

实测，而非猜测

AI 可见度 (GEO) 实务方法论

Measured, Not Guessed: A Field Methodology for AI Visibility (GEO) (英文原题)

作者 AUTHOR

Jun Liu (刘骏)

单位 PRACTICE

JUN & GEO ·
AccessSmith

地点 LOCATION

中国浙江慈溪

编号 DOCUMENT

WP-01

发布 PUBLISHED

2026 年 10 月 20 日

版本 VERSION

1.0 · 中文版 (英文版为
原文)

许可 LICENSE

CC BY-SA 4.0

篇幅 LENGTH

约 10,078 字

■ 目录 CONTENTS

§ 执行摘要 (Executive Summary)

1 问题所在：AI 答案先于网站做决定

2 测什么：曝光度 × 引用，三率，两对象

§ 本文术语定义

3 采样纪律：为什么跑三次

4 认知审计：AI 以为你在卖什么

5 修复工序

6 报告标准与方法论说明

7 局限性

R 参考文献 (References)

A 关于作者 (About the Author)

L 授权 (License)

■ § 执行摘要 (Executive Summary)

买家越来越多地先向 AI 平台——ChatGPT、Perplexity、Copilot、Gemini、Grok、Claude——询问产品推荐，然后才去访问任何一个网站。这些平台给出的答案，如今就是第一份入围名单。不在答案里的品牌，就不在考虑集合里；而在答案里、却被描述错了的品牌，处境可能比缺席还糟。

生成式引擎优化 (Generative Engine Optimization, GEO) 是测量并改善品牌在 AI 生成答案中位置的一套实务。本白皮书阐述我们在审计业务中使用的实务方法论：测什么、如何采样、如何审计 AI 系统对一个品牌的认知、修复按什么顺序排、以及结果必须如何报告，才能让任何人都能复核。

这套方法贯穿五条承诺：

- 1. 测量拆成两个相互独立的半场。** *曝光度 (Exposure)* ——AI 系统是否认识你的品牌、是否点它的名、描述是否正确——与 *引用概率* ——它是否把你自己的页面用作证据——会各自独立地失败，修复靠的也是两套不同的工作。我们曾在同一次审计中测到一个品牌提及率 100%、引用率 0%（自家实测；案例详见 §4）。
- 2. 三率分开打分，永不混合。** 提及率 (Mention Rate)、引用率 (Citation Rate)、推荐率是三种不同的事件，成因各不相同。Lily Ray (Algorythmic) 的已发表研究说明了这种区分为什么重要：在她的引用审计复测中，当 AI 答案引用了厂商自己的页面时，这些厂商在 2026 年 6 月仍有 69% 的情形被排除在答案的推荐之外，到 2026 年 9 月升至 83%（已发表研究）。
- 3. 单次采样是噪声。** AI 答案每次都是重新生成的。Kevin Indig 的研究 (Growth Memo) 发现：同一个 prompt 在同一个平台上跑三次，引用保持一致的比例只有 2.2%（已发表研究）。因此我们的标准是：固定题库、每道题在每个平台采样三次、以出现率报告。
- 4. 认知先于可见度。** 在曝光工作开始之前，我们先审计 AI 系统认为这个品牌是什么——身份、品类、商业模式——逐条核对。为一个被描述错的品牌赢得更多提及，只会把错误重复给更多买家。
- 5. 每份报告都带方法论说明。** 样本量、题目集、平台、日期与局限性，就印在数字旁边。没有方法的分数是广告，不是测量。

全篇有一条边界：我们从不承诺某个 AI 平台会推荐一个品牌。没有任何从业者能诚实地做出这种承诺。能承诺——且能被检验的——是测量、一张排好优先级的修复清单，以及在固定仪器上的复测。

