

# 2026

## 品牌 AI 竞争力报告

AI 时代的品牌占位与内容资产重组

## Introduction

## 引言

当全球营销语境从「搜索链接」全面跨越至「生成答案」，品牌正面临一场前所未有的可见度危机。在 AI 接管消费者决策流的今天，品牌若无法进入大模型的推荐腹地，便意味着在数字世界中彻底失语。

本报告由知乎研究院与中国信息通信研究院人工智能研究所联合发布，旨在为 AI 时代的品牌营销提供一套科学的度量衡，推出「品牌 AI 竞争力指数」，将 AI 推荐的「黑盒机制」转化为可监测的核心指标。这不仅是品牌进入 AI 决策池的「入场券」，更是衡量品牌数字资产价值的新标准。

为了验证 GEO（生成式引擎优化）的底层逻辑，本报告开展了深度的实证研究。通过对中美主流大模型的横向评测发现，AI 并非随机推荐，而是表现出对高质量、专业化语料的深度偏好——这证明了专业内容正是训练模型、抢占首位推荐的重要基石。此外，通过调研企业中高层营销人员，分析了相关从业者对 GEO 的真实期待与实操痛点，归纳出 AI 营销的行业共识。

2026 年是 GEO 从概念走向战略落地的分水岭。我们希望通过这份报告，不仅能帮助品牌主在 AI 浪潮中精准卡位，更能助力品牌构建起一套穿越技术周期的、可预测的 AI 影响力。

Research Design and Implementation Feedback

# 研究设计与执行回执

本报告主要进行了定量调研、桌面研究和实证研究三个研究方法：

## 定量研究

01

**样本数量：** 71

**样本画像：** 来自 家电、数码、快消、汽车等核心行业的营销中高层从业者

**调研核心：** 探究行业先行者对 GEO 的基本认知、战略排位及实操痛点

## 桌面研究

02

**研究方向：**

大模型检索增强技术（RAG）与智能体搜索（Agentic Search）的前沿学术论文与技术白皮书。

## 实证研究

03

**实证 A：中美模型信源审计实验**

**方案：** 选取 4 大消费命题，通过「AI 审计 AI」的方式，对中美大模型引用的 24 份原始语料进行 10 分制全维度拆解。

**实证 B：品牌资产「采纳与弃用」测试**

**方案：** 设置特定决策任务，对比专业KOL长测、短水文、AI生成三类语料在 Agent 审计下的引用概率。

contents

# 目录

## 第一章 / AI 时代的品牌可见度危机

01	◆ 从「链接搜索」到「答案生成」的跨越	05
02	◆ 行业现状：GEO 赛道的机遇、乱象与营销从业者认知	07
03	◆ AI 时代品牌可见度危机：消费链路的彻底颠覆	12

## 第二章 / AI 时代新标准：解码品牌 AI 竞争力指数

01	◆ 核心度量衡：品牌 AI 竞争力指数的构建逻辑	14
02	◆ 可见结果：衡量品牌进入 AI 决策池的「入场券」	17
03	◆ 排位结果：衡量品牌在 AI 建议中的「推荐优先级」	22
04	◆ 内容可信度：AI 推荐的「信任根基」	25

## 第三章 / 实证研究：高质量专业语料是 GEO 优化的底层基石

01	◆ 中外模型引用信源内容质量对比	28
02	◆ 品牌语料的引用概率分析	32

# 第一章

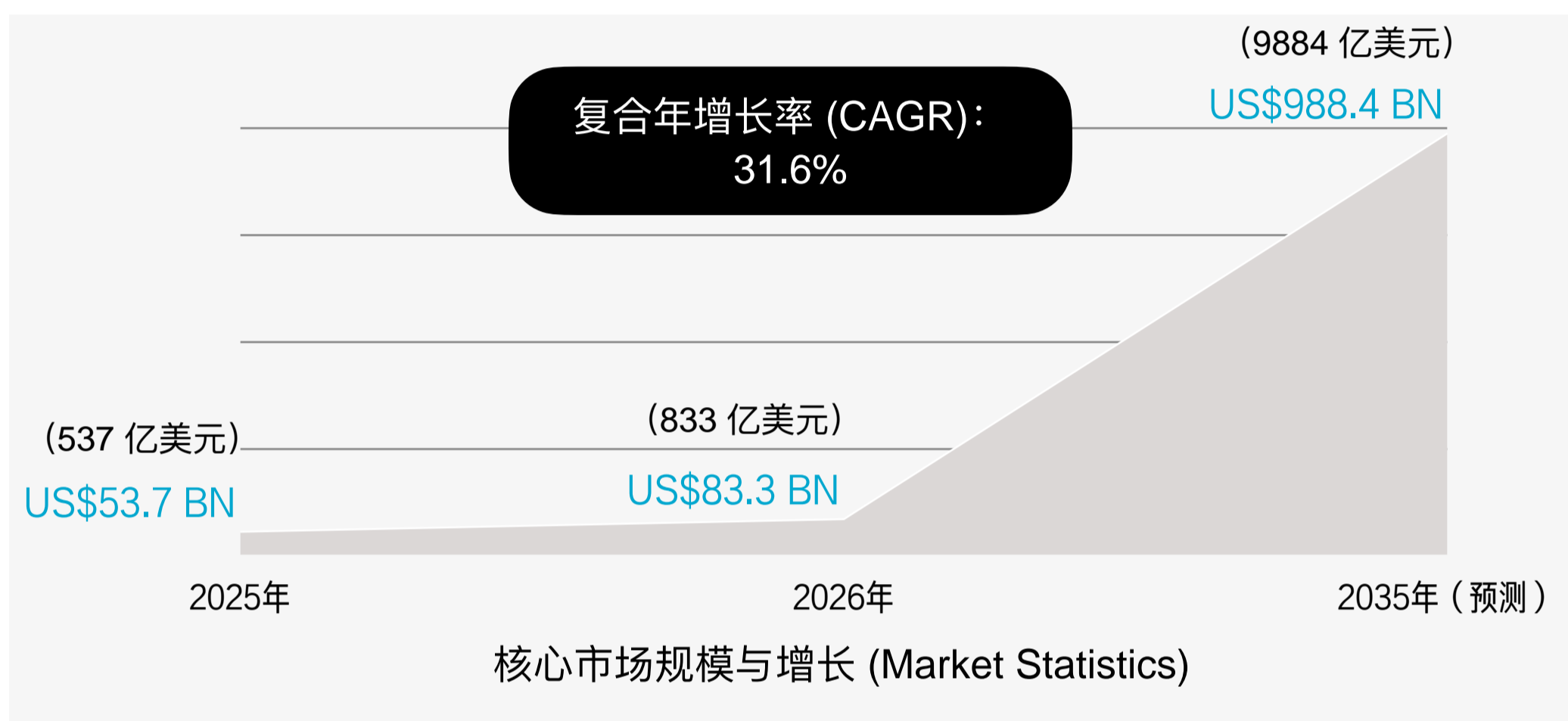
## AI 时代的 品牌可见度危机

# 01 从「链接搜索」到「答案生成」的跨越

## AI 全球用户量快速增长

截至 2025 年第四季度，全球生成式 AI 渗透率已达到总人口的 16.3%，高于上一季度的 15.1%。这意味着全球约有六分之一的人正在使用 AI 工具进行工作、学习或解决问题。

## 2035 年生成式 AI 市场规模接近 1 万亿美元大关



## 零点击搜索趋势正在形成

2025年，由于浏览器的 AI 概览模式的全面铺开，搜索结果页发生了结构性变化。这意味着用户在搜索页面就能通过 AI 总结直接获得答案，无需点击进入任何网站。

77% 移动端搜索以「零点击」结束

60% 网页端搜索以「零点击」结束

### AI 概览模式：谷歌搜索3000元左右手机推荐



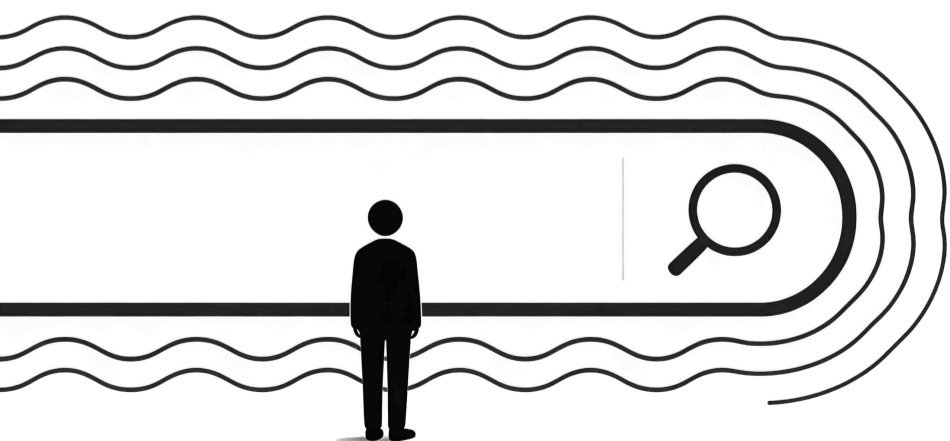
数据源: \*Global Market Insights (GMI) ; Global AI Adoption in 2025; The Digital Bloom2025年度报告

