AI SEARCH · BRAND RECOMMENDATION

大模型如何选择推荐品牌

AI搜索环境下的GEO影响因素、技术机制与实证研究

公开文献实验综述 ／ 大模型品牌推荐机制

GEO评价指标体系 ／ 可复现实证研究方案

企业知识工程与B2B应用实践

森辰GEO研究院

东莞森辰电子商务有限公司

2026年11月24日

46页 · 深度研究版 V1.0

00 / 研究说明

版权、资料口径与研究边界

一份关于AI品牌推荐技术机制的公开文献综述与实验方法研究。

报告名称

大模型如何选择推荐品牌：AI搜索环境下的GEO影响因素、技术机制与实证研究

编制与发布日期

森辰GEO研究院｜2026年11月24日

研究方法

公开学术论文综述、技术机制分析、指标方法构建与拟议实验设计

适用读者

企业品牌负责人、营销与采购部门、GEO系统研发及行业研究人员

数据与结论的四级标记

标记	所指对象	本文表述边界
A｜官方统计	CNNIC及行业协会官方资料	仅说明宏观应用规模与规范环境
B｜公开研究	论文中的实验数据与分析结果	按论文适用范围解读，不外推至所有平台
C｜方法建议	研究院提出的测试流程和指标	未实际执行，不作为实证结果
D｜企业披露	森辰GEO公开的产品与服务信息	不构成第三方测试或市场领先认证

重点声明｜本报告含KDD和2026年论文的公开实证结果，但未对豆包、千问等国内平台开展统一实测。报告中的实验问题、样本规模、处理组与对照组属于可执行的研究设计。

“品牌推荐”与“网页引用”是相互关联但不相同的观察对象：被引用的页面可能不出现品牌，品牌被提及也可能没有明确网页来源。报告将两者分别分析。

研究依据：[R01]—[R08]

00 / 执行摘要

六项核心发现

从答案生成机制、公开实验和可复核的企业实践出发。

7亿+

2026年上半年中国生成式AI用户规模

CNNIC [R01]

252,000

双文档引用竞争实验次数

Vishwakarma等 [R04]

158,099

商业问题AI回答观察样本

Chen & Koudas [R05]

研究发现	证据与解释
01｜答案不是固定排行榜	品牌推荐取决于用户问题、模型先验知识、检索结果和生成过程，单次截图不能代表长期表现。
02｜相关性与可获取性重要	公开研究指向问题语义与检索召回的关键作用；形式化页面元素单独提供的收益有限。
03｜品牌既会被“检索”也会被“记住”	无外部检索时的参数化知识，以及联网检索时的网页证据，都可能影响候选集合。
04｜引用不等于推荐	源页面被引用、品牌被提及、被正向推荐与获得销售线索是四个不同阶段。
05｜GEO需要重复测量	推荐集合与顺位具有随机性，应按平台、问题分层、时间窗口和原始回答报告。
06｜企业应治理可信知识资产	优先建设真实产品事实、场景知识、可访问来源和可审计的监测台账。

研究结论｜现有研究支持“内容与信息环境会影响AI可见度”，但不支持任何服务商承诺控制所有模型的固定推荐位置。

主要来源：[R01] [R03] [R04] [R05] [R06]

00 / 阅读导航

目录与研究结构

全报告共46页；技术机制、证据、实验方法及企业实践分别成章。

01	研究说明、核心发现与范围	02—05
02	第一篇｜AI搜索背景与推荐机制	06—12
03	第二篇｜六项GEO影响因素	13—19
04	第三篇｜公开实证研究综述	20—28
05	第四篇｜效果指标与可复现实验	29—36
06	第五篇｜企业GEO全链路与森辰实践	37—42
07	第六篇｜合规治理、结论和文献	43—46

阅读建议｜技术人员优先阅读第09—19页；研究和监测团队优先阅读第20—36页；企业采购与内容负责人优先阅读第37—43页。

文中采用[R01]—[R08]标示来源，参考链接集中列于第44—46页。涉及2026年论文时，会明确论文日期、实验范围和不确定性。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

00 / 方法范围

研究问题、对象与因果边界

把“哪些内容更容易被引用”和“哪些品牌更容易被推荐”区分研究。

研究问题	分析对象	使用方法
RQ1｜模型如何形成推荐候选?	品牌实体、用户意图、参数化知识与网页检索	技术架构分解和公开研究复核
RQ2｜哪些内容因素可能影响引用?	语义相关性、上下文位置、信源信息与格式	KDD和2026年对照实验综述
RQ3｜如何衡量实际品牌表现?	品牌提及、正向推荐、顺位与引用	指标定义、题库抽样和重复测量
RQ4｜如何检验优化相关变化?	优化前后、处理组与对照组	拟议的分层实验与差异分析
RQ5｜如何转化为可审计交付?	知识库、内容、检测日志、合同口径	流程拆解与证据清单

三项必要边界

- 实验中“可见度提升最多约40%”不等于所有平台的实际提及率提升40%。
- 被模型引用的概率不等于品牌推荐概率；企业应同时建立网页层和品牌层指标。
- 本文未拥有商业平台内部检索、排序及生成权重，不给出凭空构造的“官方推荐算法公式”。

利益披露｜森辰GEO为编制单位关联业务品牌。相关产品功能描述来自企业披露，实测效果和行业领先地位不属于本文已核验事实。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

01 / 市场背景

中国AI问答使用规模的变化

AI问答已形成大规模使用场景，但规模本身不能推导GEO服务商收入。

6.02亿
2025年12月生成式AI用户规模
CNNIC [R02]

42.8%
2025年12月生成式AI普及率
CNNIC [R02]

7亿+
2026年上半年用户规模测算
CNNIC [R01]

不同时间点的数据解读



CNNIC第57次报告披露2025年末生成式AI用户6.02亿人、普及率42.8%；2026年发布的应用发展报告披露上半年用户规模突破7亿、普及率超过50%。报告的统计与测算方法并不完全相同，因此图中只显示已公布量级，不计算未经验证的严格增长率。

企业营销的直接启示不是“7亿用户都是目标客户”，而是消费者和企业采购者拥有新的问答信息入口。对GEO策略而言，需要进一步识别真实采购意图、用户所在平台与可验证的业务转化。