■ 1 问题所在：AI 答案先于网站做决定

经典的购买路径走的是搜索结果页：十条链接、一个排名、一次点击。测量跟着页面走——排名位置、展现量、点击率。AI 答案从三方面打破了这个模型。

第一，答案就是终点。 买家向 AI 平台问“X 应该考虑哪个供应商”，收到的是一个综合出来的推荐，而不是一份需要逐条看下去的名单。品牌被点名或不被点名；少数几个被推为选项。入围名单在任何网站被访问之前就已经写好了，而且搜索引擎本来会展示的大多数品牌，根本不会出现。

第二，答案可能对你自信地搞错。 AI 系统不会去某个注册表里查询品牌。它们从自己能读到的、关于这个品牌的文本中重建出一个身份。痕迹稀薄之处，重建就用先验来填补空白——最近的品类、发音相近的名字、最常见的商业模式。结果是一种排名报告看不见的失败模式：一个品牌可以同时高度“可见”且被描述错误（§4 记录了一个实测案例）。

第三，被使用不等于被选中。 一个答案可能引用某厂商的页面来取规格参数，却在正文里推荐它的竞争对手。引用仪表盘会把这记为一次胜利。Lily Ray 反复进行的引用审计量化了这个差距：页面被引用的厂商，在 2026 年 6 月有 69%、2026 年 9 月有 83% 的情形被排除在答案的推荐之外；且在同一研究计划期间，对厂商自排名榜单文章的引用下降了 38%（均为已发表研究，Algorythmic）。一种把引用数一数就称之为可见度的测量实务，是在拿三率中最弱的那个信号安慰客户。

让这三个问题雪上加霜的是，被测对象本身不稳定。AI 答案是生成的，不是检索出来的。Kevin Indig 的表述——prompt 追踪是民调，不是排名跟踪（已发表研究，Growth Memo）——是正确的心智模型，而他 2.2% 的三跑引用一致率就是其硬证据。任何不从一开始就为这种不稳定性做设计的方法论，测的都是噪声，报的却当作事实。

本白皮书其余部分，就是我们为应对上述问题而建立的方法论：一个两段式测量框架（§2）、一套采样纪律（§3）、一项先于可见度工作运行的认知审计（§4）、一套修复工序（§5），以及一个让每个数字都可审计的报告标准（§6）——并把这套方法自身的局限直白地写明（§7）。

■ 2 测什么：曝光度 × 引用，三率，两对象

本文术语定义

本框区分什么是借来的、什么是本文定义的。一个术语属于本领域：本文遵循它，不声称拥有它。框中其余术语都是本方法论自己的仪器，在此定义，以便本文每个数字都能按这些定义复算。

遵循组——既有术语。

GEO（生成式引擎优化，Generative Engine Optimization）

——测量并改善品牌在 AI 生成答案中位置的学科。该术语源自 Aggarwal 等人开启的研究传统（GEO: Generative Engine Optimization, 2023）。本文按其既有含义使用 GEO。

定义组——本文术语。

在本方法论中，以下术语的含义严格如下：

上篇——曝光度（Part One — AI Exposure）。

审计的前半：AI 系统是否认识品牌——身份识别是否正确、在买家询问品类时是否点名、描述是否准确。通过品牌认知度、声量份额与品牌情感读取（§2.1）。

下篇——引用概率 (Part Two — Citation Probability)。

审计的后半：AI 系统作答时是否把品牌自己的页面用作证据。通过引用率与呈现质量读取 (§2.1)。

采样答案 (Sampled Answer)。

一道题在一个平台上跑一次所返回的一条答案，是本文所有率的分母：每个测量格（一对象 × 一平台）含一组固定的 5–10 道题，每题采样三次 (§3) ——每格 15–30 条采样答案——每个率都是跨这些跑次的出现率。