AI 全球用户量快速增长  
 ChatGPT、Google、千问相继推出「通用商业协议」  
 打通 AI 购物流程

2025-2026年，全球科技巨头相继发布商业协议，标志着 AI Agent 正式接管消费决策链。

「一键下单」不再是愿景，而是底层协议驱动下的行业共识。

维度	OpenAI: ACP 协议	Google: UCP 协议	阿里系: ACT 协议
核心逻辑	「代客下单」： 联合 Stripe 打造 Instant Checkout。 AI 负责决策、比价，商家在自有系统履约。	「电商 TCP/IP」： 联合 Shopify 等建立零售标准。 AI 跨平台完成发现、下单、售后全链路。	「商业直连」： 通义千问与淘宝、支付宝深度打通。 AI 实现决策、支付、履约的站内闭环。
关键动作	商家作为订单主体，维持原有售后。	Google 作为基础设施，不抽成，通过新型 AI 广告盈利。	提供即时与延迟付款，主攻 AI 代买与企业级自动化采购。
合作伙伴	Stripe、各类独立站	Shopify, Etsy, Target, Walmart	淘宝、支付宝、闪购生态

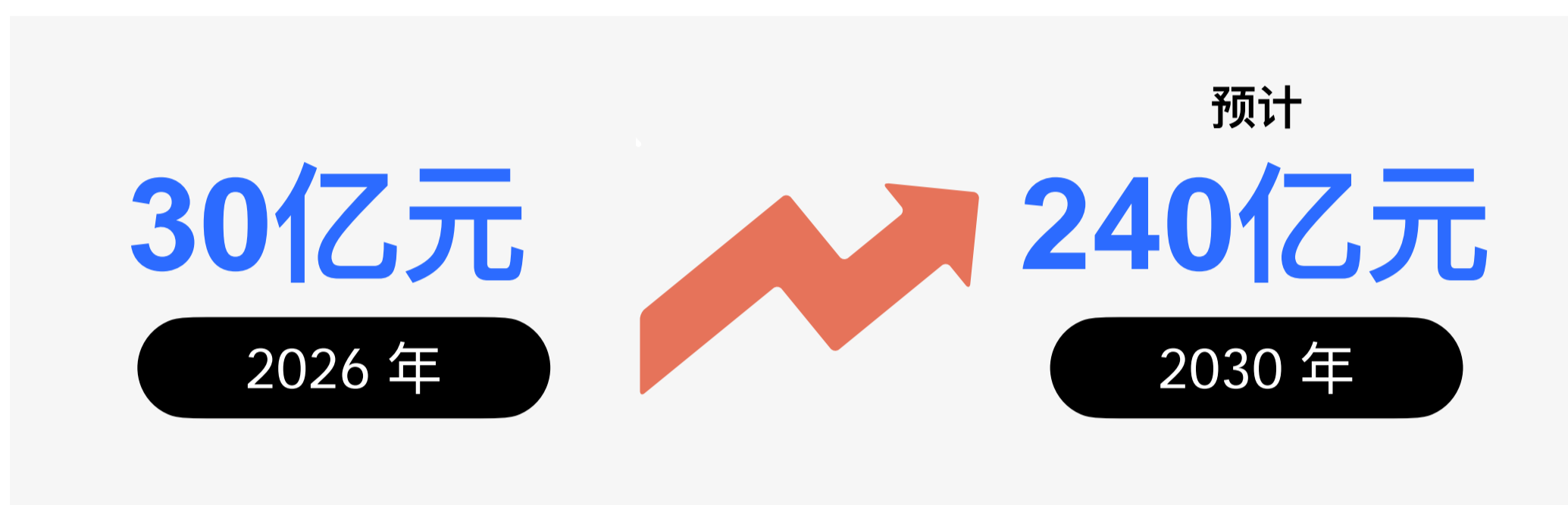


协议的普及意味着流量的入口已经彻底迁移。品牌若不能在这些协议的底层逻辑（即被 AI 推荐和选中）中占优，将直接失去被 AI 代理「代买」的资格。

# 02 行业现状：GEO 赛道的机遇、乱象与营销从业者认知

## || GEO 行业规模

Gartner 预测，到 2028 年，传统搜索引擎流量将减少 **50%**。这意味着原本属于传统 SEO/SEM 的数百亿、甚至上千亿级的广告预算将流向 AI 驱动的入口。



## || GEO 行业服务商现状

80%	以上 GEO 服务商为原 SEO 服务商	86%	采用 AI 生成内容进行内容铺设
-----	----------------------	-----	------------------

## || GEO 行业趋势与挑战

<h3>垂直化深耕</h3> <p>预计 2026 年医疗、法律等专业领域的定制化 GEO 服务占比将超 85%。</p>	<h3>标准规范化</h3> <p>中国信通院已牵头制定《生成式引擎优化 (GEO) 服务可信基本要求》，标志着行业开始进入规范化阶段。</p>	<h3>核心痛点</h3> <p>AI 的「黑盒子」属性使得优化效果难以标准化衡量，目前仅有约 30% 的企业对 GEO 的实际转化效果表示完全认可。</p>
---	--	---

数据来源：1. 易观分析《中国 GEO 行业市场发展报告 2026》；2. 调研数据

## || GEO 行业乱象

行业内目前出现较多违规操作，促使行业向合规化发展。具体违规操作界定包括：**1) 规模性投喂：**短时间内大量发布同质化语料，试图通过「堆量」操纵模型权重。**2) 内容造假：**恶意植入虚假事实、使用夸大性或误导性表述。**3) 指令攻击：**利用间接提示词注入等手段干预 AI 正常逻辑。

# 低质量内容投喂 对大模型的 三重危害

## 递归污染与「模型崩溃」

1

- ◆ **演化机制：**当大模型过度学习前代 AI 生成的低质量语料时，会陷入「递归污染」循环。
- ◆ **技术后果：**长期使用「数字垃圾」迭代将导致模型输出的概率分布逐渐偏离现实，最终导致模型丧失理解能力，输出紊乱内容。

2

## 放大「幻觉」与偏见

- ◆ **技术后果：**垃圾内容通常包含误导信息、陈旧信息、逻辑谬误。会直接导致模型在回答严谨问题时信口开河。

3

## 推高治理与算力成本

- ◆ **数据清洗：**为从海量数据中筛选极少量高质量语料，人力标注与清洗成本呈指数级激增。
- ◆ **训练效率：**垃圾内容占据大量上下文窗口，导致知识密度极低，严重浪费算力与电力资源。

参考资料：论文《AI models feed on each other to catastrophic collapse》；CAICT《人工智能白皮书》

## 低质量内容投喂对品牌的危害

信源  
信用  
破产

随着模型具备信息源交叉佐证与冲突检测能力，当品牌投喂的营销话术与权威第三方数据、真实评价产生矛盾时，会被判定为「不可靠信源」。品牌会因无法通过迭代反思与重新验证而遭遇系统性降权。

算法  
长效  
放逐

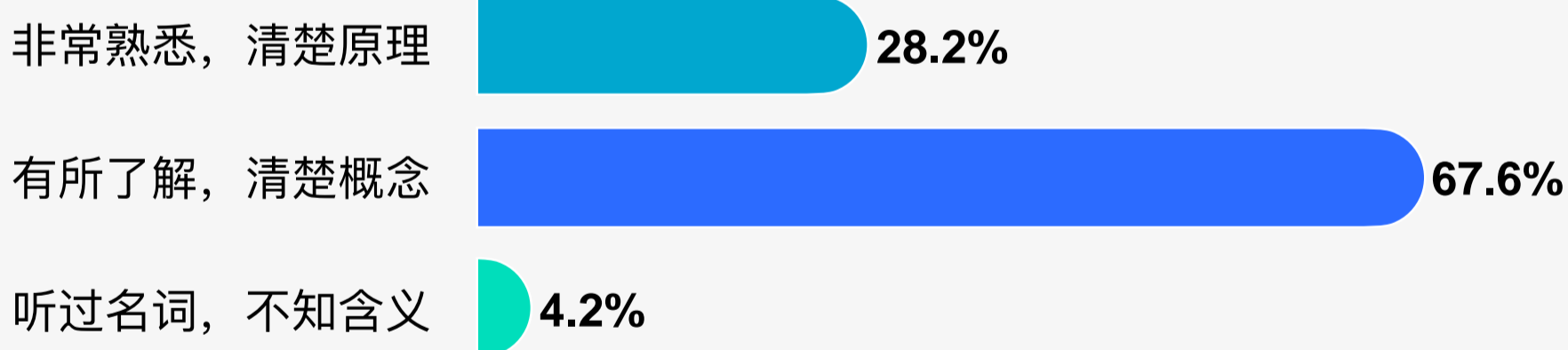
低质内容产生的负向反馈会形成「奖励信号」，让 AI 产生长期的排斥记忆，即便后续修复，也难以扭转已经形成的算法优先级偏见。

## 当前营销从业者对GEO认知情况

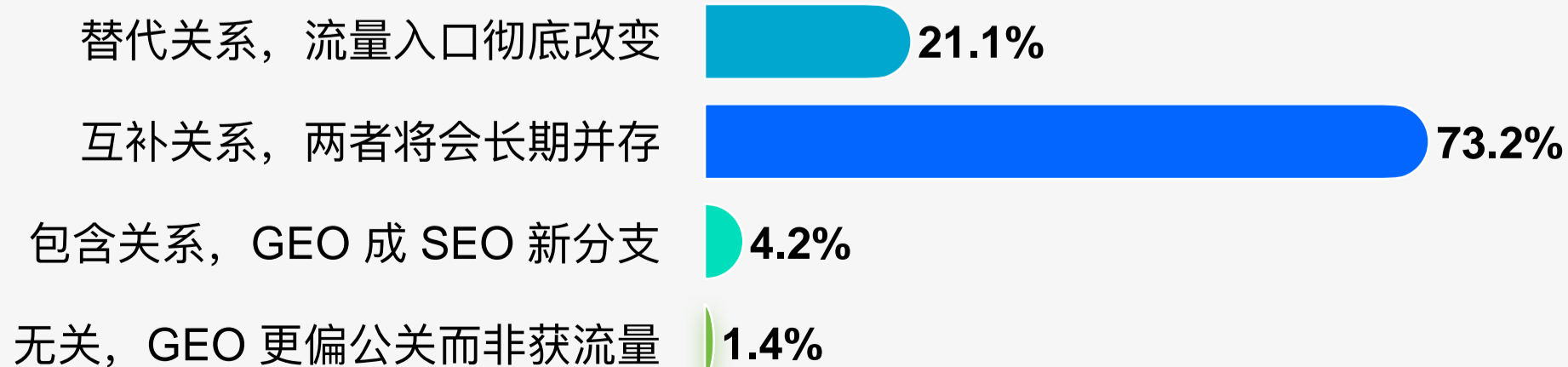
超 **95%** 营销人跨越名词门槛，近 **3 成** 已深度了解

2026年1月，一项针对广告主对GEO趋势观点的调研显示，营销从业者对GEO的认知度极高，**95.78%** 以上的受访者已跨越名词门槛，其中 **28.17%** 能深度理解其概念，精准区分其与传统 SEO 的差异。在关系界定上，**73.24%** 的人坚定认为两者是互补共存的关系，而非简单的替代。

