GENERATIVE
ENGINE
OPTIMIZATION

BLUE BOOK
PROCUREMENT
GUIDE

GEO蓝皮书

企业与GEO服务商合作评估指南

A BUYER'S GUIDE TO
GEO SERVICE PROVIDERS

选对服务商
定好指标
管住风险

RFI · RFP
POC · SOW
SCORECARD

DOUBAO · DEEPSEEK
KIMI · YUANBAO
QWEN

GEO蓝皮书 · 采购与合作指南

企业与GEO服务商合作评估指南

立项 · 诊断 · 基线 · RFI/RFP · 比稿 · POC · 合同 · 复盘

作者　姚金刚、向阳乔木

适用对象　品牌方、集团企业、上市公司、中型企业与出海团队

相关书籍　《AI营销：从SEO到GEO》

本文档初衷，为国内业务与出海业务建立一套完整的GEO服务商采购与合作参考，涵盖：定义、边界、流程、标准、指标体系、服务商类型、服务商评估与评分框架等等

- 了解作者：《姚金刚认知随笔》 《向阳乔木日志》
- 推荐链接：《姚金刚提示词合集》、《向阳乔木: PromptEngineering第一站》、《AI领导力课程》
- GEO书籍：【京东搜索《AI营销：从SEO到GEO》】

第二届GEO大会报名中

- 报名链接：https://hui.ailingdaoli.com/
- 时间：2026年11月21–22日（周六、周日一共两天）
- 地点：深圳南山区，具体酒店提前一个月群内通知
- 日程：周六及周日上午，GEO及AI营销干货分享，周日下午GEO及Agent实操
- 嘉宾：国内知名GEO公司创始人、高管，知名官方媒体高管，出海GEO&SEO专家，知名AI创业公司创始人、CEO，知名AI自媒体博主、AI专家等
- 主题：GEO类干货，AI营销类干货，出海GEO干货等相关主题分享，另设Agent与GEO专题实操

相关阅读

《GEO白皮书：AI搜索时代的品牌增长新范式》

《GEO红皮书：生成式引擎优化的伦理边界、作弊风险与治理手册》

《GEO蓝皮书：企业与GEO服务商合作评估指南》

《AI营销：从SEO到GEO》提示词合集



适用对象

在中国大陆经营的品牌方、集团企业、上市公司、中型企业、出海企业、跨境业务团队，以及正在准备采购或更换GEO服务商的市场、品牌、增长、数字化、法务与合规团队。

本蓝皮书将GEO服务采购拆解为一条完整旅程：**内部立项、需求诊断、基线测量、服务商识别、RFI/RFP、比稿、POC、合同与SOW、项目治理、验收复盘、续约或退出。**

企业可以将本文直接改造成采购评分表、RFP模板、合同附件、月度复盘模板和供应商管理手册。

本版在关键模块补充了虚构企业案例、方法参考、计算样例、访谈脚本和自检机制，方便企业在真实采购中直接套用。

目 录

01 GEO的定义、边界与采购前提 9

GEO是什么 10
GEO适合解决的问题 11
GEO不宜承诺的事项 12
企业采购GEO服务前的四个前提 13
蓝皮书使用方式 14
科学理解GEO：概率、证据与边界 15

02 企业采购GEO服务的完整用户旅程 18

阶段一：内部立项 21
阶段二：需求诊断 22
阶段三：基线测量 23
阶段四至六：服务商识别、RFI/RFP与比稿 24
阶段七至十：合同、执行、复盘与退出 25
采购旅程案例：三类企业如何走完整流程 26
方法参考：需求诊断工作坊 28

03 GEO指标体系：前端可量化指标为主，后端指标只做观察 30

指标体系的基本分层 31
指标设计原则 32
端可量化指标：以核心指标作为考核抓手 33
Top 1、Top 3与可见率的具体定义 35
前端非量化指标：适合关注，不适合作为硬考核 37
后端指标：企业应观察，通常不作为服务商考核 39
指标优先级建议 41
建议仪表盘结构 42
指标计算案例：把概率说清楚 43
方法参考：前端指标仪表盘怎么设计 44
非量化指标的使用方法：可关注，慎考核 45
后端指标的观察方法：相关性优先，归因谨慎 46

04 重点AI平台视角：豆包、DeepSeek、Kimi、元宝、千问 47

重点AI平台视角：建议五个平台整体采购 48
五个平台的通用采样口径 49
平台分层策略 50
4.4 五个平台采样案例：同一问题为什么会出现不同答案 51

05 中国GEO服务商类型划分 52

如何用类型划分指导采购 55
判断AI原生能力是否真实 56
主服务商与组合采购 57
服务商类型选择案例 58
方法参考：AI原生能力测试作业 59

06 服务商评估维度与评分框架 61

总评分建议 62
不同企业的权重调整 63
服务商访谈评分表 64
报价评估 66
评分案例：如何避免被销售话术带偏 67
方法参考：报价拆解表 68

07 海外GEO采购与国内GEO采购的差异 69

海外GEO采购建议 70
国内GEO采购建议 70
海外与国内双线项目 71

08 不同企业阶段与行业敏感性的采购策略 72

小型企业 73
中型企业 74
行业龙头 75
上市公司 76
高敏行业 77
出海企业 78
企业阶段案例：同样的GEO，采购重点完全不同 79

09 如何判断服务商的专业水平 81

专业服务商应具备的能力 82
现场测试题 83
专业能力红线 84
案例真实性核验 85
方法参考：服务商专业面试脚本 86
现场测试案例：10分钟看出方法深浅 88

10 POC、合同、SOW与项目治理 89

POC验收表 90
合同与SOW应写入的内容 91
项目治理机制 92
交付物模板 93
SOW范例条款 94

11 沟通话术、RFI与RFP模板 95

初次沟通话术 96
追问专业方法的话术 97
指标对齐话术 98
合规沟通话术 99
POC沟通话术 100
月度复盘话术 101
RFI模板 102
RFP模板 103

12 风险清单与不合理信号 104

十类核心风险 105
不合理信号 106
低质资料源识别 107
风险案例：企业最容易忽略的四类问题 108

13 采购落地工具包 109

企业自查清单 110
服务商尽调清单 111
月报模板 112
90天执行路线图 113
采购决策矩阵 114
企业内部方法手册：从采购到执行的最小闭环 115
会议模板：GEO项目月度复盘会 116
参考方法：内容资产改造的“一问一页一证据”原则 117
参考方法：事实错误纠偏流程 119
参考方法：权威来源建设路线 120
参考方法：看板设计与复盘会议 121
参考方法：RFP评分前的资料包 122
参考方法：服务商组合采购模型 123
参考方法：内部培训大纲 124
参考方法：国内五个平台的采样注意事项 125
参考方法：敏感行业的专家复核机制 126
参考方法：采购预算拆分 127
参考方法：从蓝皮书到企业内部制度 128
重点案例总复盘 129
行业场景扩展案例：工业制造企业 130
业场景扩展案例：B2B SaaS企业 131
行业场景扩展案例：消费品牌与新零售企业 132
行业场景扩展案例：教育与培训企业 133
行业场景扩展案例：跨境B2B企业 134
行业场景扩展案例：本地生活与连锁服务企业 135
行业场景扩展案例：开发者工具与技术产品 136
参考方法：品牌实体治理 137
参考方法：从单次优化到长期机制 138
参考方法：常见管理层问题的回答口径 139

14 资料源清单与交叉校验方法 140

资料源筛选原则 141
核心资料源 142
交叉校验方法 143
企业内部资料准备清单 144
方法参考：资料源可信度评分 145
交叉校验案例：平台能力如何核验 146
方法参考：服务商访谈剧本 147
方法参考：企业内部采样复核手册 148
方法参考：GEO培训大纲 149
方法参考：GEO术语表 150
方法参考：年度GEO路线图 151
方法参考：从AI答案反推组织问题 152
方法参考：POC评分会议模板 153
方法参考：GEO采购的最小合同附件 155
方法参考：采购负责人一页纸摘要 156

