

海外 GEO 研究长版报告（口径重算版）

版本说明：本版基于 2026-04-21 的目录内容重算，统一使用清洗后的搜索层口径与 features_all_platforms_72.csv 的引用影响力口径。旧版报告中若存在与本版不一致的数字，以本版为准。

执行摘要

这份研究要回答的不是“AI 会不会搜索”，而是更接近实战的问题：

1. 什么样的 Prompt 更容易触发 AI 联网搜索？
2. 触发搜索后，AI 会优先相信什么样的来源？
3. 同样被引用的页面里，什么样的内容会被 AI 深度吸收，而不是只在来源列表里出现一次？

整套资料的底座是：

- 602 条实验 Prompt
- 3 个平台：ChatGPT、Google AI Overview / Gemini、Perplexity
- 搜索层清洗后 21,181 行结果，其中有效引用 21,143 行
- 引用影响力层 23,745 行、72 维特征
- 抓取成功的引用页面 18,151 个，成功率 76.44%

本版重算后，最重要的结论是：

- 三平台几乎都会触发搜索，但搜索策略差异很大：ChatGPT 平均每条 Prompt 引用 6.88 个来源，Google 为 12.06，Perplexity 为 16.35。
- 搜索广度不等于吸收深度。按抓取成功页面计算，ChatGPT 的单条引用平均影响力为 0.2713，显著高于 Google 的 0.0584 与 Perplexity 的 0.0646。
- 对 AI 真正有效的，不是“短而精的观点文章”，而是“长、结构化、可抽取、强对题”的证据页。高影响力页面平均 1,943 词，低影响力页面只有 170 词。
- AI 偏好高度集中：官网 + 新闻 + 行业垂类 是绝对主战场；在可识别样本里，US 来源占 82.70% - 86.76%，英文来源占 82.90% - 95.07%。

如果只看一句话，这份研究的核心判断是：

GEO 的核心不是“让 AI 看见你”，而是“让你的页面足够像一个可被模型拆解、搬运、重组的证据容器”。

一、研究目标与数据逻辑

1.1 本研究在拆哪条链路

这份资料研究的是 AI 搜索时代的完整内容链路：

用户提问 -> 是否触发搜索 -> 选择哪些来源 -> 如何使用这些来源 -> 最终哪些页面真的影响了回答

因此整套分析分成两层：

- **搜索触发层**：观察 Prompt 如何触发联网搜索，以及三平台选了哪些来源站点
- **引用影响力层**：观察被引用页面的结构、语义与体裁，如何影响它在 AI 回答中的真实作用

1.2 数据资产清单

模块	文件/范围	作用
Prompt 设计	01-prompt/	602 条实验输入
搜索触发与信源	02-data/*results_with_prompt.csv	记录是否触发搜索、引用域名、网站类型、DR、国家、语言等
引用影响力特征	02-data/features_all_platforms_72.csv	每条引用一行，共 72 维特征
分析脚本	03-pipeline/	解析、抓取、特征提取、统计分析
长版报告	04-repet/	Markdown / HTML / PDF 导出

1.3 为什么要重算

这次重写不是简单润色，而是做了一次口径统一，主要修正四类问题：

- `chatgpt_results_with_prompt.csv` 混入了 16 行重复表头，需要先清洗
- ChatGPT 搜索层里 `A_news`、`A_technology` 被写成 `Anews*`、`Atechnology*`，旧脚本会漏统计
- ChatGPT 搜索层只覆盖到 587 个 Prompt，仍缺 15 个 Prompt 输出，不能继续按 602 口径硬写
- 影响力分析应基于 `fetch_ok = True` 的成功抓取页面，否则会把抓取失败页面一起拉低均值

因此，本版报告明确使用两套口径：

- 搜索层：基于清洗后的 21,181 行结果与 587 / 602 / 602 个平台级 Prompt 观测值
- 影响力层：基于 23,745 行特征表中的 18,151 条成功抓取记录

二、实验设计：602 条 Prompt 如何搭起来

2.1 四层实验架构

层级	数量	目的
A 层	432	主实验层，控制任务类型、触发强度、时效性、行业与子任务
B 层	60	风格对照层，比较自然提问、要求来源、角色型 Prompt
C 层	60	语言对照层，对比英文与中文 Prompt
D 层	50	极端与真实场景层，覆盖高风险、模糊、多约束、长决策问题

