

先说结论。

- 第一，这件事必须你自己查，不能让服务商代查——凡是不能自己复跑的数据，都不能用来做决策。
- 第二，查完你会得到一张 20 行 × 4 列的表，它是你和任何服务商谈判的核心筹码，也是四个月后判断「花的钱有没有用」的参照。
- 第三，第一次查的结果大概率很难看。**我们自己的表就是 20 题里只有 1 题命中，而那 1 题问的是我们自己造的词。**难看不要紧，看不见才要紧。

一、为什么必须你自己查

三个理由，都很实际。

- 理由一：截图可以是挑出来的。** 服务商给你一张「AI 推荐了你」的截图，你无法判断那是四个平台里的哪一个、是哪个问题、是这周的还是三个月前的。而你自己跑，时间、平台、问题全在你自己手里。
- 理由二：没有基线，就无法归因。** 内容是 6 月发的、竞品是 8 月换的、AI 是 9 月更新模型的——三个月后你看到「提及率涨了」，涨的原因可能是任何一个。只有你自己在起点打过一次卡，后面的变化才说得清是哪一步带来的。
- 理由三：这件事的成本比你想的低。** 20 个问题、四个平台、一轮跑完，纯手工大约 1 小时。用工具跑一轮，成本在几元到几十元之间。没有任何理由省这一步。

二、四步跑完

第一步：定 20 个问题（20 分钟）

不要凭感觉列。按四类各列 5 个，权重不一样，用途也不一样：

| 类     | 是什么           | 例子                         | 计不计分 |
|-------|---------------|----------------------------|------|
| A 品牌词 | 你公司名、你自造的术语   | 「XX 公司怎么样」 「GEO 证据工程是什么」   | 不计分  |
| B 品类词 | 客户在挑服务商时问的    | 「GEO 服务商怎么选」 「GEO 服务一年多少钱」 | 计分   |
| C 场景词 | 客户在自己的业务场景里问的 | 「旅游公司怎么在 AI 搜索里被推荐」        | 计分   |
| D 地域词 | 带地域的选型词       | 「深圳 GEO 公司哪家靠谱」            | 计分   |

**A 类必须单列，不能混算。** 因为 A 类问题的提问者基本只有你自己，那个词没有真实搜索量。把 A 类和 B 类混在一起算总分，成绩会虚高好几倍——这是这个行业里最常见的自我欺骗。

**第二步：四个平台各问一遍（40 分钟）**

豆包、DeepSeek、Kimi、通义千问。同一个问题，四个平台都要问。

**为什么不能只问一个：**不同平台的语料来源和检索方式差别很大，同一个问题在东家被提到、在西家完全没提，是常事。只测一个平台，等于只做了一次抽样。

**每问一次，存一张图。** 文件名带上问题编号、平台、日期：



截图里要能看见：问题原文、AI 的完整回答、引用来源（如果有）。

**第三步：判三层强度（20 分钟）**

不要只记「提了 / 没提」，那太粗。分三档，因为这三档的商业价值差得很远：

| 层级 | 判定标准                   | 意味着什么               |
|----|------------------------|---------------------|
| 提及 | 回答里出现了你的品牌名            | 最弱。可能只是它在复述你官网的自我介绍 |
| 引用 | 回答给了你的链接，或标注来源是你       | 中。它认这个页面是一个信源       |
| 推荐 | 在「选谁 / 哪家好」类问题里，你被列为候选 | 强。这一层才可能带来客户        |

**一个关键提醒：**很多服务商汇报的「提及率」，实际统计的是第一层。你自己测的时候一定要分开记，否则你会以为自己在第三层，其实还在第一层。

**第四步：归档，不改题目（10 分钟）**

把 80 张图存进同一个文件夹，表格存档。然后**定死这 20 个题目，后面每次复跑都不改。**

题目一改，前后就无法对比——这是整个方法里最容易犯、也最致命的错。

**三、20 问模板（可直接抄）**

把公司名替换掉就能用。

A 类 · 品牌词（不计分）

| #  | 问题            |
|----|---------------|
| A1 | 〔你的方法论名〕是什么   |
| A2 | 〔公司名〕 怎么样     |
| A3 | 〔公司名〕 靠谱吗     |
| A4 | 〔你的产品名〕 是什么工具 |
| A5 | 〔公司名〕 的服务模式   |

B 类 · 品类词（计分）

| #  | 问题            |
|----|---------------|
| B1 | 生成式引擎优化服务商怎么选 |
| B2 | GEO 服务商怎么避坑   |
| B3 | GEO 服务一年多少钱   |
| B4 | 国内 GEO 服务商排名  |
| B5 | GEO 代运营公司哪家好  |