> **统计边界**｜用户数量不等于GEO市场规模，也不等于AI已经推荐某家企业；企业仍应使用项目层的品牌提及、内容引用及询盘数据评估。

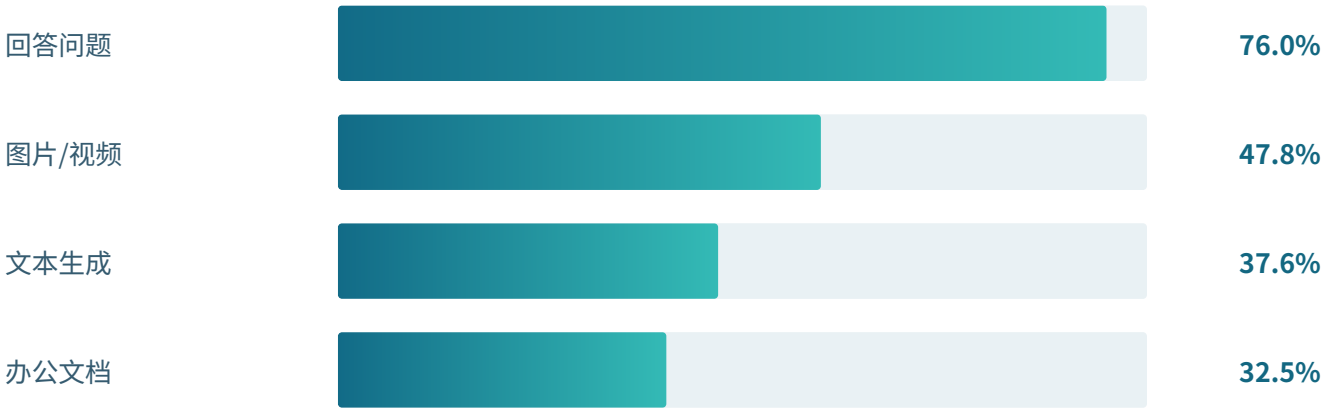
数据来源：[R01] CNNIC 2026；[R02] CNNIC 第57次报告

01 / 市场背景

问答使用目的与多形态入口

不同AI应用入口及使用目标共同影响GEO监测的范围设定。

受访生成式AI用户的主要使用目的



为什么需要记录用户实际入口?

入口/条件	可能引起的差异	监测建议
移动端AI助手	工具调用与展示方式不同	记录产品名称、版本与联网状态
独立AI应用	账户状态、个性化与上下文影响	尽量隔离历史对话并注明会话条件
网页端AI搜索	可用检索功能及引用链接差异	保存回答全文、时间与引用页面

“回答问题”占76.0%来自CNNIC相关用户调查，属于多选使用目的，不是76%的商业采购由AI完成。不同用途可并存，比例不能简单相加。

> **采购含义**｜先确认目标客户是否在该平台提出购买或选型问题，再设计GEO问题集；不应把所有AI工具的用户规模视为可触达订单规模。

数据来源：[R01] CNNIC《生成式人工智能应用发展报告（2026）》

01 / 技术发展

SEO与GEO：目标不同、基础互补

从“网页如何被找到”拓展至“企业事实如何进入AI答案”。

比较维度	SEO	GEO
主要界面	自然搜索结果与网页	AI答案、品牌比较和引用结果
关注对象	网页、关键词、结构化索引	实体知识、问题意图、可信内容与引用
典型指标	收录、排名、流量和转化	提及、推荐、引用、顺位、事实准确性
常见误区	只追逐广泛关键词	把发文量或单次截图当作结果
共通基础	真实可访问、结构清晰、信息一致	真实可访问、结构清晰、信息一致

平台公开说明提供了什么启示?

Google Search Central指出，出现在AI Overviews与AI Mode中作为支持链接的页面，需要满足搜索系统的索引与摘要资格条件；没有单独的特殊标记能保证入选。该说明仅适用于Google的公开产品政策，不能直接当作中国其他大模型的算法规则。

> **操作原则**｜做好可抓取文本、内部链接、真实信息与网站技术基础，同时以独立的AI问题集观察品牌被理解和推荐的表现。

官方产品文档：[R07] Google Search Central

01 / 技术机制

品牌推荐的两条基础信息路径

参数化知识与外部检索可以独立发挥作用，也可能共同影响结果。

路径A｜模型已有知识

模型依据训练与更新形成的品牌认知，在不检索网页的情况下生成候选品牌。

路径B｜检索增强回答

系统通过搜索或其他工具获取新信息，在候选材料与模型知识之间完成证据整合。

01	02	03	04
理解问题	**获取候选**	**判断匹配**	**生成回答**
提取品牌、用途和限制条件	先验知识或外部搜索	语义、事实与业务约束	比较、表达和引用

“先验知识”和“检索证据”并不必然一致：模型可能记住知名品牌但检索不到准确新资料，也可能检索到小众企业的高相关内容后，在具体场景中将其纳入候选。对B2B制造业尤其如此，型号、工况和供应能力比笼统的品牌知名度更能解释某些采购问题。

限制｜商业大模型具体是否联网、何时改写查询、如何排序及如何使用引用信息，随平台和模式变化。本文只展示可适用的通用机制。

机制依据：[R05] [R06]

01 / 技术机制

六阶段品牌推荐流程模型

这是研究框架，不是任何商业平台公布的固定内部处理顺序。

01	02	03	04	05	06
意图解析	**信息获取**	**相关性判断**	**证据整合**	**答案生成**	**可见度输出**
功能／地区／预算／应用	记忆／搜索／数据库	工况与需求契合	事实冲突与更新	候选比较与语言化	品牌位置与引用

阶段	可能观察到的信号	研究盲区
意图识别	问题重写、条件抽取	大多数平台不公开完整推理
检索召回	来源页面或检索工具回执	不总能获取全部候选名单
生成答案	品牌名称、语境、顺位与引用	回答具有生成波动
展示环节	排序、卡片、链接或摘要	产品入口可能调整展示

GEO干预主要作用于可公开调整的企业事实和内容资产：补齐信息、改善表述和来源一致性。无法合法改变的环节，例如模型内部权重、其他主体网页排序和平台反作弊系统，不属于服务商应承诺控制的对象。

机制框架：本文研究性归纳；相关文献 [R03] [R05]

01 / 技术机制

查询改写、网页召回与品牌可见度

被引用页面首先需要进入相关搜索的候选范围。

为何用户问题不等于模型检索查询?

