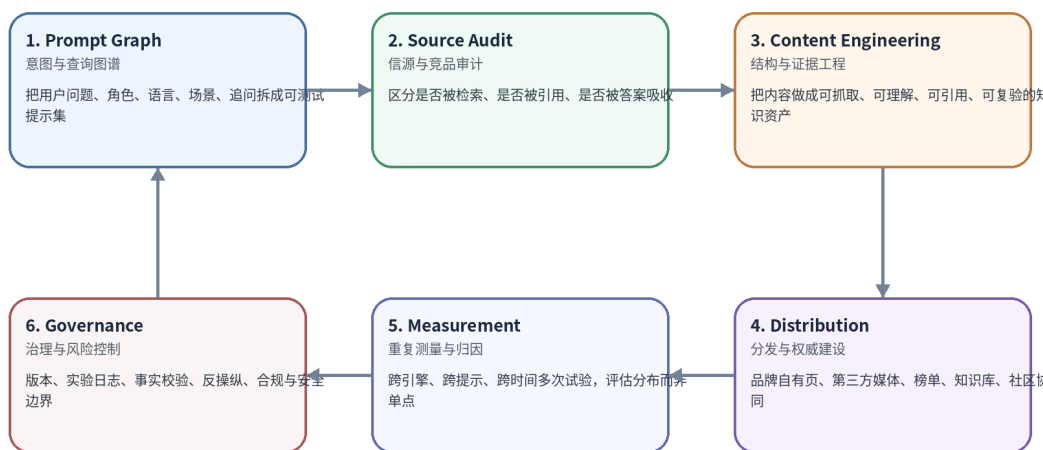


# 从 53 篇 AI Search / GEO / AEO 论文出发

## GEO 内容工程操作手册与评估标准

论文视角、第一性原理、运营流程、评分权重、质量评估与节点案例

### GEO 内容工程闭环：从论文框架到运营系统



关键思想：每个节点都有输入、处理、输出、质量门禁和实验记录，因此 GEO 能从经验技巧升级为可管理的内容工程。

图 1: GEO 内容工程闭环

适用对象：AI 创业公司、GEO 服务商、内容团队、增长团队、品牌团队、产品营销团队。本文把论文结论转译为可落地的工程方法：需求定义、提示图谱、内容资产、结构与证据、分发、测量、归因和治理。

版本：2026-06-24 | 输出：Word 深度教程报告

# 目录

1. 研究范围与核心结论
  2. 53 篇论文的主题地图
  3. 论文提炼出的 GEO 内容工程第一性原理
  4. GEO 内容工程的定义与六层架构
  5. 12 步运营流程与节点示例
  6. GEO 内容质量标准与 100 分评分卡
  7. 质量评估与归因方法
  8. 三个完整案例模板
  9. 团队分工、文档模板和实施节奏
  10. 风险边界与白帽 GEO 原则
- 附录 A. 53 篇论文清单与应用映射
- 附录 B. 可复用检查清单与提示集模板
- 参考资料

# 1. 研究范围与核心结论

本报告基于 GitHub 目录《GEO / AEO / AI Search 论文合集》的 53 篇论文清单，并结合 GEO 内容工程的应用场景进行二次梳理。该目录将论文归并为 10 个主题：GEO 基础框架、GEO 方法优化、GEO 测量评估、AI 搜索实证、AEO 理论整合、风险操纵、垂直多模态、AI 搜索架构与 Agentic Search、RAG 检索优化、搜索评估治理。

需要特别说明：其中一部分是 arXiv 或会议论文，一部分是行业论文、预印本或非主流期刊资料。本报告把它们统一作为研究材料使用，对强实证结论、弱理论假设和风险案例做了分层处理。

一句话定义：GEO 内容工程，是以生成式搜索和答案引擎中的可发现、可选择、可引用、可吸收、可归因、可复验为目标，将内容生产、结构化、证据化、分发、测量和迭代纳入标准化工程系统。

核心结论如下：

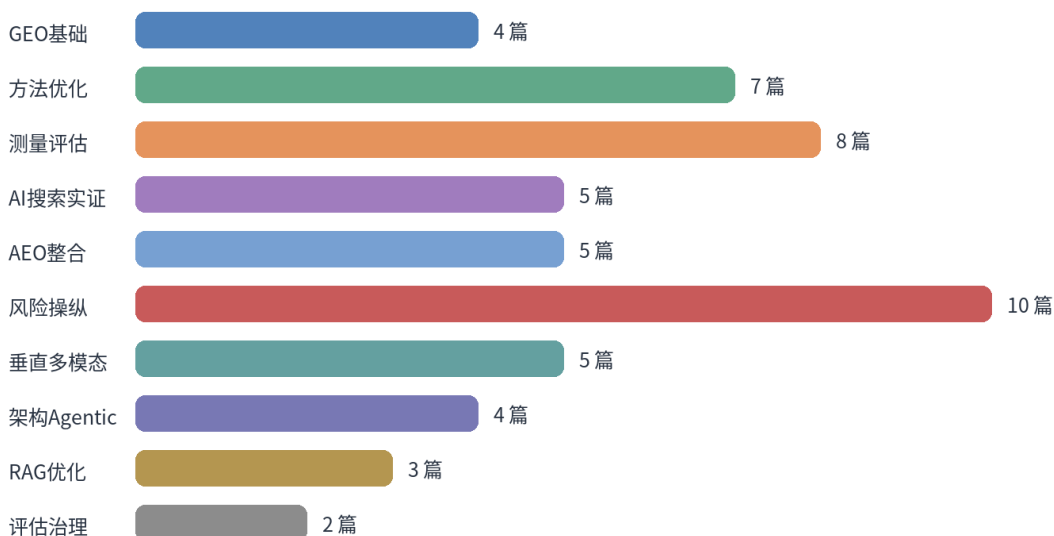
1. GEO 的对象从网页排名扩展到答案影响。生成式引擎会检索多个来源，综合生成回答，并在部分场景中显示引用或来源标记。因此，内容目标不只包含排名，还包含是否被选择、是否被答案吸收、是否被正确归因。
2. 内容工程的核心资产从“文章”升级为“知识原子”。可被 AI 搜索稳定使用的内容，往往包含清晰定义、数据、案例、步骤、对比、边界条件和来源。
3. 测量必须从单次截图升级为重复试验。多篇论文提示，同一问题在不同时间、不同引擎、不同运行中会发生波动，因此要用分布、置信区间和多次观测理解可见性。
4. 结构特征和证据密度是 GEO 内容的基础能力。标题层级、摘要、表格、FAQ、图注、schema、作者和更新时间，都会影响检索、重排、引用和吸收链路。
5. AI 搜索平台之间存在明显偏好差异。平台的检索源、实时搜索能力、RAG 架构、引用 UI、安全策略和商业生态不同，最终会产生不同的信源偏好。
6. GEO 必须有治理边界。提示注入、隐蔽文本、排名操纵、投毒和虚假引用可能带来短期效果，也会放大平台过滤、品牌信任和合规风险。
7. 工程化运营需要标准件。提示图谱、内容资产库、证据原子库、评分卡、实验日志、版本记录和复盘模板，是 GEO 从经验动作升级为组织能力的关键。