2.2 A 层：主实验骨架

A 层由以下变量交叉组成：

- `TaskType`：事实、对比、聚合、解释/推理
- `TriggerLevel`：Low / Medium / High

- `TimeSensitivity` : Stable / Recent / Realtime
- `Industry` : Technology / Commerce / Local / Healthcare / Finance / News
- `SubType` : Info / Decision

它的价值不在于“多”，而在于可以把问题拆成可比较的控制变量。

2.3 B/C/D 层：为什么重要

- `B 层` 用来回答：同一个问题，换一种说法会不会改变搜索强度？
- `C 层` 用来回答：中英文环境会不会改变平台的联网倾向？
- `D 层` 用来回答：遇到真实世界里高风险、模糊、多约束问题时，平台会怎么做？

三、处理流程与方法口径

3.1 搜索层处理流程

搜索层使用三份 CSV：

- `chatgpt_results_with_prompt.csv`
- `Google_results_with_prompt.csv`
- `perplexity_results_with_prompt.csv`

每行记录一个被引用域名，并附带如下字段：

- `是否触发搜索`
- `引用域名`
- `最终评级(Final_DR)`
- `网站类型`
- `国家(Country)`
- `语言(Language)`

在 Prompt 级统计时，本版使用：

- 同一 `文件名` 下任一行 `是否触发搜索 = 是`，即视为该 Prompt 触发搜索
- 同一 `文件名` 下有效 `引用域名` 的行数，作为该 Prompt 的引用数

3.2 引用影响力层处理流程

引用影响力分析的逻辑是：

1. 从三平台回答页解析引用 URL
2. 批量抓取引用网页正文
3. 为每条引用提取 72 维特征
4. 计算每条引用在 AI 回答中的 `influence_score`

全量特征表共有 23,745 行，抓取成功 18,151 行，平台分布如下：

平台	原始特征行数	抓取成功行数	抓取成功率
ChatGPT	4,494	3,323	73.97%
Google	8,476	6,385	75.33%
Perplexity	10,775	8,443	78.36%
合计	23,745	18,151	76.44%

3.3 72 维特征里最关键的几类

虽然表里有 72 个字段，但可以抽象成五类：

- **页面结构**：词数、标题数、段落数、列表密度、表格/图片/链接等
- **内容质量**：信息密度、句长、词汇丰富度、标题与问题匹配度等
- **内容体裁**：是否含数字、定义、对比、how-to、Q&A、代码
- **语义对齐**：Embedding 相似度、LLM 相关性评分、内容质量评分
- **结果变量**：引用次数、首次出现位置、覆盖段落比例、文本重叠度

3.4 影响力分数怎么定义

本项目使用以下加权分数描述“某条引用对最终回答的真实影响力”：

```
influence_score =  
    0.20 × min(ref_count / 3, 1)  
+ 0.15 × (1 - first_position_ratio)  
+ 0.20 × paragraph_coverage_ratio  
+ 0.25 × tfidf_cosine  
+ 0.20 × (bigram_overlap + trigram_overlap) / 2
```

这意味着：

- 被引用多次、出现更靠前、覆盖更多回答段落的页面，分数更高
- 与回答文本重合度更高、语义更贴近的页面，分数也更高

3.5 一个重要的方法论约束

由于 `ref_count`、`position`、`coverage`、`tfidf`、`ngram overlap` 本身就是 `influence_score` 的定义组件，因此本版不把它们当“原因变量”来讲发现。真正用来解释影响力的，是：

- 页面结构
- 内容体裁
- 语义对齐
- 独立的 LLM 评分

四、搜索触发层：三平台到底怎么搜

4.1 平台总览

平台	观测 Prompt 数	触发搜索 Prompt 数	触发率	平均引用数	中位引用数	单条 Prompt 最大引用数
ChatGPT	587	579	98.64%	6.88	6	21
Google	602	600	99.67%	12.06	12	37
Perplexity	602	602	100.00%	16.35	17	27

第一层结论非常明确：

- 三平台基本都会触发搜索
- 但 Perplexity 和 Google 更像“宽搜型”平台
- ChatGPT 搜得更收敛，平均引用宽度明显更窄

如果从策略上总结：

- **ChatGPT**：少而精，后续对单条引用使用更深
- **Google**：搜索覆盖广，对来源要求敏感
- **Perplexity**：几乎无条件联网上下文补全，引用最密集