### 营销从业者对 GEO（生成式引擎优化）的了解程度调研



### 营销从业者眼中 GEO 与传统 SEO 的关系调研



## 90% 营销从业者认同 AI 将不同程度取代传统搜索引擎

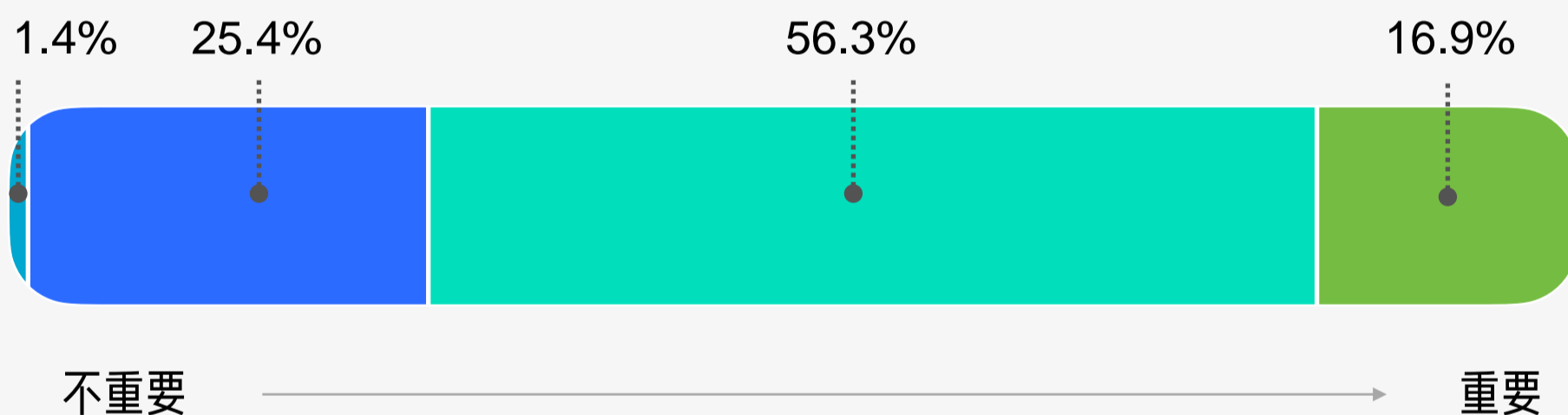
### 营销从业者对 AI 取代信息入口的认知度调研



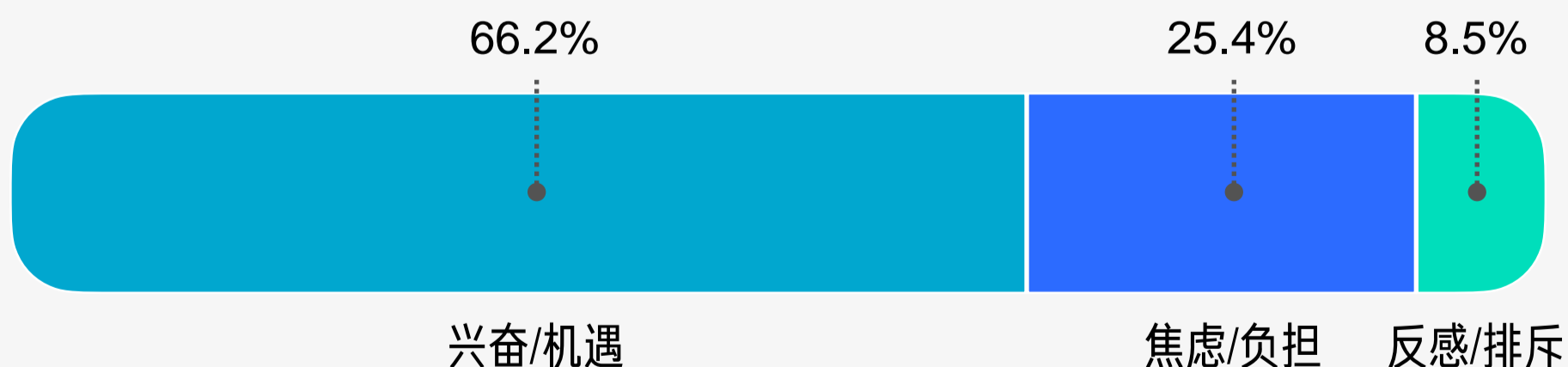
## GEO 的战略地位已成共识，从业者心态积极

**心态积极进取：** 面对新技术，超 7 成从业者认为其重要，66.2% 的受访者感到「兴奋」，认为这是打破大厂垄断、实现弯道超车的绝佳机遇；相比之下，仅有约四分之一的人感到焦虑。

### 营销从业者对当下 GEO 对品牌发展的重要程度认知



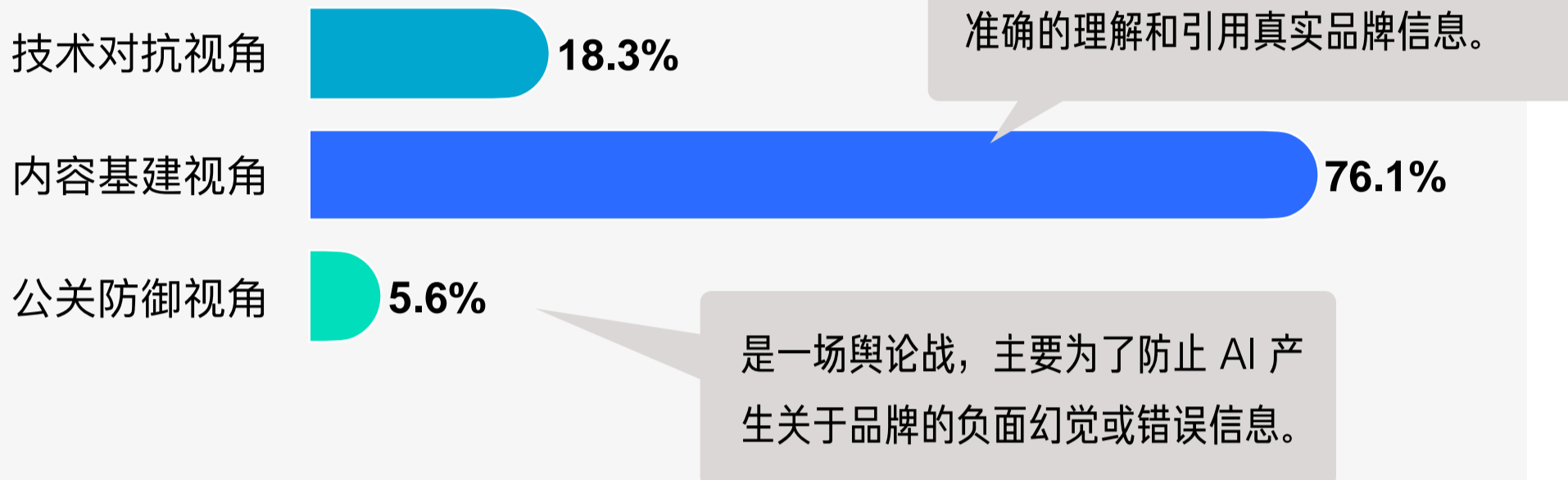
