测

编号	来源	引用的要点
B1	艾瑞咨询《2026 年 GEO 生成式引擎优化行业研究报告》	国内 GEO 市场 2025 年约 6 亿元、2030 年预计破 500 亿元; 超 41%用户几乎完全转向 AI 搜索、47.8%大幅增加使用频率、仅 10.6%以传统搜索为首选; 超 80%用户购买前用 AI 搜索辅助决策; 增长来源: 存量端为 SEO 与 PR 预算转移, 增量端为知识库构建、网页优化、API 对接
B2	Similarweb AI Search 《AI Search Stats 2026》(统计窗口 2025.06—2026.05)	ChatGPT 全球网页流量份额 76%→53%, Gemini <9%→27—28%, Claude 2%→近 9%; ChatGPT 引用率 1.6%→6.8%; 品类差异: 旅游约 23%、汽车约 20%、专业服务<4%; 全球 AI 平台月均访问约 95 亿次、同比+70%; 约 26%的 ChatGPT 回答含广告; 主页型引荐占比 26—29%→62—63%

编号	来源	引用的要点
B3	Leapd 《How ChatGPT, Google AI Overviews, and Perplexity Source Information in 2026》	三平台取源机制差异； ChatGPT 引用中 Wikipedia 约 7.8%； AI Overviews 中 Reddit 约 2.2%、 LinkedIn 为 B 端关键源、 多模态可显著提升选中率、 语义完整性相关系数约 0.87； Perplexity 顶级来源中 Reddit 约 46.7%、 约 82%引用来自 30 天内内容、 单条回答平均引用约 21.87 条； FAQ schema 与 ChatGPT 引用权重 正相关约+40%； 统计数据约+22%、直接引述约+37%
B4	第三方引用审计（2026）	ChatGPT 与 Perplexity 引用域名 重叠率仅约 11%； 引用率约 0.59% 对 13.05%

A.3 C 级：行业分析与服务商披露

以下来源未经独立验证，本白皮书仅作参考性引用，具体百分比采取保守呈现。

编号	来源	引用的要点
C1	行业分析文章 《2026 年 GEO 生成式引擎优化全解析》	GEO 四步实施路径 （品牌语义校准、关键词矩阵、 结构化内容、数据监测）； 白酒品牌案例：优化前 AI 首位推荐率 不足 50%，优化后核心产品线 首位提及率显著提升； 见效周期 2—4 周观察、3 个月稳定； 国内市场"数十亿元"量级、 年增速超 100%、 超 70%企业纳入年度营销预算
C2	国内 GEO 服务商测评类文章 （火山引擎开发者社区等）	国内 GEO 服务市场的供给结构与选型维度

说明：本白皮书对三级证据分层呈现，读者在对外引用时应保留其证据等级标注。学术与平台观测数据具有较高可信度；服务商披露数据仅供参考，不应作为效果预期依据。

附录 B GEO 自查清单（32 项）

按第四章五层体系编制。建议每季度自查一次，勾选项低于 20 项时不建议启动大规模内容投入。

B.1 L1 事实层（8 项）

序号	检查项	勾选
1	已建立书面的"品牌事实基准表"并指定维护责任人	☐
2	公司全称、简称、英文名在全渠道一致	☐
3	业务描述中明确写了"不做什么"	☐
4	在售产品线、型号、关键参数全渠道一致	☐
5	已清理停产产品、旧价格、旧地址等过时信息	☐
6	核心专家/创始人的姓名、职务、资历已公开且一致	☐
7	已排查并处理非官方的同名账号与废弃渠道	☐
8	已明确品牌名的标准写法与需被关联的变体	☐

B.2 L2 内容层（8 项）

序号	检查项	勾选
9	内容以"可独立引用的信息块"组织，而非长文	☐
10	小标题使用用户真实提问的问句形式	☐
11	每个信息块结论前置	☐
12	关键结论后附具体数字	☐
13	关键结论后附外部来源出处	☐
14	已清除关键词堆砌段落	☐
15	已为核心观点准备"可引述句"	☐
16	已产出对比型内容（在合规客观前提下涉及竞品）	☐

