制造业出海 GEO 白皮书：让 AI 搜索正确找到你的工厂

GEO White Paper for Export Manufacturers

写给出海制造企业的负责人与外贸、市场团队：GEO 是什么、AI 回答怎么选来源、怎样用可复核的方法量测，以及一条 90 天的落地路线。

版本	v1.0（2026-11-27）
适用对象	出海制造企业的负责人、外贸与市场团队
性质	公开资料整理与方法框架；不含客户案例与客户名称，不承诺排名、AI 推荐或询盘数量
数据口径	外部数字均注明来源与存取日期；推论、估算与演算示例另行标示

摘要

海外买家开始先问 AI 助手、再决定点进哪家官网。GEO（Generative Engine Optimization，生成式引擎优化）要解决的问题很具体：**当买家问「哪里有可靠的某某工厂」时，AI 的回答是否提到你、是否说对、是否把买家带到你的官网**。

- **GEO 不是买排名**。Google 官方指南的立场是：针对其生成式 AI 功能做优化，本质上仍是 SEO；可以忽略「切块」、llms.txt 这类额外文件与不真实的提及 [2]。
- **AI 回答会变，单次截图不是量测**。有研究以同一提示词重复测试，发现同一份品牌清单几乎不会重复出现 [4]；因此要看固定题库上的「出现率」，并附置信区间。
- **来源偏好因平台而异**。各平台最常被引用的来源并不相同 [6]；与品牌被提及次数相关性最高的是第三方网络提及，而非外链数量 [5]（相关不等于因果）。
- **量测要能被复核**。固定题库、固定条件、足够样本、Wilson 区间、保留不改动的对照组，并留存原始回答。
- **落地分四段、约 90 天**。基线、修复、内容与佐证、复测。每段有交付件与验收方式，涉及产品事实的内容由企业自己核准。
- **合规是边界，不是附录**。不刷评论、不买帖、不伪造提及；结构化数据与页面可见内容一致；AI 生成内容必须人工审核。

> **读法建议**
>
> 只想了解概念，读第 1、2 节；负责验收或写合同，读第 3、5 节；准备开工，读第 4 节并先备齐第 6 节清单。

目录

01 1 GEO 是什么，与 SEO 的关系
02 2 AI 回答怎么选来源
03 3 怎么量测：固定题库、Wilson 区间、对照组
04 4 90 天落地路线
05 5 合规边界
06 6 开工前清单：企业这一侧要备齐的 8 项
07 资料来源
08 免责声明

1 GEO 是什么，与 SEO 的关系

GEO 一词出自 Aggarwal 等人 2023 年的论文（KDD 2024）。论文把「生成式引擎」定义为先检索多个来源、再由大模型综合作答的搜索形态，并在自建基准上显示：对内容做某些改写，可以提高其在回答中的可见度，最高约 40% [1]。这是研究者在受控环境中量到的结果，衡量的是单一回答里的份额，不是流量、点击或订单，不能直接当作企业官网的预期成效。

面向	SEO	GEO
目标	页面在搜索结果中的位置与点击	被 AI 回答提及、引用、正确描述
单位	关键词与页面	买家的问题（意图）与回答中的来源
结果稳定性	排名相对稳定，可按日追踪	回答随机变动，需要重复抽样
主要手段	技术可抓取、内容相关、站内外链接	官网事实准确且可抓取、全网口径一致、可被引用的第三方佐证
验收口径	排名、曝光、点击	固定题库上的出现率、引用率、事实准确率，附区间

两者不是替代关系。Google 在指南中写明，从 Google Search 的角度，优化生成式 AI 搜索就是优化搜索体验，因此仍是 SEO [2]。实务上的含义是：GEO 的基础工程（可抓取、可索引、内容真实有用）与 SEO 重叠；GEO 多出来的是「回答层」的量测，以及对第三方提及与实体一致性的管理。

对出海制造业意味着什么

采购商的提问通常带着条件：品类、材质、认证、起订量、产地、交期。AI 的回答往往只列出少数几家，信息缺漏或前后不一致的工厂，更容易被漏掉或被写错。工厂能做的，是把可被核对的事实摆在官网上，并让同样的事实在名录、认证机构页面、展会资料中保持一致。