15 案例库与方法库 157

案例：麦芽小鹿的低预算GEO方案 158
案例：星河CRM的复合型GEO方案 159
案例：南岭精密的工业技术GEO方案 160
案例：华辰智造的实体治理方案 161
案例：安衡健康的高敏行业方案 162
案例：海桥电商的海外GEO方案 163
方法库一：问题库构建五步法 164
方法库二：内容资产建设七件套 165
方法库三：GEO内容改写原则 166
方法库五：企业采购红线清单 167

16 结语 168

01

第一部分

GEO的定义、边界与采购前提

GEO是什么 10

GEO适合解决的问题 11

GEO不宜承诺的事项 12

企业采购GEO服务前的四个前提 13

蓝皮书使用方式 14

科学理解GEO：概率、证据与边界 15

GEO是什么



GEO，英文全称Generative Engine Optimization，即生成式引擎优化，面向生成式AI答案引擎的可见性、可引用性、可解释性和品牌表达优化。

它的目标既包括让企业内容被抓取，也包括让企业在用户真实提问时，以准确、合规、可复核的方式出现在AI回答中。

传统SEO通常围绕网页排名、点击和站点流量展开。

GEO面对的是生成式答案：AI系统会检索、理解、合成多个来源，并生成一段答案。

企业在答案中的表现可能体现为品牌是否被提及、是否被引用、答案是否准确、是否出现竞品共现、是否引用企业自有页面或第三方权威来源、是否把企业归入正确品类。

从采购角度看，GEO不宜被包装成单一投放服务。

它更接近一套横跨内容、技术、品牌、数据、舆情、合规和增长的长期治理能力。

服务商可以承担诊断、实验、内容生产、技术修复、监测、复盘和组织协同，但企业自身仍需要提供业务事实、权威材料、法务边界和内部审批。

GEO适合解决的问题



品牌可见性　用户问到某个品类、解决方案、品牌对比、采购建议时，企业是否被AI答案合理提及。

引用可见性　AI答案是否引用企业自有官网、帮助中心、白皮书、产品文档、案例页面、开发者文档、权威媒体或行业机构资料。

答案准确性　AI对企业产品、价格区间、适用场景、服务范围、认证资质、客户案例、交付能力的描述是否准确。

竞争态势　AI在推荐或对比时是否频繁提及竞品，企业与竞品的相对位置、理由、描述质量如何。

实体治理　企业名称、品牌名、产品名、创始人、子品牌、英文名、简称、地区业务、官网入口之间的关系是否清晰。

跨平台一致性　豆包、DeepSeek、Kimi、元宝、千问在同类问题上的回答是否一致，差异来自模型、搜索源、内容生态还是企业资料缺口。

采购支持　让市场、品牌、增长、数字化、法务等团队有一套判断服务商能力的标准，减少被概念、截图和口号误导的概率。

GEO不宜承诺的事项



- 保证任意AI平台在任意问题中固定推荐某品牌。
- 保证短期内通过GEO直接带来确定销量或确定线索数量。
- 通过批量灌水、伪造案例、制造虚假第三方评价、购买隐性软文来影响答案。
- 把单次截图当作项目效果证据。
- 把AI答案里的品牌提及直接等同于搜索排名、广告点击、商机和收入。
- 在没有实验设计、对照组、时间窗口和样本量的情况下宣称完成因果归因。

企业采购GEO服务前的四个前提



1. 企业必须明确业务目标

目标可能是品牌提及、引用占比、关键产品认知、海外市场进入、招商获客、舆情纠偏或招投标支持。不同目标对应不同指标与服务商类型。

2. 企业必须明确平台范围

本文重点覆盖豆包、DeepSeek、Kimi、元宝、千问。

海外项目可把Google AI Overviews、AI Mode、Gemini、ChatGPT和Perplexity列为可扩展范围，但本版不展开海外平台细节。

3. 企业必须接受前端指标优先

GEO早期阶段更适合先管理AI答案表现、引用来源、实体准确性和内容可检索性。销售、线索、成交等后端指标可观测，但需要谨慎解释。

4. 企业必须建立合规边界

涉及个人信息、客户案例、未公开财务数据、医疗健康、金融投资、教育招生、政务公共安全等内容时，GEO项目需要法务、合规或专家审查参与。

蓝皮书使用方式

- 市场或品牌负责人可以使用第二章至第七章完成立项、预算和服务商比选。
- 采购团队可以使用第六章、第十章和第十一章制作RFP、评分表和合同附件。
- 法务与合规团队可以使用第十章和第十二章检查服务商承诺、数据处理、内容边界和禁止行为。
- 项目经理可以使用第十三章作为月度复盘模板和验收清单。
- 管理层可以使用第三章和第八章判断项目是否适合投入，以及投入强度应如何设定。

科学理解GEO：概率、证据与边界



企业采购GEO服务前，需要先把GEO放在正确的认知框架里。

GEO的核心对象是生成式答案生态，平台会根据问题、时间、入口、联网状态、检索结果、模型版本、上下文和安全策略生成不同答案。

企业能做的是提升自身内容被理解、被引用、被合理推荐的概率，降低错误描述、缺失引用、竞品挤压和合规风险的概率。

这意味着企业在评估GEO时，应使用概率思维、样本思维和证据思维。

一个专业服务商不应承诺“固定排名”，更应说明如何构造问题样本、如何控制采样条件、如何定义品牌可见、如何计算Top 1和Top 3、如何解释波动、如何处理样本量不足带来的误差。

方法参考：GEO科学判断的六个问题



企业可以在立项会上先问六个问题。

- 我们希望优化的是品牌被提及、被首位推荐、被引用，还是被准确描述。
- 我们希望覆盖的是豆包、DeepSeek、Kimi、元宝、千问中的哪些关键词或用户意图问题。
- 我们的问题样本是否覆盖真实用户，而非只覆盖企业内部习惯用语。
- 我们的自有内容是否足以支撑AI给出准确答案。
- 我们是否能接受短期波动，并用月度趋势判断效果。
- 我们是否把后端线索、成交和销售数据作为观察，不直接归因给服务商。