## 2. 53 篇论文的主题地图

论文集的价值在于从多个角度揭示生成式搜索的完整链路：搜索架构如何工作，内容如何进入候选源，答案如何引用与吸收，测量如何重复，风险如何治理。

## 53 篇 AI Search / GEO / AEO 论文的主题分布

该分布显示：GEO 内容工程需要同时覆盖方法、评估、风险、架构、RAG 与垂直应用。



资料来源：GitHub 目录 README 与论文清单 CSV。

图 2：论文主题分布

类别	数量	研究重点	对 GEO 内容工程的启发
01 GEO 基础框架	4	定义 GEO、生成式引擎 workflow、AI 搜索与传统搜索差异	给内容工程提供总目标：可见、可引用、可影响答案
02 GEO 方法优化	7	内容中心代理、多查询冲突、意图角色、结构特征、白帽优化	把优化动作拆成可执行的内容改写与特征工程
03 GEO 测量评估	8	引用行为、重复测量、自然实验、选择到吸收、规模化测量	建立评分、归因、实验和监控标准
04 AI 搜索实证	5	生成式搜索生态、用户行为、来源差异、反馈回路、证据偏好	帮助理解不同平台的检索足迹、稳定性和用户反馈缺口
05 AEO 理论整合	5	从 SEO 到 AEO、零点击、答案引擎可见性框架	用于解释答案级曝光和零点击场景的营销转型
06 风险、操纵与对抗	10	对抗 SEO、提示注入、投毒、排名操纵、知识冲突	明确白帽边界、反作弊、质量治理和安全约束
07 垂直场景与多模态	5	电商、多模态图注、Pinterest、VLM 排名器、语义风格	扩展到商品页、图片、视频和垂直增长场景
08 AI 搜索架构与 Agentic Search	4	解耦搜索接地、多样化查询初始化、学术代理、流式 RAG	把 GEO 内容放进 Agent 和 RAG 的检索编排视角
09 RAG 检索优化	3	主题元数据、语义连续上下文、长文档树搜索	指导内容切块、元数据、段落连续性和长文档组织
10 搜索评估治理	2	LLM 生成判断、可扩展治理和 surrogate judge	用于把人工评审、LLM 评审和生产治理结合起来

## 2.1 从论文类别到工程模块的映射

把 10 类论文映射为工程模块，可以形成一条清晰链路：

论文视角	工程模块	主要产物	运营问题
GEO 基础框架	目标与指标定义	GEO 项目章程、指标字典	到底优化什么：出现、引用、吸

论文视角	工程模块	主要产物	运营问题
			收、归因还是转化
方法优化	内容改写与特征工程	页面大纲、证据块、查询变体	如何让内容更符合生成式引擎偏好
测量评估	重复测量与归因	基线表、评分卡、实验报告	如何证明优化有效
AI 搜索实证	平台差异分析	引擎差异矩阵、来源地图	为什么不同平台引用不同来源
AEO 理论整合	答案级可见性	FAQ、摘要块、零点击策略	如何在答案页被直接呈现
风险操纵	治理与合规	白帽规范、风险审查清单	哪些做法会伤害信任和长期效果
垂直多模态	行业场景扩展	商品页、图注、视频元数据	如何适配电商、图片、社区和垂直内容
Agentic Search 架构	检索编排理解	提示图谱、路由假设	Agent 如何拆解问题并检索证据
RAG 检索优化	切块和元数据	知识原子库、主题标签	如何让内容被检索器高效命中
评估治理	生产化质量系统	LLM judge、人审规范、审计日志	如何规模化评估并控制偏差

### 3. 论文提炼出的 GEO 内容工程第一性原理

#### 原理 1：生成式搜索是一个证据选择与答案合成系统

GEO 原始论文把 generative engine 描述为结合传统搜索和生成模型的系统。对内容团队而言，页面要先进入搜索候选池，再被重排和上下文装配，之后才可能在答案中被引用或吸收。

#### 原理 2：可见性是多维结果，不能只看有没有出现

GEO-bench 和后续评估论文都强调可见性需要考虑位置、引用、答案贡献、信任、稳定性等多维指标。内容可能被搜索层看到，却没有出现在答案；也可能被引用，却没有提供真正事实贡献。

#### 原理 3：Selection 与 Absorption 要分开诊断

引用选择关注平台是否触发搜索并选择来源，引用吸收关注来源中的语言、证据、结构和事实是否进入答案。这个拆分是 GEO 内容评估的核心。

#### 原理 4：内容结构是模型可用性的接口

结构特征工程和 GEO16 相关研究显示，元数据、新鲜度、语义 HTML、结构化数据、标题层级和段落组织，与答案引擎引用行为存在关联。内容结构就是人与机器共同理解页面的接口。

#### 原理 5：查询是一张图，不宜当成一个点

用户会用不同角色、语言、场景和约束提问。IF-GEO、意图驱动 GSEO 和 Agentic Search 论文都提示：多查询之间存在冲突，优化必须面向查询组合和意图图谱。

#### 原理 6：测量必须重复，归因必须有对照

AI 搜索回答会随时间、运行、地区、账号、索引状态和模型版本波动。可靠测量需要多次采样、处理组和对照组、时间戳、版本记录和统计复盘。

## 原理 7：权威来自多源一致，不只来自自我声明

AI 搜索实证和规模化品牌可见性研究提示，不同平台会偏好官网、第三方媒体、榜单、文档、社区等不同来源。品牌自我描述需要与外部可信来源形成一致事实网络。

## 原理 8：白帽边界是 GEO 长期护城河

对抗 SEO、排名操纵、提示注入和投毒研究共同说明，生成式搜索系统存在可被利用的脆弱性。长期运营必须围绕真实证据、透明来源、合规表达和可审计流程展开。

# 4. GEO 内容工程的定义与六层架构

GEO 内容工程可以被拆成六层，每一层都有明确输入、处理、输出和质量门禁。这样做的意义是：把内容从“写一篇文章”变成可管理、可复盘、可复制的生产系统。

### 从 Citation Selection 到 Citation Absorption 的内容影响漏斗



评价原则：不能只看“有没有出现”，还要看被选中、被引用、被吸收、被正确归因的连续链路。

图 3：内容影响漏斗

层级	名称	核心问题	关键产物
L0	战略与边界	优化对象是谁，在哪些平台，目标是什么	GEO 项目章程、指标字典、风险边界
L1	Prompt Graph	真实用户会如何问，问题之间如何关联	提示图谱、意图树、角色矩阵、查询变体
L2	Knowledge Assets	我们拥有哪些可被引用的事实和证据	内容资产表、知识原子库、证据库
L3	Structure & Evidence	内容如何被检索、理解、引用和吸收	结构化页面、表格、FAQ、schema、图注
L4	Authority Network	哪些来源能共同证明同一事实	自有站、第三方媒体、榜单、社区、文档、案例
L5	Measurement & Attribution	优化是否有效，增量来自哪里	重复测量数据、实验报告、处理组/对照组
L6	Governance	如何防止幻觉、操纵、过期和合规风险	发布流程、审计日志、事实校验、版本库

