製造業出海 GEO 白皮書：讓 AI 搜尋正確找到你的工廠

GEO White Paper for Export Manufacturers

寫給出海製造企業的負責人與外貿、市場團隊：GEO 是什麼、AI 回答怎麼選來源、怎樣用可複核的方法量測，以及一條 90 天的落地路線。

版本	v1.0（2026-11-27）
適用對象	出海製造企業的負責人、外貿與市場團隊
性質	公開資料整理與方法框架；不含客戶案例與客戶名稱，不承諾排名、AI 推薦或詢盤數量
資料口徑	外部數字均註明來源與存取日期；推論、估算與演算範例另行標示

摘要

海外買家開始先問 AI 助手、再決定點進哪家官網。GEO（Generative Engine Optimization，生成式引擎優化）要解決的問題很具體：**當買家問「哪裡有可靠的某某工廠」時，AI 的回答是否提到你、是否說對、是否把買家帶到你的官網**。

- **GEO 不是買排名**。Google 官方指南的立場是：針對其生成式 AI 功能做優化，本質上仍是 SEO；可以忽略「切塊」、llms.txt 這類額外文件與不真實的提及 [2]。
- **AI 回答會變，單次截圖不是量測**。有研究以同一提示詞重複測試，發現同一份品牌清單幾乎不會重複出現 [4]；因此要看固定題庫上的「出現率」，並附信賴區間。
- **來源偏好因平臺而異**。各平臺最常被引用的來源並不相同 [6]；與品牌被提及次數相關性最高的是第三方網路提及，而非外鏈數量 [5]（相關不等於因果）。
- **量測要能被複核**。固定題庫、固定條件、足夠樣本、Wilson 區間、保留不改動的對照組，並留存原始回答。
- **落地分四段、約 90 天**。基線、修復、內容與佐證、複測。每段有交付件與驗收方式，涉及產品事實的內容由企業自己核准。
- **合規是邊界，不是附錄**。不刷評論、不買帖、不偽造提及；結構化資料與頁面可見內容一致；AI 生成內容必須人工稽核。

> **讀法建議**
>
> 只想瞭解概念，讀第 1、2 節；負責驗收或寫合同，讀第 3、5 節；準備開工，讀第 4 節並先備齊第 6 節清單。

目錄

01 1 GEO 是什麼，與 SEO 的關係
02 2 AI 回答怎麼選來源
03 3 怎麼量測：固定題庫、Wilson 區間、對照組
04 4 90 天落地路線
05 5 合規邊界
06 6 開工前清單：企業這一側要備齊的 8 項
07 資料來源
08 免責宣告

1 GEO 是什麼，與 SEO 的關係

GEO 一詞出自 Aggarwal 等人 2023 年的論文（KDD 2024）。論文把「生成式引擎」定義為先檢索多個來源、再由大模型綜合作答的搜尋形態，並在自建基準上顯示：對內容做某些改寫，可以提高其在回答中的可見度，最高約 40% [1]。這是研究者在受控環境中量到的結果，衡量的是單一回答裡的份額，不是流量、點選或訂單，不能直接當作企業官網的預期成效。

面向	SEO	GEO
目標	頁面在搜尋結果中的位置與點選	被 AI 回答提及、引用、正確描述
單位	關鍵詞與頁面	買家的問題（意圖）與回答中的來源
結果穩定性	排名相對穩定，可按日追蹤	回答隨機變動，需要重複抽樣
主要手段	技術可抓取、內容相關、站內外連結	官網事實準確且可抓取、全網口徑一致、可被引用的第三方佐證
驗收口徑	排名、曝光、點選	固定題庫上的出現率、引用率、事實準確率，附區間

兩者不是替代關係。Google 在指南中寫明，從 Google Search 的角度，優化生成式 AI 搜尋就是優化搜尋體驗，因此仍是 SEO [2]。實務上的含義是：GEO 的基礎工程（可抓取、可索引、內容真實有用）與 SEO 重疊；GEO 多出來的是「回答層」的量測，以及對第三方提及與實體一致性的管理。

對出海製造業意味著什麼

採購商的提問通常帶著條件：品類、材質、認證、起訂量、產地、交期。AI 的回答往往只列出少數幾家，資訊缺漏或前後不一致的工廠，更容易被漏掉或被寫錯。工廠能做的，是把可被核對的事實擺在官網上，並讓同樣的事實在名錄、認證機構頁面、展會資料中保持一致。