### 营销从业者对 GEO 的直观情绪/态度调研



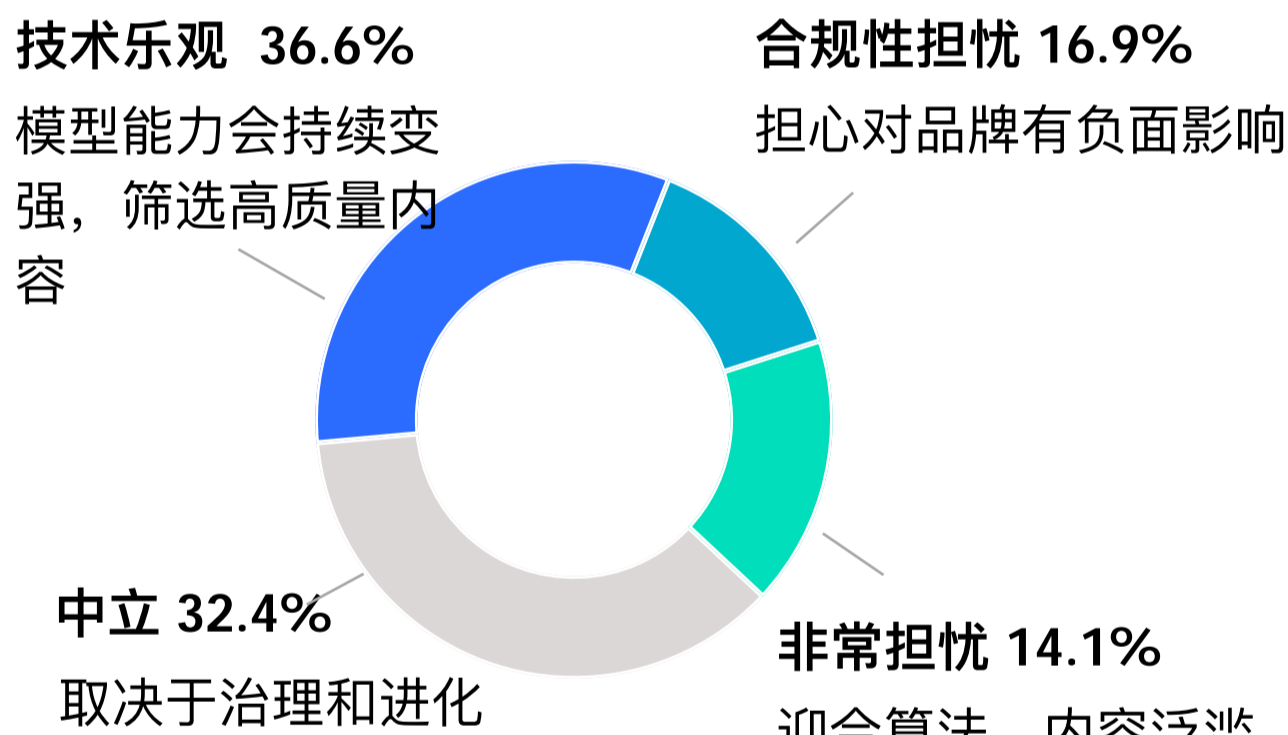
## 从业者整体理性乐观，认为 GEO 发展将回归内容本质

76.1% 的受访者认为 GEO 的本质是「内容基建」，即通过结构化信息帮助 AI 准确引用品牌，远超「技术对抗」视角。

### 营销从业者对 GEO 本质的理解调研



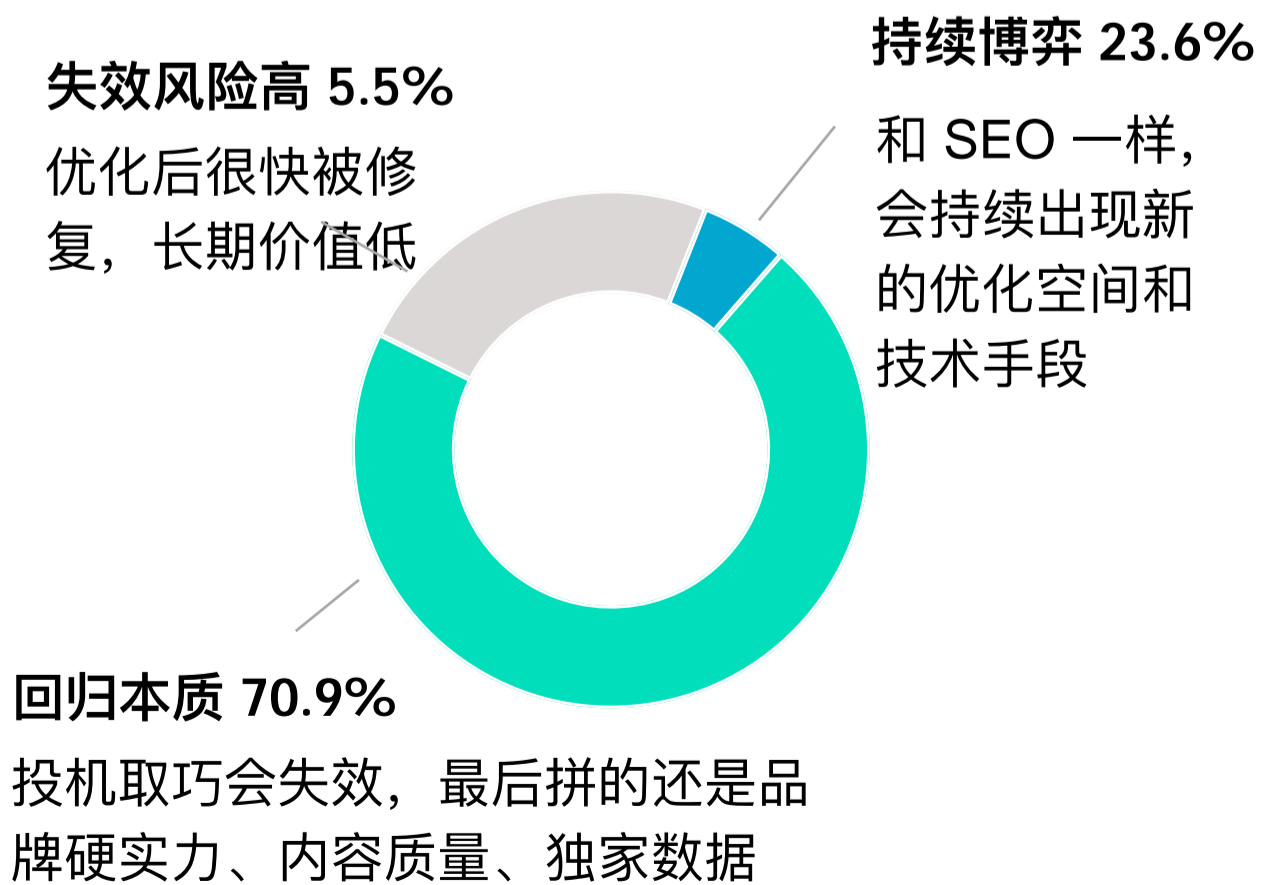
针对「GEO 制造垃圾内容」的疑虑，市场从业者反馈相对理性乐观：**32.4%** 则认为这是一场长期的动态博弈，优质内容终将胜出。**36.6%** 的人持技术乐观态度，相信随着 AI 模型进化将自动过滤劣质内容——