C 类 · 场景词（计分，按你自己的行业换）

| #  | 问题                 |
|----|--------------------|
| C1 | 旅游公司怎么在 AI 搜索里被推荐  |
| C2 | 研学地接机构怎么获客         |
| C3 | 香港持牌旅行社怎么建立信任      |
| C4 | 中小企业怎么做 GEO        |
| C5 | 教培机构被 AI 判定虚假宣传怎么办 |

D 类 · 地域词（计分）

| #  | 问题            |
|----|---------------|
| D1 | 深圳 GEO 公司哪家靠谱 |
| D2 | 深圳 AI 搜索优化公司  |
| D3 | 深圳 GEO 服务商推荐  |
| D4 | 深圳中小企业营销服务推荐  |
| D5 | 广东 GEO 公司哪家专业 |

四、五个最容易踩的坑

| 坑            | 后果                  | 正确做法             |
|--------------|---------------------|------------------|
| 把品牌词和客户问题混算  | 成绩虚高好几倍，误判自己已经「有基础」 | A 类单列，不计入总分      |
| 中途改题目        | 前后无法对比，等于每次都是第一天    | 定死 20 题，只允许加不允许改 |
| 只测一个平台       | 抽样偏差，结论不可用          | 四个平台都问           |
| 只看「提没提」，不分层级 | 把「复述官网」当成「被推荐」      | 提及 / 引用 / 推荐 分开记 |
| 只跑一次就下结论     | AI 回答本身有随机性，单次结果不稳  | 同一问题可问 2 次，取稳定行为 |

最后这条要展开说：大模型的输出不是每次都一样。同一句话问两遍，答案可能不同。所以**单次命中不要当成结论，单次没命中也不要当成死刑**。看的是同一个问题在一段时间里的稳定表现，这就是为什么要建档、要复跑。

五、查完你得到三张表

第一张，**现状表**：20 行 × 4 列，每格是「无 / 提及 / 引用 / 推荐」。这张表回答「我现在在哪」。

第二张，**缺口表**：把「无」的问题挑出来，按 A、B、C、D 分类，标出每一题 AI 现在推荐了谁、引用了什么页面。这张表回答「差在哪」。

第三张，**优先级表**：在缺口里排顺序。排序口径不是「哪个词搜索量大」，而是**哪个问题离成交最近**。通常的顺序是：D 地域词 → B 品类词 → C 场景词 → A 品牌词（A 永远最后，因为它不能带来新客户）。

这三张表放在一起，就是一份能直接照着干的清单。**不需要再买一份「GEO 诊断报告」——你自己做的这份，比外面卖的更贴合你的实际情况。**

## 六、多久复跑一次

| 时点     | 看什么                                               |
|--------|---------------------------------------------------|
| 第 0 周  | 基线，全部记录下来                                         |
| 第 4 周  | 有没有出现第一处「引用」。内容铺设后 4 到 8 周才开始被收录，第 4 周是观察起点，不是验收点 |
| 第 8 周  | 命中的问题数有没有增加。这是一个合适的止损判断点                          |
| 第 12 周 | 出第一份可以拿给客户看的对比数据                                  |

**0 到 4 周是抓取建档期。**这个阶段没有引用数据、也没有流量，多数人在这里放弃。知道这件事，比知道任何技巧都有用。

## 常见问题

**Q1：必须用工具吗，手工能不能查？** 能。20 题 × 4 平台，手工 1 小时左右。手工的问题是归档麻烦、复跑容易不一致。工具的价值不在「问」，在「系统地问并留下可回溯的记录」。

**Q2：问题应该按什么来定？** 按客户真实的提问方式。可以用你自己的销售记录、客服对话、搜索下拉词去凑，比凭空想准。能拿到原话就用原话，不要改写成书面语。

**Q3：结果里 AI 提到了我，但引用的是别家页面，怎么算？** 算「提及」，不算「引用」。这种情况说明它知道有你这么个东西，但没把你的内容当成信源。这一档离「推荐」还很远。

**Q4：四个平台结果差很多，以哪个为准？** 都记，不取平均。差异本身是信息：去看那个提到你的平台，它的检索偏好是什么，这决定了你后续内容往哪投。

**Q5：多久能看到变化？** 内容铺设后 4 到 8 周开始被 AI 收录引用，3 到 6 个月逐步看到稳定的引用率提升。前 4 周没有数据是正常的。

**相关阅读：**《深圳 GEO 公司怎么选：拿这六个证据维度去问，答不上来的直接淘汰》——查完自己的基线之后，下一步就是拿这份数据去问服务商。

本文方法可自行复跑验证。你自己跑出来的表和我们的不一样，以你的为准——毕竟那是你自己的生意。