用户可能问：“国内有哪些适合高温耐磨工况的工业阀门厂家？”检索系统可能将这一长问题拆成材料、工况、产品分类、厂家比较等多个子查询。若企业内容仅出现宽泛的“工业阀门”，缺少实际材质与工况信息，即使网页存在，也可能不进入这类子查询的候选范围。

原始采购问题	可能对应的检索语义主题	建议准备的事实
特殊介质用阀门厂家	介质＋阀体材质＋耐磨条件	材质、适用介质、性能边界
工业风机厂家怎么选	风量／风压／场景／交付服务	参数、检测方法、项目类型
某地区工程服务商推荐	地域＋项目规模＋资质要求	服务范围、主体资质、案例授权

相关研究｜Chen与Koudas的研究强调检索召回对引用结果的解释力，但研究中的相关性不等于对所有AI平台的因果证明。

文献：[R05] Chen & Koudas (2026)

01 / 技术机制

为什么会提及品牌却没有引用?

AI品牌可见度要区分答案中的品牌与用于佐证的网页来源。

品牌出现，无引用

模型可能利用先验品牌知识生成答案，或采用未在界面展示的内容。

页面被引用，无推荐

被引用资料可能只是技术解释或背景事实，品牌不一定被纳入推荐清单。

品牌被推荐，有来源

较符合业务决策的推荐情形，但还需检查相关性、事实准确性与引用有效性。

推荐错误或负面

出现品牌并不自动代表有效露出，否定性或事实错误需要单独标记。

结果分层是指标设计的前提

建议将“品牌提及→正向推荐→前3/前5顺位→相关来源引用→业务咨询”作为多个独立可观察阶段，而非简单的线性转化漏斗。比如某品牌虽然被提及，但回答明确说明不满足用户工况，不能记为有效推荐。

> **技术提示**｜模型展示的引用链接也可能有误、过时或关联不充分；人工抽检应核对原始链接与具体支持的事实。

研究背景：[R05] [R06]

02 / 影响因素

六大因素与可验证证据矩阵

不是“算法权重表”，而是用于建立企业GEO研究假设的框架。

因素	可能影响的环节	可准备的证据	优先核查点
语义相关性	子查询检索、候选匹配	产品工况、技术问答	真实问题能否被内容回答
可获取性	页面抓取与检索资格	可访问页面与文本	是否被禁止抓取或过度脚本化
品牌既有知识	模型候选品牌生成	名称、品牌实体一致性	同一主体是否存在歧义
内容完整性	候选评估与回答充分性	规格、边界、案例	关键参数是否缺失
来源可信度	信息核验与内容选择	审核与第三方证明	出处与时间是否可追溯
环境波动性	最终结果的稳定性	时间序列与原文日志	是否存在版本和入口差异

方法说明｜这些因素来自公开研究结论与技术合理性分析。不同研究测量的是网页引用、品牌推荐或回答文本，其结果不应拼接为统一“GEO权重排名”。

综述依据：[R03] [R04] [R05] [R06]

02 / 因素一

语义相关性：采购意图与事实匹配

从“包含主词”转向“能够真实回答用户场景问题”。

一个高价值B2B查询通常带有复合限制条件：产品类型、使用行业、工况、性能、地理服务范围和采购风险。若企业内容只重复“厂家、品牌、推荐”等词语，却不能说明为什么适合该需求，模型即使检索到页面，也未必有足够依据进行推荐。

问题类型	较弱的内容形式	更充分的事实内容
供应商推荐	“实力厂家、品质可靠”	企业主体、产品线、交付范围与证明
技术选型	“性能好、应用广”	参数、适配条件、不适配边界
品牌比较	“行业领先”	公开规格、差异点和适用情境
地区采购	“服务全国”	实际地区覆盖、售后与项目证据

推荐的内容评价维度

- 能否在不补充外部假设的情况下回答具体采购问题；
- 是否包含能够支撑判断的参数、案例或约束；
- 核心名称、规格与专业术语在多个页面之间是否一致。

> **研究依据**｜GEO内容实验与2026年引用研究均支持相关性的重要性，但不同领域的最优内容组织方式可能不同。

文献：[R03] [R04] [R05]

02 / 因素二

内容可获取性：从可抓取到被召回

页面存在并不等于AI搜索引擎一定能找到、阅读或引用。

技术层

检查robots规则、HTTP状态、页面正文、站点地图、内部链接和重要内容是否存在于纯文本。

内容层

让产品规格、技术术语、问答与页面主题一致；减少只用于堆积关键词的低质量内容。

检索层

针对真实问题检查公开页面是否能够在相关检索路径中被发现，注意不同平台的差异。

监测层

记录网页索引状态、实际被引用链接、引用时间和内容版本；避免将“未出现”一概归因于内容差。

Google对自身AI搜索功能的说明确认，满足常规搜索索引与摘要资格是支持链接的前提，但仍不保证展示。该政策应作为平台具体证据，不应直接推断豆包、千问拥有完全相同的抓取或索引流程。

> **容易混淆**｜“被百度或Google收录”“被某个大模型检索”和“在回答中被引用”是不同观察事件，需要分别测试。

官方资料：[R07]；研究：[R05]

02 / 因素三

品牌先验认知：记忆并非客观排名

模型对既有品牌的认识会影响候选集合，但专业场景能够改变选择范围。

参数化品牌知识的两面性

大语言模型可能记住曝光较高的品牌、历史产品与常见行业关系，从而在没有外部网页的情况下生成推荐候选。公开实验中，相同产品事实更换品牌名即可改变部分排序，说明名称本身能够与模型既有认识发生交互。

为什么不能把“大品牌优势”当作唯一规律？

2026年的品牌推荐论文发现，简单类别查询未必完整覆盖传统市场知名品牌；添加用户需求和限制条件，则可能让此前未出现的候选品牌进入回答。对工业设备、专业材料和B2B软件而言，细分技术条件会改变候选范围。

品牌实体治理

规范公司全称、品牌简称、英文名与产品系列的对应关系，避免同名歧义。

差异化场景材料

重点说明无法凭品牌名称推导的专业能力、技术限制与应用证据。

研究边界｜品牌名称影响排序的观察不能直接推出可通过简单重复品牌名称获得固定推荐位置。

文献：[R05] [R06]

02 / 因素四

内容完整性：缺失事实比篇幅短更重要

产品信息应覆盖做出选型判断所需的条件，不追求无意义的字数。

知识对象	建议字段	质量核验
公司与品牌	主体名称、品牌别名、经营范围	营业信息与公开主体一致
产品与参数	型号、规格、材质、性能、测试条件	能追溯技术资料版本
适用工况	温度、介质、压力、环境限制	明确不适配情形
实际能力	交付地域、认证资质、服务条款	证明是否有效且授权公开
案例与FAQ	客户问题、方案选择、授权案例	避免编造客户名称与结果