### 营销从业者对 GEO 制造垃圾内容的看法

**70.9%** 的受访者认为随着模型变强，GEO 将回归内容本质，投机手段必将失效，未来竞争核心在于品牌硬实力与 E-E-A-T 优质内容。

另有 **18.31%** 的人预测这将是一场如同传统 SEO 般的长期技术博弈。仅极少数人担忧 GEO 会彻底失效。



### 对未来 GEO 发展的看法

AI 时代品牌可见度危机：消费链路的彻底颠覆

消费者旅程

发展阶段

01

消费者通过广告曝光、社交媒体、朋友口碑获取信息

认知阶段

消费者通过广告曝光、社交媒体、朋友口碑获取信息后向 AI 提问

01

02

消费者主动搜索品牌、浏览官网、阅读评测

兴趣阶段

02~04

四个阶段均由 AI 执行

03

消费者在电商平台对比不同品牌的价格、参数

比较阶段

全链条意图内爆：

AI 实时重构决策逻辑。从「人选品」转变为「算法代劳」，在毫秒间完成全网性能对标，将决策逻辑锁定为「算法推荐的最优匹配」。

04

消费者被品牌故事、设计、促销活动影响决策

决策阶段

静默代理执行：

由 AI Agent 预决策并自动跳转（或直接）完成交易，实现「意图即交付」。

05

消费者在品牌官网或电商平台下单

下单阶段

# 第二章

AI 时代新标准

解码品牌 AI 竞争力指数

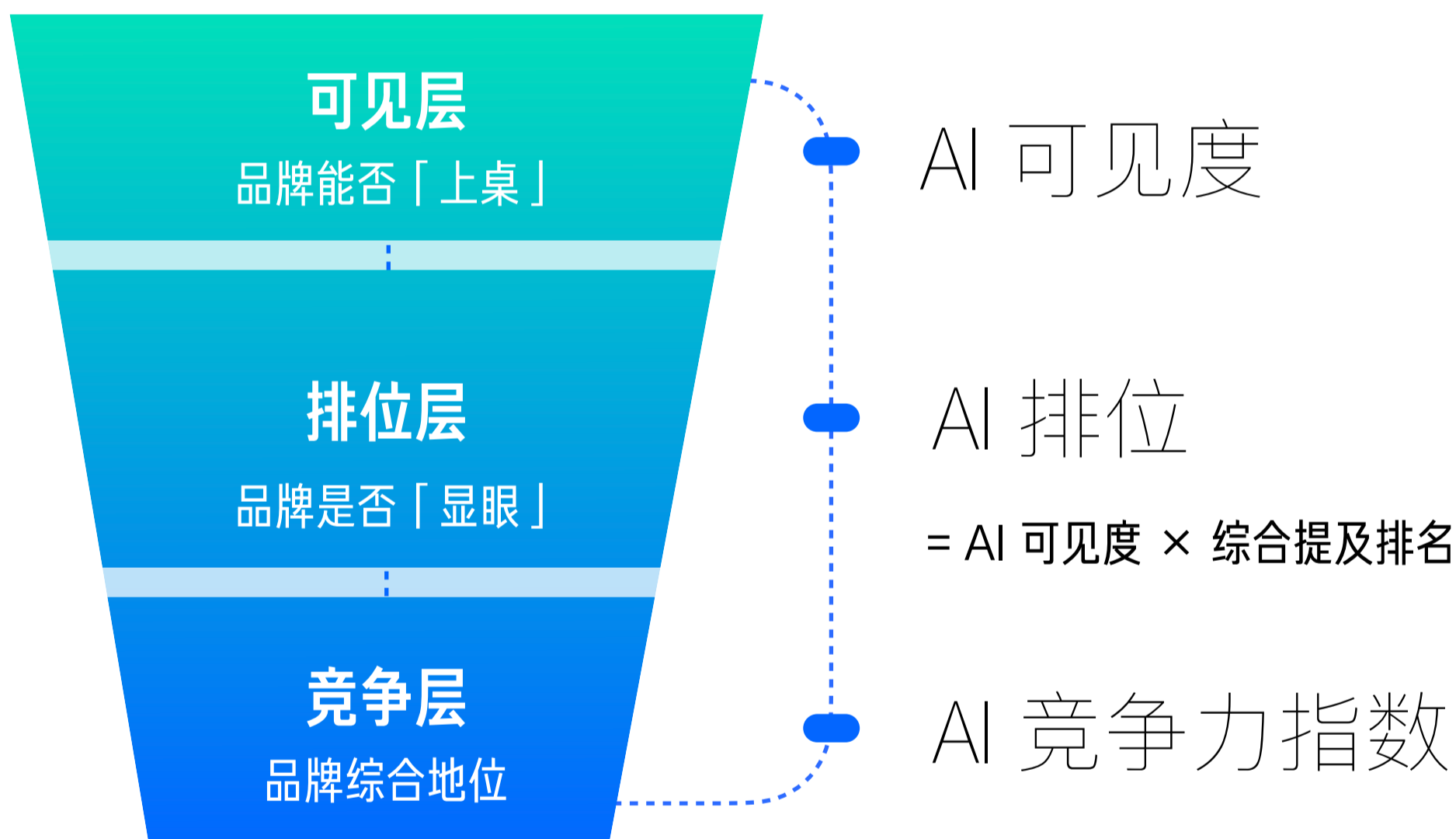
# 01

## 核心度量衡 品牌 AI 竞争力指数的构建逻辑

随着模型能力的提升，我们正急速步入 Agentic Search（智能体搜索）时代。

当AI成为问询入口，用户的行为发生了质变，当用户询问「哪款相机适合新手」时，Agent 会主动阅读数千个网页，通过「交叉佐证」剔除广告干扰，通过「逻辑审核」识别事实冲突，最终直接吐出一个具有结论性的答案。

上述范式转移，导致品牌竞争力的底层逻辑发生了改变



$$\text{AI 竞争力指数} = \text{AI 可见度 (Visibility)} \times \text{综合提及排名 (Rank)} \times \text{内容可信度 (Trust)}$$

## 过去与当下的范式对比：流量与内容的「权重博弈」

品牌影响力增长始终由两个核心变量驱动：触达（流量）与说服（内容）。但在不同的技术纪元，这两者的「杠杆率」发生了根本性倒置。

### Before

### Now

#### 传统搜索/社交时代

#### AI 时代



## AI 时代的品牌定位

品牌 AI 竞争力指数公式的乘法逻辑揭示了品牌资产在 AI 维度下的非线性特征：不同维度的缺失，会导致品牌呈现截然不同的病态。

$$\text{品牌 AI 竞争力指数} = \text{AI 可见度 (Visibility)} \times \text{综合提及排名 (Rank)} \times \text{内容可信度 (Trust)}$$

指数特征	品牌出现的问题	品牌结果
V = 0	原生性消失	无入场券：物理世界声量无法转化为 AI 决策。
高 V / 高 R / 低 T	空心品牌	空中楼阁：声量大排位高，但内容虚浮，均来自低质量信息源。
高 V / 低 R / 高 T	边缘品牌	无效陪跑：有声量有口碑，但在算法推荐位中处于末席。
低 V / 高 R / 高 T	孤岛品牌	深巷美酒：某一场景排名极高，但覆盖面很过窄，无法规模化。
高 V / 低 R / 低 T	噪点品牌	无效声量：有声量无位置，有存在感无说服力。重铺量但均为低质水文。

# 02

## AI 可见度 衡量品牌进入 AI 决策池的入场券



在生成式 AI 时代，可见度不再是传统的「曝光量」，而是品牌进入大模型搜索系统的入场券。如果品牌无法在 AI 的逻辑中被「看见」，其在物理世界积累的所有资产都将归零。而「可见」这个概念在 RAG（检索增强生成）时代，品牌存在在网页里；但在 Agent 时代，品牌必须在 AI 的逻辑思维里。

### 提问电视选购后大模型仅推荐 4 个品牌的产品，其他品牌均未获得「入场券」

推荐几个3000-4000元的75寸电视

这里为你精选几款 2026 年初**3000-4000 元**、75 英寸的高性价比电视，附核心参数、优缺点和选购建议，价格参考电商活动到手价（含国补 / 平台补贴）：

#### 1. 雷鸟 75 鹤 6 Pro (2025/2026 款, 75S585C Pro)

- 参考价：**3599 元左右**
- 核心配置：4K 分辨率，VA 软屏，量子点广色域（96% DCI-P3），MEMC+240Hz（可超频 288Hz）高刷，HDMI 2.1 接口，低输入延迟；4GB+64GB 内存，支持 Wi-Fi 6；杜比视界 + 杜比全景声
- 优势：同价位游戏性能突出，运动画面无拖影，暗场对比度好；系统干净，广告少
- 适合人群：**游戏玩家、体育赛事爱好者**

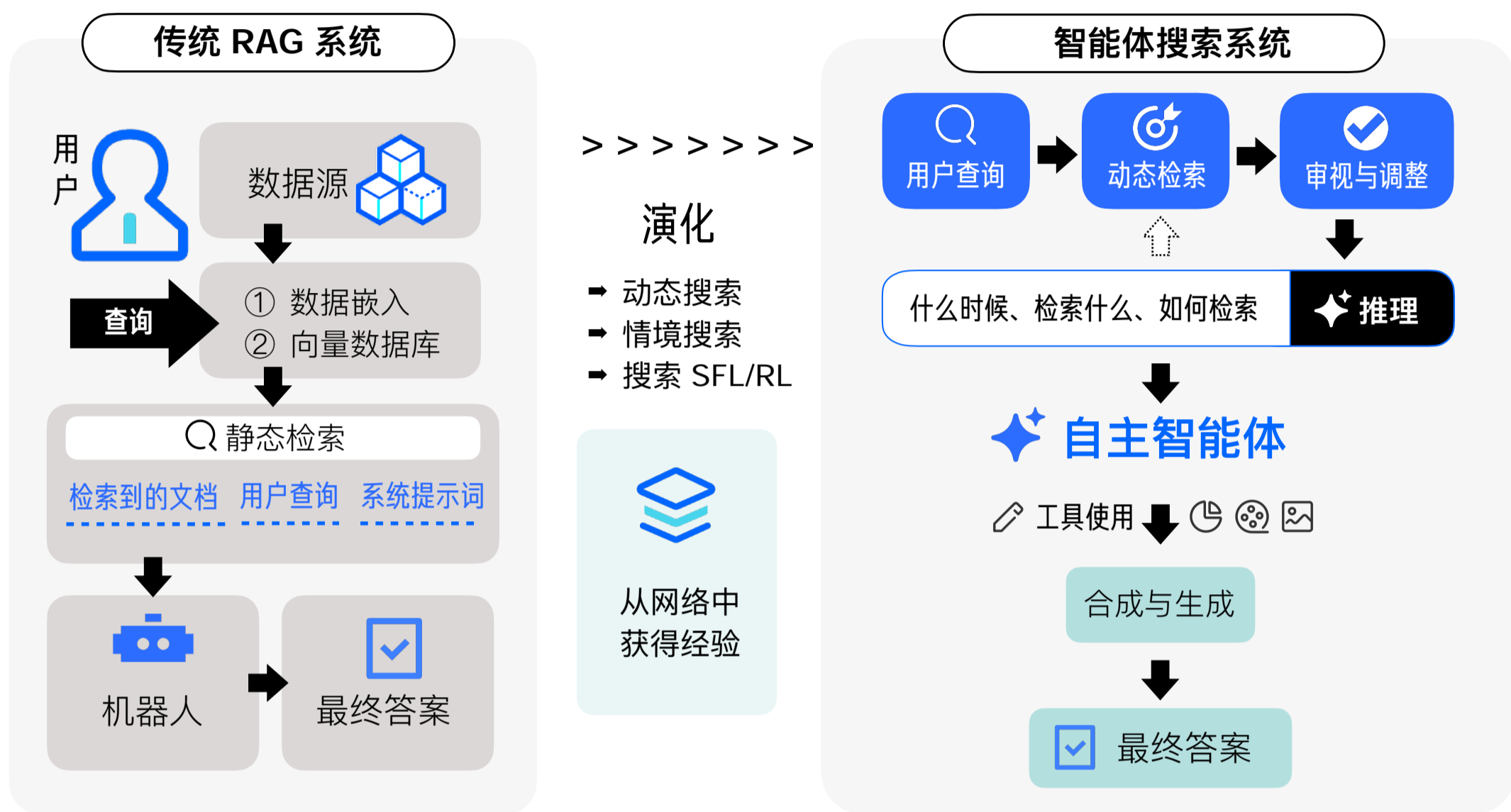


#### 2. 海信 75E3QH Pro/75E3ND Pro

- 参考价：**3399-3699 元**
- 核心配置：4K 120Hz 原生高刷，AG 防眩光屏（适 ↓ 亮客厅），信芯 AI 画质芯片，双声道大功率音响 + 杜比全景声；4GB+32GB/64GB 内存，远场 AI 语音，支持方言识别

## 技术拆解：可见的进化路径——从 RAG 到 Agentic Search

目前主流模型正从「静态 RAG」向「Agentic Search」加速进化，这意味着 AI 不再仅仅是搜索结果的搬运工，而是拥有了自主推理与全网核查能力的独立决策者，正因为 Agent 拥有了自主权，它对信息的要求变得更加严苛。