B.3 L3 信号层（6 项）

序号	检查项	勾选
17	已确认 AI 爬虫可访问关键页面（robots 与渲染方式）	☐
18	已部署 FAQ schema	☐
19	已部署 Organization / Product / Person 等 schema	☐

序号	检查项	勾选
20	内容署名到具体的人并附资历	☐
21	页面标注发布时间与更新时间	☐
22	已建立定期更新节奏并保留更新记录	☐

B.4 L4 分发层（5项）

序号	检查项	勾选
23	已产出至少一份含自有数据的可被引用材料	☐
24	百科类条目信息准确且符合平台规范	☐
25	在目标客户活跃的社区/问答平台有真实存在	☐
26	已进入相关行业目录、评测平台或协会名录	☐
27	有多模态内容（视频/图文）覆盖核心问题	☐

B.5 L5 监测层（5项）

序号	检查项	勾选
28	已完成首次基线测试（30 问题×3 平台×3 次）	☐
29	已建立五指标记录表并按月复测	☐
30	已记录每次优化动作的时间点以便归因	☐
31	销售/客服来源问卷中已加入"AI 助手推荐"选项	☐
32	已设定"答案正确率"的底线值与触发暂停机制	☐

附录 C 基线测试记录表模板

配合第五章 5.3 节使用。建议用表格软件维护，每月复测一次。

C.1 测试执行规范

1. 问题集 30 个，按品类型 30%、场景型 30%、对比型 25%、品牌型 15% 配比
2. 每个问题在每个平台各问 **3 次**，分别记录（应对生成随机性）
3. 使用退出登录或无痕环境，避免个性化历史影响
4. 保持问题原文、提问次数、环境三者在各期之间完全一致

C.2 记录字段

字段	填写说明	示例
问题编号	Q01—Q30，跨期保持不变	Q07
问题原文	完整问句，不可修改	"预算 50 万的中小制造企业做 AI 质检该怎么起步"
问题类型	品类型/场景型/对比型/品牌型	场景型
平台	本期测试的平台名称	豆包
提及次数	3 次中被提及的次数	2
首位提及次数	3 次中被列为第一个推荐的次数	1
被提及的竞品	列出全部 ，这是竞争基线	A 公司、B 公司、C 公司
描述准确性	准确 / 部分错误 / 严重错误	部分错误
错误内容	如有错误，原样记录	"称我方提供 X 服务"（实际不提供）
引用来源	回答中标注的全部来源域名	wikipedia.org、zhihu.com、…
测试日期	YYYY-MM-DD	2026-08-15
备注	异常情况说明	第 3 次提问未给出推荐

C.3 月度汇总表

指标	计算方式	本期	上期	变化
① 提及率	总提及次数 ÷ (问题数 × 3)	—	—	—
② 首位推荐率	首位提及次数 ÷ (推荐类问题数 × 3)	—	—	—
③ 引用份额	我方相关来源次数 ÷ 全部来源次数	—	—	—
④ 答案正确率	准确描述数 ÷ 全部涉及描述数	—	—	—
⑤ AI 引荐流量	流量工具中 AI 来源访问量	—	—	—