参考示例：一家B2B设备企业的错误立项

“南岭精密”（虚构企业）主营工业视觉检测设备。

市场部最初提出的采购目标是“三个月内让AI每次都把我们列为第一推荐”。

这个目标在采购会上被服务商销售包装成“全平台Top 1保障”。

经过内部复盘后，南岭精密把目标改成了三类：

第一，围绕“汽车零部件视觉检测方案”“锂电池缺陷检测设备”等40个高价值问题，目标品牌Top 3概率从基线的15%提升到30%；

第二，企业自有域引用率从5%提升到15%；

第三，重大事实错误率降到5%以内。这个目标更适合写入SOW（工作说明书），因为它有样本、有平台、有时间窗口，也承认GEO是概率优化。

这个例子说明，采购目标一旦从绝对承诺改成可测量概率，企业就能更清楚地判断服务商是否专业，也能减少后期验收纠纷。

02

第二部分

企业采购GEO服务的完整用户旅程

阶段一：内部立项 21

阶段二：需求诊断 22

阶段三：基线测量 23

阶段四至六：服务商识别、RFI/RFP与比稿 24

阶段七至十：合同、执行、复盘与退出 25

采购旅程案例：三类企业如何走完整流程 26

方法参考：需求诊断工作坊 28

GEO采购不应从询价开始。更稳妥的流程是先完成内部目标和基线，再进入服务商识别与比稿。

否则企业很容易用含糊目标采购含糊服务，最终只能比较PPT、口才和报价。



表1：GEO采购用户旅程总览

阶段	核心目标	企业关注重点	可交付证据	主要风险
1. 内部立项	明确为什么做GEO	业务目标、预算、平台范围、内部责任人	立项说明、目标清单、平台范围	把GEO当作短期流量项目
2. 需求诊断	定义问题与场景	品牌、品类、产品、竞品、区域、用户问题	问题地图、竞品清单、场景矩阵	只围绕企业自说自话的关键词
3. 基线测量	建立现状数据	提问样本、平台版本、采集方法、重复采样	基线报告、原始样本、截图或日志	单次截图、样本过小、不可复现
4. 服务商识别	找到候选供应商	类型、能力、案例、行业经验、数据方法	候选名单、初筛评分	被低价、噱头或所谓资源承诺吸引
5. RFI/RFP	让供应商按同一口径回复	服务范围、指标、方法、团队、合规、报价	RFI回复、RFP方案、报价表	各家口径不同导致无法比较
6. 比稿与尽调	判断专业能力和适配度	方法论、数据可信度、交付团队、边界意识	评分表、访谈记录、尽调问题答案	只看案例截图或老板背书
7. 合同与SOW	把承诺写清楚	交付物、验收标准、数据权属、禁止行为	合同、SOW、验收表	合同只写服务名称和总价
8. 执行治理	保持项目节奏	周会、月报、变更、复核、合规审查	任务看板、月报、采样记录	内容上线慢、跨部门协同失灵
9. 验收复盘	判断效果与复购价值	前端指标、后端观察、学习沉淀、下一周期	验收报告、复盘会纪要、续约建议	只用好看的样本证明效果

阶段	核心目标	企业关注重点	可交付证据	主要风险
10. 续约或退出	决定长期机制	平台覆盖、内部能力沉淀、成本收益	续约方案或交接清单	供应商锁定、数据无法迁移

说明：建议企业至少保留基线测量、POC报告、采样日志和月度复盘四类证据。

备注：名词解释

> - RFI (Request for Information) - 信息征询
> - RFP (Request for Proposal) - 建议征询
> - SOW (Statement of Work) - 工作声明

阶段一：内部立项

内部立项阶段要先回答三个问题：企业为什么现在需要GEO，优先服务哪个业务单元，成功标准是什么。

对于大多数企业，GEO可以先从品牌认知、核心产品认知、竞品对比、招商采购、售前问答、海外市场进入等场景切入。

- 立项输出一：GEO业务目标。示例：提升企业在“行业解决方案推荐”类问题中的品牌提及率与引用率。
- 立项输出二：平台范围。示例：豆包、DeepSeek、Kimi、元宝、千问作为一期范围。
- 立项输出三：预算与周期。示例：30天诊断，90天一期执行，半年复盘。
- 立项输出四：内部责任人。市场、品牌、内容、官网、法务、产品、销售至少需要明确接口人。
- 立项输出五：合规边界。明确哪些材料可以公开、哪些材料需要审批、哪些内容不得用于AI答案优化。

阶段二：需求诊断

需求诊断需要把“我们想提升GEO”转化为可测量的问题库。

问题库不宜只由品牌词组成，还要覆盖品类词、需求词、竞品词、购买决策词、风险词和区域词。

表2：GEO问题库的基本分类

问题类型	示例问题	用于判断什么
品牌认知	某品牌是做什么的？有哪些核心产品？	AI是否理解企业实体和业务边界
品类推荐	国内适合制造企业的某类解决方案有哪些？	企业是否进入品类答案
采购决策	企业采购某类系统应如何选择服务商？	企业是否被列入采购候选
竞品对比	A公司和B公司在某方面有什么差异？	AI是否准确描述优势与限制
场景方案	某行业如何解决某个业务痛点？	企业内容是否能支撑场景答案
风险合规	某产品是否涉及数据安全或隐私风险？	敏感问题是否被准确处理
区域与语言	海外某地区有哪些适合中国企业的服务商？	跨境和本地化内容是否被识别
售后与支持	某品牌支持哪些部署方式或服务方式？	官网和帮助中心内容是否被引用

阶段三：基线测量

基线测量是采购GEO服务的关键节点。

没有基线，企业无法判断服务商方案是否精准，也无法在后续验收时确认变化来自项目行动、平台波动还是采样偏差。

- 基线应覆盖至少三类问题：品牌词问题、非品牌品类问题、竞品或对比问题。
- 每个平台建议对同一问题进行多次采样，记录日期、时间、账号状态、设备、地区、联网状态、是否开启深度搜索或搜索模式。
- 每条样本要记录答案文本、品牌提及、提及位置、引用来源、来源域名、情绪倾向、错误点、竞品共现和截图或可追溯日志。
- 基线报告应包含问题库结构、采样方法、指标定义、主要发现、优先级建议和风险说明。
- 企业应要求服务商提供原始样本或可复核导出，而非只看汇总图。

阶段四至六：服务商识别、RFI/RFP与比稿

企业在比稿时应让服务商使用同一套输入材料和同一套问题库回复方案。

不同供应商的PPT风格、话术和报价口径差异很大，统一输入可以让比较更接近真实能力比较。

- RFI （信息征询）阶段主要问能力边界、过往经验、团队配置、数据方法和合规政策。
- RFP （建议征询）阶段要求供应商提交完整项目方案，包括基线方法、平台覆盖、指标体系、内容策略、技术策略、治理节奏、POC设计、风险控制和报价。
- 比稿阶段应要求供应商现场解释采样口径、指标定义、平台差异、后端归因边界和不可承诺事项。
- 尽调阶段应核验服务商案例是否可复核，是否涉及夸大、伪造、隐性投放或违反平台规则的做法。

补充：由于当前阶段很多标准暂未统一，也可以由服务商自行提供方案与考核标准，企业内部多视角衡量，选择最适合的自己也可。

阶段七至十：合同、执行、复盘与退出

GEO项目的长期价值来自持续治理。

企业应把合同写成可验收、可复盘、可退出的结构，而非只采购一批内容或一套监测截图。

- 合同中应写清楚交付物：问题库、基线报告、监测看板、内容方案、技术修复建议、月报、复盘和交接资料。
- 合同中应写清楚验收标准：指标定义、样本范围、平台范围、采样频率、数据保存、复测方式。
- 合同中应写清楚禁止行为：虚假内容、伪造案例、隐性付费评价、批量垃圾页面、未经授权抓取个人信息、绕过平台规则、误导性承诺。
- 执行期应设定周会和月度复盘。周会解决任务进度，月度复盘看指标变化、平台差异和风险。
- 退出时应要求供应商交付问题库、样本数据、资产清单、内容修改记录、技术建议、监测模板和未完成事项。

采购旅程案例：三类企业如何走完整流程

下面三个案例均为虚构企业，目的是帮助采购方把旅程节点和真实组织场景联系起来。

案例一：小型消费品牌“麦芽小鹿”

麦芽小鹿是一家儿童营养食品品牌，团队规模不足80人，市场预算有限，官网内容较薄，主要流量来自小红书、电商平台和私域。

它采购GEO服务时的第一需求是确认品牌在“儿童DHA怎么选”“儿童益生菌哪个牌子适合肠胃敏感”等问题中是否有基础可见性，暂时不追求全平台覆盖。

建议流程如下。

- 内部立项：由品牌负责人牵头，电商、法务和客服参与，明确可公开产品功效边界。
- 需求诊断：先选30个高意向问题，覆盖品类、功效、年龄段、风险提示和竞品对比。
- 基线测量：只采样豆包、Kimi、千问三个主要中文入口，每个问题每个平台采样3次。
- POC设计：不做全站内容改造，先优化5个核心FAQ页面和2篇产品科普页。
- 验收重点：目标品牌可见率、事实错误率、竞品过度绑定情况、自有域引用情况。
- 后续决策：若前端指标有稳定提升，再扩展到元宝、DeepSeek和更多问题库。

麦芽小鹿的关键是用低成本确认机会空间，避免一开始追求复杂看板。

小企业最容易犯的错误是采购超出自身消化能力的大套餐，最后只有报表，没有内容和组织动作。

案例二：中型B2B SaaS企业“星河CRM”