2 AI 回答怎麼選來源

2.1 檢索加綜合，並且會「擴寫」問題

Google 說明，AI 功能會從其搜尋索引取回相關網頁，再據此生成回答，並附上鍊接；其中的 query fan-out 指模型自行生成一組並行的相關子查詢，去取回更多結果 [2]。所以買家的一句話，背後可能是多條子查詢：你的頁面要能回答的，不只是原句，還有它的各個側面（規格、認證、適配、價格條件）。

2.2 先確認爬蟲沒有被擋在門外

OpenAI 的文件說明，OAI-SearchBot 與 GPTBot 是兩個獨立的 robots.txt 設定：可以允許前者以出現在搜尋結果中，同時不允許後者用於訓練 [3]。ChatGPT-User 則是使用者在對話中觸發的抓取。檢查 robots.txt、CDN 與防火牆是否誤擋搜尋類爬蟲，是成本最低的第一步。

2.3 不同平臺偏好的來源不同

平臺（Profound 樣本，2024-08 至 2025-06）	佔全部引用比例最高的前三個來源
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%

來源：Profound 公開文章，樣本為其監測到的約 6.8 億條引用 [6]。樣本來自該公司的監測範圍，且是 2025 年上半年以前的資料，平臺行為會改變，請當作方向而非現狀。

另一個常被引用的觀察來自 Ahrefs：以 75,000 個品牌分析，品牌的網路提及與其在 AI Overviews 中的可見度相關係數（Spearman）為 0.664，品牌錨文字 0.527，反向連結數只有 0.218 [5]。Ahrefs 自己也提醒這是相關性。可以借用的是方向：第三方的、可被查證的品牌提及，比單純堆外鏈更值得經營。

> **推論（非研究結論）**
>
> 對工廠而言，較穩妥的做法是同時經營三類來源：官網的原始資料（規格、認證、產能、聯絡方式）、行業名錄與認證機構頁面、展會與媒體上可查證的報道。不要把可見度押在單一來源型別上，因為平臺偏好會變。

3 怎麼量測：固定題庫、Wilson 區間、對照組

3.1 為什麼單次截圖不是量測

SparkToro 與 Gumshoe 的測試以同一提示詞重複詢問 AI，約 3,000 次，結論是同一份品牌清單重複出現的機率不到百分之一，順序一致的機率更低 [4]。這意味著「排名第幾」幾乎沒有意義，可比較的是「在一批問題裡，出現了多少比例」。

3.2 題庫：來源可追溯，版本要凍結

- **來源可追溯**：站內搜尋詞、Search Console 查詢改寫成對話式問題、銷售詢盤與客服記錄、展會問答、客戶書面確認的清單。中國廣告協會 T/CAAAD 046—2026 也要求樣本有可說明、可追溯的來源 [7]。
- **分層**：品牌題（核對事實）與非品牌題（觀察選型推薦）分開統計；再按意圖分為知識、推薦、比較、驗證四類。
- **措辭變體**：每個意圖寫 3 至 5 種自然說法，並按市場語言各備一套。
- **版本凍結**：驗收用的核心題集在啟動時由雙方確認，驗收開始後不再變更；確需變更須書面確認並記錄對結果的影響 [7]。

3.3 指標定義

指標	定義	備註
提及率	提到品牌的有效回答數 ÷ 有效回答總數	不同工具分母不同，跨工具不可直接比較

指標	定義	備註
推薦率	品牌進入事先定義的「推薦組」的回答數 ÷ 推薦/比較題的有效回答數	推薦組須事先寫明，例如前三名
引用率	至少含一個目標網址的有效回答數 ÷ 有效回答總數	區分「被抓取」與「被顯示為引用」
事實準確率	正確的事實項 ÷ 被核對的事實項	對照企業核准的事實表，逐項記錄錯誤原句
平均位置	僅作診斷	位置極不穩定，不建議當 KPI

請求失敗、降級回答與拒答要單獨計數，不能默默當作「沒提到」。

3.4 樣本量與 Wilson 區間

出現率是二項比例，應附信賴區間。Wilson 區間在小樣本與接近 0%、100% 時比常用的正態近似更穩 [10]。最壞情況（比例 50%、簡單隨機抽樣）下，95% 信賴、誤差 ±10 個百分點約需 97 次，±5 個百分點約需 385 次（演算：$n \geqslant 1.96^2 \times 0.25 \div \text{誤差}^2$）。AI 回答並非完全獨立的隨機樣本，實際需要的次數通常更多。