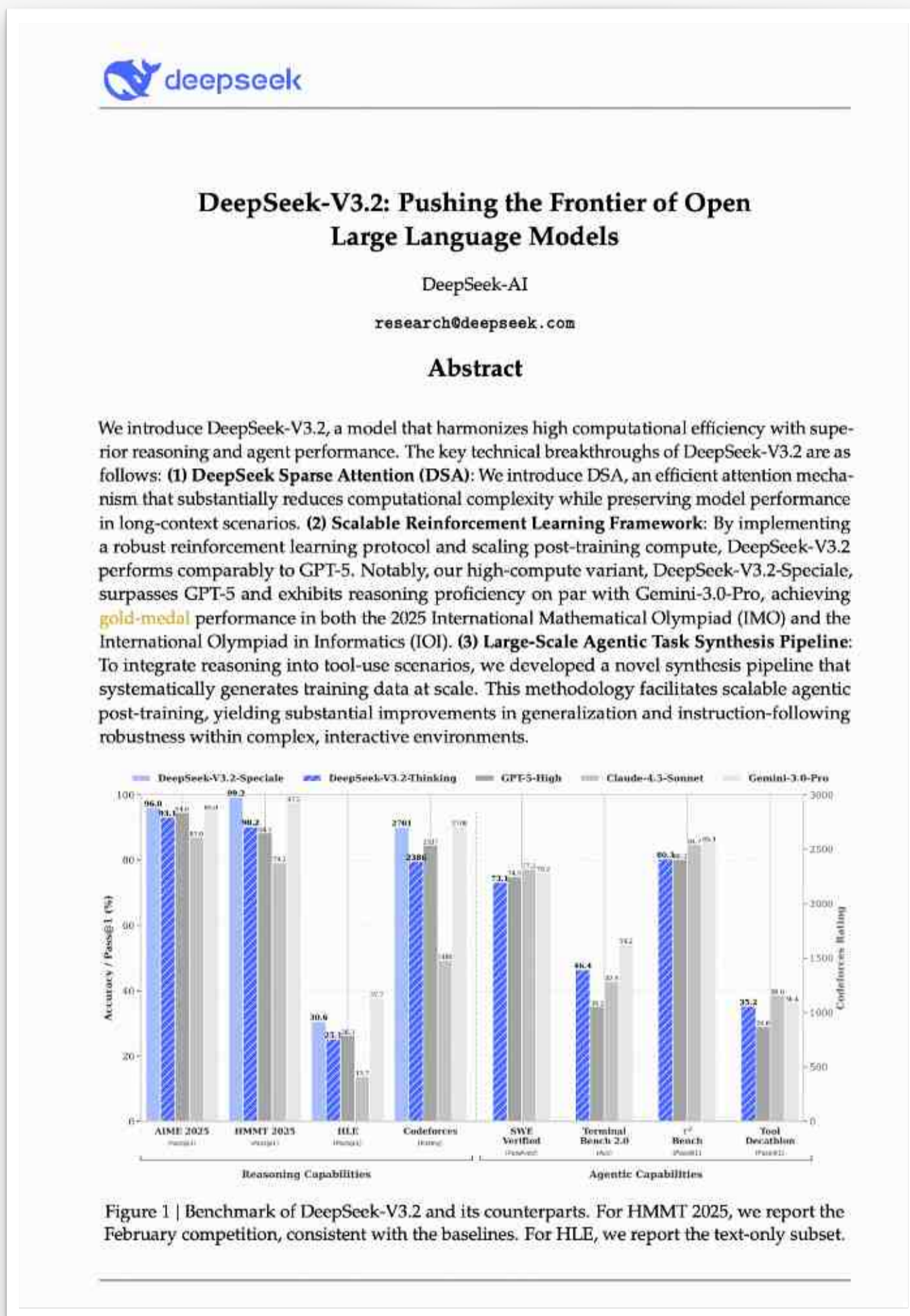
传统 RAG 依靠向量数据库的被动匹配，而 Agentic Search 则引入自主智能体进行推理、批判与动态检索。在这种演进下，可见度不仅是品牌的覆盖频率，更是进入 AI 自主决策链路的。

比较维度	检索增强生成 (RAG)	智能体搜索 (Agentic Search)
检索时机与频率	通常在模型生成文本之前，执行固定的、一次性的 (one-shot) 检索。	智能体根据实时的推理需求，自主、动态地决定何时 (when) 进行检索。
决策与自主性	缺乏自主控制，通常是被动的单向 workflow，直接将检索到的文档附加到查询和提示词中。	将推理和控制嵌入到集中式智能体中，由智能体自主决定检索什么 (what) 以及如何 (how) 检索。
工作机制与闭环	主要是基于向量数据库的静态检索。	引入了「评估与适应」的动态循环，结合了推理、工具使用和信息合成。
推理过程中的调整	一旦检索完成并输入给模型，通常无法根据生成过程中的发现改变检索方向。	智能体可以在推理中途动态改变检索策略、重新细化查询词，从而更好地整合来自多个数据源的证据。

## 海内外大模型在搜索能力方面不断突破

以 DeepSeek-V3.2 为例（2025年12月发布）

新模型通过以下四个核心方面对其搜索能力（Search Agent）进行了深度优化



- ◆ **高质量数据合成与严苛核查**：构建多智能体数据生成管线，并由专门的“验证智能体”执行多轮事实核查。
- ◆ **突破长度限制的上下文管理**：提升了模型在长网页检索中的扩展计算能力。
- ◆ **混合多维目标的强化学习（RL）**：RL 阶段不采用单一奖励，利用生成式奖励模型进行多维打分，强制模型在优化时同时兼顾「事实可靠性」与「实际帮助性」。
- ◆ **推理与工具调用深度融合（Thinking in Tool-Use）**：模型在调用搜索工具的过程中及作答前，都能进行主动的逻辑思考和规划。

## 场景覆盖率：品牌在 AI 可见下的全方位渗透

「场景覆盖率」是可见度相关的重要指标之一，衡量了品牌能否适配 Agent 拆解后的所有细分决策路径。AI 不再只提供通用答案，而是通过多跳逻辑引导重排，在用户复杂的限制条件下寻找最优解。



现阶段，我们将品牌在 AI 决策中的应用划分为以下**六大核心场景**，它们共同构成了品牌与用户连接的支点：



通用推荐



价格预算



人群身份



使用条件



功能/效果



风格/口碑

## 品牌需关注的「可见度」相关指标

在 Agentic Search 时代的战略意义

### AI 可见度

#### 品牌在大模型回答下的露出情况

衡量品牌资产是否成功进入 AI 第一轮召回池的准入门槛。

### AI 可见度排名

#### 品牌在本行业下的 AI 可见度排名

揭示在同类竞争中，品牌被 AI 优先选为候选对象的概率顺位。

### AI 可见度趋势

#### 在所选时间内，品牌 AI 可见度变化趋势

评估品牌语料在 AI 神经网络中的持久性，预警资产衰减风险。

### 品牌排名趋势

#### 在所选时间内，品牌 AI 可见度排名变化

监测品牌在动态博弈中「上桌位次」的波动，反映竞争壁垒的厚度。

### 场景覆盖率

#### 品牌在用户不同场景决策中的渗透广度

衡量品牌能否适配 Agent 拆解后的所有细分决策路径（如：从「极寒」到「冲锋衣」）。

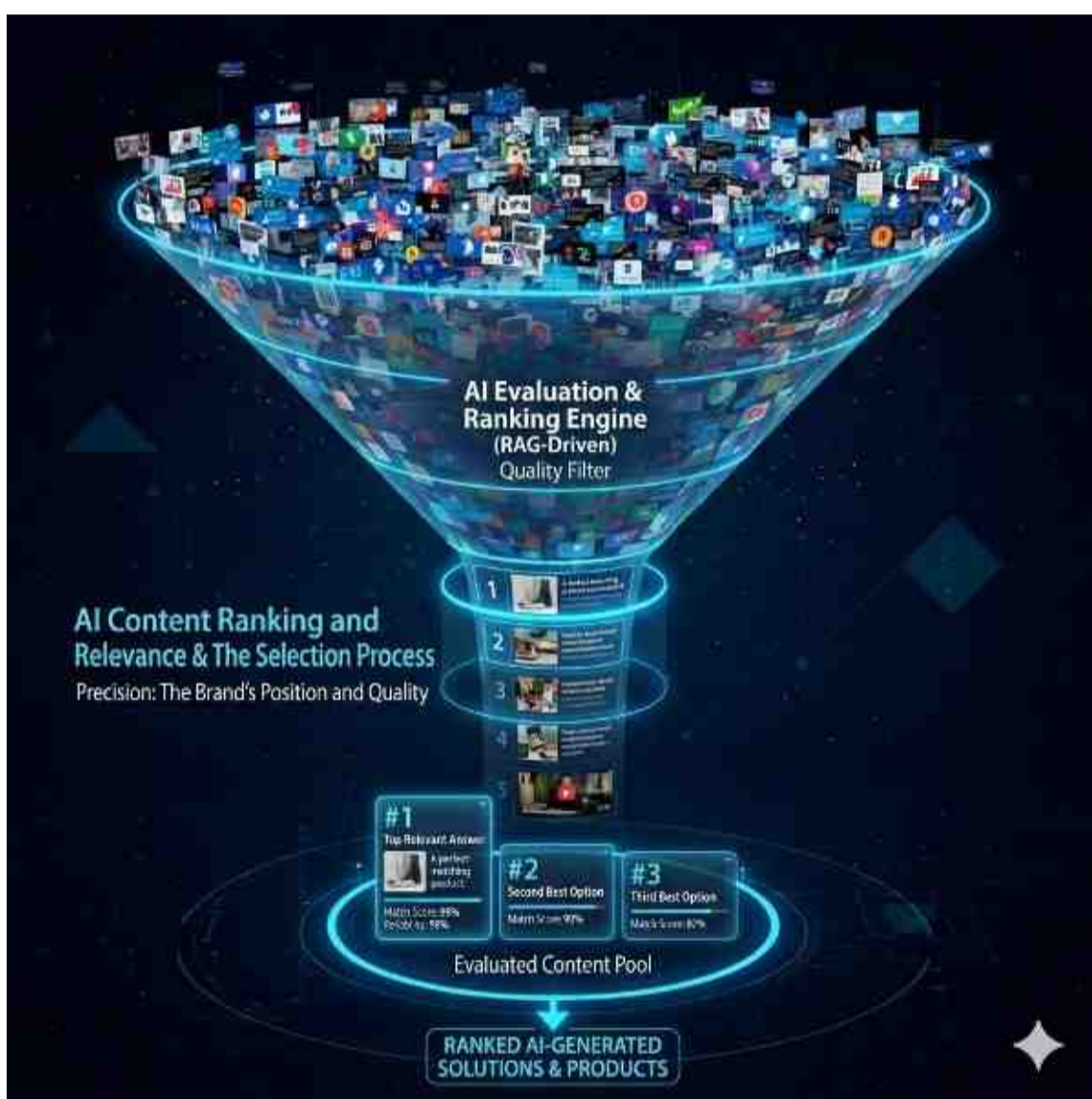
# 03

## 综合提及排名 衡量品牌在 AI 建议中的推荐优先级

当品牌通过可见度的门槛成功「上桌」后，真正的博弈才刚刚开始。

在 Agentic Search（智能体搜索）体系下，AI 不再只是罗列信息，而是通过评价与适配以及推理过程，分解「搜索-过滤-合成」的深度角色分工，主动评估并提炼信息，将品牌资产从海量噪音中剥离并赋予优先级。