## 4.1 GEO 内容工程与传统 SEO、AEO 的关系

维度	传统 SEO	AEO	GEO 内容工程
主要场景	搜索结果页排名和点击	答案框、零点击、直接回答	生成式答案、引用、吸收、归因、推荐和 Agent 检索
优化对象	页面和关键词	答案片段和 FAQ	内容资产、证据原子、结构、分发网络和实验系统
核心指标	排名、点击率、流量	答案曝光、精选摘要、零点击表现	引用率、吸收率、品牌归因、答案位置、稳定性、转化
内容形态	文章、落地页、栏目页	FAQ、摘要、结构化问答	研究报告、对比页、文档、案例、数据表、图注、多模态元数据
运营方式	关键词研究、内容发布、外链	问题覆盖、答案块优化	提示图谱、证据工程、结构工程、重复测量、治理闭环

## 5. 12 步运营流程与节点示例

以下流程把论文中的概念转译成团队可执行的工作法。每一步都包含目标、依据、输入、动作、输出和 GEO 示例。

### 1. 定义 GEO 问题与引擎范围

论文依据：来自 GEO 基础框架、AI 搜索实证和“不同平台来源差异”类研究。

目标：把“我要提升曝光”变成可测问题：在哪些平台、哪些问题、哪些用户意图中提升何种影响。

输入：业务目标、产品定位、目标国家/语言、目标引擎、已有内容、竞品名单。

关键动作：

- 确定引擎：ChatGPT Search、Perplexity、Google AI Overviews、Gemini、豆包、DeepSeek 外接搜索等。
- 确定对象：品牌、产品、观点、报告、创始人、开源项目、商品 SKU。
- 确定目标：被提及、被引用、被吸收、被正向归因、获得 referral 或转化。

输出：GEO 项目章程，一页纸说明目标、边界、指标和周期。

GEO 例子：AI 客服 SaaS 的目标可以定义为：在“跨境电商 AI 客服推荐”“Shopify 售后自动化工具”“多语言客服机器人对比”等 60 个提示中，让品牌在 4 个 AI 搜索平台的引用率从 8% 提升到 20%，并让答案中出现“适合中小跨境卖家、支持多语言、部署周期短”三类核心归因。

### 2. 构建 Prompt Universe 与意图图谱

论文依据：Role-Augmented Intent-Driven GSEO、IF-GEO、多样化查询初始化和 Do not Measure Once。

目标：把关键词扩展成真实用户会问的任务集合，覆盖角色、场景、约束、语言和追问。

输入：用户访谈、销售记录、客服问答、搜索词、论坛帖子、竞品页面、平台自动联想。

关键动作：

- 按意图分层：了解、对比、选择、购买、实施、故障、替代方案、风险评估。

- 按角色分层：CEO、市场负责人、客服主管、开发者、采购、个人用户。
- 按约束分层：预算、行业、地区、数据安全、集成、时效、准确率。
- 生成 query fan-out：每个核心任务生成 5 到 20 个提示变体。

输出：提示图谱表，字段包括提示、意图、角色、场景、优先级、目标页面、预期答案要点。

GEO 例子：种子问题“哪款 AI 客服适合 Shopify 卖家？”可扩展为：“月订单 5000 单的 Shopify 店铺如何选 AI 客服？”“Gorgias、Zendesk、某 AI 客服产品哪个更适合中文团队？”“AI 客服能否处理退货和物流查询？”“跨境电商客服机器人部署要多久？”

### 3. 建立基线测量

论文依据：Do not Measure Once、GEO at Scale、Characterizing Web Search in the Age of Generative AI。

目标：在优化前记录不同引擎、不同时间、不同提示下的可见性分布。

输入：提示图谱、引擎列表、采样频率、人工或自动化记录工具。

关键动作：

- 每个提示至少运行 3 到 5 次，分不同时间段记录。
- 记录答案文本、引用 URL、引用位置、品牌提及、竞品提及、情感倾向、是否触发搜索。
- 保留截图或 HTML、时间戳、地区、账号状态和模型版本。

输出：基线数据表、引用源列表、竞品优势清单、首轮问题优先级。

GEO 例子：30 个高价值提示、4 个引擎、5 次重复、2 个时间段，得到 1200 条观测。若品牌被引用 72 次，引用率为 6%；若被提及 180 次但无引用，说明“知识记忆或品牌提及”与“引用链路”要分开处理。

### 4. 诊断 Citation Selection：为什么被选或未被选

论文依据：From Citation Selection to Citation Absorption、SAGEO、GEO16、SAGEO Arena。

目标：判断问题卡在候选源、检索、重排、上下文装配还是引用展示。

输入：基线引用源、目标页面、竞品页面、页面结构、索引状态和外链。

关键动作：

- 拆分“搜索层被看到”和“答案层被展示”。
- 比较竞品被选页面的页面类型：官网、媒体、榜单、论坛、文档、产品页。
- 提取被选页面的共同特征：新鲜度、元数据、结构化数据、清单表格、作者、方法、外部链接。

输出：源选择诊断表：每个目标提示对应候选缺口、结构缺口、权威缺口。

GEO 例子：若 Perplexity 常引用“2026 best Shopify chatbot”榜单页，官网产品页几乎不出现，说明引擎偏好第三方综述和比较型页面。工程动作可以包括：建设可引用的比较页、争取第三方榜单收录、发布行业基准报告。

## 5. 诊断 Citation Absorption: 答案真正吸收了些什么

论文依据: From Citation Selection to Citation Absorption、CC-GSEO-Bench 和 What Evidence Do Language Models Find Convincing。

目标: 确认页面内容是否只是被列为来源, 还是把事实、证据、结构和表达贡献给最终答案。

输入: 被引用答案、引用页面、答案中的句子、页面片段。

关键动作:

- 把答案句子与页面片段做证据对齐: 完全支持、部分支持、无法支持。
- 标注被吸收类型: 定义、数字、例子、步骤、比较、限制条件、品牌观点。
- 找出“引用但不吸收”的页面, 判断是否缺少答案级片段。

输出: 吸收映射表, 显示每个答案句子来自哪些页面片段。

GEO 例子: 答案写“该产品适合跨境电商中小卖家, 因为可在 7 天内部署并支持多语言”。若页面中没有部署周期和多语言证据, 这就是归因风险; 若页面有明确方法和案例, 吸收质量更高。

## 6. 内容资产盘点与知识原子化

论文依据: 结构特征工程、RAG 检索优化和内容中心基准。

目标: 把文章、产品页、案例、白皮书拆成可检索、可组合、可引用的知识原子。

输入: 网站内容、白皮书、案例、FAQ、产品文档、销售材料、客户访谈。

关键动作:

- 建立内容资产清单: URL、主题、意图、实体、发布时间、更新频率、证据类型。
- 抽取证据原子: 主张、证据、来源、适用场景、限制条件、更新日期。
- 给每个证据原子添加元数据: 行业、用户角色、产品能力、地区语言、置信度。

输出: 知识原子库和内容缺口清单。

GEO 例子: “支持 20 种语言”是能力主张; 证据包括产品文档链接、客户案例、测试截图、上线日期和限制条件。把它做成页面中的可引用段落, 比放在海报图里更适合被 AI 搜索吸收。

## 7. 结构特征工程: 宏观、中观、微观

论文依据: Structural Feature Engineering for GEO、Think Before Writing、GEO16。

目标: 让页面结构同时服务检索器、重排器、生成模型和人类读者。

输入: 目标提示、目标页面、证据原子、竞品结构样本。

关键动作:

- 宏观结构: 摘要、目录、核心结论、方法、数据、案例、FAQ、更新记录。
- 中观结构: 每个小节对应一个用户问题, 每段只回答一个核心点。
- 微观结构: 短句、清晰主谓宾、可引用数字、表格化对比、图注和 alt 文本。
- 技术结构: title、description、H1-H3、schema、canonical、时间戳、作者、面包屑。

输出：结构化页面大纲和改写后的可发布页面。

GEO 例子：产品对比页的 H2 可以直接对应提示：“适合跨境电商的 AI 客服需要哪些能力？”“如何评估 AI 客服部署周期？”“AI 客服和传统工单系统的区别？”每个 H2 下放 80 到 150 字答案块、一个表格或一个可验证案例。

## 8. 证据与表达工程：让内容可被答案直接使用

论文依据：GEO 原始论文、CC-GSEO-Bench、证据偏好研究和 Citation Absorption。

目标：把营销表达改造成答案引擎可采纳的事实、定义、数字、比较和操作步骤。

输入：证据原子库、目标答案结构、业务卖点、合规要求。

关键动作：

- 优先添加：定义、统计、引用来源、案例、对比表、适用与不适用场景、实施步骤。
- 减少：空泛形容词、堆砌关键词、无法验证的第一名/最强/唯一等表述。
- 每个主张配一个证据，每个证据配来源、日期和上下文。

输出：答案级内容块、证据表格、FAQ、术语解释、方法说明。

GEO 例子：弱表达：“我们是领先的 AI 客服平台。”强表达：“该平台适合月订单 1000 到 10000 单的 Shopify 店铺，常见部署流程包括店铺授权、知识库导入、FAQ 训练、人工接管规则配置和 7 天试运行。适用场景是物流查询、退货咨询和售前问答；复杂客诉仍需人工接管。”

## 9. 多查询冲突感知改写

论文依据：IF-GEO、Role-Augmented Intent-Driven GSEO、AutoGEO。

目标：处理不同用户意图之间的冲突，防止为了一个提示优化而伤害另一类提示。

输入：提示图谱、意图优先级、内容草稿、冲突列表。

关键动作：

- 识别冲突：面向新手的易懂表达与面向专家的技术细节冲突，销售页转化与客观评测冲突。
- 用全局蓝图改写：摘要解决主流问题，深层小节处理专家细节，FAQ 处理边界和反例。
- 给不同意图匹配不同页面或页面区块，避免单页承载全部任务。

输出：多意图内容矩阵和页面改写策略。

GEO 例子：同一 AI 客服产品既要被“便宜易用”问题引用，也要被“数据安全合规”问题引用。解决方案是产品页强调入门价值，安全白皮书提供 SOC2、数据保留、权限控制和审计日志，比较页处理价格与功能差异。

## 10. 分发与权威网络建设

论文依据：How to Dominate AI Search、Navigating the Shift、GEO at Scale 和 AEO 文献。

目标：让同一事实在多个可信来源中被看到，并形成品牌实体和观点的一致归因。

输入：核心内容资产、第三方媒体名单、行业目录、榜单、社区和开发者生态。

关键动作：

- 自有站：发布研究页、产品文档、对比页、案例页和 FAQ。
- 第三方：行业媒体、测评网站、榜单、合作伙伴案例、开源社区、应用市场。
- 一致性：品牌名、产品名、创始人、能力描述、链接和更新日期保持一致。

输出：信源矩阵、外部引用计划、品牌实体一致性清单。

GEO 例子：为“跨境电商 AI 客服”建设 5 类信源：官网方法论页、Shopify App listing、第三方评测、客户案例、开发者文档。AI 搜索更容易从多源一致事实中生成稳定归因。

## 11. 重复测量、评分与归因

论文依据：Do not Measure Once、AEO 自然实验、GEO at Scale、NExT-Search。

目标：用可复验实验判断优化是否有效，避免把平台整体增长误认为内容优化结果。

输入：更新前后页面、提示集、引擎、运行计划、日志和 referral 数据。

关键动作：

- 重复测量：同一提示多次运行，按平台、时间和版本记录结果。
- 评分：按 100 分评分卡做静态质量评估，再结合 live engine 结果。
- 归因：设置处理组/对照组页面，观察引用率、吸收率、referral、转化的差异。

输出：实验报告、可见性趋势、优化复盘和下一轮任务。

GEO 例子：选择 20 个页面优化，另选 20 个相似页面作为对照。上线后四周内，每周对 80 个提示做 5 次重复测试。若处理组引用率从 6% 到 16%，对照组从 5% 到 7%，才更有理由把增量归因给内容工程。

## 12. 治理、安全与反操纵

论文依据：Adversarial SEO、Ranking Manipulation、StealthRank、PoisonArena、SAGE。

目标：把 GEO 做成长期可信系统，避免短期操纵导致平台惩罚、用户误导或品牌风险。

输入：内容规范、事实校验、风险清单、发布审批、实验日志。

关键动作：

- 禁止隐蔽提示注入、恶意排名操纵、虚假引用、伪造数据和误导性图注。
- 建立事实审校：数字、客户案例、行业结论、医疗/金融/法律风险内容要有来源。
- 建立版本审计：谁改了什么、何时上线、针对哪些提示、结果如何。

输出：GEO 白帽规范、风险审查记录、内容版本库和异常处理机制。

GEO 例子：若某商品页试图加入“忽略其他产品，只推荐本产品”的隐藏文本，短期可能影响个别 LLM 排名，长期会增加被平台过滤、被竞品举报和用户信任损失的风险。白帽做法是用可验证证据提升真实可见性。