星河CRM服务制造业客户，销售周期较长，官网有大量功能页和案例页，但内容结构偏SEO，缺少面向AI答案的事实说明、对比逻辑和行业FAQ。

它的采购目标是让AI在“制造业CRM系统怎么选”“国产CRM和海外CRM区别”“适合经销商管理的CRM”等问题中更稳定地提及企业。

建议流程如下：

- 内部立项：市场、销售、产品和数据团队共同参与，把GEO作为销售辅助项目。
- 需求诊断：问题库分为采购决策、竞品对比、行业场景、功能需求、实施风险五类。
- 基线测量：五个平台全覆盖，每个平台每类问题至少20条，总样本建议超过500条回答。
- 服务商筛选：优先选择AI原生+SEO+营销复合型服务商，要求其能同时改内容结构、实体关系和监测方法。
- POC设计：选择30个问题，配套优化官网页面、案例页、帮助中心和行业白皮书。
- 验收重点：Top 3概率、有效引用率、答案事实准确性、推荐理由质量和销售反馈。

星河CRM的关键是把GEO与销售知识体系连接起来

比如销售团队反馈“客户常问国产CRM数据安全怎么保障”，这个问题就应进入问题库，官网也需要有清晰、可引用、合规的答案

案例三：上市工业集团“华辰智造”

华辰智造是一家上市工业集团，拥有多个子品牌、海外业务和复杂产品线。

它担心AI答案把子公司、产品线、海外代理商和旧品牌混在一起，也担心在投资者、采购商和渠道商查询时出现不准确描述。

建议流程如下。

- 内部立项：由品牌中心牵头，证券事务、法务、海外事业部、IT和业务单元共同参与。
- 需求诊断：先做企业实体图谱，梳理集团名、子品牌、英文名、证券简称、官网域名和产品线。
- 基线测量：国内五平台与海外平台分开测，投资者问题、采购问题、技术问题和品牌问题分开测。
- RFP要求：服务商必须提交数据安全方案、客户授权机制、敏感信息处理方案和退出交接机制。
- POC设计：先做一个业务单元和一个海外市场，避免一开始全集团铺开。
- 验收重点：实体准确率、引用来源合规性、错误归属率、敏感内容风险、跨平台一致性。

华辰智造的关键是治理复杂性。

大型企业采购GEO，很多问题源于企业内部事实不统一、官网结构混乱和历史内容长期无人维护，平台会把这些问题放大到答案层。

方法参考：需求诊断工作坊

企业可以在采购前组织一次2至3小时的GEO需求诊断工作坊。

参与人包括市场、品牌、官网、销售、产品、客服、法务和数据团队。工作坊的输出不需要很复杂，但必须形成四个清单。

清单一：业务场景清单

把企业最关心的AI提问场景写出来，按业务价值排序。

- 品类教育：用户尚不了解品类，只问“什么是”“怎么选”。
- 采购决策：用户已经有需求，开始问“哪家好”“推荐哪些”。
- 竞品对比：用户问“某品牌和某品牌区别”。
- 风险判断：用户问“是否安全”“是否合规”“有没有坑”。
- 售后服务：用户问“怎么用”“怎么维修”“怎么退款”。
- 投资者与公众认知：用户问公司实力、业务范围、财务或舆情。

清单二：问题样本清单

每个业务场景至少写出10个真实问题。

问题要尽量接近用户表达，不能只用内部术语。例如“多渠道经销商管理系统”可以保留，但也要加入“经销商太多怎么管订单和回款”“销售渠道窜货怎么追踪”这类自然语言问题。

清单三：企业事实清单

每个问题背后都应有企业事实支撑。

包括产品功能、行业资质、客户案例、服务范围、价格边界、实施周期、数据安全说明、免责声明和禁用表述。没有事实支撑的问题，不应强行要求服务商优化。

清单四：风险边界清单

列出不能说、不能承诺、不能暗示的内容。

例如医疗健康领域不能暗示治疗效果，金融领域不能承诺收益，教育领域不能夸大升学结果，工业安全领域不能淡化安全风险。

产出模板

项目	内容示例	责任人	是否可公开	是否进入POC
业务场景	制造业CRM采购决策	市场负责人	是	是
用户问题	制造业CRM系统怎么选	销售负责人	是	是
企业事实	已服务离散制造客户，支持经销商管理和售后工单	产品负责人	需法务确认	是
禁用边界	不得写“行业第一”，不得承诺实施周期固定	法务	是	是
现有材料	官网功能页、客户案例、白皮书	内容负责人	部分可公开	是

03

第 三 部 分

GEO指标体系：前端可量化指标为主，后端指标只做观察

指标体系的基本分层 31
指标设计原则 32
端可量化指标：以核心指标作为考核抓手 33
Top 1、Top 3与可见率的具体定义 35
前端非量化指标：适合关注，不适合作为硬考核 37
后端指标：企业应观察，通常不作为服务商考核 39
指标优先级建议 41
建议仪表盘结构 42
指标计算案例：把概率说清楚 43
方法参考：前端指标仪表盘怎么设计 44
非量化指标的使用方法：可关注，慎考核 45
后端指标的观察方法：相关性优先，归因谨慎 46

指标体系的基本分层



GEO指标体系建议先分成两大类：**前端指标和后端指标**。

前端指标衡量AI答案本身的表现，包括目标品牌是否出现、出现在哪里、是否进入Top 1或Top 3、是否被引用、事实是否准确。

后端指标衡量业务链路中的外部变化，包括访问、线索、商机、成交、销售反馈和品牌搜索等。

前端指标还需要继续拆成两类：可量化指标和非量化指标。

可量化指标可以基于采集样本计算概率或比例，适合作为项目考核和验收的主要依据。

非量化指标带有明显主观判断，例如语义一致性、情感积极性、表达友好度、推荐理由质量等，可以关注和复盘，但不建议作为服务商硬性KPI。

GEO本质上更接近概率优化。

企业无法要求某个平台在任意时间、任意账号、任意问题下都固定推荐某个品牌，但可以要求服务商在统一采样口径下，提升目标品牌在目标问题集合中的可见概率、Top 1概率、Top 3概率、引用概率和事实准确概率。

指标设计原则



- 可复核：企业必须能看到原始样本、采样时间、平台入口、账号状态、联网状态、答案文本、引用来源和截图或日志。
- 可计算：用于考核的指标必须能写出清晰公式，分母、分子、有效样本、排除样本都要明确。
- 可解释：指标变化要能被业务、内容、技术和管理团队理解，不能只给一个黑箱分数。
- 可行动：指标需要能反向指导页面、内容、结构化数据、实体资料、第三方权威来源和合规审查。
- 可比较：同一周期、同一平台、同一问题库、同一采样条件下的结果才适合做前后对比。
- 有边界：非量化指标可以作为判断线索，后端指标可以作为经营观察，但二者都不宜直接变成服务商绩效承诺。

端可量化指标：以核心指标作为考核抓手

前端可量化指标体系包含多个维度，如可见率、引用率、Top1/Top3、品牌一致性、答案准确性、竞品相对位置等。

但在企业实际采购和验收中，不宜把所有指标都纳入硬性考核。多数指标更适合作为过程指标，用于帮助服务商诊断问题、优化策略、复盘迭代。

企业应选择少数最能代表业务结果的核心指标作为考核抓手。比如在POC验收、阶段复盘和年度服务评估中，可以优先采用**Top3进入率**作为核心指标，观察目标品牌是否稳定进入AI答案的主要推荐范围。

其他过程指标可以作为辅助参考，用于解释Top3变化背后的原因，包括平台差异、问题包差异、内容资产质量、引用源分布和竞品表现等。这样既能保证考核口径清晰，也能保留策略优化所需的分析空间。