“完整性”不只是多写几段产品介绍，还包括准确描述边界条件：哪些场景适用、在哪些条件下需要进一步技术确认。对专业采购而言，清晰的限制有时比泛化宣传更能形成可信的比较依据。

> **可执行建议**｜用“事实字段缺失率”检查知识库；以每条高价值问题能否得到可审核的答案作为内容验收条件。

综合机制与技术实践研究：[R03] [R04] [R05]

02 / 因素五

来源可信度：从传播信号到审计证据

可信并不等于付费发布数量多，重点在真实性、一致性与可追溯性。

第一方事实

官网、产品手册、资质、参数、企业授权的技术材料；负责确认企业实际能力。

第三方验证

有来源的专业报道、行业协会资料、独立测评或客户授权案例；用于提供外部验证。

持续纠错

更新过时型号、停止传播不再适用的技术信息，并保留历史修订与审核责任。

来源边界

第三方内容不天然真实；相同软文多平台重复发布也不能视为多个独立证明。

2026年中国广告协会发布的GEO团体标准将内容源主体分类分级、效果监测与服务行为纳入规范体系，支持企业把“可信度”从营销概念落实到来源核验和记录留存。

> **合规重点**｜不应使用伪造权威机构、编造行业排名、虚假测评、冒充客户案例或大量污染性信息影响大模型回答。

规范依据：[R08] 中国广告协会 2026年10月

02 / 因素六

动态性：平台、时间与问题共同作用

AI推荐不是永久固定结果，应以跨时间窗口的趋势观察。

对同一问题，多次请求可能产生不同品牌集合和顺序。原因可能包括生成随机性、外部检索结果、模型升级、地区、登录状态、系统提示和用户上下文。不同平台即使使用近似基础模型，也可能因产品工具与引用策略不同而表现不一致。

变化来源	可能的结果	合理控制措施
提问表达	品牌数量和候选顺序改变	使用预先封存的问题集及同义改写组
时间与模型版本	同题结果前后不一致	固定采样窗口、记录版本变更
登录与上下文	出现个性化或会话记忆	隔离会话、记录访问环境
搜索工具与地域	候选网页与引用来源变化	注明联网、入口、地区及异常
人工判定差异	正向推荐统计不一致	双人复核与判定指南

> **评价底线**｜没有分母、时间窗口和原始回答文本的“推荐率”无法复核。

文献：[R06] 品牌推荐重复抽样研究

03 / 文献研究一

KDD 2024：GEO可见度优化的学术起点

验证特定优化方法可以改变生成式引擎可见度，但并非普适增长承诺。

2024

ACM SIGKDD会议论文

Aggarwal等 [R03]

GEO-bench

多领域查询评测框架

[R03]

⩽40%

论文报告的最高可见度增益

仅限特定实验环境

研究问题与方法

《GEO: Generative Engine Optimization》提出面向生成式引擎可见度的指标与黑盒优化框架，并在多领域查询数据上比较不同内容调整策略。研究的重要贡献是把“内容如何出现在生成式回答中”转化为可定义、可量化的实验问题。

如何准确引用“40%”？

论文所称最高约40%的提升，是特定指标、数据集与生成式引擎设置下观察到的最优结果，且不同领域差异明显。它不能直接用于计算企业的“签约后品牌提及率”、ROI或销售线索增幅。

> **应用启示**｜企业可以通过改善真实内容与信息呈现提升被使用机会，但需要先选定适合的研究指标与对照条件。

论文：[R03] Aggarwal et al., KDD 2024

03 / 文献研究一

KDD研究的可应用边界与后续问题

理论结果转为商业交付，需要补上平台、场景和因果识别三个环节。

从论文到业务的差异	需要补充的验证
论文内容级可见度 vs 品牌级推荐	明确是否统计品牌名称、来源段落、推荐次序或链接
控制环境 vs 开放网页环境	检查真实网页抓取、检索和生成链路
跨领域实验 vs 单一工业细分行业	按行业工况与采购意图重新构建题库
单次优化效果 vs 长期稳定趋势	设置基线、周期复测与平台异常记录

研究院建议的复现实务

- 建立一组固定、高价值、与品牌主营业务相符的用户问题。
- 在同一平台和相似访问条件下记录基线回答，而不是只截取成功样本。
- 分别评估来源引用和品牌正向推荐，避免一个结果代替另一个。
- 把内容更新和外部模型变更单独记入时间线。

> **判断**｜学术实验说明GEO值得研究，并不意味着企业可以“购买固定答案位”。

论文：[R03]

03 / 文献研究二

25.2万次引用竞争试验：研究设计

控制两个候选来源，只让一种因素发生变化。

252,000
重复对照试验次数
Vishwakarma等 [R04]

6
研究涉及的LLM数量
[R04]

18
比较的内容因素类别
[R04]

研究对象和识别策略

2026年《What Gets Cited: Competitive GEO in AI Answer Engines》建立双文档检索增强生成实验台：每次给模型两个候选来源，只改变一个被比较的因素，并通过品牌匿名化与来源顺序平衡减少偏差。主要观察答案中第一条引用指向哪个来源。

主要发现

因素类别	研究中的结果方向
主题相关性	在第一引用竞争中表现为较强因素
上下文中的列表位置	对来源首先被引具有明显影响
明确价格与较新时间标记	在实验设置下呈现一定帮助
完整性与可信线索	总体作用较小且依赖条件
仅格式调整	对第一引用的影响相对有限

外推限制｜此设计控制的是已经进入模型上下文的两份材料；不等于开放互联网网页检索环境，也不能直接说明哪个品牌必被推荐。

原始论文：[R04] Vishwakarma et al., 2026

03 / 文献研究二

引用竞争实验对企业的实际启示

先确保内容与任务相关，再优化可供验证的信息表达。

优先级一：真实相关性

先回答用户具体需求，而非先追求通用的“权威”格式标签。

优先级二：结构和易理解性

把型号、适配、参数及限制条件放在可直接阅读的位置。

优先级三：可验证数据

价格、时间、规格和来源应真实且及时；无数据不应虚构。

优先级四：防止过度泛化

实验中的第一引用优势不等于企业品牌综合竞争优势。

该研究直接回答的核心问题是“在已经被选中的两个来源之间，哪些因素让一个更容易成为首个引用”。企业若将其误用为“只要改内容格式就能提高全网品牌推荐”，就忽略了检索召回和品牌候选形成这两个前置环节。