指标	计算方式	本期	上期	变化
本期执行的优化动作	按五层标注（L1—L5）与完成时间			

附录 D 平台适配速查表

重要提示：以下为 2026 年年中的外部观测结果，非平台公布规则，可能随时变化。使用前建议按第五章方法自测验证。

平台	取源机制	偏好来源	优化重点	不适合
ChatGPT	训练数据为主 + 选择性检索	百科式权威 (Wikipedia 约 7.8%) ； 第三方评测与 评论平台	强化 Bing 侧可见性； 争取第三方评测收录； 完善百科条目准确性； 建设主题集群	指望靠高频更新 快速见效
Google AI Overviews	实时索引调用， 整合自然搜索	社区来源 (Reddit 约 2.2%) ； B 端场景中 LinkedIn； 多模态内容	投资图文视频结合； 保持页面定期更新； 优先保证语义完整性 (相关系数约 0.87)	纯文字、 长期不更新的内 容
Perplexity	每次查询 实时检索	社区来源占比极高 (Reddit 约 46.7%) ； 新鲜内容 (约 82%来自 30 天 内)	提高发布频率； 问题式标题结构； 建设社区与评测存在	低频更新策略
国内主流平台 (豆包、DeepSeek、 通义、夸克、 百度 AI 搜索、Kimi 等)	缺少系统性 第三方研究 (详见 2.4 节)	推测对应生态： 百科类条目、知乎、 小红书、垂直社区、 行业媒体、 政务与协会信息源	基础动作跨平台通用； 中文实体识别 (简称、别称、英文名、 错写变体) 是特有重点； 必须自测验证	直接照搬 国际方法论

D.1 跨平台通用的四类信号（资源有限时优先做这四项）

1. **语义完整性**——每个内容块能自成答案，不写"详见下文"
2. **结构化标记**——FAQ、Organization、Product、Person 等 schema
3. **E-E-A-T 信号**——署名到人并附资历、标注更新时间、引用权威外部来源
4. **原创数据与可引述句**——统计数据约+22%、直接引述约+37%

附录 E 术语表

术语	本白皮书中的含义
GEO 生成式引擎优化	Generative Engine Optimization。通过优化信息的可理解性、结构、可信度与分布，使品牌与内容在生成式 AI 回答中被正确理解、引用与推荐的方法体系。
五层优化体系	本研究院提出的 GEO 工作框架：事实层（L1）、内容层（L2）、信号层（L3）、分发层（L4）、监测层（L5）。下层不成立时，上层投入会被浪费。见第四章。
五指标框架	本研究院提出的 GEO 度量体系：提及率、首位推荐率、引用份额、答案正确率、AI 引荐流量。见第五章。
提及（Mention）	AI 在回答正文中提到品牌，不一定附带链接。 在零点击场景下，提及本身即产生商业价值。
引用（Citation）	AI 在回答中标注的信息来源，通常附链接。
首位推荐率	在推荐类问题中，品牌被列在第一位或被单独推荐的比例。 本白皮书认为这是商业价值最高的 GEO 指标。
答案正确率	回答中关于品牌的事实性描述正确的比例。 本白皮书建议将其设为不可妥协的底线指标。
零点击	用户在 AI 回答中获得所需信息、不再点击来源链接的现象。 GEO 面临的基本前提之一。
语义完整性	一段内容能自成答案、不依赖上下文的程度。 观测显示其与被引用概率高度相关（相关系数约 0.87）。
信息块	GEO 时代的内容单位——一段能独立回答一个问题的内容，取代 SEO 时代的“页面”。
可引述句	一句话说清、可被原样搬进 AI 回答的完整陈述。 直接引述可提升可见性约 37%。
E-E-A-T	Experience、Expertise、Authoritativeness、Trustworthiness，经验、专业性、权威性、可信度。生成式引擎判断内容可信度的重要信号组。
schema（结构化数据）	嵌入页面的机器可读标记，如 FAQ、HowTo、Product、Organization。 观测显示 FAQ schema 与 ChatGPT 引用权重正相关。
关键词堆砌	SEO 时代围绕目标关键词高密度重复的做法。 在生成式引擎中为负效果（学术实验约-7.8%），应当停止。
GEO-bench	KDD 2024 论文构建的 GEO 效果评测基准，覆盖多领域用户查询与相应网页来源，是目前唯一公开的可复现评测集。

关于求智启新人工智能研究院

我们是谁

求智启新人工智能研究院（QIUZHI QIXIN AI RESEARCH INSTITUTE）总部位于杭州浙江大学校友企业总部经济园，由**浙江大学管理学院实践教授郭峻峰与企业 AI 落地专家陈灵军**联合创办。

研究院汇聚高校教授、企业管理者、产品与项目专家、财务与人力资源顾问、供应链顾问、AI 绘画设计专家、医学教育专家、程序员提效与企业系统改造顾问等多领域实战力量，面向企业家、高管团队、业务部门、职能岗位与产业平台，提供从 AI 认知培训、岗位提效工作坊，到 Agent 数字员工、知识库、GEO 营销增长与企业 AI 转型陪跑的系统化赋能。