提及率 (Mention Rate)。

采样答案中品牌被点名出现的比例：点名品牌的答案数 ÷ 采样答案数。

引用率 (Citation Rate)。

采样答案中品牌自己的页面出现在答案来源中的比例：引用品牌页面的答案数 ÷ 采样答案数。

推荐率 (Recommendation Rate)。

采样答案中答案正文把品牌推为一个选项的比例：推荐品牌的答案数 ÷ 采样答案数。分母为该格全部采样答案，不只是提及品牌的那些答案。当推荐性措辞（如“最适合”“首选”“我们推荐”，或一个明确的选定）出现在品牌提及前后 ±200 字符的窗口内时，记为一次推荐。

声量份额 (Share of Voice, SOV)。

业界既有营销术语；但本文所用的分母口径由本文定义。在不点任何品牌名的品类级买家问题中，声量份额是该品牌出现次数占同一批采样答案中全部品牌出现次数的比例：本品牌出现次数 ÷ 全部品牌（本品牌 + 具名竞品）的出现次数。

2.1 两个半场，分开测、分开修

一次 GEO 审计拆成两部分，因为两部分失败的原因不同，修复靠的工作也不同。

上篇——曝光度 (Exposure)。 AI 系统是否认识这个品牌、在买家询问品类时是否点它的名、描述是否正确？曝光度通过三个子指标读取：

- **品牌认知度**——系统是否识别出品牌真实的身份：它做什么、在哪里、卖给谁。
- **声量份额 (Share of Voice, SOV)** ——在不点任何品牌名的品类级买家问题中，这个品牌相对于竞争对手出现的频率。
- **品牌情感**——当品牌被提及时，相关表述的倾向。

下篇——引用概率。 当系统作答时，它是否把品牌自己的页面用作来源？这一篇通过引用率本身读取，并通过**呈现质量**——被引材料在答案中如何被呈现——来读取。

市场竞争不是第三篇，而是一个横切的附篇：上面每一个数字，都与具名竞争对手在相同问题、相同跑次下的同一数字并列报告。一个没有竞品基线的率，是一个没有刻度的数。

两个半场在经验上相互独立。在 §4 详述的匿名演示测量中，一个品牌在同一批答案里测得提及率 100%、引用率 0%（自家实测）。如果审计报告的是一个混合的“可见度分数”，这种结构——被认识、却从未被用作证据——就会被掩盖，随之被掩盖的还有整套修复策略。

2.2 三率，独立打分

在两个半场内，每条答案都分别记录三个事件：

- **提及率（Mention Rate）**——采样答案中品牌被点名出现的比例。
- **引用率（Citation Rate）**——采样答案中品牌自己的页面出现在答案来源中的比例。
- **推荐率**——采样答案中答案正文确实把品牌推为一个选项的比例。操作上，当推荐性措辞（如“最适合”“首选”“我们推荐”，或一个明确的选定）在品牌提及前后 ± 200 字符的窗口内出现时，记为一次推荐。

三率的失败方式各不相同，这正是它们永不被平均成一个分数的原因。提及率低是认知问题：平台没有把这个名字和这个品类关联起来。引用率低是内容问题：品牌的页面没有以平台能整段取用的形式陈述事实。推荐率低是信任问题：第三方没有在平台读取的地方为这个品牌背书。三个问题，三张修复清单 (§5)。

2.3 两个对象：品牌与每个产品

每一个率都测两次——一次针对**品牌**，一次针对**每个产品或 SKU** 分别测。AI 系统可能很了解一个品牌，却从不把买家真正会买的那个具体产品推出来；反过来，一个产品可能在流通，而其制造者的身份被弄得面目全非。两个对象的修复手段不同：品牌缺口主要靠第三方信号修，产品缺口主要靠页面级的可引用性修。把两个对象压成一个测量，就掩盖了该拉哪根杠杆。

2.4 六个平台，按平台分别报告

测量覆盖买家真正在问的六个平台：ChatGPT、Perplexity、Copilot、Gemini、Grok、Claude。我们测的是作为产品的平台——买家实际接触的那个界面，连同它的浏览、信源选择与呈现层——而不是原始的模型端点。结果按平台分别报告，永不混合成跨平台平均值：一个品牌可能在一个平台上是稳定的答案、在另一个平台上完全缺席，而这两个事实的平均值对任何人的任何决策都没有帮助。