4.2 分层看：A/B/C/D 哪一层最能拉高搜索强度

平台	A 层	B 层	C 层	D 层
ChatGPT	6.75	7.13	7.40	7.02
Google	12.15	13.45	9.55	12.66
Perplexity	16.41	16.52	16.18	15.86

这里三个值得注意的点：

- Google 对 B 层风格对照 最敏感，说明“怎么提问”对它的联网强度影响很大
- ChatGPT 在 C 层语言对照 上反而略高，意味着中文并未削弱其搜索意愿
- Perplexity 几乎对任何层级都保持高强度联网，说明其默认策略就是“先搜再说”

4.3 B 层：换个问法，AI 会不会搜得更多

平台	自然提问	要求来源	角色型 Prompt
ChatGPT	7.30	6.15	7.95
Google	14.05	15.90	10.40
Perplexity	15.70	17.15	16.70

这部分的结论不是“要求来源一定最强”，而是：

- Google 和 Perplexity 明显吃“要求来源”的刺激
- ChatGPT 则更偏好角色型 Prompt，对“强行要求来源”没有同样强的提升

因此，Prompt 包装对搜索强度的影响是平台相关的，不应简单概括为一个统一结论。

4.4 C 层：英文与中文，平台反应一样吗

平台	中文 Prompt 平均引用数	英文 Prompt 平均引用数
ChatGPT	7.77	7.03
Google	7.53	11.57
Perplexity	15.93	16.43

这组结果很有价值：

- Google 明显更偏向英文上下文
- Perplexity 也略偏英文，但差距不大
- ChatGPT 在这组样本里反而中文略高

这说明“英文一定更能触发 AI 搜索”并不严谨。更准确的说法是：

英文对 Google 的加成最大；对 ChatGPT，这种加成在本样本中并不成立。

4.5 D 层：真实世界问题，平台会怎么处理

场景	ChatGPT	Google	Perplexity
高风险	6.0	13.9	16.0
模糊问题	7.9	8.9	13.1
多约束任务	3.4	12.6	17.7
长决策问题	9.2	14.5	17.4
宏观趋势类	8.6	13.4	15.1

这一组里最反常识的现象是：

- ChatGPT 面对 **多约束任务** 时平均只引用 **3.4** 个来源，明显低于其它场景
- Perplexity 在同类问题上达到 **17.7**，几乎是完全相反的策略

可以把它理解成：

- ChatGPT 更容易把复杂约束问题压缩成自身推理任务
- Perplexity 更倾向把复杂任务拆成广泛的信息检索与拼装

五、来源偏好：AI 更容易选中什么样的网站

5.1 行业维度：不同平台最爱搜哪些领域

平台	Commerce	Finance	Healthcare	Local	News	Technology
ChatGPT	6.88	6.86	7.70	5.40	6.25	7.49
Google	11.94	10.51	13.92	12.96	11.71	11.83
Perplexity	16.36	17.06	16.24	16.96	15.68	16.17

平台偏好并不完全相同：

- ChatGPT 最强的是 Healthcare 和 Technology
- Google 最强的是 Healthcare 和 Local
- Perplexity 最强的是 Finance 和 Local

如果从内容策略上理解：

- 医疗、科技、金融这类高信息密度领域，更容易触发大范围来源整合
- 本地类内容在 ChatGPT 上明显弱，说明缺少明确地理锚点时，它更容易回退到内部知识

5.2 网站类型：三平台的候选池高度集中

平台	官网	新闻	行业垂类	其他高频类型
ChatGPT	34.22%	31.17%	22.13%	测评类 4.51%、blog 4.04%
Google	46.35%	18.99%	22.00%	blog 5.04%、测评类 2.45%
Perplexity	44.07%	16.07%	18.99%	unknown 4.47%、噪声值“成功”3.86%