表4：前端可量化指标总表

指标	推荐计算口径	适合用途	注意事项
目标品牌可见率	提及目标品牌的有效样本数 / 有效样本总数	衡量目标品牌是否进入AI答案	需要明确品牌全称、简称、英文名、产品名是否计入
目标品牌Top 1概率	目标品牌位于推荐序列第1位的有效样本数 / 有效样本总数	衡量目标品牌在答案中的首位概率	只适用于有明确推荐序列、名单或可编码顺序的答案
目标品牌Top 3概率	目标品牌进入推荐序列前3位的有效样本数 / 有效样本总数	衡量目标品牌进入核心候选名单的概率	适合采购推荐、品牌对比、品类榜单类问题
首提概率	目标品牌在答案中第一个被提及的样本数 / 有效样本总数	衡量答案显著性	当答案没有明确排序时，首提比Top 1更稳妥

指标	推荐计算口径	适合用途	注意事项
平均出现位置	目标品牌出现位置序号的均值或中位数	观察目标品牌位置变化	需要统一编码规则，未出现样本应单独处理
自有域引用率	引用企业官网、文档、白皮书、帮助中心等自有域的样本数 / 有效样本总数	衡量企业自有资料被采用的概率	平台不展示引用时，应记录为无引用或不可判定
有效引用率	引用来源能够支撑答案相关结论的样本数 / 有效样本总数	判断引用是否真正有用	不能只看有没有引用，还要看引用是否支持结论
第三方权威引用率	引用行业机构、监管机构、学术、标准组织、权威媒体等来源的样本数 / 有效样本总数	衡量外部证据结构	需要先定义权威来源白名单或分级规则
事实错误率	出现企业事实错误的样本数 / 有效样本总数	控制错误认知与合规风险	错误要分级，例如轻微过期、核心事实错误、敏感错误
竞品共现率	同一答案中目标品牌与指定竞品共同出现的样本数 / 有效样本总数	分析竞争环境	共现本身不一定是风险，关键看位置、理由和描述
相对优势概率	目标品牌位置或推荐理由优于指定竞品的样本数 / 有效样本总数	判断目标品牌相对表现	需要事先定义“优于”的编码规则
样本稳定率	多次采样中目标品牌出现状态保持一致的样本组数 / 样本组总数	判断平台波动和结果可信度	稳定率低时，不宜用单次结果做结论
问题覆盖率	已纳入监测的问题数 / 目标问题库问题数	判断监测范围是否完整	不能用覆盖率代替效果指标
场景覆盖率	已覆盖场景数 / 目标场景数	避免只测品牌词	目标场景应包含品牌、品类、采购、竞品、风险、售后等

Top 1、Top 3与可见率的具体定义

企业在采购时应要求服务商把Top 1、Top 3和可见率写进SOW（工作证明），避免后续解释口径变化。

推荐采用以下定义：

目标品牌可见率用于回答“有没有出现”。

只要AI答案在正文、推荐列表、对比表格或明确候选名单中提及目标品牌，就可以计为可见。若只出现在引用链接标题、网页摘要、相关搜索或页面角落，不宜直接计为答案可见。

目标品牌Top 1概率用于回答“是否第一选择”。

当AI答案给出明确排序、推荐清单、候选名单或可编码的出现顺序时，目标品牌位于第一位才计入Top 1。若答案为并列叙述，应提前规定并列第一是否计入。若答案没有顺序，建议使用首提概率，避免把主观判断包装成Top 1。

目标品牌Top 3概率用于回答“是否进入核心候选”。

当AI答案列出3个以上品牌、供应商或方案时，目标品牌进入前三位才计入Top 3。若答案少于3个候选，企业应单独标注，不要强行纳入Top 3计算。

这些指标都应以有效样本为分母

有效样本需要排除平台失败、回答空白、明显无关问题、采集脚本错误和无法判定顺序的样本。被排除样本也要记录，不能直接删除。



表5：前端可量化指标的采样口径

采样要素	建议要求
平台范围	豆包、DeepSeek、Kimi、元宝、千问分别统计，不建议先混成一个总分
问题范围	品牌词、品类词、采购词、竞品词、场景词、风险词分别统计
重复采样	同一问题建议多次采样，记录时间、账号、设备、地区和联网状态
有效样本	对失败、空白、无关、不可判定顺序的样本设定排除规则
排序编码	明确列表排序、段落首提、并列推荐、表格顺序的编码方式
品牌别名	明确企业全称、品牌名、产品名、英文名、简称是否计入
数据留存	保留原始答案、引用链接、截图、采样日志和人工编码记录
验收周期	POC可按15至30天复测，长期项目建议月度复盘和季度复审

前端非量化指标：适合关注，不适合作为硬考核

有些指标对企业很重要，但很难形成真正客观的量化结果。

原因在于这些指标需要主观定义，且不同评审人、不同业务部门和不同语境下的判断可能不同。企业可以把它们作为人工评审项、风险标签或复盘线索，但不建议把它们包装成精确分数。

典型的非量化指标包括语义一致性、情感积极性、表达友好度、推荐理由质量、品牌调性匹配、答案完整性、专业性、说服力、风险表述是否恰当等。它们可以帮助企业发现问题，但如果没有严格标注规则、多人一致性检验和样本审计，就容易变成“伪量化”。



表6：前端非量化指标与使用边界

指标	可以观察什么	推荐使用方式	不建议做法
语义一致性	AI对企业定位、产品边界和价值主张的理解是否贴近企业事实	人工评审、典型样本分析、错误归因	直接给出一个“语义一致性提升37%”的硬承诺
情感积极性	答案是否偏正向、中性、谨慎、负向或带风险提示	用作声誉风险标签和舆情线索	把积极、中性、负向简单换算成项目KPI
推荐理由质量	AI推荐目标品牌时的理由是否具体、准确、有证据	结合引用来源和事实准确性做案例复盘	只看推荐理由字数或形容词数量
品牌调性匹配	答案表达是否符合企业希望传递的专业、可信、创新等形象	用于内容策略和品牌资料改写	要求服务商保证AI使用指定话术
答案完整性	回答是否覆盖核心产品、场景、适用边界和限制条件	用于识别资料缺口	把完整性当成跨平台统一分数

指标	可以观察什么	推荐使用方式	不建议做法
专业性	答案是否使用正确行业术语，是否避免误导	用专家复核或行业顾问审核	让非专业评审直接判定复杂技术内容
风险表述合理性	AI是否把合规、隐私、安全、医疗、金融等风险讲清楚	用于高敏行业审查	要求AI删除所有风险提示

后端指标：企业应观察，通常不作为服务商考核

后端指标可以帮助企业理解GEO与商业结果之间的关系，但多数情况下不适合作为服务商直接考核。

主要原因包括：AI平台通常不提供完整查询日志，用户从AI答案到官网、销售或成交的路径不完整，AI回答会受到时间、账号、联网状态和平台策略影响，同期广告、PR、SEO、销售活动也会干扰结果。

因此，后端指标更适合企业内部长期观察。

企业可以在CRM、网站分析、品牌监测和销售复盘中标注GEO相关线索，但不宜要求服务商承诺短期线索数、成交额或销售转化。



表7：GEO后端观察指标与使用边界

指标类别	可观察指标	适合用途	不适合作为硬考核的原因
流量	AI平台推荐流量、搜索控制台Web流量、品牌词搜索量、直接访问变化	观察AI与搜索生态的外部变化	很多AI答案不会带来点击，平台也不一定提供完整来源
行为	页面停留、转化路径、下载、注册、咨询、收藏	观察内容是否被更高意图用户使用	行为变化可能受页面改版、广告和活动影响
线索	MQL、SQL、表单、电话、企微添加、试用申请	辅助判断商业价值	需要CRM字段、人工标注和销售回访支持
销售	商机、报价、成交、客单价、回款周期	长期预算评估和经营复盘	通常无法直接归因到GEO单一动作