2 AI 回答怎么选来源

2.1 检索加综合，并且会「扩写」问题

Google 说明，AI 功能会从其搜索索引取回相关网页，再据此生成回答，并附上链接；其中的 query fan-out 指模型自行生成一组并行的相关子查询，去取回更多结果 [2]。所以买家的一句话，背后可能是多条子查询：你的页面要能回答的，不只是原句，还有它的各个侧面（规格、认证、适配、价格条件）。

2.2 先确认爬虫没有被挡在门外

OpenAI 的文件说明，OAI-SearchBot 与 GPTBot 是两个独立的 robots.txt 设置：可以允许前者以出现在搜索结果中，同时不允许后者用于训练 [3]。ChatGPT-User 则是用户在对话中触发的抓取。检查 robots.txt、CDN 与防火墙是否误挡搜索类爬虫，是成本最低的第一步。

2.3 不同平台偏好的来源不同

平台（Profound 样本，2024-08 至 2025-06）	占全部引用比例最高的前三个来源
ChatGPT	Wikipedia 7.8%、Reddit 1.8%、Forbes 1.1%
Google AI Overviews	Reddit 2.2%、YouTube 1.9%、Quora 1.5%
Perplexity	Reddit 6.6%、YouTube 2.0%、Gartner 1.0%

来源：Profound 公开文章，样本为其监测到的约 6.8 亿条引用 [6]。样本来自该公司的监测范围，且是 2025 年上半年以前的数据，平台行为会改变，请当作方向而非现状。

另一个常被引用的观察来自 Ahrefs：以 75,000 个品牌分析，品牌的网络提及与其在 AI Overviews 中的可见度相关系数（Spearman）为 0.664，品牌锚文字 0.527，反向链接数只有 0.218 [5]。Ahrefs 自己也提醒这是相关性。可以借用的是方向：第三方的、可被查证的品牌提及，比单纯堆外链更值得经营。

> **推论（非研究结论）**
>
> 对工厂而言，较稳妥的做法是同时经营三类来源：官网的原始资料（规格、认证、产能、联系方式）、行业名录与认证机构页面、展会与媒体上可查证的报道。不要把可见度押在单一来源类型上，因为平台偏好会变。

3 怎么量测：固定题库、Wilson 区间、对照组

3.1 为什么单次截图不是量测

SparkToro 与 Gumshoe 的测试以同一提示词重复询问 AI，约 3,000 次，结论是同一份品牌清单重复出现的概率不到百分之一，顺序一致的概率更低 [4]。这意味着「排名第几」几乎没有意义，可比较的是「在一批问题里，出现了多少比例」。

3.2 题库：来源可追溯，版本要冻结

- **来源可追溯**：站内搜索词、Search Console 查询改写成对话式问题、销售询盘与客服记录、展会问答、客户书面确认的清单。中国广告协会 T/CAAAD 046—2026 也要求样本有可说明、可追溯的来源 [7]。
- **分层**：品牌题（核对事实）与非品牌题（观察选型推荐）分开统计；再按意图分为知识、推荐、比较、验证四类。
- **措辞变体**：每个意图写 3 至 5 种自然说法，并按市场语言各备一套。
- **版本冻结**：验收用的核心题集在启动时由双方确认，验收开始后不再变更；确需变更须书面确认并记录对结果的影响 [7]。

3.3 指标定义

指标	定义	备注
提及率	提到品牌的有效回答数 ÷ 有效回答总数	不同工具分母不同，跨工具不可直接比较

指标	定义	备注
推荐率	品牌进入事先定义的「推荐组」的回答数 ÷ 推荐/比较题的有效回答数	推荐组须事先写明，例如前三名
引用率	至少含一个目标网址的有效回答数 ÷ 有效回答总数	区分「被抓取」与「被显示为引用」
事实准确率	正确的事实项 ÷ 被核对的事实项	对照企业核准的事实表，逐项记录错误原句
平均位置	仅作诊断	位置极不稳定，不建议当 KPI

请求失败、降级回答与拒答要单独计数，不能默默当作「没提到」。

3.4 样本量与 Wilson 区间

出现率是二项比例，应附置信区间。Wilson 区间在小样本与接近 0%、100% 时比常用的正态近似更稳 [10]。最坏情况（比例 50%、简单随机抽样）下，95% 置信、误差 ±10 个百分点约需 97 次，±5 个百分点约需 385 次（演算：$n \geqslant 1.96^2 \times 0.25 \div \text{误差}^2$）。AI 回答并非完全独立的随机样本，实际需要的次数通常更多。