2.5 每个测量格内的答案质量分

率说的是品牌是否出现；一套次级评分标准说的是出现得好不好。在每个测量格（一对象 $\times$ 一平台）内，采样答案还按三个子维度打 0–100 分——**相关性**（答案是否真在回答买家的问题）、**权威性**（它所依据的信源有多强，按一个等级阶梯：顶端是政府/学术与主要专业信源，其次是普通网站与社区内容，底端是匿名或未知来源）、**时效性**（信息是否现行）。曝光度半场的测量格最重相关性；引用半场的测量格最重权威性。这套评分是诊断性的——它告诉修复团队，一个弱的测量格究竟是内容问题、信源问题还是过期问题——它永不替代三率，三率始终是头条数字。

■ 3 采样纪律：为什么跑三次

一次测量的诚实程度，取决于它的采样。由于 AI 答案每次运行都是重新生成的，一个 prompt 的单次采样，与其说是一次读数，不如说更接近抛一次硬币。证据来自 Kevin Indig：同一个 prompt 在同一个平台上跑三次，引用保持一致的比例只有 2.2%（已发表研究，Growth Memo）。建立在单次运行上的“可见度分数”，记录的只是硬币落在了哪一面。

我们的标准有五个要素。

固定题库。 每个测量格用一组固定的 5–10 个真实买家问题来测。问题来自 prompt 调研——买家真正在用的问法，采集自搜索与社区数据——并在进入题库前经客户访谈验证。问题的质量胜过问题的数量：八个真问题、采样得当，比一百个编出来的 prompt 采样一次告诉你的更多。题库随后冻结，因为固定的仪器才是月度间对比诚实的前提——两次读数之间，问题不许变。

每题每平台采样三次。 题库中的每道题在六个平台中的每一个上跑三次。率以跨跑次的 *出现率* 报告——一个品牌在某道题上的位置是 3 次中出现 0 次、1 次、2 次还是 3 次——而不是一个单次的是或否。

为什么是三次，而不是三十次。 三次采样足以分开修复清单真正要处理的三种状态——从不、有时、总是——而成本低到中小品牌负担得起月度测量。这套方法刻意用小数精度换节奏与趋势，并且在每份报告里明说（§6）。需要更细分分辨率的从业者可以提高跑次；不可接受的是把单次运行当作一个率来报告。

快检就标注为快检。 单次运行的查看有正当用途——快速看一眼某道题今天是怎么被回答的。在我们的工作区里它们存在，并且在出现之处一律标注为快检（spot check），从不冒充测量，也从不进入任何率。

随时间的重复胜过单点的规模。 同一冻结题库按计划复测（§5）。任何单次读数都含有生成噪声；在不变的仪器上、跨多次读数的趋势，才是企业可以据以行动的信号。这就是字面意义上的民调纪律：固定问卷、声明样本、重复多轮、报告轮次之间的变动。

■ 4 认知审计：AI 以为你在卖什么

可见度工作假设平台知道这个品牌是什么。这个假设失败得足够频繁，因此我们先审计它——在一个专设的步骤里，先于任何曝光或引用工作运行：**品牌认知审计**。

4.1 一个实测案例

以下案例来自我们自己的审计工作，已匿名；数字为自家实测。对象是一家位于浙江慈溪的家电 OEM（原始设备制造商）——一家成熟的制造商，以自有品牌向海外买家销售，拥有网站、认证，以及一条已生产多年的产品线。按传统标准它完全可被找到：搜它的名字，你就能找到它。