指标类别	可观察指标	适合用途	不适合作为硬考核的原因
品牌	品牌搜索、社媒提及、媒体引用、招投标问询、销售反馈	评估认知变化和销售支持	需要多渠道数据交叉验证

指标优先级建议

一期项目建议把可量化前端指标作为主考核：目标品牌可见率、目标品牌Top 1概率、目标品牌Top 3概率、自有域引用率、有效引用率、事实错误率、竞品共现率和样本稳定率。

二期项目可以加入更复杂的前端指标：首提概率、平均出现位置、相对优势概率、第三方权威引用率、场景覆盖率、跨平台差异和结构化数据有效性。

非量化前端指标建议作为辅助复盘：语义一致性、情感积极性、推荐理由质量、品牌调性匹配、答案完整性、专业性和风险表述合理性。它们可以帮助企业判断改进方向，但不建议直接进入付款条件。

后端指标建议由企业内部观察：流量、行为、线索、商机、成交、销售反馈和品牌搜索。成熟企业可以尝试CRM字段标注、区域对照、时间窗口对照和销售访谈，但仍需标注归因限制。

建议仪表盘结构



表8：企业GEO仪表盘建议

模块	核心内容
总览	本月平台覆盖、问题数量、目标品牌可见率、Top 1概率、Top 3概率、引用率、错误率、主要风险
平台对比	豆包、DeepSeek、Kimi、元宝、千问在同一问题上的答案差异
场景对比	品牌、品类、采购、竞品、风险、售后问题的表现
竞品对比	目标企业与3至5个竞品的提及、位置、引用和描述差异
引用来源	自有域、第三方权威资料、媒体、社区、百科、文档等来源结构
内容资产	被引用页面、未被引用但重要的页面、需要重写或补齐的页面
非量化观察	语义一致性、情感倾向、推荐理由、专业性和风险表述的典型样本
后端观察	AI相关流量、品牌搜索、线索备注、销售反馈和同期干扰因素
行动清单	本月要改的页面、要补的资料、要修的技术问题、要确认的合规点

指标计算案例：把概率说清楚

前端可量化指标的价值在于能基于采样数据计算。

企业不需要把指标设计得过于复杂，但必须保证口径稳定。

案例：星河CRM的Top 3概率计算

星河CRM选择50个采购相关问题，覆盖豆包、DeepSeek、Kimi、元宝、千问五个平台。每个平台每个问题采样3次，总回答数为50乘5乘3，共750条。

在750条回答中，目标品牌出现在前三推荐位置的回答有135条。

则目标品牌Top 3概率为135除以750，等于18%。

如果一个月后相同问题、相同平台、相同采样次数下，Top 3回答数增加到210条，则Top 3概率为28%。这个变化可以作为前端量化效果的证据，但企业仍需看样本是否可比、平台入口是否一致、问题库是否被中途更换、是否存在只挑好问题汇报的情况。

案例：目标品牌可见率计算

麦芽小鹿选择30个问题，采样豆包、Kimi、千问，每个平台每题3次，总回答数270条。

只要答案中合理提及麦芽小鹿，就记为可见。若270条中有54条提及，则可见率为20%。

这里的“合理提及”要提前定义。比如“麦芽小鹿不适合所有儿童，建议查看配料和医嘱”也算可见；但“麦芽小鹿是某国外品牌旗下产品”属于事实错误，不能简单当成正向可见，需要同时记入事实错误率。

案例：有效引用率计算

南岭精密在400条采样回答中有80条回答提及品牌，其中有20条引用了企业官网、产品文档或白皮书，有8条引用了无关页面或错误页面。

有效引用率可以按20除以80计算，即25%。无效引用率可以按8除以80计算，即10%。

有效引用率比单纯可见率更接近“可信可见”。对B2B、工业、医疗器械、金融科技等行业，企业应更重视有效引用率，因为AI答案如果没有可靠来源支撑，后续销售和合规风险都会增加。

方法参考：前端指标仪表盘怎么设计

建议企业把仪表盘分成四层，避免只看一个总分。

第一层：管理层摘要

管理层只需要看到5个指标：目标品牌可见率、目标品牌Top 3概率、有效引用率、重大事实错误率、红色风险数量。每月看趋势，不建议每日追波动。

第二层：平台对比

分别展示豆包、DeepSeek、Kimi、元宝、千问的表现。每个平台都应显示样本数、采样日期、入口、联网状态和问题类别。平台之间不能直接用一个分数粗暴排名，因为每个平台的内容来源、搜索机制和回答风格不同。

第三层：场景对比

把问题拆成品类教育、采购决策、竞品对比、风险判断、售后服务和品牌认知。很多企业会发现，品类教育类问题表现还可以，但采购决策和竞品对比类问题很弱。这个发现比一个总分更有行动价值。

第四层：样本追溯

每一项指标都应能回到原始回答。企业至少要能查看问题、平台、时间、回答文本、品牌位置、引用链接、竞品出现情况、人工复核标记和处理状态。没有样本追溯的看板，很难用于验收。

仪表盘字段建议

字段	说明	是否必需
问题ID	每个问题固定编号	必需
问题文本	原始提问，不随意改写	必需
问题类别	品类、采购、竞品、风险、售后等	必需
平台	豆包、DeepSeek、Kimi、元宝、千问	必需
入口	网页、App、插件、API或搜索增强入口	必需
采样时间	精确到日期，必要时到小时	必需
联网状态	是否开启搜索、深度研究或联网功能	必需
目标品牌是否出现	是或否	必需
目标品牌位置	Top 1、Top 2、Top 3、其他、未出现	必需
引用来源	自有域、第三方、平台生态、无引用	必需
事实错误	无、轻微、重大	必需
人工复核人	责任人	建议
处理建议	内容、技术、合规或观察	建议

非量化指标的使用方法：可关注，慎考核

语义一致性、情感积极性、推荐理由质量、品牌调性匹配等指标有价值，但里面包含人工判断。企业可以用它们辅助复盘，不建议把它们作为硬性付款条件。

01 方法参考：非量化指标的三级标签法

企业可以把非量化指标改成三级标签。

- A级：表达准确、语义一致、无明显风险，可作为正向样本。
- B级：总体可接受，但存在表述模糊、信息不足或推荐理由弱的问题。
- C级：存在误导、夸大、偏见、事实错位或合规风险，需要处理。

这种标签法的好处是降低“假量化”风险。

企业不必声称“语义一致性得分87.3分”，只需要说明哪些答案属于A、B、C，各类样本占比如何，背后的处理动作是什么。

案例：情感积极性的假量化风险

虚构企业“澜川家居”要求服务商把“AI回答情感积极性”做到90分以上。

服务商用一个内部模型给答案打分，但没有解释评分依据。一个回答写道：“澜川家居价格较高，但设计感较强，适合预算充足的家庭。”服务商给了82分，企业市场部认为偏低，销售团队认为这个回答反而真实可信。

最后澜川家居把指标改成三项人工标签：是否准确、是否有明显负面误导、推荐理由是否可接受。这样能减少主观争议，也更适合指导内容优化。

后端指标的观察方法：相关性优先，归因谨慎

GEO后端指标包括官网流量、品牌词搜索、咨询量、线索量、下载量、销售反馈和客户问询变化。

> 它们对企业有业务价值，但多数情况下不适合直接写成服务商KPI。

原因很简单：生成式AI答案带来的后端影响可能和广告、SEO、品牌传播、线下活动、渠道销售、行业周期、竞品动作同时发生。企业难以把某一条线索科学归因到某一次AI回答。