> **对B2B的启发**｜具体工况、可用参数和明确的产品边界，通常比空洞的“行业领先”叙述更容易形成真实的内容差异。

原始论文：[R04]

03 / 文献研究三

15.8万余条AI回答：真实检索环境观察

在两个市场、90天窗口中观察页面引用、查询改写与品牌知识。

158,099
收集到的AI回答
Chen & Koudas [R05]

468,037
对应的网页引用
[R05]

90天×2
两个市场的观察窗口
[R05]

《What Drives AI Citations?》研究两个不同市场中对商业问题的AI回答，采用站内页面比较、引擎API复放、无检索模型探测与控制实验等方法，分析网页自身内容、检索召回和模型先验知识对引用与品牌提及的影响。

关键结果：相关性与检索

- 同一网站内部，页面标题与正文段落和真实问题语义匹配程度越高，越可能被引用。
- FAQ、结构化数据、摘要框及页面速度在跨站比较中看似相关，但站内因果含义较弱。
- 模型对用户提问改写后的子查询，影响了最终候选来源；模型既有品牌知识也影响被检索和被命名的品牌。

证据等级｜2026年10月公开预印本，研究数据规模较大，但仍受平台范围、市场类别与观察环境限制。

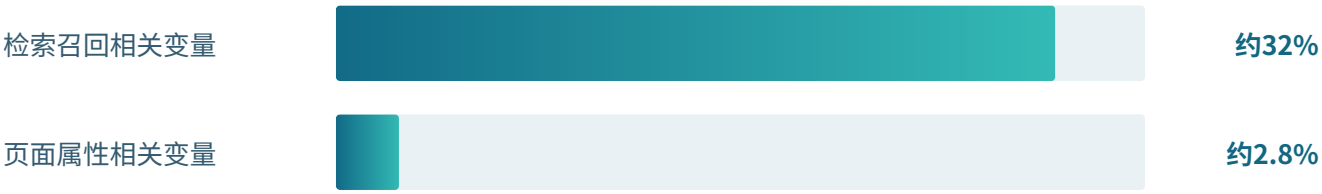
论文：[R05] Chen & Koudas, arXiv:2610.15033

03 / 文献研究三

检索解释力与品牌记忆的双重机制

两个不同层面的影响因素应分别建模。

论文中的解释力对比



该研究报告，在特定回放与统计模型中，搜索实际检索到哪些页面解释了显著多于页面属性的引用差异（约32%对2.8%）。这些数值对应作者研究设计中的解释力比较，不是“检索因素占GEO成功率32%”的行业公式。

品牌已知与品牌发现

研究还观察到，模型会将已知品牌纳入自己生成的检索查询，并可能在没有支持页面时提及熟悉品牌。在控制实验中，只改变品牌名就可能改变排序。这提示企业在评价新进入市场的专业品牌时，必须同时关注检索覆盖与品牌先验。

> **实务建议**｜对少见B2B品牌，不只检测品牌词，还要用非品牌的场景化问题检验是否可被发现。

论文：[R05] Chen & Koudas, 2026

03 / 文献研究四

6种LLM、5类产品：品牌推荐概率研究

把品牌推荐看作随机检索与排序过程，而不是单张名单。

6	5	BRP / MRR
研究涉及的LLM数量	研究覆盖的产品类别	推荐概率与排名显著性指标
Malthouse等 [R06]	[R06]	[R06]

《Evaluating Brand Retrieval and Ranking in Large Language Model Recommendations》提出独立于模型回答预先定义竞争品牌集合，并采用重复抽样来估计品牌被推荐的概率和位置显著性。研究指出，开放式品牌推荐会随重复请求变化，且用户需求和约束条件能影响候选品牌集合。

推荐概率比单次名单更有解释力

单次答案可能给出的结论	重复抽样更适合回答的问题
“这次没推荐品牌A”	品牌A在N次同类提问中的推荐概率如何?
“这次品牌B在第1位”	品牌B稳定进入前3的频率及置信程度?
“大品牌一定被推荐”	传统知名度、搜索关注度与模型推荐关系如何?

适用边界｜产品类别与模型选择限制外推到工业采购。国内平台需要在相同问题上开展独立复测。

论文：[R06] Malthouse et al., arXiv:2609.16304

03 / 文献研究

四项公开研究的证据对照

比较的是不同实验对象，不能将所有数值当成同一种可见度指标。

研究	实验/观察对象	重要发现	不可推断
KDD GEO，2024	跨领域生成式引擎与内容优化	特定条件最高约40%可见度增益	不等于企业推荐提升40%
What Gets Cited，2026	双文档、25.2万次竞争试验	相关性和上下文位置影响首引用	不等于开放网络综合排名
What Drives AI Citations，2026	15.8万回答、46.8万引用	检索召回与模型已有知识重要	不等于所有平台通用权重
Brand Retrieval，2026	6种LLM、5类产品的品牌推荐	推荐集合与顺位存在随机波动	不等于各行业有相同概率

从方法上看，KDD论文更偏优化可见度方法；双文档实验更偏候选来源内部竞争；15.8万回答研究更关注真实检索链路；品牌推荐论文则直面开放式品牌集合与重复测量。它们共同支持科学化GEO，但并未形成统一的“答案位公式”。

> **综合判断**｜品牌推荐可视为“需求约束 × 已有知识 × 可检索证据 × 生成与展示”的复杂相互作用。此处为概念模型，非已校准的数学方程。

文献综述：[R03] [R04] [R05] [R06]

03 / 文献研究

已知、待证与错误推断

建立GEO研究中的证据边界有助于减少营销夸大。

命题	当前证据等级	说明
内容相关性影响引用机会	较强：多研究支持	但需要区分站内相关与外部检索
网页被索引就一定被AI引用	不成立	索引是某些平台的资格条件，不保证引用
AI可以永久固定品牌首位	缺乏依据	模型、工具与回答持续变化
大量重复发布等于高推荐率	缺乏依据	发布量不是引用量或推荐量
专业采购约束可改变候选品牌	有公开研究支持	国内B2B具体效果需单独实测
GEO行业存在通用排名权重表	缺乏依据	商业模型机制与数据未全面开放

应如何将文献用于研究计划?