我们用买家的语言向 AI 系统询问，这个品牌是什么、做什么。三个系统给出了三个不同的答案，而且三个全错：

- **行业错了。** 一个系统把该品牌描述为工业洗衣设备的制造商——供自助洗衣店和酒店使用的商用机器。这家工厂做的是家用电器。它的品类挨着洗衣；系统伸手够到了文献最多的相邻品类，就停在那里。
- **名字错了。** 第二个系统把品牌名写成了一个不同的、发音相近的名字，把这家公司归给了另一个完全不同的企业。它随后产出的每一条事实都属于那个错误的实体——比隐形更糟，因为买家的询盘现在瞄准的是别人。
- **商业模式错了。** 第三个系统被用中文提问时，猜测该品牌是一家卖小家电的网店——是零售商，不是制造商。它买家最在乎的那一个事实不是缺失了；是被替换掉了。

同一次测量得出了该品牌的演示分数：**提及率 100%、引用率 0%、声量份额 54.5%**（自家实测）。合在一起读，图景是精确的：每个系统都知道这个名字，没有一个把公司自己的页面用作来源（在总共收集的 12 条答案中引用率为 0%），于是各自从碎片和先验出发重建这家公司——并即兴发挥、发挥错了。如果我们只报告头两个数字，客户会去庆祝一个可见度位置，而它的实质，是三个正在流通的错误。

4.2 方法

认知审计有三个步骤。

1. **直接问。** 向每个平台提出关于品牌本身的买家式问题——这家公司做什么、总部在哪里、卖给谁、是制造商还是销售商——用其买家使用的语言。
2. **逐条核实。** 把每个答案拆成单条事实主张，对照事实真相给每一条评级：**正确、错误或编造**。平均值会掩盖错误；一张主张台账不会。
3. **追溯每个错误的来源。** 对每一条错误主张，确定其可能的来源——同名公司、过期的目录条目、过时的页面、纯粹的品类假设。修复因来源而异：同名冲突需要消歧信号；过期条目需要在条目处更正；编造的事实需要把真实事实以可引用的形式发表到系统真正在读的地方。

4.3 顺序不容商量

认知修复第一，曝光第二，推荐最后。为一个被答案描述错误的品牌赢得更多提及没有价值——每一次新提及都是把错误再重复给另一个买家。“身份优先”不是风格偏好；§4.1 就是反序操作产出的东西。

■ 5 修复工序

只有测量没有修复，那是一份天气预报。我们的项目按固定顺序运行——**审计 → 修复 → 监控（Audit → Fix → Monitor）**——而修复工作沿着与测量同一条线拆分：曝光修复与引用修复是两种不同的手艺。

5.1 曝光修复：第三方信号

平台对一个品牌是什么的了解，大多来自品牌自家网站之外的地方。因此曝光修复作用于第三方痕迹：

- **凡被收录处，实体事实保持一致**——在目录、注册库、市场平台与公开实体记录中，使用相同的名字、品类、地点与描述，让重建（§4）有前后一致的材料可用。
- **在平台读取的地方赢得提及**——行业媒体、行业出版物、对比与评测内容，以及真实的社区讨论。这些同样是滋养推荐句（§1）的信源，这正是为什么曝光工作与推荐工作是从两侧看的同一笔投入。
- **在源头消歧与更正**——当同名冲突或过期条目正在喂养一个错误时，更正在那个条目处完成，而不是在品牌自己的页面上争辩。

我们不买版面、不养小号讨论、不制造评价。撇开伦理不谈，平台恰恰会给这类信号打折——对厂商自排名榜单文章的引用下降 38%（已发表研究，Lily Ray, Algorythmic），正是这种打折看得见的边缘。