把这三类加总后：

- ChatGPT: 87.52%
- Google: 87.34%
- Perplexity: 79.12%

结论非常直接：

官网、新闻、行业垂类不是“会被引用的一类来源”，而是几乎定义了 AI 搜索的默认候选池。

5.3 地区与语言：谁在被优先看见

在排除 `unknown` 和 `WW` 后，三平台在可识别国家样本中的 `US` 占比如下：

平台	US 占比
ChatGPT	85.89%
Google	86.76%
Perplexity	82.70%

在排除 `unknown` 后，三平台在可识别语言样本中的英文占比如下：

平台	英文占比
ChatGPT	95.07%
Google	91.98%
Perplexity	82.90%

因此，对希望“先进入 AI 候选池”的内容团队来说，最稳妥的现实结论是：

- 首发环境优先英文
- 分发环境优先美国或强全球化域名
- 非英文、非美区内容不是没有机会，但竞争门槛更高

5.4 域名权威度：高 DR 仍是硬门槛

平台	平均 Final_DR	中位 Final_DR
ChatGPT	584.60	592
Google	541.15	526
Perplexity	558.33	542

这说明一件很现实的事：

在“能不能先被选中”这一层，域名权威度仍然非常重要。

它不是决定引用深度的唯一因素，但明显决定了内容是否有资格进入第一轮候选。

5.5 高频域名：AI 到底在重复看谁

搜索层合并后的 Top 15 域名如下：

排名	域名	次数
1	youtube.com	560
2	en.wikipedia.org	352
3	reddit.com	315
4	reuters.com	287
5	linkedin.com	187
6	nytimes.com	174
7	pmc.ncbi.nlm.nih.gov	167
8	facebook.com	151
9	forbes.com	146
10	finance.yahoo.com	146
11	deloitte.com	134
12	theguardian.com	124
13	wsj.com	122
14	investopedia.com	121
15	weforum.org	121

这个结果反映出三类典型来源：

- 解释型/百科型来源：如 Wikipedia
- 新闻型来源：如 Reuters、NYT、WSJ
- 平台型/聚合型来源：如 YouTube、Reddit、LinkedIn

需要强调的是：高频出现不等于高影响力。它更像是在说，这些来源经常能进入候选池。

六、引用影响力层：什么页面会被 AI 深度吸收

6.1 平台层面的“引用深度”差异

以下分析仅基于抓取成功页面：

平台	抓取成功引用数	平均影响力	中位影响力
ChatGPT	3,323	0.2713	0.2611
Google	6,385	0.0584	0.0515
Perplexity	8,443	0.0646	0.0333

这是本研究最重要的一张表之一。

它说明：

- ChatGPT 虽然引用少，但会更集中地使用少数来源
- Google 与 Perplexity 更像“广引薄用”
- 对 GEO 来说，“被 ChatGPT 引一次”和“被 Google 引一次”不能被视为等价事件

6.2 长度与结构：高影响力页面长什么样

按 `influence_score` 的 Top 25% 与 Bottom 25% 对比：

指标	Top 25%	Bottom 25%	倍数
词数	1,943.30	169.82	11.44x
标题总数	10.59	0.85	12.50x
段落数	47.49	8.34	5.69x
列表密度	0.428	0.048	8.94x
回答-引用语义相似度	0.570	0.247	2.31x
LLM 相关性评分	3.535	1.856	1.90x
LLM 内容质量评分	3.404	2.289	1.49x

这张表说明高影响力页面并非只是“更长”，而是更像一个可拆解的信息容器：

- 标题分层更清楚
- 段落更多
- 列表更密
- 与问题和回答的语义贴合更高

6.3 多长最合适：不是越短越好

页面词数区间	样本数	平均影响力	中位影响力
<=100	5,281	0.0546	0.0055
101-300	1,597	0.0846	0.0360
301-600	2,130	0.1129	0.0590
601-1000	2,417	0.1117	0.0674
1001-3000	4,742	0.1258	0.0838
>3000	1,984	0.1457	0.0990

这组结果回答了一个常见误解：

AI 并不偏爱“短小摘要页”，它更偏爱“信息足够充分、可以切出多个可复用片段”的长内容页。

如果从制作成本与收益平衡来看：

- 1000-3000 词是最稳妥的实操区间
- >3000 词仍然更强，但维护成本明显更高

6.4 哪些特征最能预测高影响力

在成功抓取页面中，与影响力相关性最高的独立特征包括：

特征	与影响力的相关系数 r
LLM 相关性评分	0.4322
回答-引用 Embedding 相似度	0.3561
LLM 内容质量评分	0.2917
问题-引用 Embedding 相似度	0.2548
页面词数	0.1995
含定义句式	0.1934
含数字/统计	0.1842
标题总数	0.1751
含对比内容	0.1741