演算範例（假設數字，僅示範計算）	觀察到的出現率	95% Wilson 區間
20 次回答中出現 2 次	10.0%	約 2.8% 至 30.1%
100 次回答中出現 20 次	20.0%	約 13.3% 至 28.9%
300 次回答中出現 60 次	20.0%	約 15.9% 至 24.9%

同樣是 20%，樣本從 100 增加到 300，區間才從約 16 個百分點寬收窄到約 9 個百分點寬。區間互相重疊的兩個數字，不能說誰高誰低。

3.5 對照組：把改動的效果和平臺變化分開

- **保留不改動的對照**：約兩成題目對應的頁面、或一批相似頁面，在觀察期內不修改；改動組與對照組的差，才是較接近改動效果的估計。
- **記錄條件**：平臺與模型版本、查詢日期時間、請求原文、測試帳號狀態、地區、是否聯網與上下文、重複次數、原始回答、引用連結與截圖。這也是 T/CAAAD 046—2026 對現狀評估記錄的最低要求 [7]。
- **量測面要與買家看到的一致**：網頁版與 API 的行為可能不同，驗收用的量測面須事先約定。
- **不隱匿異常**：該標準要求不得選擇性隱匿異常樣本、失敗樣本與口徑變更，也不得用恐嚇性、誇大性表述促成交易 [7]。

3.6 流量與商機層

GA4 可在「管道組」中自建 AI 助手管道，並排在 Referral 之上，否則會先被歸入 Referral [11]。伺服器日誌可區分訓練、搜尋收錄與使用者觸發三類爬蟲 [3]。詢盤表單加一欄「您怎麼找到我們」，能補上沒有帶來源資訊的流量。這些是觀察到的來源，不證明因果。

4 90 天落地路線

以下是一個單市場、單品類起步的參考路線，時間為估算，實際取決於企業能否及時核准產品事實、以及網站後臺的修改許可權。

階段	做什麼	交付件	驗收方式
第 1–2 週 基線	確認市場與品類；建立並凍結題庫；在選定引擎上做首次抽樣；檢查爬蟲放行、sitemap、標題與跳轉	題庫 v1、基線報告、問題清單	題庫經雙方書面確認；基線含樣本數與區間
第 3–6 週 修復	修正官網的事實缺漏與不一致；補齊規格、認證、聯絡方式；Organization、Product 等結構化資料與可見內容一致；名錄與目錄資料對齊	修改清單與上線記錄	逐項抽查，上線前保留原值
第 7–10 週 內容與佐證	按題庫缺口撰寫採購問答與規格頁；在可查證的行業名錄、認證機構與展會資料上補齊一致的資訊；全部內容由企業核准	內容與佐證清單	產品事實經企業核准；外部合作留存連結與是否付費的記錄
第 11–13 週 複測	沿用凍結題庫複測；與對照組比較；整理詢盤來源；決定下一輪範圍	複測報告、下一輪建議	區間與對照組並列呈現，含異常樣本

不預設「第幾週出現多少提升」。若複測的區間重疊，結論寫成「未見可區分的變化」，而不是挑一個對自己有利的數字。

5 合規邊界

- **不造假**：不刷評論或評分、不買帖、不用不真實的提及。Google 的 AI 優化指南明確把「不真實的提及」列為可忽略、不應追求的做法 [2]；買賣連結以操縱排名屬於其垃圾內容政策點名的行為 [8]。
- **結構化資料與頁面一致**：標記的內容必須是讀者看得到的真實資訊 [9]。
- **AI 生成內容須人工稽核**：為操縱排名而大量生成低價值頁面屬於「規模化內容濫用」[8]。產品事實、認證與數字只能來自企業核准的資料。
- **報告要誠實**：方法、樣本數、區間、失敗樣本與口徑變更都要寫明 [7]。
- **資料最小化與跨境**：只收集執行服務所需的資料；處理與傳輸方式應事先告知，並符合適用的個人資訊保護規定。具體條款以書面合同與法律意見為準。
- **不承諾結果**：AI 與搜尋平臺的回答與政策會改變，任何服務都不應承諾排名、AI 推薦或詢盤數量。

6 開工前清單：企業這一側要備齊的 8 項

1. 目標市場、語言與首要品類，以及一位能核准產品事實的技術或業務視窗。
2. 官網後臺的修改許可權，或願意配合修改的網站建置商聯絡方式。
3. 認證與資質清單（證書編號、有效期、發證機構），以及哪些可以公開。
4. 過去 6 至 12 個月的真實詢盤與買家常問問題（可去除身分資訊）。
5. 起訂量、交期、可供應地區、付款方式的對外口徑，且全站一致。
6. 現有的名錄、展會、平臺資料頁網址，用來核對口徑是否一致。