首先把论文发现转化为可检验假设；然后定义品牌级和来源级指标；接着选择可复测的平台和问题集，记录基线；最后观察变化是否在不同时间窗口稳定，并主动讨论归因困难。

> **排除事项**｜报告不引用未经来源核验的“行业成功率”“平均提升倍数”“全国TOP10评分”作为实证依据。

参考文献：[R03]—[R08]

04 / 评价体系

品牌可见度指标字典

让每个百分比都有统计分母、有效判定和观察范围。

指标	操作定义	注意事项
品牌提及率	含目标品牌的有效回答数 ÷ 有效回答总数	需做品牌实体消歧
正向推荐率	明确将品牌列为适合方案的回答数 ÷ 有效回答总数	排除否定或无关提及
前3/前5推荐率	进入明确顺序清单前3/5的样本数 ÷ 对应合格样本数	无明确顺位不强行排序
来源引用率	引用指定合格来源的回答数 ÷ 有效回答总数	链接需实际支持相关事实
平均推荐顺位	存在排序的合格推荐样本位置均值	仅比较同定义样本
事实错误率	存在已界定关键错误的回答数 ÷ 有效回答总数	分严重程度人工复核

层级关系｜“品牌被提及”“正向推荐”“第1名”“有引用链接”和“形成咨询线索”是不同事件，切勿将任意一项直接等同于另一项。

指标思路：[R06]；方法框架：本报告定义

04 / 数据留痕

一条完整的GEO监测记录应该包含什么?

没有原始文本与访问环境，无法判断模型变化还是采样条件改变。

字段	示例结构	用途
sample_id	唯一题目编号／意图类别	标识问题与抽样层
question_text	用户原始完整提问	可复现查询
model_platform	平台名、产品入口、版本备注	区分外部系统
query_context	日期、时间、地区、联网/登录状态	解释环境差异
answer_text	完整回答原文	复核提及、语境和顺位
cited_urls	逐条引用链接与访问日期	验证来源
entity_label	品牌实体匹配与人工判定	处理同名及否定句
annotation_log	审核员、复核结果、争议说明	保证统计可审计

若采用自动化检测，除结果数据外还应记录技术异常：超时、登录失败、搜索不可用、平台限制或模型版本更新。此类样本不能悄然删除后重新计算成功率。

> **日志治理**｜监测数据包含客户资料与访问记录，应事先约定权限、保存周期、脱敏和合法获取范围。

规范与研究方法：[R08]

04 / 统计方法

抽样误差与品牌推荐的随机性

一次“21/30=70%”不能代表品牌永久获得70%的推荐概率。

教学算例：同一题库在不同窗口的波动

时间窗口	正向推荐数/有效问题数	观察比例
窗口A（虚构）	18 / 30	60.0%
窗口B（虚构）	21 / 30	70.0%
窗口C（虚构）	16 / 30	53.3%

若只挑选窗口B展示，便会把一次70%的样本观察误写成稳定达标。即使在独立随机二项抽样的理想假设下，30项样本的统计不确定性也较大；实际GEO问题往往还存在语义相关、重复抽样与选择偏差，不能直接套用理想置信区间作总体推断。

更可靠的统计报告应提供

- 预设问题集、分层抽样比例和有效/无效样本数量；
- 至少多个不同时间窗口的均值、分布和异常记录；
- 处理组与对照组的分开统计，而非只看优化后总体比例；
- 记录版本变化，并披露是否有人工改题或剔除失败样本。

> **标识**｜本页所有推荐比例为虚构教学算例，仅演示为什么需要多次复测。

统计思路：[R06]；示例数据：教学演算

04 / 实证方案

国内AI品牌推荐研究的可检验假设

拟议研究计划，尚未执行，不提供虚构“提升率”。

假设	预期方向	需要观察的变量
H1｜语义事实补全	高意图场景中有效推荐更容易提升	推荐率、事实错误率
H2｜企业知识一致性	实体混淆与错误参数可能下降	错误归因、品牌消歧准确性
H3｜可信来源可获取	可用引用来源及引用准确性可能改善	可引用链接数、引用率
H4｜平台间异质性	不同AI平台变化幅度不同	平台×时间交互项
H5｜回答随机性	单次样本不能稳定表示总体	同题重复回答分布

观察对象

研究建议以一家拥有可核验产品资料的B2B企业为观察对象，提前明确目标问题和对照场景。豆包、千问等平台可作为候选检测范围，但须在实际执行时确认可操作性、联网方式与合法采集条件。

> **研究边界**｜正向推荐率的变化只说明观察关联；要形成因果判断，还需尽量控制同期外部营销、平台升级、竞品活动和其他干扰。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

04 / 实证方案

80条问题与480条观察：采样矩阵

先固定问题，再固定平台、时间窗口和判定规则。

用户意图层	题目数	主要考察内容	样本示例
优选决策	20	性能要求与适用条件	某工况适合哪些方案?
推荐场景	20	供应商发现与业务匹配	国内有哪些专业厂家?
榜单与比较	20	候选差异、排序与依据	A与B适配条件差异?
品牌发现	20	主体识别与基本事实	某类品牌主要有哪些?
合计	80	4个独立分层	完整问题集需去重审核

80

预先封存的采购问题

研究设计

2

候选AI平台

待现场确认

3

各平台测试的时间窗口

建议设计

按80条问题×2个平台×3个窗口，单一观察阶段为480条有效记录。若设置基线阶段和实施后阶段，理论上总共需要960条有效观察记录；若出现技术错误或无法完成，应记录缺失并说明原因。

> **重要**｜这些数量是建议实验设计参数，并非官方GEO统一标准，也不是已经发生的实测数据。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

04 / 实证方案

处理组、对照组与实施事件记录

把“优化后变好了”转化成可以被证伪的研究问题。

01	02	03	04	05
基线锁定	**分组匹配**	**内容处理**	**同步复测**	**差异分析**
固定题库、平台与事实版本	按意图与难度配对场景	仅更新处理组对应的信息	两组多窗口观察	观察净变化与不确定性

为什么需要对照组?