这表明真正强的不是单个机械 SEO 指标，而是两类能力叠加：

- 强语义对齐
- 强结构化表达

6.5 布尔体裁特征：AI 明显更爱什么

特征	True 平均影响力	False 平均影响力	提升
含代码	0.1747	0.0988	+76.88%
含数字/统计	0.1171	0.0725	+61.55%
含定义句式	0.1252	0.0795	+57.33%
含对比内容	0.1389	0.0894	+55.28%
含 how-to	0.1296	0.0918	+41.20%
含 Q&A 格式	0.0947	0.1005	-5.74%

这一节非常适合直接指导内容生产：

- 定义
- 数字
- 对比
- 步骤

这四类内容，不是“看起来更像干货”，而是客观上更容易进入 AI 的搬运与重组流程。

而纯粹的 Q&A 页面 没有天然优势，这本身就是一个反常识发现。

七、AI 在“怎么用引用”这件事上的偏好

7.1 语义角色：什么角色的内容最值钱

先看高频语义角色及其平均影响力：

语义角色	样本数	平均影响力
evidence	6,190	0.1235
reference	4,094	0.0529
background	2,464	0.0801
example	1,291	0.1047
definition	1,099	0.1531
statistical_data	1,048	0.1120
opinion	799	0.0938
comparison	778	0.1524
procedure	121	0.0717

如果只看主流角色，可以得出一个很清楚的排序：

- `definition`、`comparison` 最强
- `evidence`、`statistical_data` 次之
- `reference` 很弱，说明很多页面只是被列在来源里，并没有被深度消化

7.2 AI 对引用的使用方式

使用方式	样本数	平均影响力
factual_basis	9,511	0.1224
supplementary	5,673	0.0678
paraphrase	1,245	0.1443
reference	1,037	0.0444
structural_guide	193	0.0618
background	119	0.0697
example	41	0.1294
direct_quote	19	0.1633

需要注意：

- `direct_quote` 虽然均值高，但样本只有 19，不应过度解读
- 样本量最大且最稳定的高价值方式是 `factual_basis` 和 `paraphrase`
- `reference` 的均值只有 0.0444，说明很多来源只是“被点到”，并没有实质作用

7.3 站点身份与深度吸收并不是一回事

按 `domain_type` 看平均影响力，样本量较有代表性的几类如下：

站点类型	样本数	平均影响力
encyclopedia	527	0.2144
commercial	11,779	0.1028
nonprofit	2,009	0.0971
academic_publishing	86	0.1118
academic	1,024	0.0815
government	892	0.0769
news_media	1,546	0.0726

这个结果很值得讲清楚：

- 新闻媒体很容易进入候选池
- 但平均影响力未必最高
- 百科型、结构化解释型来源虽然频次不一定第一，却更容易被深度吸收

这意味着：

“先到高权威新闻站”解决的是进入候选池的问题；“把页面写成定义清楚、结构清楚的解释页”解决的是被深度吸收的问题。

八、平台差异：三大 AI 搜索不是一套玩法

8.1 各平台最强驱动因素不同

按平台分别看主要驱动因素，与影响力相关性最高的特征如下：

ChatGPT

特征	r
llm_relevance_score	0.5370
emb_answer_cit_cosine	0.4210
emb_question_cit_cosine	0.3664
llm_content_quality	0.3354
cit_has_definition	0.2271
cit_heading_total	0.1940

ChatGPT 最像“深读型平台”：

- 极重视内容是否真的对题
- 也重视页面能否作为定义与证据来源被复用

Google

特征	r
emb_answer_cit_cosine	0.5789
emb_question_cit_cosine	0.5516
cit_has_definition	0.4027
cit_word_count	0.3503
llm_relevance_score	0.3486
cit_heading_total	0.3172

Google 的信号最像“强语义对齐 + 清晰定义页”：

- 对问题与答案的语义匹配特别敏感

- 标题、词数、定义结构都更重要

Perplexity

特征	r
llm_relevance_score	0.3884
emb_answer_cit_cosine	0.3288
cit_heading_total	0.2577
emb_question_cit_cosine	0.2416
cit_word_count	0.2245
llm_content_quality	0.2191

Perplexity 仍然重视相关性，但更强调页面的“可拆解性”：

- 标题数
- 页面长度
- 能否覆盖更多子问题

8.2 对内容策略的直接含义

如果必须把三平台区别开来，可以用一句话概括：

- **ChatGPT**：重“深度吸收”，适合做定义、证据、背景整合型页面
- **Google**：重“对题与定义”，适合做标题-关键词-正文强对齐页面
- **Perplexity**：重“覆盖广度”，适合做综合型、模块化、可拼装页面