5.2 引用修复：值得引用的页面

引用修复作用于品牌自己的网站，目标是让它的事实可被整段取用：

- **答案优先的改写。** 事实以自包含、可引用的段落重新陈述——系统可以整段拿走、不用重组页面的一段话。大多数企业网站栽在这里：事实存在，却被涂抹在语境、弹窗、图片或 PDF 里，于是系统读了它，仍然去别处找同一事实更干净的表述。
- **正确的事实，只发表一次，并保持现行。** 每条核心事实的权威表述——公司做什么、在哪里、从何时起、持有什么认证——以纯文本住在网站上；别处流通的错误版本，就地更正。
- **机器能用的结构。** 有效的结构化数据（schema）、为 AI 读者留出干净的抓取通道，以及一份 llms.txt 清单——在我们的实务中，它归档在“卫生项”而非“战略项”之下。我们自己的前后对比测量——在那些 llms.txt 是唯一改动的网站上——显示它对引用率的直接效果接近于零（自家实测）。我们照样上线它：它便宜、不破坏任何东西，而且一个为机器读者认真给自己贴标签的网站，通常也是一个页面值得读的网站。我们拒绝做的，是把它当作可见度杠杆来卖。
- **在问题所在处做深度。** 页面围绕测量题库中真实的买家问题来新建或加深，让可引用的表述恰好存在于买家真正在问的那些问题上。

5.3 修复清单的纪律

清单上的每个修复项都写明三件事：它背后的证据（它所回应的答案与主张）、它意在撬动的率，以及它的优先级。无法与一个实测缺口挂钩的项，不上清单。清单是有限的，在开工前定价；不存在没有数字附着的开放式项目。

5.4 监控：同一台仪器，按期运行

修复上线后，冻结题库按同一标准每月复跑——相同的问题、每平台三次运行、按平台报告。每个周期报告什么动了、什么没动、下一步做什么。AI 答案会随着平台更新模型与信源而漂移；一次性修复会悄无声息地衰减，只有按期的复测能在衰减还便宜可纠时抓住它。

■ 6 报告标准与方法论说明

一个测量实务的成色，由它愿意展示什么来定义。我们的报告标准有四条规则。

规则 1——报率，不报分。 每份报告的头条是三率（提及、引用、推荐），按平台、按对象（品牌与每个 SKU）列出，并附竞品数字并列。报告不给混合的可见度分数，因为混合恰好掩盖了客户采取行动所需要看清的结构（§2.2）。

规则 2——证据随附。 采样答案的全文被保存并随报告一同交付，包括每个答案引用的信源。报告中的任何数字都能追溯到它所来自的答案，任何分类都能由真人阅读者复核。

规则 3——修复清单是报告的一部分。 发现与排好优先级的修复清单（§5.3）一同到达，每项都挂钩它所针对的率。一份止于诊断的报告，只做完容易的那一半。

规则 4——每份报告都带方法论说明。 说明与数字印在一起，就在报告本身里，而不是放在另一份文档里。我们的标准模板：

方法论说明。 问题：每个测量格一组固定的 [5–10] 个真实买家问题，来自 prompt 调研并经客户访谈验证；题库在报告期之间不变更。采样：每道题在每个平台运行 3 次；所有率均为跨运行的出现率。平台：ChatGPT、Perplexity、Copilot、Gemini、Grok、Claude——按买家使用的产品形态测量，按平台报告，从不平均。周期：[开始日期] – [结束日期]。打分：提及、引用、推荐独立打分；当推荐性措辞出现在品牌提及 ± 200 字符范围内时，记为一次推荐。证据：答案全文保留并随本报告提供。局限：样本不是平台行为的普查；平台模型与信源选择会无通知变更；结果描述的是所测的问题、语言与周期，不构成对未来答案的保证。本报告不承诺——任何 GEO 实务也无法承诺——任何平台会推荐某个品牌。

最后几句是刻意的。这套实务的红线——在每份报告和每个项目里都声明——是：我们承诺测量、修复与复测。我们不承诺推荐、不承诺在 AI 答案中的排名、不承诺任何具体的平台行为，因为这些没有一样是我们有权承诺的。