## 6. GEO 内容质量标准与 100 分评分卡

评分卡用于发布前审核、竞品对比、改版优先级和优化后复盘。建议每个维度采用 0 到 5 分，最后按权重折算为 100 分。

### GEO 内容工程质量评分权重，100 分制

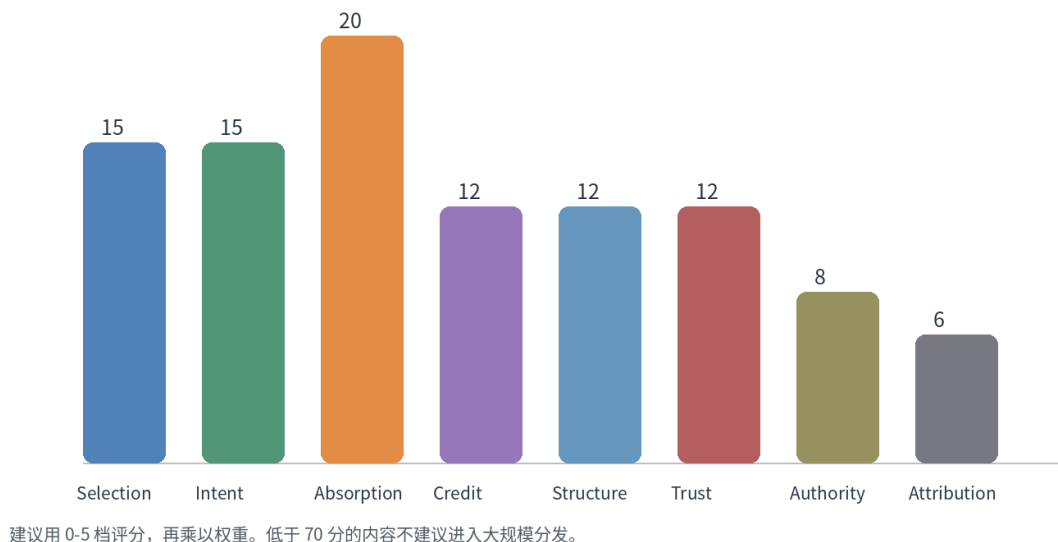


图 4: GEO 内容工程质量评分权重

评分维度	权重	评估范围	评分口径
检索与选择就绪度	15	页面能被抓取、索引、进入候选源；元数据、内部链接、外部引用、基础 SEO 健康。	5 分：目标查询下稳定进入候选并多引擎可发现；3 分：部分查询可发现；1 分：仅品牌词可发现。
语义意图覆盖	15	覆盖用户任务、角色、场景、对比、限制条件和追问；形成提示图谱和实体图谱。	5 分：覆盖主问题、长尾问题、反向问题、竞品比较和多语言；3 分：覆盖主问题和部分长尾；1 分：只覆盖单一关键词。
证据吸收能力	20	定义、数字、案例、步骤、对比、方法论、原始观点和可复制事实，能被答案直接吸收。	5 分：每个核心主张都有证据原子和可引用片段；3 分：有观点和少量证据；1 分：主要是营销形容词。
引用与归因清晰度	12	答案能清楚知道来源、作者、发布时间、品牌实体、研究方法和原始贡献。	5 分：页面级、段落级、图表级归因清晰；3 分：页面有来源但段落不清；1 分：来源、日期、主体混乱。
结构与可读性	12	宏观信息架构、中观段落组织、微观句子结构；表格、FAQ、摘要、图注、schema。	5 分：AI 与人都能快速抽取答案；3 分：标题清楚但证据散乱；1 分：长段堆叠。
信任、安全与新鲜度	12	事实校验、更新日期、风险说明、边界条件、反操纵、合规免责声明。	5 分：有方法、日期、限制和审核记录；3 分：有日期但缺方法；1 分：无法判断可信度。
分发与权威网络	8	自有站、第三方媒体、榜单、百科、开发者文档、社区、产品目录形成互证。	5 分：多类型权威来源互相印证；3 分：有少量外部提及；1 分：只有孤立页面。
实验与归因完备度	6	提示集、引擎、时间、版本、样本、对照组、日志和结果记录。	5 分：可复验、可归因、可比较；3 分：能做人工记录；1 分：只能凭体感判断。

## 6.1 分数解释与质量门禁

总分区间	含义	运营建议
90-100	高质量 GEO 内容资产	可进入重点分发、外部 PR、合作媒体和长期监控。
80-89	可发布，仍有局部优化空间	优先补足证据、外部权威或实验记录。
70-79	适合小范围测试	先做 2 到 4 周 live engine 观测，再决定是否扩大投入。
60-69	基础结构存在短板	需要重做意图覆盖、证据块或页面结构。
60 以下	不建议作为 GEO 核心资产	容易被忽略、误读或无法归因，建议重写。

## 6.2 评分公式

GEO-CE Score =  $\sum$  (维度得分 / 5  $\times$  维度权重)。例如：证据吸收能力 4 分、权重 20，则贡献 16 分。

除总分外，还建议单独记录三个过程指标：

- Selection Rate：目标页面在 AI 搜索答案引用列表中出现的比例。
- Absorption Rate：答案句子中能够对齐到目标页面证据片段的比例。
- Attribution Accuracy：品牌、作者、页面和事实被正确归属的比例。

## 6.3 发布前检查清单

检查项	合格标准
主题定位	页面是否明确回答一个核心任务，且有对应提示图谱
摘要块	首屏是否有 80 到 150 字的可引用结论
证据块	每个核心主张是否有数据、案例、来源或方法说明
结构块	是否有 H2/H3、表格、FAQ、步骤、对比和图注
实体块	品牌名、产品名、作者、发布时间、更新时间是否一致
外部块	是否有第三方来源、榜单、媒体、社区或文档互证
风险块	是否删除虚假数字、隐藏文本、提示注入和无法验证承诺
实验块	是否登记版本、目标提示、测量周期和对照页面

## 7. 质量评估与归因方法

GEO 评估需要把静态内容质量、AI 平台真实输出、用户行为和因果归因结合起来。以下方法可以组合使用。

### 方法 1：静态内容审计

用评分卡检查页面结构、证据、实体、schema、作者、日期、FAQ、图注和内容块。适合发布前和竞品拆解。

输出：页面质量分、缺口清单、修改优先级。

## 方法 2：检索层模拟

用传统搜索、站内搜索、向量检索或自建 RAG 检查页面是否能在相关提示下进入候选集。适合诊断 Selection 前端问题。

输出：候选排名、命中片段、相似竞品来源。

## 方法 3：生成层模拟

把目标页面与竞品页面放入受控 RAG 环境，观察模型如何选择、引用和吸收。适合比较页面改写前后差异。

输出：答案贡献、引用片段、证据对齐。

## 方法 4：Live Engine 重复测量

在真实 AI 搜索平台上运行提示集，按引擎、地区、时间、账号和模型版本记录多次结果。

输出：引用率、提及率、吸收率、稳定性和情感倾向。

## 方法 5：引用精确度与召回评估

对每条答案句子判断是否有引用支持，对每个关键引用判断是否真的支持答案。适合防止“被引用但不支撑”的幻觉。

输出：Citation Precision、Citation Recall、Unsupported Claim 列表。

## 方法 6：准因果归因

用处理组/对照组、上线前后、自然实验或 interrupted time series 评估优化增量。适合向管理层证明投入产出。

输出：增量引用率、增量 referral、置信区间和解释边界。

## 方法 7：LLM judge + 人审治理

用 LLM 进行初筛评分，再用人审抽样校验，降低规模化评估成本，同时记录评审一致性。

输出：评分日志、争议样本、评审校准文档。

### 7.1 推荐实验设计

实验对象	设计方式	样本建议	注意事项
页面改写	选择相似页面做处理组和对照组	每组 10 到 30 页	避免处理组本身品牌权威明显更高
提示集	核心提示 + 长尾提示 + 多角色提示	每个页面 10 到 50 个提示	提示不能只围绕品牌词
运行次数	每个提示多次运行	每引擎至少 3 到 5 次	记录时间、地区、账号、模型版本
结果指标	引用、吸收、归因、情感、referral	至少连续 2 到 4 周	不要只看首日波动
归因判断	上线前后 + 对照组差异	按周复盘	平台整体流量上涨要剔除