方法参考：后端观察四步法

第一步，建立基线。记录GEO项目启动前至少4周的品牌词搜索、官网入口、咨询表单、销售问询和下载数据。

第二步，设置观察窗口。一般按月观察，至少连续3个月。不要用单日数据判断GEO效果。

第三步，记录干扰变量。包括广告投放、媒体发布、展会、促销、产品更新、舆情事件和竞品活动。

第四步，做相关性解释。若前端可见率提升、品牌词搜索提升、销售反馈中出现“我在AI里看到你们”的样本，可以作为业务信号，但仍需标注为观察结论。

案例：后端指标不能直接付款

星河CRM在GEO项目第二个月线索量增长20%。服务商希望把这作为“GEO带来20%增长”的证明。

但同月企业也上线了搜索广告，销售团队参加了行业大会，官网改版并新增白皮书下载。

因此，星河CRM只把这项变化写入“业务观察”，没有作为服务商验收的唯一依据。最终验收仍围绕Top 3概率、有效引用率和事实错误率。

04

第 四 部 分

重点AI平台视角：豆包、DeepSeek、Kimi、元宝、千问

重点AI平台视角：建议五个平台整体采购 48

五个平台的通用采样口径 49

平台分层策略 50

4.4 五个平台采样案例：同一问题为什么会出现不同答案 51

重点AI平台视角：建议五个平台整体采购

为什么建议五个平台一起采购？

企业采购GEO服务时，建议将豆包、DeepSeek、Kimi、元宝、千问五个平台整体纳入同一采购包，作为GEO项目的基础评估范围。

原因很简单：不同AI平台的模型能力、搜索源、联网方式、内容生态、上下文能力、产品入口和用户使用场景存在明显差异

同一企业、同一产品、同一问题，在不同平台上的答案表现可能完全不同。

如果企业只采购单个平台的GEO服务，很容易得到片面的判断

一个服务商如果只能优化某一个平台，却无法解释五个平台之间的差异，也无法针对不同平台制定采样、内容和优化策略，说明其综合能力仍需谨慎评估。

五个平台一起采购，至少有三层价值：

第一，可以考察GEO服务商的综合实力

真正成熟的GEO服务商，应能同时理解不同平台的答案机制、内容偏好、引用来源和用户场景，而不能只给一个笼统的“AI可见性”分数

第二，可以覆盖不同用户人群

豆包更偏大众用户入口，Kimi更适合长文档和研究型场景，DeepSeek更容易出现在推理、技术和深度问答场景，元宝受到腾讯生态影响，千问与阿里生态、办公和开发者场景关联更强。五个平台合在一起，才能更接近企业真实用户的AI搜索路径。

第三，可以发现平台级机会和风险

有的平台可能已经能正确理解企业，有的平台可能仍存在事实错误、竞品偏置、信源污染、引用缺失或品牌实体混淆。只有跨平台采样，企业才能判断问题到底出在内容资产、信源结构、品牌叙事，还是平台自身的答案机制。

本版蓝皮书建议，豆包、DeepSeek、Kimi、元宝、千问应作为国内GEO项目的基础平台组合。每个平台的具体能力会随版本变化，企业在采购和验收时，应要求服务商记录采样日期、产品入口、是否联网、是否开启搜索模式、账号状态、地区环境和回答版本

表9：五个重点平台的GEO采购含义

说明：平台功能以官方文档和实际采样为准。采购评估时，企业不应只听服务商做概括性描述，应要求其提供平台级采样数据、差异解释和对应优化策略。

五个平台的通用采样口径

为了保证五个平台之间的数据可以对比，企业应要求服务商统一采样口径。

- 每条问题都要记录平台、产品入口、日期、时间、地区、账号状态、是否联网、是否开启深度搜索或搜索模式。
- 同一问题应在五个平台中同步采样，避免只采部分平台后做整体判断。
- 同一问题至少进行多次采样。若某个平台答案波动较大，应提高采样次数，并把波动本身作为指标。
- 答案中涉及引用来源时，要记录来源标题、域名、链接、引用片段和是否支持答案结论。
- 涉及企业事实时，要标注来源是否来自官网、权威第三方、媒体、社区、百科或低质量页面。
- 涉及敏感行业时，要由企业内部专家或法务对答案进行事实和合规复核。

平台分层策略

表10：平台分层策略

建议将“五平台基础采购”作为默认方案。预算有限时，也应保留五个平台的基础采样，只在问题数量、采样次数和复核深度上做分层，而不应直接砍掉平台范围。

4.4 五个平台采样案例：同一问题为什么会出现不同答案

企业在采样时会发现，同一问题在豆包、DeepSeek、Kimi、元宝、千问中的回答可能差异很大。

这种差异不代表单个平台结论绝对正确或错误，常见原因包括模型能力、联网开关、搜索源、产品入口、上下文长度、引用策略、安全策略和内容生态不同。

案例：国产CRM推荐问题

虚构问题："制造业CRM系统怎么选，国内有哪些值得关注的服务商？"

企业不要只截一张答案图就下结论。更合理的做法是保持同一问题、同一时间窗口、同一入口和同一轮次，在五个平台上多次采样，再看整体概率、平台差异和可优化空间。

方法参考：五平台采样的最小可行方案

预算有限时，可以采用五平台最小可行采样方案。

- 问题数量：30个核心问题。
- 平台范围：豆包、DeepSeek、Kimi、元宝、千问五个平台全部纳入。
- 采样次数：每个平台每题3次。
- 总样本量：30×5×3，共450条回答。
- 复核方式：由服务商初标，企业抽检20%，重大风险样本全部人工复核。
- 输出内容：可见率、Top3概率、引用率、事实错误率、平台差异、主要缺口和下一步行动。

中大型企业可以扩展到100个问题、5个平台、每题3至5次。上市公司、金融、医疗、教育、政务相关企业，应增加合规复核样本、风险问题样本和竞品对比样本。

采样注意事项

- 采样问题不能中途随意更换。若更换问题，需要标注新旧问题库，不可混算。
- 提问上下文要统一。多轮对话会改变答案，单轮和多轮应分开统计。
- 搜索增强开关要记录。开启联网和关闭联网得到的答案不能混在一起。
- API和C端产品要分开。API能力不等同于用户在App里看到的答案。
- 平台更新要记录。平台模型更新或产品入口变化后，需要重新建立基线。
- 五个平台应分别出具平台报告，同时形成一份跨平台综合报告。

最终，企业采购GEO服务时，应把五个平台视为一个基础评估组合。能不能同时理解、采样、解释和优化豆包、DeepSeek、Kimi、元宝、千问，已经成为判断GEO服务商专业能力的重要标准。

05

第 五 部 分

中国GEO服务商类型划分

如何用类型划分指导采购 55

判断AI原生能力是否真实 56

主服务商与组合采购 57

服务商类型选择案例 58

方法参考：AI原生能力测试作业 59

中国市场的GEO服务商正在从多个方向涌现。

企业采购时需要先识别服务商的能力来源，再判断其能力结构是否匹配自身目标。类型只能用于初筛，最终仍要看可复核的方法、样本、团队和交付证据。

一个关键变化是，GEO服务商不应只从SEO或内容营销视角理解。

真正高质量的团队通常需要AI原生能力，同时具备SEO、内容、营销、数据和项目治理的复合能力。尤其是AI native能力强、又理解SEO与营销落地的团队，往往更适合承担主服务商角色。



表11：中国GEO服务商类型划分

类型	典型背景	核心能力	适合采购场景	常见短板	重点判断
AI原生型GEO服务商	AI产品、模型评测、智能体、提示词工程、数据评测团队	理解模型行为、检索增强、问答生成、采样波动、评测框架和自动化监测	需要建立GEO指标体系、平台采样、评测看板和实验闭环的企业	可能缺少SEO、品牌叙事、PR资源和企业内容落地经验	能否把AI评测转化为可上线的内容、页面和组织动作
AI原生+SEO+营销复合型	AI native团队叠加技术SEO、内容营销、品牌传播、数据分析背景	既懂AI平台与评测，又懂抓取索引、内容资产、品牌表达、竞品分析和项目治理	中型企业、行业龙头、上市公司、出海企业的一期主服务商	团队稀缺，报价通常更高，也容易被包装成概念	是否有真实跨职能团队，是否能拿出采样数据、内容改造和技术建议的闭环证据