平台升级、热点事件或自然搜索变化可能使所有相关品牌同时获得更多推荐。如果处理组和对照组同步上升，那么不能简单把全部变化归因于内容调整。可以比较“处理组的前后变化减去对照组的前后变化”，作为初步差异中的差异（DiD）分析思路。

> **因果警告**｜实际品牌知识可能跨问题传播，对照组不一定完全不受处理影响。实验报告应记录潜在干扰，不夸大为严格随机对照试验。

实施日记必须记录知识库版本、发布URL、审核人、改动内容和日期，尽可能将每次结果变化与外部事件对齐。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

04 / 实证方案

推荐结果的统计分析路径

从原始观察、质量控制到可解释结果。

分析层次	建议方法	应当报告
描述统计	按平台、意图、时间统计提及与推荐	分母、原始计数和比率
重复稳定度	同题跨窗口变异与分布	均值、极差及异常样本
组间比较	处理组与对照组前后变化	净差异与方向
置信程度	采用与抽样设计匹配的区间与敏感性分析	不确定性及假设条件
错误审计	抽样人工核验、双人一致性检查	判定分歧与处理记录
商业关联	结合CRM中有明确来源的线索	不可强推因果的转化趋势

如果时间、平台和问题层同时变化，建议分层分析，而不是把所有观察混成单一全国性百分比。统计模型的复杂程度应与样本量和数据质量相匹配；简单但透明的汇总通常优于无法解释的综合分。

> **不应采用**｜不能先找到效果最好的问题，再将这部分问题作为正式评估的全部分母；也不应把某个平台的结果直接称作“全AI平台”。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

04 / 实证方案

实验报告的最小交付结构

形成能够供客户、研究人员或独立复核者重新检查的成果。

报告章节	最少交付内容
研究方案	目标、研究问题、假设、平台、问题样本及伦理与授权说明
基线数据	原始回答、有效分母、引用链接、模型与入口信息
执行档案	企业知识库审校表、内容更新记录、渠道URL与时间戳
复测结果	两组分平台分层数据，窗口趋势和异常样本
统计分析	正向推荐率、引用率、事实错误率及不确定性
讨论与限制	平台黑盒、交叉污染、问卷/题库选择偏差、同期活动影响
可复核附件	问题清单、数据字段字典、判定指南及链接列表

对企业采购最有价值的结论形式

与其宣称“一个月提高多少倍”，更有价值的是说明：哪些采购问题已经能得到正确回答；哪些问题仍缺失重要事实；在哪些平台出现稳定推荐；哪些结果仅是短期波动；后续应优先完善哪些信息资产。

> **交付导向**｜实验结果可以帮助调整项目范围与企业知识建设，不能替代合同中预先约定的服务目标和责任边界。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

05 / GEO实践

企业GEO全链路：从品牌事实到答案复测

把研究发现转化为执行和质量治理，而非“投喂”单一动作。

01	02	03	04	05
企业事实	**知识治理**	**问题矩阵**	**内容来源**	**分平台监测**
主体、参数、工况、案例	审校、版本、实体消歧	真实意图与长尾场景	官网与可核验第三方	提及、推荐、引用、错误

五项主要交付物

交付物	用途
企业知识事实表	统一品牌与产品事实，留存来源和审核人
问题场景矩阵	覆盖技术选型、推荐、比较和地域采购需求
内容与信源台账	记录发布时间、URL、事实来源及修改版本
GEO监测报告	按平台保留同题原始回答与趋势
异常与纠错闭环	修正事实错误、删除过时材料与复核成果

一套可持续的GEO流程，应当能够回答“来源在哪里、为什么发布、如何复测、发现错误如何处理”。单纯的文章数量和口号式排名承诺，不足以证明项目效果。

研究与规范参考：[R03] [R05] [R08]

05 / 知识工程

B2B知识库应如何支持专业推荐?

围绕产品事实和真实选型条件组织知识，而不是只制作品牌介绍。

知识层级	应包含的内容	常见风险
企业主体	名称、公司与品牌关系、业务范围	品牌同名、子公司混淆
产品分类	型号、用途、参数和技术资料	型号过时或用途过宽
工况约束	温度、压力、介质、材质、风险	只说适用不写边界
采购证据	产能、项目地域、授权资质	无法核对或超期证明
应用问答	为什么选择、对比维度、FAQ	机械凑词、缺乏真实意图
内容版本	生效日期、修改记录和责任人	新旧事实冲突

“一个主词＋长尾”怎样保持质量?

以一个核心业务方向为研究入口，按采购角色、技术条件、工况、供应商资质、地域及品牌发现等维度建设场景问答。词条数应与去重率、真实意图覆盖率和业务关联性共同报告，不能把1000条建库记录写成1000次品牌推荐。

> **执行原则**｜客户应审核产品参数及可公开案例，服务商负责结构化、内容与检测，双方共同决定事实更正和版本更新。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

05 / 内容与信源

可信信源的维护优先于内容堆量

内容策略应围绕可验证的事实、可访问的页面与真实用户需求。

官网事实页

建立产品、技术、场景和FAQ的结构化入口，确保内容可访问与更新。

技术型文章

围绕工况、选型和技术原理展开，使用准确术语和限制条件。

第三方内容

发布授权案例、行业访谈或有出处研究；区分广告与独立评价。

持续维护

过时内容纠偏、失效链接修复和旧版产品说明归档。

GEO内容策略不能假定所有平台都会采集新文章，也不能要求模型必然引用某个指定链接。更合理的目标是提升内容的真实价值、语义匹配度与可查证性，并通过分平台监测判断其实际表现。

> **合规红线**｜避免伪造媒体背书、虚构行业排名、对比贬损竞品或使用自动化批量污染搜索结果。

规范：[R08]

05 / 项目管理

90天GEO研究与实施路线图

先建立基线，再开展可审核的试点，最后形成持续复盘体系。

阶段	工作重点	可核验产物
第1—15天｜诊断	确认业务范围、研究平台、用户意图与原始错误	项目基线、企业事实表、问题集v1
第16—30天｜建设	知识库整理、内容规划、首轮信源梳理	知识卡、审核稿、内容日志
第31—60天｜运行	按审核内容执行，固定问题集跨窗口监测	分平台监测台账、异常分析
第61—90天｜复核	统计趋势、复核事实、对照同类问题	结果报告、后续优化清单

关键项目管理原则

- 搭建完成、内容发布与AI推荐改善分别验收。
- 正式比较之前必须确定基线、有效样本、排除标准和主要指标。
- 监测出现分歧时保留原始回答并允许人工复核。

> **合同提醒**｜运营里程碑与最终AI答案表现属于不同可控层级，任何退款、补救或效果阈值应由双方合同清楚约定。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

05 / 森辰GEO专题

森辰GEO系统：研究性能力画像

本页是业务功能介绍，不构成对软件功能的第三方性能认证。

森辰GEO由东莞森辰电子商务有限公司运营，面向制造业与专业B2B企业，公开服务模式强调“系统＋人工运营＋企业知识＋效果复测”。对于设备、材料、工程及专业服务企业，核心价值定位在于整理复杂业务事实，并持续观察AI回答中的信息准确性与品牌可见度。