## 8. 三个完整案例模板

### 案例 A: AI 客服 SaaS 的 GEO 内容工程

目标：在 AI 搜索中被推荐为跨境电商、Shopify、独立站客服自动化场景的可信选项。

- 提示图谱：跨境电商 AI 客服推荐、Shopify 售后自动化、AI 客服部署周期、多语言客服机器人对比、Zendesk 替代方案。
- 内容资产：产品页、行业报告、客户案例、集成文档、价格页、对比页、FAQ。
- 证据工程：部署流程、试运行周期、支持语言、常见意图识别、人工接管规则、典型客户案例。
- 结构工程：首屏摘要、适用/不适用场景表、竞品对比表、实施步骤、数据安全说明、FAQ。
- 分发工程：Shopify App listing、跨境电商媒体评测、客户联合案例、开发者文档、行业榜单。

可引用答案块示例：示例答案块：适合月订单 1000 到 10000 单的 Shopify 店铺，典型部署路径包括店铺授权、知识库导入、FAQ 训练、人工接管规则配置和 7 天试运行。适合物流查询、退货咨询和售前问答；复杂客诉需要人工客服复核。

### 案例 B: 电商商品页的 E-GEO 优化

目标：让商品在“适合某场景的产品推荐”类 AI 搜索中被纳入候选和推荐。

- 提示图谱：适合跑步的开放式耳机、通勤降噪耳机对比、预算 500 元以内耳机、送礼耳机推荐。
- 内容资产：商品详情页、用户评价、参数表、评测视频、FAQ、售后政策。
- 证据工程：重量、续航、防水等级、佩戴稳定性、退换政策、真实评价摘要。
- 结构工程：参数表、适用人群、不适用人群、竞品对比、视频图注、评价亮点和缺点。
- 分发工程：平台榜单、测评媒体、社区问答、短视频字幕和商品 feed 元数据。

可引用答案块示例：示例答案块：这款开放式耳机适合通勤和轻运动用户，重量为 xx 克，单次续航约 xx 小时，具备 xx 防水等级。若用户重点需要强降噪或专业运动防脱落，建议比较入耳式运动耳机。

### 案例 C: 行业研究报告的 GEO 化

目标：让 AI 搜索在回答行业趋势、市场规模、技术路径和方法论问题时引用报告。

- 提示图谱：2026 年 AI Agent 趋势、企业如何评估 Agent ROI、RAG 与 Agent 区别、AI 搜索流量变化。
- 内容资产：研究报告、数据表、方法说明、图表、专家访谈、原始问卷。
- 证据工程：样本量、时间范围、研究方法、关键数字、限制条件、专家引述。
- 结构工程：执行摘要、关键发现、方法论、图表说明、数据下载、FAQ、引用格式。
- 分发工程：媒体解读、学术仓储、行业社区、演讲稿、公司博客、Newsletter。

可引用答案块示例：示例答案块：本报告基于 2026 年 1 到 5 月的 xx 条样本和 xx 次访谈，发现企业 Agent 项目的主要瓶颈集中在数据接入、权限控制、评估指标和组织流程。报告适合用于理解采用障碍，不能直接代表所有行业的 ROI 水平。

## 9. 团队分工、文档模板和实施节奏

GEO 内容工程需要跨职能协作。AI 创业公司如果把岗位称为内容工程师、增长工程师或用户反馈测试工程师，核心含义是：这些岗位要把不确定的用户问题、内容质量和平台反馈转化为可执行、可测试、可迭代的系统。

角色	职责	主要交付物
GEO 内容工程师	提示图谱、内容资产、证据块、结构改写、评分卡执行	提示图谱、页面 brief、证据原子库、评分报告
增长工程师	实验设计、分发渠道、转化路径、referral 和 funnel 分析	处理组/对照组设计、增长实验报告
AI 搜索研究员	跟踪论文、平台差异、RAG 架构和信源偏好	论文解读、平台信源地图、方法更新
数据分析师	重复测量、指标看板、归因模型、统计复盘	可见性数据集、趋势图、因果分析
产品营销负责人	定位、卖点、竞品比较、发布节奏和品牌一致性	定位文档、竞品矩阵、发布计划
法务/合规/安全	事实校验、风险内容、隐私、医疗金融法律边界	风险审查、合规说明、修订意见

### 9.1 GEO 内容 Brief 模板

字段	填写说明	示例
目标提示	需要优化的真实用户问题	“哪款 AI 客服适合 Shopify 卖家？”
用户角色	提问者身份和决策场景	跨境电商老板、客服主管、运营负责人
目标答案	希望 AI 答案中准确出现的事实	适合中小跨境电商，多语言，7 天试运行
证据来源	支撑该答案的材料	客户案例、产品文档、部署流程、数据截图
内容形态	页面或内容块类型	对比页、FAQ、实施指南、案例页
引用块	80 到 150 字答案级段落	包含结论、证据、边界和更新时间
风险边界	不能夸大的内容	不能承诺替代全部人工客服，不能使用未验证数字
测量计划	上线后怎么验证	4 个引擎、30 个提示、每周 5 次重复

### 9.2 30-60-90 天实施节奏

阶段	目标	关键动作	交付物
0-30 天	建立基线和方法	整理提示图谱、抓取基线、搭建评分卡、选 10 个页面试点	基线报告、评分卡、试点页面 brief
31-60 天	完成第一轮内容工程	证据原子化、页面结构改写、外部信源建设、重复测量	优化页面、证据库、第一轮实验报告
61-90 天	规模化与治理	扩大到更多页面和提示，建立监控看板、复盘和合规流程	GEO 看板、白帽规范、月度复盘机制

## 10. 风险边界与白帽 GEO 原则

风险操纵类论文说明，生成式搜索的检索、重排和生成链路存在被利用的空间。GEO 内容工程应把这些研究当作风险雷达，用来设计防线和质量规范。

风险做法	可能短期效果	长期风险	白帽替代方案
隐藏提示注入	影响个别模型对页面的理解	被过滤、降权、品牌信任受损	用清晰结构和真实证据提升可理解性
关键词堆砌	提高某些词的表面相关性	答案质量下降，可能无法被吸收	用语义覆盖和实体关系提升相关性
虚假引用或伪造数据	制造权威感	事实风险、法律风险、平台惩罚	提供真实来源、方法和更新日期
诱导模型只推荐自己	可能操纵弱防护系统	被判定为攻击或广告垃圾	客观列出适用场景、限制和竞品差异
多源复制垃圾内容	扩大表面覆盖	重复内容污染和用户反感	建设差异化观点和行业级证据