截至本白皮书编制时，研究院已服务 50 余家企业，交付 200 余场培训，累计授课学员 3.1 万余人次。研究院编著的《AI 智能体实操指南》由电子工业出版社出版，发行 1.2 万册，并被多所高校采购为教材。

GEO 在研究院服务体系中的位置

GEO 是研究院四阶段服务体系中的**第四阶段：长期增长**。这个位置安排本身反映了我们的判断——**GEO 不是一个可以孤立开展的营销动作，它依赖前三阶段打下的基础。**

表 F-1 四阶段服务体系与本白皮书的对应

阶段	服务产品	周期	与 GEO 的关系
第一阶段 认知扫盲	企业 AI 战略领导力专题课	2—3 小时	理解入口迁移与零点击的含义
第二阶段 单点提效	AI 效能革命工作坊	1—2 天	内容生产能力是 GEO 内容层的前提
第三阶段 整体转型	AI 智能化升级陪跑营	1—3 个月	知识库建设直接对应 GEO 事实层与内容层
第四阶段 长期增长	GEO 生成式搜索优化	持续	本白皮书全部内容

与 GEO 直接相关的服务

- **GEO 基线诊断**——按第五章方法完成首次基线测试，产出现状认知、竞品名单、错误清单与渠道清单
- **品牌事实库建设**——对应五层体系的 L1 事实层，是所有后续工作的前提
- **知识库搭建与内容改造**——对应 L2 内容层与 L3 信号层
- **GEO 持续优化陪跑**——按月复测、迭代与汇报支持，解决第五章讨论的度量与归因难题

研究院的白皮书序列

本白皮书是研究院公开研究序列的第四份，四者构成一个完整链条：

白皮书	回答的问题
《企业人工智能落地实践指南——行业白皮书（2026）》	做什么场景 ——18 个行业 82 个场景的前后成本对照与落地前提
《FDE 企业 AI 前端部署工程师白皮书（2026）》	谁来把它做成 ——岗位定义、能力模型与七步现场交付法
《少年 CEO：青少年 AI 创新与创业蓝皮书（2026）》	能力从哪里长出来 ——少年思维的结构性优势与 LEAP 方法论
《 GEO 生成式引擎优化白皮书（2026） 》	做成之后怎么被看见 ——平台机制、九种玩法、五层体系与度量

联系我们

项目	信息
官方网站	qzqxai.com (可查阅专家团队、服务体系、交付案例与代表客户的完整信息)
地址	杭州市浙江大学校友企业总部经济园 B2 幢 606
合作负责人	陈灵军
电话 / 微信	13148398839
合作方向	企业内训 · 岗位提效工作坊 · AI 转型陪跑营 · Agent 与知识库定制 · GEO 营销增长

关于本白皮书

GEO 是一个成型不足三年、且平台规则持续变化的领域。本白皮书提出的框架属于阶段性归纳，其中的具体数字更应被视为时点快照而非长期结论。我们建议读者把**方法框架**作为长期参考，把**具体数据**作为需要自行验证的起点。

我们欢迎来自一线的批评、补充与案例贡献——**尤其欢迎"做了但没效果"的经历**。这个领域的公开材料中营销成分较高，失败经验的稀缺程度远高于成功案例，而它们对读者的价值往往更大。相关反馈可通过官网 **qzqxai.com** 或上述联系方式与我们取得联系。

《GEO 生成式引擎优化白皮书（2026）》

编著：求智启新人工智能研究院

版本：2026 年 8 月 第一版

本白皮书所载信息截至 2026 年 8 月，仅供研究参考，不构成任何营销投放建议、服务商选型建议或效果承诺。文中引用的平台观测数据为外部统计推断而非平台公布规则，可能随版本更新而变化；服务商披露的案例数据未经独立验证。本白皮书不鼓励任何试图操纵生成式引擎输出的行为。