九、分行业与分问题类型：哪些题更值得做

9.1 行业维度

按成功抓取样本计算，六大行业的平均影响力如下：

行业	样本数	平均影响力
A_technology	2,252	0.1272
A_healthcare	2,379	0.1021
A_commerce	2,243	0.0994
A_finance	1,831	0.0965
A_news	2,204	0.0948
A_local	2,223	0.0916

科技、医疗、商业内容更容易出现高影响力引用，原因通常是：

- 定义与术语密度高
- 可验证数据多
- 结构化比较更常见

9.2 问题类型维度

问题类型	样本数	平均影响力
comparison	1,711	0.1093
list	289	0.1081
how_to	563	0.1016
other	6,024	0.1015
what_is	3,594	0.0987
why	3,712	0.0986
which	2,070	0.0946
opinion	188	0.0854

这张表说明：

- **comparison** 型问题最适合让页面吃到高影响力引用
- **opinion** 型问题最弱

原因也很直观：

- 对比类问题天然需要结构化差异、指标和证据
- 观点类问题更容易被 AI 自己总结，页面的“搬运价值”反而下降

十、16 条最值得对外讲的核心洞察

1. 研究底座为 **602** 条 Prompt、**21,143** 条有效搜索层引用、**23,745** 条 citation-level 记录、**72** 维特征。
2. 三平台几乎都会搜索：ChatGPT **98.64%**、Google **99.67%**、Perplexity **100%**。
3. 搜索宽度差很大：平均每条 Prompt 引用数分别为 **6.88**、**12.06**、**16.35**。
4. 搜索广度不等于引用深度：ChatGPT 平均影响力 **0.2713**，分别是 Google 的 **4.64x**、Perplexity 的 **4.20x**。
5. Google 最吃“要求来源”的 Prompt，B 层中该类平均引用数达到 **15.90**。
6. ChatGPT 在 B 层里更吃角色型 Prompt，平均引用数 **7.95**。
7. 英文对 Google 的提升最明显：**11.57 vs 7.53**。
8. ChatGPT 在 C 层里中文略高于英文：**7.77 vs 7.03**。
9. 模糊问题不会逼平台更努力搜索，D 层里模糊问题整体均值只有 **9.97**。
10. ChatGPT 面对多约束任务最保守，平均只引用 **3.4** 个来源。
11. **官网 + 新闻 + 行业垂类** 占三平台来源的 **79.12% - 87.52%**。
12. 可识别国家中，US 来源占 **82.70% - 86.76%**；可识别语言中英文占 **82.90% - 95.07%**。
13. 被引用来源的中位 **Final_DR** 在 **526 - 592** 之间，说明高权威域名仍是候选池门槛。
14. 高影响力页面平均 **1,943** 词，低影响力页面仅 **170** 词。
15. 含数字、定义、对比、how-to 的页面平均影响力提升分别为 **+61.55%**、**+57.33%**、**+55.28%**、**+41.20%**。