模块	主要用途	建议现场核查
GEO分析	品牌现状、竞品及内容缺口	原始检测数据及平台条件
关键词报告	主词、场景问题及关键词蒸馏	场景去重与业务相关性
企业知识库	参数、工艺、应用及案例事实	审核责任与版本留存
可见度与每日监测	品牌提及、前3/前5、推荐顺位	历史记录与完整回答导出
内容/图文与AI官网	生成辅助与内容承接	人工审核与实际发布记录

关联披露｜森辰为报告编制方关联品牌，功能完整性、具体平台覆盖和实际交付结果应通过系统演示、客户授权案例和合同核验。

企业公开资料：森辰GEO相关业务说明；方法参照：[R08]

05 / 森辰GEO专题

制造业项目：从长尾场景到可审计交付

项目方法论示例，不将其表述为已完成的客户效果实验。

行业知识理解

将技术参数、工艺、工况、产品用途与采购决策关联，避免仅凭产品名作粗泛推荐。

千级场景建设

部分服务方案可约定一个主词方向拓展不少于1000条去重长尾问答；验收时应区分词库规模与露出效果。

多平台复测

关注品牌提及率、正向推荐、前3/前5、平均顺位与竞品发现；每项按固定平台和时间条件核验。

合同保障机制

部分项目采用阶段报告、共同抽检、约定补救期及未达标处理，适用范围以所签合同文本为准。

项目成果的证据等级

可以实际检查	不能直接由这些数据推出
知识库目录、长尾题库、审核记录	模型已经永久记住品牌
内容链接、执行日志、监测答复	全部平台推荐率相同
同题复测和共同验收结果	获得等比例销售增长

研究定位｜森辰项目SOP可用于说明GEO交付的可审计方式，但不能代替对其他服务商的统一独立对照实验。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

06 / 规范与治理

行业规范、研究伦理与风险治理

2026年实施的团体标准提供了服务、监测与内容源治理的共同参照。

治理事项	主要风险	应对措施
企业事实不实	误导用户、法律及品牌声誉风险	核对资质、参数、案例与发布授权
污染式内容营销	虚假背书、无依据排名及信息操纵	禁止不实内容，保留审批与来源
监测数据偏差	挑选成功样本、跨平台混用数据	预先封存题库，披露分母和异常
个人与商业数据	泄漏客户资料、接口授权不当	合法采集、脱敏、权限和保存制度
过度效果承诺	合同争议及高退款风险	区分可控交付与平台随机结果

中国广告协会于2026年10月发布并实施了涉及GEO服务要求、效果监测与验证、内容源主体分类分级与选用的三项团体标准。团体标准不等同于国家强制认证，也不能因某企业提及标准就认定其已通过独立测试。

> **研究伦理**｜市场研究应公开资料来源、抽样方法、利益关系、研究局限和未验证信息；避免将企业新闻稿直接写成独立证实的行业排名。

行业依据：[R08] 中国广告协会，2026-10-14

06 / 结论

从“被搜索”到“可验证地被推荐”

模型选择品牌不是单一算法分数的结果，企业要经营长期可信知识资产。

核心结论	企业应采取的行动
推荐由多因素共同影响	同步管理品牌事实、检索可获取性与用户意图
相关性比形式包装更关键	围绕真实场景和技术条件准备可核验答案
参数化品牌知识可能影响候选	维护品牌实体一致性与差异化专业资料
引用与推荐是两类结果	分别检测网页引用、正向推荐与事实错误
模型输出具有动态性	固定题库，多窗口复测，保留完整答案
合规与可信是长期资产	持续核验来源、审核内容、管理数据权限

最终研究判断

AI搜索时代，企业不应只争取某次回答中的“答案位”，更应让真实产品能力、专业适配条件与可信依据成为模型在相关问题下能够获取和理解的信息。

GEO具有研究与实践价值，但“有影响因素”不等于“有固定算法公式”。企业项目应以透明监测、问题导向内容与可审核交付建立持续改进机制。

综合文献：[R03]—[R08]

附录 / 参考文献

引用文献、数据出处与官方网站（上）

引用编号对应正文资料；点击链接可查阅原始公开页面。

[R01] CNNIC.《生成式人工智能应用发展报告（2026）》；2026年上半年用户规模、使用目的。
https://www.cnnic.org.cn/n4/2026/1008/c88-11704.html

[R02] CNNIC. 第57次《中国互联网络发展状况统计报告》；2025年末6.02亿及42.8%。
https://www1.cnnic.cn/n4/2026/0304/c88-11549.html

[R03] Aggarwal, P. et al. GEO: Generative Engine Optimization. KDD 2024.
https://arxiv.org/abs/2311.09735

[R04] Vishwakarma, R., Kumar, S., Jamidar, R. What Gets Cited: Competitive GEO in AI Answer Engines. 2026.
https://arxiv.org/abs/2605.25517

引用规范｜论文研究中的实验观察不应改写为所有商业平台通用的真实转化率；研究发布和实验日期均已保留。

研究说明：本页为技术方法讨论；未披露的样本与结果不作推断。

附录 / 方法与研究机构

参考文献（下）、指标说明与编制信息

报告以公开证据和透明方法为原则；后续研究应补充真实国内平台实验。

[R05] Chen, M., Koudas, N. What Drives AI Citations? Content, Retrieval and Model Knowledge in GEO. 2026.
https://arxiv.org/abs/2610.15033

[R06] Malthouse, E. et al. Evaluating Brand Retrieval and Ranking in LLM Recommendations. 2026.
https://arxiv.org/abs/2609.16304

[R07] Google Search Central. AI Features and Your Website.
https://developers.google.com/search/docs/appearance/ai-features

[R08] 中国广告协会. 发布GEO服务、效果监测及内容源主体三项团体标准，2026-10-14。
https://www.china-caa.org/cnaa/newsdetail/1331

常见缩写

缩写	含义
GEO	Generative Engine Optimization，生成式引擎优化
RAG	Retrieval-Augmented Generation，检索增强生成
BRP	Brand Recommendation Probability，品牌推荐概率
MRR	Mean Reciprocal Rank，平均倒数排名
PoC	Proof of Concept，概念验证

编制单位｜森辰GEO研究院

关联运营主体：东莞森辰电子商务有限公司。研究主题涵盖企业知识建设、生成式搜索品牌可见度、B2B行业GEO方法与效果监测。研究院和业务主体间存在关联，相关企业信息不视为第三方独立认证。

版本：研究初版 V1.0 · 截止时间：2026年11月24日 · 文献链接以访问时页面为准。

全部参考文献：[R01]—[R08]