■ 7 局限性

一份不声明自身局限的方法论文，就是一份销售文件。我们的局限如下。

样本不是总体。 每个测量格一组固定的 5–10 个问题、采样三次，是刻意用精度换节奏的选择（§3）。这些数字支撑方向、优先级与趋势——修复工作真正需要的决策。它们不支撑关于某平台总体行为的小数点级主张，我们也不做这种主张。

平台是移动靶。 模型被重新训练，检索信源在变，答案形态在不通知的情况下被重新设计。一套方法论要活过这些，只能靠冻结自己的仪器，并在平台行为发生不连续跳变时诚实地重设基线；跨越这种跳变的结果并不严格可比，报告在这种情况下发生时明说。

采集界面如实披露。 在我们当前的实现中，平台答案通过 API 访问采集，而对某些平台，这种访问经由中间的（代理）模型端点，而不是平台自家的原生第一方 API。代理端点在浏览行为与信源选择上，可能与消费级产品界面不同。我们把弥合这个差距——六个平台全部接入原生第一方——视为持续进行的工作，并在此披露当前状态，而不是让读者想当然。读者若把我们的数字与在消费级界面上取得的测量相比较，应把这一差异记在心里。

语言与市场范围。 买家问题以品牌真实买家所用的语言与市场框架提出。在一种语言中测得的结果不会自动迁移到另一种语言；§4 的认知案例中就有一个错误，只在用中文提问时出现。一个品牌在每种销售语言里，都需要它自己的题库。

分类涉及判断。 推荐检测（在提及周围固定窗口内的推荐性措辞）与情感读取都是启发式方法；边界情形可能被误分类。这正是每份报告都保留答案全文的原因（§6，规则 2）：任何数字都能由真人从同一份证据重新推导，任何分类都能被质疑。

案例研究只是一个案例。 §4 的数字来自单次匿名演示测量。它们说明这套方法能检测到什么；它们不是市场统计数据，也不以市场统计数据的面目呈现。

■ R 参考文献 (References)

- Indig, K. — Growth Memo。关于 prompt 追踪一致性的研究：同一个 prompt 在一个平台上跑三次，引用保持一致的运行占 2.2%；"prompt 追踪是民调，不是排名跟踪"。（已发表研究。）
- Ray, L. — Algorythmic。引用审计复测，2026 年 6 月与 2026 年 9 月：页面被引用的厂商在 69%（6 月）与 83%（9 月）的情形下被排除在答案推荐之外；对厂商自排名榜单文章的引用下降 38%。（已发表研究。）
- Liu, J. — JUN & GEO 测量实务。匿名认知审计案例（提及率 100%、引用率 0%、声量份额 54.5%）；llms.txt 前后对比测量，显示其对引用率的直接效果接近于零。（自家实测，2026 年。）

■ A 关于作者 (About the Author)

Jun Liu（刘骏）是一名 GEO 顾问，常驻中国浙江慈溪，是 GEO 服务实务机构 AccessSmith 的创始人。他在中国从事了 22 年产品管理与 13 年电子商务，主导的产品品类曾做到数亿级的营收规模。他以 JUN & GEO 之名工作——"Jun"是他的名字；GEO，即生成式引擎优化（Generative Engine Optimization），是他所从事的学科。他在 jungeo.com 发表 GEO 研究，并在微信公众号上发表中文 GEO 笔记。项目合作与问题：hi@accesssmith.com。

■ L 授权 (License)

© 2026 Jun Liu (刘骏)。本作品——《实测，而非猜测：AI 可见度 (GEO) 实务方法论》(*Measured, Not Guessed: A Field Methodology for AI Visibility (GEO)*, v1.0) ——依知识共享署名-相同方式共享 4.0 国际许可协议 (CC BY-SA 4.0) 授权。您可以为任何目的 (包括商业目的) 分享与改编本材料，条件是给出适当的署名，并以相同许可分发您的贡献。许可全文：<https://creativecommons.org/licenses/by-sa/4.0/>