演算示例（假设数字，仅示范计算）	观察到的出现率	95% Wilson 区间
20 次回答中出现 2 次	10.0%	约 2.8% 至 30.1%
100 次回答中出现 20 次	20.0%	约 13.3% 至 28.9%
300 次回答中出现 60 次	20.0%	约 15.9% 至 24.9%

同样是 20%，样本从 100 增加到 300，区间才从约 16 个百分点宽收窄到约 9 个百分点宽。区间互相重叠的两个数字，不能说谁高谁低。

3.5 对照组：把改动的效果和平台变化分开

- **保留不改动的对照**：约两成题目对应的页面、或一批相似页面，在观察期内不修改；改动组与对照组的差，才是较接近改动效果的估计。
- **记录条件**：平台与模型版本、查询日期时间、请求原文、测试账号状态、地区、是否联网与上下文、重复次数、原始回答、引用链接与截图。这也是 T/CAAAD 046—2026 对现状评估记录的最低要求 [7]。
- **量测面要与买家看到的一致**：网页版与 API 的行为可能不同，验收用的量测面须事先约定。
- **不隐匿异常**：该标准要求不得选择性隐匿异常样本、失败样本与口径变更，也不得用恐吓性、夸大性表述促成交易 [7]。

3.6 流量与商机层

GA4 可在「渠道组」中自建 AI 助手渠道，并排在 Referral 之上，否则会先被归入 Referral [11]。服务器日志可区分训练、搜索收录与用户触发三类爬虫 [3]。询盘表单加一栏「您怎么找到我们」，能补上没有带来源信息的流量。这些是观察到的来源，不证明因果。

4 90 天落地路线

以下是一个单市场、单品类起步的参考路线，时间为估算，实际取决于企业能否及时核准产品事实、以及网站后台的修改权限。

阶段	做什么	交付件	验收方式
第 1–2 周 基线	确认市场与品类；建立并冻结题库；在选定引擎上做首次抽样；检查爬虫放行、sitemap、标题与跳转	题库 v1、基线报告、问题清单	题库经双方书面确认；基线含样本数与区间
第 3–6 周 修复	修正官网的事实缺漏与不一致；补齐规格、认证、联系方式；Organization、Product 等结构化数据与可见内容一致；名录与目录资料对齐	修改清单与上线记录	逐项抽查，上线前保留原值
第 7–10 周 内容与佐证	按题库缺口撰写采购问答与规格页；在可查证的行业名录、认证机构与展会资料上补齐一致的信息；全部内容由企业核准	内容与佐证清单	产品事实经企业核准；外部合作留存链接与是否付费的记录
第 11–13 周 复测	沿用冻结题库复测；与对照组比较；整理询盘来源；决定下一轮范围	复测报告、下一轮建议	区间与对照组并列呈现，含异常样本

不预设「第几周出现多少提升」。若复测的区间重叠，结论写成「未见可区分的变化」，而不是挑一个对自己有利的数字。

5 合规边界

- **不造假**：不刷评论或评分、不买帖、不用不真实的提及。Google 的 AI 优化指南明确把「不真实的提及」列为可忽略、不应追求的做法 [2]；买卖链接以操纵排名属于其垃圾内容政策点名的行为 [8]。
- **结构化数据与页面一致**：标记的内容必须是读者看得到的真实信息 [9]。
- **AI 生成内容须人工审核**：为操纵排名而大量生成低价值页面属于「规模化内容滥用」[8]。产品事实、认证与数字只能来自企业核准的资料。
- **报告要诚实**：方法、样本数、区间、失败样本与口径变更都要写明 [7]。
- **数据最小化与跨境**：只收集执行服务所需的资料；处理与传输方式应事先告知，并符合适用的个人信息保护规定。具体条款以书面合同与法律意见为准。
- **不承诺结果**：AI 与搜索平台的回答与政策会改变，任何服务都不应承诺排名、AI 推荐或询盘数量。

6 开工前清单：企业这一侧要备齐的 8 项

1. 目标市场、语言与首要品类，以及一位能核准产品事实的技术或业务窗口。
2. 官网后台的修改权限，或愿意配合修改的建站商联系方式。
3. 认证与资质清单（证书编号、有效期、发证机构），以及哪些可以公开。
4. 过去 6 至 12 个月的真实询盘与买家常问问题（可去除身份信息）。
5. 起订量、交期、可供应地区、付款方式的对外口径，且全站一致。
6. 现有的名录、展会、平台资料页网址，用来核对口径是否一致。