7. 誰負責回覆來自 AI 管道的詢盤，回應時限多久。
8. 書面確認的驗收方式：題庫版本、引擎與量測面、重複次數、觀察期。

資料來源

存取日期 2026-11-27。「已開啟核對」表示編寫時已開啟原文核對所引內容；「二手」表示未能開啟原文，僅依公開轉述，請讀者自行查證。

[1] Aggarwal 等，GEO: Generative Engine Optimization，arXiv:2311.09735（KDD 2024）。已開啟核對摘要。
https://arxiv.org/abs/2311.09735
用於第 1 節的定義與「最高約 40%」的口徑。

[2] Google Search Central，Optimizing for generative AI features on Google Search（最後更新 2026-07-10）。已開啟核對。
https://developers.google.com/search/docs/fundamentals/ai-optimization-guide
用於第 1、2、5 節：仍是 SEO、query fan-out、可忽略 llms.txt 與不真實的提及。

[3] OpenAI，Overview of OpenAI Crawlers。已開啟核對。
https://developers.openai.com/api/docs/bots
用於第 2、3 節：OAI-SearchBot、GPTBot、ChatGPT-User 的區別。

[4] SparkToro，AI 推薦清單高度不一致的研究（二手）。作者站點對自動訪問返回 403，未能開啟；數字依 Search Engine Land 報道轉述。
https://sparktoro.com/blog/new-research-ais-are-highly-inconsistent-when-recommending-brands-or-products-marketers-should-take-care-when-tracking-ai-visibility/
用於第 3.1 節；請以原文為準。

[5] Ahrefs，AI Overviews 品牌可見度相關性研究（75,000 個品牌）。已開啟核對。
https://ahrefs.com/blog/ai-overview-brand-correlation/
用於第 2.3 節；相關係數為相關性，非因果。

[6] Profound，AI platform citation patterns（約 6.8 億條引用，2024-08 至 2025-06）。已開啟核對。
https://www.tryprofound.com/blog/ai-platform-citation-patterns
用於第 2.3 節的比例表；樣本為該公司監測範圍。

[7] 中國廣告協會，T/CAAAD 046—2026《生成式引擎優化服務要求》（2026-10-13 發布）。已開啟核對第 7.2.1、7.2.3、7.4.2 至 7.4.4 與 12.1.2 條。
https://digitalad.china-caa.org/TechLab/pdf/TCAAAD%20046-2026.pdf
用於第 3、5 節；團體標準，非強制性法規。

[8] Google Search Central，Spam policies for Google web search。已開啟核對。
https://developers.google.com/search/docs/essentials/spam-policies
用於第 5 節：買賣連結、規模化內容濫用。

[9] Google Search Central，General structured data guidelines。
https://developers.google.com/search/docs/appearance/structured-data/sd-policies
用於第 5 節：標記須與可見內容一致。

[10] Wilson score interval（統計方法說明）。
https://en.wikipedia.org/wiki/Wilson_score_interval
用於第 3.4 節；樣本量與區間數字為本文演算，非調查資料。

[11] Google Analytics 說明，自訂管道組（Custom channel groups）。
https://support.google.com/analytics/answer/13051316
用於第 3.6 節；管道按順序歸類，第一個符合者勝出。

免責宣告

本白皮書由 wantone 萬通搜尋顧問服務（營運單位：台灣網通社 Taiwan Mobile Net，統一編號 38469314）編寫，僅供一般參考，不構成法律、稅務、投資或任何專業意見。

本文不承諾任何搜尋排名、AI 助手的推薦或引用、詢盤數量或營收。AI 與搜尋平臺（包括 Google、OpenAI、Perplexity 等）的回答、來源偏好與政策會隨時間、模型與提問方式改變，不在我們的控制範圍內。

文中的外部數字取自所列公開來源，口徑、樣本與時間各不相同，且部分為相關性研究；標示為「推論」「估算」或「演算範例」者是本文自行推導或假設的數字，不是調查結果。二手來源未經我們逐一重新驗證。

文中外部連結指向第三方網站，其內容與可用性不由我們負責。本文引用的團體標準、平臺指南與統計數字可能已更新，使用前請自行核對最新版本。

本文不含任何客戶案例或客戶名稱。版權歸 wantone 與相關原作者所有；引用請註明出處。聯絡：contact@search.com.tw。