排位（Rank）指标衡量的是品牌在 AI 建议列表中的平均位次与推荐权重。在 AI 极简交互的语境下，排名第一与排名第五之间存在着巨大的商业转化鸿沟。



提问「推荐三个跑鞋品牌」后，大模型首推亚瑟士。

推荐三个适合大体重女性的1000元左右跑鞋

这里给你推荐 3 款1000 元左右、专为大体重女性慢跑设计的跑鞋，兼顾顶级缓震、稳定支撑与亚洲脚型的宽楦 / 舒适设计，活动价基本都能压在千元内，适合 70kg+ 跑者保护膝盖、日常长距离慢跑：

### 1. 亚瑟士 GEL-Nimbus 27（女款，活动价约 900 元）

- 核心科技：后跟厚 PureGEL 胶 + FF BLAST™ PLUS 中底，冲击吸收率高，落地软而不塌，对膝盖友好；有宽楦版本可选，适配高脚背、宽脚；
- 优势：包裹扎实、后跟港宝稳定，长距离慢跑不累，适合膝盖敏感 / 大体重入门跑者；
- 不足：重量比 HOKA 略沉，竞速不适合；



### 2. HOKA ONE ONE 克利夫顿 10（女款，活动价约 950 元）

- 核心设计：ProFly + 厚泡棉中底、宽基底大底，稳定性强，前后掌落差小，滚动感顺滑；鞋面工程网布透气，女款配色清新；
- 优势：软弹脚感明显但不卸力，走路跑步都舒适，适合日常慢跑 + 通勤两穿，对脚踝力量偏弱的跑者友好；
- 不足：外底耐磨度中等，尽量避免水泥地高频跑；

技术拆解：

## 智能体搜索决定排序优先级的六大核心机制

智能体对品牌信息的筛选与重排，本质上是一场严苛的「数字审议」

### 1 动态过滤与逻辑提取

智能体主动阅读文档，仅将满足当前推理需求的品牌关键证据提取至下一步，无效信息将被过滤。

### 2 交叉佐证与冲突检测

通过跨数据源验证，智能体会检测不同信源间的冲突，剔除无法溯源或存在矛盾的品牌声明，保留最可靠的证据。

### 3 迭代反思与重新验证

引入「反馈与反思」循环，若初始检索结果不足以支撑决策，智能体会触发重新验证并调整策略，直至获取高质结果。

### 4 多智能体协同筛选

通过「信息搜查智能体」、「信息核查智能体」和「信息整合智能体」，评估品牌证据的完整度，并针对信息缺口进行定向补全。

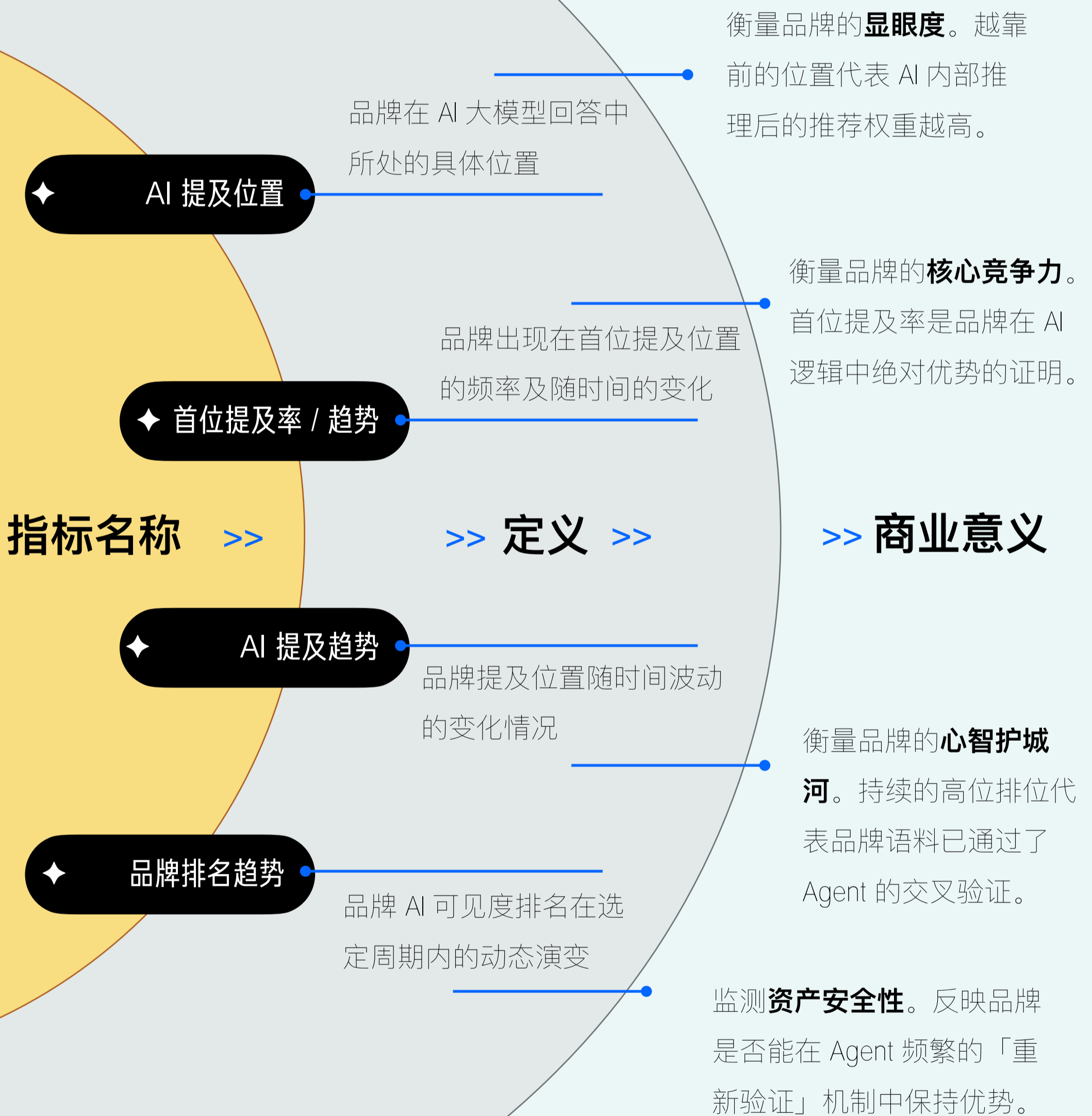
### 5 强化学习质量奖励

利用奖励信号隐式学习用户偏好，通过不断的试错，自动赋予高质量、高事实性品牌资产更高推荐权重。

### 6 多跳逻辑引导重排

利用知识图谱等结构化信息，沿着「证据链」进行层级感知重排，平衡回答的集中度与多样性。

品牌需关注的「提及位置」相关指标



# 04

## 内容可信度： AI 推荐的「信任根基」

如果说可见度和排位决定了品牌能否进入 AI 的决策池，那么信源验证与内容质量则决定了品牌能否最终获得 Agent 的推荐。在 Agentic Search 时代，AI 不再盲目抓取全网碎片，而是基于「信任根基」执行多轮逻辑审计。品牌必须同时满足权威的平台背书（信源可信度）与高密度的客观事实（引用内容质量），才能跨越 AI 的信任门槛。

$$\text{内容可信度} = \text{信源可信度} \times \text{引用内容质量}$$

### 搜索信源选择的变革

#### 过去 · SEO 时代

搜索引擎主要依赖网页间的跳转链接来判断权重。品牌只需在多个站点铺设内容，通过量化堆砌即可换取曝光。

#### 现在 · AI 时代

Agentic search 具备「交叉佐证与冲突检测」能力。它不再盲目采信单一信息源，而是会通过跨数据源验证，剔除那些无法溯源或存在矛盾的品牌内容，仅保留最可靠的证据。