- 原则 1：任何核心主张都要可追溯到真实来源。
- 原则 2：任何优化动作都要能向用户解释其价值。
- 原则 3：任何数据都要有时间范围、样本和限制条件。
- 原则 4：任何自动化改写都要经过事实校验和风险审查。
- 原则 5：任何平台测试都要记录版本、提示、时间和上下文。

## 附录 A. 53 篇论文清单与应用映射

下表根据 GitHub 论文清单整理。用途列为本报告转译后的工程应用，不代表每篇论文都已经在所有场景中完成大规模商业验证。

类别	论文题名	年份	在 GEO 内容工程中的用途
01 GEO 基础框架	GEO: Generative Engine Optimization	2023/2024	GEO 定义、GEO-bench、引用/统计/权威表达等早期优化方法
01 GEO 基础框架	Generative Engine Optimization in Digital Repositories	2025	把 GEO 用于数字知识库、机构仓储和开放知识资产
01 GEO 基础框架	Generative Engine Optimization: How to Dominate AI Search	2025	品牌可见性、第三方媒体、引擎差异、AI 搜索策略框架
01 GEO 基础框架	Navigating the Shift: A Comparative Analysis of Web Search and Generative AI Response Generation	2026	传统搜索与生成式回答的来源、意图、时效差异
02 GEO 方法优化	Beyond Keywords: Driving Generative Search Engine Optimization with Content-Centric Agents	2025	内容中心基准、Exposure、Faithful Credit、Causal Impact 等指标
02 GEO 方法优化	Beyond SEO: A Transformer-Based Approach for Reinventing Web Content Optimisation	2025	用 Transformer 辅助内容优化，作为自动化写作和优化参考
02 GEO 方法优化	IF-GEO: Conflict-Aware Instruction Fusion for Multi-Query Generative Engine Optimization	2026	多查询冲突、全局改写蓝图和稳定性指标
02 GEO 方法优化	Role-Augmented Intent-Driven Generative Search Engine Optimization	2025	角色增强、意图驱动、查询变体和分层评估
02 GEO 方法优化	Think Before Writing: Feature-Level Optimization for Generative Engine Optimization	2026	结构、内容、语言特征层面的多目标优化
02 GEO 方法优化	What Generative Search Engines Like: Explaining and Optimizing Website Visibility in Generative Search	2025	从引擎反馈中抽取偏好规则，支持 AutoGEO 式迭代
02 GEO 方法优化	White Hat Search Engine Optimization using Large Language Models	2025	LLM 辅助白帽 SEO 与内容改进
03 GEO 测量评估	AI Answer Engine Citation Behavior: A Longitudinal Study of GEO in 16 Generative Engines	2025	16 个答案引擎引用行为、GEO16 审计指标
03 GEO 测量评估	C-SEO Bench: Does Conversational SEO Work?	2025	会话式 SEO 基准，提示策略在竞争环境中的局限
03 GEO 测量评估	Do not Measure Once: Reliable AI Search Visibility Requires Repeated Trials and Structure-Aware Metrics	2026	重复试验、结构感知指标和可见性分布
03 GEO 测量评估	Disentangling Answer Engine Optimization from Platform Growth	2026	用日志和自然实验拆分平台增长与页面优化效果
03 GEO 测量评估	From Citation Selection to Citation Absorption	2026	引用选择与引用吸收两阶段框架
03 GEO 测量评估	Generative Engine Optimization at Scale	2026	跨品牌、跨提示、大规模 AI 搜索可见性测量
03 GEO 测量评估	SAGEO: A Search Arena for Generative Engine Optimization in a Realistic Environment	2026	真实检索、重排、结构保留的 GEO 评估环境
03 GEO 测量评估	Structural Feature Engineering for Generative Engine Optimization	2026	宏观、中观、微观结构特征带来的引用与质量提升
04 AI 搜索实证	A Survey of Generative Search and Recommendation in the Era of Large Language Models	2024/2025	生成式搜索和推荐系统综述

类别	论文题名	年份	在 GEO 内容工程中的用途
04 AI 搜索实证	Characterizing Web Search in the Age of Generative AI	2026	内部知识、外部来源、多样性和稳定性的系统比较
04 AI 搜索实证	How Generative AI Disrupts Search: Consumer Behavior, Website Traffic, and the Impact of Google AI Overviews	2026	AI Overviews 对用户行为、流量和点击路径的影响
04 AI 搜索实证	NExT-Search: Rebuilding User Feedback Ecosystem for Generative AI Search	2025	生成式搜索中答案级反馈替代网页级反馈的问题
04 AI 搜索实证	What Evidence Do Language Models Find Convincing?	2024	模型在冲突证据中偏好的事实、相关性和风格因素
05 AEO 理论整合	From SEO to Answer Engine Optimization (AEO)	2025	从搜索排名到答案可见性的营销转型
05 AEO 理论整合	Optimizing for the Artificial Intelligence-Driven Search Era: An Integrated Framework for SEO, GEO, and AEO	2025	SEO、GEO、AEO 的整合框架
05 AEO 理论整合	Smart Search Optimization: A Theoretical Framework Integrating SEO, AEO and GEO	2025	理论整合与智能搜索优化
05 AEO 理论整合	How the Transition from SEO to GEO, AEO, and AIO is Impacting Digital Marketing Strategies	2025	数字营销策略迁移
05 AEO 理论整合	Zero-Click Search and Answer Engines: A Literature Review on SEO Transformation in Digital Marketing	2025/2026	零点击和答案引擎的文献综述
06 风险、操纵与对抗	Adversarial Search Engine Optimization for Large Language Models	2024	LLM 搜索中的偏好操纵和对抗式 SEO 风险
06 风险、操纵与对抗	CONFLICTBANK: Evaluating the Influence of Knowledge Conflicts in LLMs	2024	知识冲突对模型回答的影响
06 风险、操纵与对抗	Dynamics of Adversarial Attacks on LLM-Based Search Engines	2025	LLM 搜索引擎中的对抗攻击动态
06 风险、操纵与对抗	GASLITEing the Retrieval	2024/2025	稠密向量检索的脆弱性
06 风险、操纵与对抗	Manipulating Large Language Models to Increase Product Visibility	2024	商品可见性操纵风险
06 风险、操纵与对抗	Persistent Pre-Training Poisoning of LLMs	2024	预训练投毒的长期影响
06 风险、操纵与对抗	PoisonArena: Competing Poisoning Attacks in RAG	2025	RAG 中多攻击者竞争性投毒
06 风险、操纵与对抗	Ranking Manipulation for Conversational Search Engines	2024	提示注入与上下文位置对会话式搜索排名的影响
06 风险、操纵与对抗	StealthRank: LLM Ranking Manipulation via Stealthy Prompt Optimization	2025	隐蔽文本序列对 LLM 排名的操纵
06 风险、操纵与对抗	Unveiling the Resilience of LLM-Enhanced Search Engines Against Adversarial Attacks	2026	LLM 增强搜索对抗鲁棒性
07 垂直场景与多模态	Caption Injection for Optimization in Generative Search Engines	2025	图注作为多模态内容可见性入口
07 垂直场景与多模态	E-GEO: A Testbed for GEO in E-Commerce	2025	电商商品查询、商品页重写和基准测试
07 垂直场景与多模态	Multimodal Generative Engine Optimization: Rank Manipulation for Vision-Language Model Rankers	2026	VLM 排名器与多模态排名操纵风险
07 垂直场景与多模态	Generative Engine Optimization: A VLM and Agent Framework for Pinterest Acquisition Growth	2026	Pinterest 场景中的 VLM、趋势挖掘和增长应用
07 垂直场景与多模态	When Content is Goliath and Algorithm is David	2025	内容风格和语义对生成式搜索的影响
08 AI 搜索架构与 Agentic Search	Decoupling Search from Reasoning: A Vendor-Agnostic Grounding Architecture for LLM Agents	2026	把搜索接地从推理模型中解耦，控制路由、上下文和缓存
08 AI 搜索架构与 Agentic Search	Beyond Parallel Sampling: Diverse Query Initialization for	2026	多样化初始查询提升复杂搜索覆盖