16. 纯 Q&A 页面没有加成，平均影响力反而下降 5.74%。

十一、方法归因复盘：这些模式为什么成立

11.1 为什么长内容更强

不是因为 AI 天生偏爱长文，而是因为长页面能提供更多可切片的信息单元：

- 更多标题
- 更多段落
- 更多列表
- 更多子主题

这让模型可以在不同回答段落中重复使用同一起来源的不同片段。

11.2 为什么“定义 + 数字 + 对比 + 步骤”最有效

这四类内容之所以强，并不是形式主义，而是因为它们最符合模型的搬运需求：

- 定义 解决概念锚定
- 数字 提供可验证事实
- 对比 提供结构化判断框架
- 步骤 提供直接可复用的操作顺序

11.3 为什么新闻高频但未必最深

新闻网站容易成为“发生了什么”的初始来源，因此很容易被列入候选池；但 AI 在真正组织答案时，往往还需要：

- 定义型页面
- 百科型解释页
- 结构化对比页

因此新闻频率高，不代表它对最终答案的贡献也最高。

11.4 为什么 Q&A 页面没占到便宜

很多团队会误以为“把内容写成问答”就天然适合 AI。数据说明并非如此。

Q&A 格式本身不构成优势，因为 AI 真正需要的是：

- 信息块清晰
- 证据密度高
- 结构化强
- 与问题高度相关

如果问答页没有这些东西，它只是“看起来像 AI 内容”，但不是真正高价值的 AI 引用页。

十二、面向 GEO 的实操建议

12.1 内容生产

优先写以下类型的页面：

- 定义清晰的解释页
- 含数字与证据的分析页
- 含对比结构的选择页
- 含步骤的 how-to 页

更实用的页面骨架通常长这样：

1. 开头直接定义主题
2. 给出 3-5 个关键数字或事实
3. 做一个结构化对比
4. 给出步骤或判断框架
5. 补充边界条件与适用场景

12.2 篇幅与结构

最稳妥的建议不是“越长越好”，而是：

- 目标篇幅至少 1000+ 词

- 至少 **6-10** 个清晰小节
- 小节标题尽量贴近用户可能会问的子问题
- 重点信息尽量列表化

12.3 分发环境

如果目标是先进入候选池：

- 优先权威域名或强垂类站点
- 优先英文与美区环境
- 优先官网、新闻媒体、行业垂类

如果目标是被深度吸收：

- 页面本身要有定义、数字、对比、步骤
- 标题与正文必须强对题
- 不要把所有信息压缩成一段“短总结”

12.4 平台化打法

- 面向 **ChatGPT**：写深度解释与证据整合页
- 面向 **Google**：强化标题关键词对齐与定义结构
- 面向 **Perplexity**：覆盖更多子问题，让页面适合被拆成多个答案片段

十三、开源时必须写清楚的 caveats

为了避免外部用户误读，这些 caveats 必须保留：

1. ChatGPT 搜索层只覆盖 **587** 个 Prompt，缺少 **15** 个 Prompt 输出。
2. `chatgpt_results_with_prompt.csv` 有 **16** 行重复表头，统计前必须清洗。
3. `A_news` 与 `A_technology` 在 ChatGPT 搜索层里原始命名为 `Anews*`、`Atechnology*`，需要归一化。
4. **国家(Country)** 和 **语言(Language)** 存在大量 `unknown / WW`，所以地区与语言结论要注明“可识别样本口径”。
5. **网站类型** 中存在少量噪声值，如 `成功`，公开仓库里应视为待进一步标准化的脏值。

6. `llm_semantic_role` 与 `llm_influence_type` 有少量超出预设 taxonomy 的长尾标签，应视为模型输出漂移，而不是正式标签体系的一部分。
7. 本仓库更适合作为一次静态研究快照，而不是实时监测系统，因为未在公开数据层保留统一采集时间戳。

十四、复现与导出

14.1 环境变量

公开版脚本已改为读取环境变量：

```
export OPENAI_API_KEY=...
export GEMINI_API_KEY=...
export DATAFORSEO_BASE64_AUTH=...
export AHREFS_API_KEY=...
export BATCH_API_TOKEN=...
export BATCH_API_BASE_URL=http://188.166.211.11:9000
```

14.2 常见重跑方式

重算影响力报告：

```
cd 03-pipeline
python3 analyze_influence.py \
  --input ../02-data/features_all_platforms_72.csv \
  --output ../04-repet/citation_influence_report.md
```

重导出长版 HTML：

```
cd 04-repet
python3 build_self_contained_html.py
```

14.3 本次长版报告的导出方式

本次长版文档导出链路为：

```
final_report.md
-> build_self_contained_html.py
-> final_report.html
-> chromium --headless --print-to-pdf
-> final_report.pdf
```

结论

如果把整份研究压缩成一句最终结论，那就是：

在 AI 搜索时代，最有价值的内容不是“最会写观点的内容”，而是“最容易被模型拆成定义、数字、对比、步骤并重组成答案的内容”。

从进入候选池到被深度吸收，这条链路至少由四件事共同决定：

- 站点是否足够强，能先被选中
- 页面是否足够对题，能通过语义筛选
- 结构是否足够清晰，能被模型抽取
- 内容是否足够像证据，而不只是观点

这也是本研究对 GEO 最实用的启发：

- 先解决“看得见”
- 再解决“用得深”
- 不要把 SEO for AI 理解成写一堆 AI 口吻内容
- 真正有效的是把页面做成高质量、强结构、强对题的证据容器