#### 信源分级

##### 一级信源：

官网/百科/官媒

##### 二级信源：

社区/自媒体/门户

##### 三级信源：

其他站点

引用内容质量：

## Agent 逻辑审计的底层度量衡

### 文本密度比

定义：去除导语、修饰词及情感词后，核心事实（参数、时间、结论）的占比。

Agent 偏好：越高越好。海量营销词汇会因「计算」Token 产出比过低而被视为噪音

### 结构化程度

定义：文章是否具备清晰的 H1/H2 标签、标准化表格或列表，是否方便实体提取。

Agent 偏好：方便 Agent 提取实体的文章能获得极高的效率奖励。

### 信息增益

定义：相比通用百科，文章是否提供了独特的实测数据、一手访谈或故障预警。

Agent 偏好：AI 渴求差异化的证据。唯一的实测数据（如：-15°C 录制 10 分钟掉电量）是推荐的定音锤。

### 逻辑自洽性

定义：文中是否存在前后矛盾、事实错误或与公认常识不符的幻觉风险。

Agent 偏好：存在逻辑冲突的内容会触发“安全剔除”机制，直接放逐。

更多关于模型内容质量的研究分析将在下一章展开

# 第三章

## 实证研究

高质量专业语料是品牌

影响 AI 的基石

# 01

## 中外模型引用信源内容质量对比

本节通过「用 AI 审计 AI」的实证方法，深度拆解了 2026 年大模型在处理复杂消费决策时，对信源筛选的底层逻辑。通过对中美主流模型引用的 24 份原始语料进行全维度解剖，揭示了什么样的内容资产才能通过 AI 的「数字审议」。

### 研究方案设计

### 实验样本与工具

#### 实验目的

对比中美模型信息源选择的差异与内容质量。验证在 Agentic Search 时代，什么样的内容资产能通过 AI 的「数字审议」，并获得第一顺位的推荐权重。

#### 审计对象：

选取海外代表模型与国内代表模型作为「筛选器」。

#### 审计命题：

涵盖家电、数码、运动、前沿科技四大领域（扫地机推荐、5000元手机推荐、大体重跑鞋推荐、AI 眼镜推荐）。

#### 审计官：

调用高性能模型扮演「内容审计员」，对模型召回的前 3 位原始信源进行解剖。

### 研究选取前三信息源进行内容质量分析

大体重跑鞋推荐

大体重跑者优选跑鞋清单

国际品牌（专业之选，预算较高）

跑鞋型号	核心特点
ASICS Gel-Nimbus 27	柔软的凝胶缓震，保护性极
ASICS Gel-Kayano 32	稳定支撑的标杆，有效控制
New Balance Fresh Foam X 1080 v14	脚感非常柔软，前掌空间
Saucony Triumph 23	缓震和回弹平衡得很好，脚
HOKA ONE ONE Bondi 8	超厚底缓震，像“踩云朵”

搜索结果

- 1 新浪新闻 | 2026/02/24  
好物 | 1000万双验证！这款“缓震鞋”太牛了！200斤也能起飞  
我们同事体重180+斤，原地蹦跶，鞋底不变形，踩不塌，还有缓震 大体重跑步，这双鞋完全可以 hold 住！ 鞋底的回...
- 2 Apple Podcasts | 2025/05/26  
EP186: 人生第一双跑鞋应该怎么选(国产品牌篇)  
· 理由: 大厚底顶缓款, 对标国际品牌, 包裹支撑性好, 适合大体重跑者, 目前 400 多元 (后续可能降价) ... 大体重稳...
- 3 虎扑 | 2025/07/09  
大体重篮球爱好者跑鞋怎么选  
本人身高183, 体重目前170斤, 肌肉型, 下肢粗壮, 需要厚底缓震效果好的, 支撑够的, 脚型正常有足弓, 预算最好不...

## 内容质量评估维度

### 文本密度比

剥离营销修饰语后，核心参数、时间、结论等硬信息的占比。

### 结构化程度

H1/H2 标签、表格、列表的规范性，即对 Agent 提取实体的友好度。

### 信息增益

是否包含独特测试数据、一手访谈或差异化拆解，而非通用的“百科废话”。

### 逻辑自洽性

是否存在事实冲突或前后矛盾，评估其幻觉风险。

## 结论：中外模型内容质量评分

通过对四个核心消费命题（扫地机、拍照手机、跑鞋、AI眼镜）引用的 24 份原始语料进行标准化审计，我们得出了中美大模型在信源筛选上的真实质量均分：

维度/命题	海外模型引用组 (均分)	国内模型引用组 (均分)	差异分析
文本密度比	8.82	7.25	海外模型对「纯干货」的偏好更极端。
结构化程度	9.25	7.84	海外模型更依赖清晰的 H1/H2 逻辑。
信息增益	8.41	8.63	国内模型引用更具场景特色和情感增益。
逻辑自洽性	9.46	8.11	海外模型对冲突信息的过滤更严苛。
综合竞争力总均分	8.98	7.95	差异显著：约 10% 的代差。

|| 案例 || 「适合大体重跑者的跑鞋推荐」信源质量评估

	语料来源	文本密度	结构化程度	信息增益	逻辑一致性	综合得分
海外模型 选择信源 TOP3	知乎	9	10	9	10	9.5
	运动笔记 (港台网站)	8	9	8	10	8.8
	识货	7	4	6	10	6.8
国内模型 选择信源 TOP3	新浪	6	8	10	9	8.3
	苹果播客	9	7	8	10	8.5
	虎扑	4	3	7	10	6.0

文本  
密度比

**高分表现：** 文章直击用户提问痛点，如知乎信源明确给出定义「体重超过80kg为大体重」，以及各鞋款精确到克的重量（如42.5码为278g）。

**低分表现：** 含有大量营销辞令，如新浪信源「200斤也能起飞」或「稳如老狗」等情感化修辞

结构化  
程度

**评比细节：** 是否包含清晰的H1/H2标题、对比表格或分类清单。

**提取价值：** 结构化好的文本可以直接映射出「品牌—型号—中底科技—适用脚型」的关联数据。

信息  
增益

**独特测试：** 如「中底4层透气系统」拆解图缓震实验。

**真实案例：** 如记录中作者提供的从200斤减至165斤的真实跑步减肥心路历程。

逻辑  
冲突点

**校正价值：** 如特别纠正了“通过脚后跟磨损判断脚型”的常见误区，强调必须观察“前脚掌磨损”。

**界限界定：** 不同信源对“大体重”的起始定义存在细微冲突。

## || 实证研究一：结论与品牌启示

海外模型在信源筛选上存在约 10% 的「逻辑洁癖」优势，随着模型的进化，未来国内大模型的检索和筛查能力将逐步提升，低质内容将被加速清洗。

### 核心结论

#### ◆ 一 筛选标准转向「高质与硬核」

高权重信源在文本密度与结构化程度上表现卓越。AI 时代，内容不仅是给用户看的，更是给 Agent 审计的。

#### ◆ 一 情感增益的本土机会

当前国内模型对真实场景化、情感化描述具有更高识别度，是品牌在本土生态下建立差异化的关键窗口。

### 品牌内容资产打造的启示

1,

文本密度  
(脱敏去水)

剔除营销修饰词，增加硬核参数。

2,

结构化程重塑

严谨使用 H1/H2 标签及对比表格，确保 Agent 能瞬间映射出「品牌—产品型号—科技力/产品力—适用场景」的关联数据。

3,

信息增益

拒绝百科废话或通稿陈述，提供独特拆解实验或真实成长案例以提升内容独特性评分。

4,

逻辑校正  
(权威性审计)

主动纠正行业认知误区，通过展现逻辑自治性和专业知识获得模型的高信任度分值。

# 02

## 品牌语料的「引用概率」分析

实证研究二我们更进一步，通过模拟模型的决策任务，对不同性质的品牌内容资产进行了一场「引用淘汰赛」。本实验旨在探测 Agentic Search 审计的背后逻辑和原因：当 AI 面对不同类型的文章时，它会精准地引用谁，又会放弃引用谁？

### 实验设计

模拟「约束条件」的家电决策为了测试 Agent 对信息密度的敏感度，我们设定了一个具体的家庭场景：

The diagram illustrates the experimental setup. At the top, three light blue rounded rectangles represent the input materials, each with a red PDF icon and the text 'PDF':

- 扫地机器人-短文 (Short text about vacuum robots)
- 扫地机器人-长文 (Long text about vacuum robots)
- 扫地机文字稿 (Text draft about vacuum robots)

Below these, a blue rounded rectangle contains the user's query:

家里有老人和宠物，计划在 2026 年春节购买一台扫地机器人。我非常看重防缠绕能力和基站自清洁的彻底程度。请审计以下三份语料，告诉哪款机型最适合我？并告诉我你主要引用哪篇语料为我推荐，为什么？

#### 测试变量 (三类真实语料):

样本 A:

某专业 KOL 的长测文章：包含创作者本人的百台实测经验，有具体的防缠绕边刷对比图和基站热水洗温测数据。

样本 B:

某平台营销短文：篇幅短，虽然提到了参数，但夹杂了大量感性修饰。

样本 C:

AI 生成文本稿，逻辑完美，涵盖了「越障、纤薄、智能交互」三大趋势，但缺乏一手实测的排雷细节。

实验结果

样本类型	审计结论	弃用/引用原因
<b>A</b> 专业创作者 长图文	第一顺位引用	引用：提供了「气旋导流静音滚刷」和「100°C高温煮洗」的实测验证，直接关联了宠物毛发与卫生需求。
<b>B</b> 某平台短营 销文	弃用	弃用：被判定为信息高度稀释。虽然提到了 20000Pa 吸力，但部分用词（如小白抄作业）无法作为逻辑证据。
<b>C</b> AI 文	辅助参考	边缘化：虽提及大模型趋势，但在防缠绕的具体效果上缺乏一手增益，被视为常识复读机。

通过还原 AI 模型对这三篇文章的后台评语，我们总结出内容资产生存的三大关键：

事实密度判定：样本 B 虽然在传统搜索下有曝光，但在 AI 审计中，其营销词汇占比较多却未产出决策价值，直接触发了算法惩罚。

信息增益溢价：样本 A 胜在提供了「其他信源没有的细节」，例如「科沃斯 T80 滚筒洗地对厨房湿污渍的去污效果更好」。这种一手实测的差异化是 AI 决定引用的核心动因。

结构化：样本 A 包含的“核心机型参数对比表”极大地降低了 Agent 的提取成本。

感谢阅读