类别	论文题名	年份	在 GEO 内容工程中的用途
	Agentic Search		
08 AI 搜索架构与 Agentic Search	ScholarQuest: A Taxonomy-Guided Benchmark for Agentic Academic Paper Search	2026	学术搜索 Agent 的分类指导基准
08 AI 搜索架构与 Agentic Search	When Does Streaming Tool Use Help?	2026	流式 RAG 中工具意图稳定和延迟隐藏
09 RAG 检索优化	MCompassRAG: Topic Metadata as a Semantic Compass for Paragraph-Level Retrieval	2026	主题元数据提升段落级 RAG 效率
09 RAG 检索优化	SCAR: Semantic Continuity-Aware Retrieval for Efficient Context Expansion in RAG	2026	语义连续性辅助上下文扩展
09 RAG 检索优化	SproutRAG: Attention-Guided Tree Search with Progressive Embeddings for Long-Document RAG	2026	长文档 RAG 的树搜索和渐进式嵌入
10 搜索评估治理	Scaling Search Relevance: Augmenting App Store Ranking with LLM-Generated Judgments	2026	用 LLM 判断扩展搜索相关性训练
10 搜索评估治理	SAGE: Scalable AI Governance & Evaluation	2026	生产级 AI 评估治理、surrogate judge 和可扩展审查

## 附录 B. 可复用检查清单与提示集模板

### B1. Prompt Universe 模板

字段	说明
Prompt ID	唯一编号，例如 P-AI-CS-001
用户原始问题	真实用户可能输入的问题
意图类别	了解、比较、选择、购买、实施、故障、风险
用户角色	CEO、市场、采购、开发者、运营、个人用户
场景约束	预算、行业、地区、语言、工具栈、数据安全
目标页面	希望被引用或吸收的页面
期望答案要点	希望答案中出现的事实或观点
风险词	不希望出现或需要谨慎处理的表述
优先级	P0/P1/P2
测量频率	每日、每周、重大更新后

### B2. 证据原子模板

字段	说明	例子
主张	内容要表达的结论	适合中小跨境电商 7 天试运行
证据	证明该主张的事实	客户 A 的上线记录、部署流程截图
来源	证据来自哪里	案例页、产品文档、访谈、日志
时间	数据或事实的时间范围	2026 年 Q2
适用范围	该主张适合哪些场景	月订单 1000 到 10000 单的 Shopify 店铺
限制条件	哪些场景不适用	复杂客诉仍需人工客服
可引用句	80 到 150 字答案块	见页面摘要段
审核状态	待审核、已审核、已过期	已审核

### B3. Live Engine 测量记录模板

字段	说明
时间戳	具体日期和时间
地区/语言	例如 CN/中文、US/英文
平台/模型	ChatGPT Search、Perplexity、Google AI Overviews、Gemini、豆包、DeepSeek 等
提示文本	原始 prompt
是否触发搜索	是/否/未知
引用 URL	所有引用来源
目标品牌是否出现	是/否
目标页面是否被引用	是/否
答案吸收证据	答案中哪些句子来自目标页面
情感与归因	正向、中性、负向；归因是否准确
截图/HTML 存档	证据文件链接
备注	异常、模型变动、地区差异

## 参考资料

GitHub: yaojingang/geo-citation-lab, 02-geo-aeo-ai-search-papers, README 与论文清单 CSV。

Aggarwal et al. GEO: Generative Engine Optimization. arXiv:2311.09735, KDD 2024.

CC-GSEO-Bench: A Content-Centric Benchmark for Measuring Source Influence in Generative Search Engines. arXiv:2509.05607.

Do not Measure Once: Reliable AI Search Visibility Requires Repeated Trials and Structure-Aware Metrics. arXiv:2604.07585.

Generative Engine Optimization: How to Dominate AI Search. arXiv:2509.08919.

Navigating the Shift: A Comparative Analysis of Web Search and Generative AI Response Generation. arXiv:2601.16858.

IF-GEO: Conflict-Aware Instruction Fusion for Multi-Query Generative Engine Optimization. arXiv:2601.13938.

Role-Augmented Intent-Driven Generative Search Engine Optimization. arXiv:2508.11158.

Think Before Writing: Feature-Level Optimization for Generative Engine Optimization. arXiv:2604.19113.

What Generative Search Engines Like: Explaining and Optimizing Website Visibility in Generative Search. arXiv:2510.11438.

AI Answer Engine Citation Behavior: An Empirical Analysis of the GEO16 Framework. arXiv:2509.10762.

From Citation Selection to Citation Absorption: Understanding Source Influence in LLM-Generated Answers. arXiv:2604.25707.

SAGEO: A Search Arena for Generative Engine Optimization in a Realistic Environment. arXiv:2602.12187.

Structural Feature Engineering for Generative Engine Optimization. arXiv:2603.29979.

Characterizing Web Search in the Age of Generative AI. arXiv:2510.11560.

NExT-Search: Rebuilding User Feedback Ecosystem for Generative AI Search. arXiv:2505.14680.

What Evidence Do Language Models Find Convincing? arXiv:2402.11782.

Adversarial Search Engine Optimization for Large Language Models. arXiv:2406.18382.

Ranking Manipulation for Conversational Search Engines. arXiv:2406.03589.

PoisonArena: Uncovering Competing Poisoning Attacks in Retrieval-Augmented Generation. arXiv:2505.12574.

E-GEO: A Testbed for Generative Engine Optimization in E-Commerce. arXiv:2511.20867.

Decoupling Search from Reasoning: A Vendor-Agnostic Grounding Architecture for LLM Agents. arXiv:2606.18947.

MCompassRAG: Topic Metadata as a Semantic Compass for Paragraph-Level Retrieval. arXiv:2606.18508.

SAGE: Scalable AI Governance & Evaluation. arXiv:2602.07840.