7. 谁负责回复来自 AI 渠道的询盘，响应时限多久。
8. 书面确认的验收方式：题库版本、引擎与量测面、重复次数、观察期。

资料来源

存取日期 2026-11-27。「已开启核对」表示编写时已开启原文核对所引内容；「二手」表示未能开启原文，仅依公开转述，请读者自行查证。

[1] Aggarwal 等，GEO: Generative Engine Optimization，arXiv:2311.09735（KDD 2024）。已开启核对摘要。
https://arxiv.org/abs/2311.09735
用于第 1 节的定义与「最高约 40%」的口径。

[2] Google Search Central，Optimizing for generative AI features on Google Search（最后更新 2026-07-10）。已开启核对。
https://developers.google.com/search/docs/fundamentals/ai-optimization-guide
用于第 1、2、5 节：仍是 SEO、query fan-out、可忽略 llms.txt 与不真实的提及。

[3] OpenAI，Overview of OpenAI Crawlers。已开启核对。
https://developers.openai.com/api/docs/bots
用于第 2、3 节：OAI-SearchBot、GPTBot、ChatGPT-User 的区别。

[4] SparkToro，AI 推荐清单高度不一致的研究（二手）。作者站点对自动访问返回 403，未能开启；数字依 Search Engine Land 报道转述。
https://sparktoro.com/blog/new-research-ais-are-highly-inconsistent-when-recommending-brands-or-products-marketers-should-take-care-when-tracking-ai-visibility/
用于第 3.1 节；请以原文为准。

[5] Ahrefs，AI Overviews 品牌可见度相关性研究（75,000 个品牌）。已开启核对。
https://ahrefs.com/blog/ai-overview-brand-correlation/
用于第 2.3 节；相关系数为相关性，非因果。

[6] Profound，AI platform citation patterns（约 6.8 亿条引用，2024-08 至 2025-06）。已开启核对。
https://www.tryprofound.com/blog/ai-platform-citation-patterns
用于第 2.3 节的比例表；样本为该公司监测范围。

[7] 中国广告协会，T/CAAAD 046—2026《生成式引擎优化服务要求》（2026-10-13 发布）。已开启核对第 7.2.1、7.2.3、7.4.2 至 7.4.4 与 12.1.2 条。
https://digitalad.china-caa.org/TechLab/pdf/TCAAAD%20046-2026.pdf
用于第 3、5 节；团体标准，非强制性法规。

[8] Google Search Central，Spam policies for Google web search。已开启核对。
https://developers.google.com/search/docs/essentials/spam-policies
用于第 5 节：买卖链接、规模化内容滥用。

[9] Google Search Central，General structured data guidelines。
https://developers.google.com/search/docs/appearance/structured-data/sd-policies
用于第 5 节：标记须与可见内容一致。

[10] Wilson score interval（统计方法说明）。
https://en.wikipedia.org/wiki/Wilson_score_interval
用于第 3.4 节；样本量与区间数字为本文演算，非调查数据。

[11] Google Analytics 说明，自订渠道组（Custom channel groups）。
https://support.google.com/analytics/answer/13051316
用于第 3.6 节；渠道按顺序归类，第一个符合者胜出。

免责声明

本白皮书由 wantone 检索优化顾问（营运单位：台湾网通社 Taiwan Mobile Net，统一编号 38469314）编写，仅供一般参考，不构成法律、税务、投资或任何专业意见。

本文不承诺任何搜索排名、AI 助手的推荐或引用、询盘数量或营收。AI 与搜索平台（包括 Google、OpenAI、Perplexity 等）的回答、来源偏好与政策会随时间、模型与提问方式改变，不在我们的控制范围内。

文中的外部数字取自所列公开来源，口径、样本与时间各不相同，且部分为相关性研究；标示为「推论」「估算」或「演算示例」者是本文自行推导或假设的数字，不是调查结果。二手来源未经我们逐一重新验证。

文中外部链接指向第三方网站，其内容与可用性不由我们负责。本文引用的团体标准、平台指南与统计数字可能已更新，使用前请自行核对最新版本。

本文不含任何客户案例或客户名称。版权归 wantone 与相关原作者所有；引用请注明出处。联系：contact@search.com.tw。