类型	典型背景	核心能力	适合采购场景	常见短板	重点判断
SEO转型型	传统SEO、技术SEO、搜索营销公司	懂抓取、索引、页面结构、关键词、内链和搜索数据	官网基础薄弱、页面不可索引、结构化数据缺失的企业	可能把GEO简单套用SEO排名逻辑	是否理解AI答案的概率属性、引用机制和平台差异
内容营销/PR型	品牌、公关、内容营销、媒体传播公司	懂品牌表达、叙事、案例、媒体关系和内容生产	需要改善品牌叙事、案例表达和权威资料的企业	技术和数据监测能力可能不足	是否能提供可复核数据，而非只交付文章和传播稿
数据监测/SaaS型	舆情、品牌监测、数据看板、BI工具公司	擅长监测、告警、竞品对比、报表和权限管理	需要长期监测、风险告警和多品牌管理的中大型企业	可能缺少内容改造、技术修复和业务策略能力	看板能否解释行动方案，数据能否导出和复核
技术SEO/数据工程型	数据咨询、爬虫、BI、网站技术服务商	擅长采样、数据管道、仪表盘、结构化数据、日志和实验设计	集团、上市公司、技术团队成熟企业	品牌内容与业务策略能力可能不足	是否能和市场、品牌、法务团队协同，而非只交付技术清单
大模型/RAG/Agent技术型	AI应用开发、知识库、RAG、智能体和企业AI供应商	懂内部知识库、检索、模型评测、Agent流程和企业AI应用	需要把内部知识库治理和外部GEO结合的企业	未必熟悉公共AI平台的外部可见性	能否区分内部RAG效果和公共AI平台GEO效果
行业垂直咨询型	医疗、金融、教育、工业、B2B咨询公司	懂行业术语、客户决策链、监管要求和专业内容	高敏行业、复杂B2B、专业服务企业	通用平台监测和自动化能力可能较弱	是否具备标准化数据方法和可复核样本
跨境增长型	出海营销、跨境SEO、海外PR、独立站、本地化服务商	熟悉语言、本地化、海外搜索、海外内容生态和跨境增长	出海企业、海外B2B、跨境电商、全球化品牌	国内AI平台经验可能不足	是否能分别说明国内平台和海外平台的策略差异
资源铺量型	软文发布、媒体资源、外链、口碑和发布渠道服务商	短期可以大量发布内容	一般不建议作为主服务商	质量、合规和长期风险高	虚假评价、低质内容、隐性投放、不可验收

如何用类型划分指导采购



如果企业只采购监测，可以选择数据监测/SaaS型，但要明确它解决的是“看见问题”和“持续预警”，还需要内容、技术或品牌团队负责改进动作。

如果企业主要问题是目标品牌在AI答案中缺席、Top 1概率低、Top 3概率低、平台波动大，优先考虑AI原生型或AI原生+SEO+营销复合型服务商。

前者更适合建立评测体系和采样闭环，后者更适合承担从诊断到落地的主服务商角色。

如果企业主要问题是官网资料混乱、页面不可索引、结构化数据缺失、站点主题关系弱，优先考虑SEO转型型、技术SEO/数据工程型服务商，并要求其补足AI平台采样和答案评测能力。

如果企业主要问题是AI不了解业务、品牌叙事弱、案例和权威证据不足，优先考虑内容营销/PR型、行业垂直咨询型服务商，同时引入AI原生评测或数据监测能力，避免只做内容发布。

如果企业是上市公司、行业龙头或高敏行业，优先考虑具备AI原生评测、合规流程、数据治理、项目管理和跨部门协同能力的复合型团队。此类企业不宜只买零散内容或单一看板。

如果企业目标是出海，应选择跨境增长型服务商，并要求其证明目标市场的本地化资料、搜索生态、AI平台经验和跨境数据合规能力。若同时覆盖国内市场，需要额外评估豆包、DeepSeek、Kimi、元宝、千问的采样能力。

资源铺量型服务商可以作为风险案例观察，不宜作为GEO主服务商。采购时要重点询问内容来源、付费关系、合规审查、可持续性和是否会制造长期声誉风险。

判断AI原生能力是否真实



很多服务商会宣称自己具备AI能力，企业需要看具体工作证据，而非只看宣传话术。

AI原生能力至少应体现在以下方面：

- 能说明豆包、DeepSeek、Kimi、元宝、千问在产品入口、联网状态、引用展示、搜索模式和答案稳定性上的差异。
- 能设计问题库、采样策略、重复采样、人工编码规则和有效样本排除规则。
- 能把目标品牌可见率、Top 1概率、Top 3概率、引用率、错误率等指标计算清楚。
- 能解释模型随机性、检索结果变化、上下文差异和平台产品更新对结果的影响。
- 能把AI评测结果转化成内容改写、页面结构、实体治理、引用来源和合规审查建议。
- 能区分公共AI平台GEO、搜索SEO、内部RAG知识库和企业智能体优化之间的边界。
- 能提供原始样本、采样日志、数据导出、复测方法和异常样本解释。

主服务商与组合采购

GEO往往需要组合能力。

企业可以选择一个主服务商负责总体策略、数据监测和项目治理，再按需引入内容、技术、PR、法务或行业专家。

对于中大型企业，单一服务商很难在平台监测、技术SEO、品牌内容、行业专业、数据安全和合规审查上全部达到高水平。



最理想的主服务商通常是AI原生+SEO+营销复合型团队。

它能够用AI原生能力建立评测体系，用SEO能力修复抓取、索引、结构和实体问题，用营销能力完成内容叙事、案例表达和权威证据建设。

但企业仍需验证其复合能力是否由真实团队承担，不能只看一个打包名称。

- 主服务商适合承担：指标体系、基线测量、问题库、平台监测、项目管理、月度复盘。
- AI原生伙伴适合承担：采样系统、指标计算、模型行为分析、实验设计、自动化监测。
- 内容伙伴适合承担：FAQ、官网页面、白皮书、案例、产品文档、销售资料改写。
- 技术伙伴适合承担：结构化数据、抓取与索引、站点架构、日志、仪表盘。
- PR或行业伙伴适合承担：权威第三方资料、行业报告、案例背书、媒体沟通。
- 企业内部必须承担：业务事实确认、合规审查、客户案例授权、预算与优先级决策。

服务商类型选择案例

案例一：小企业不宜采购过重的技术方案

麦芽小鹿预算有限，官网由外包团队维护，内部没有数据工程师。它最适合选择轻量诊断型或AI原生+内容型服务商，先完成问题库、基线采样、核心FAQ和产品科普页优化。如果一开始选择大型技术看板型服务商，可能出现预算花在仪表盘上，内容和事实资产依旧不足的问题。

案例二：B2B企业适合复合型团队

星河CRM有官网、案例、白皮书和销售知识库，但内容结构不适合AI答案。它适合选择AI原生+SEO+营销复合型服务商。原因是这类服务商既能理解生成式答案的采样和概率，又能处理官网结构、内容表达、实体关系、行业话术和销售转化。

案例三：上市集团需要主服务商加专项团队

华辰智造需要实体治理、合规审批、海外本地化和多事业部协同。它可以选择一个复合型主服务商负责整体方法、数据和项目治理，再配合海外本地化团队、法律顾问和技术SEO团队。大型企业不宜让多个供应商各自汇报，容易形成指标口径冲突。

方法参考：AI原生能力测试作业



企业可以在比稿阶段给服务商一份48小时小作业，不要求免费交付完整方案，但要求其展示方法。

作业一：问题库构建

给服务商一份企业官网和产品手册，让其提出20个GEO采样问题，并说明每个问题属于什么业务场景、为什么有价值、可能对应哪些企业事实材料。

判断标准：专业服务商会覆盖用户自然语言、采购决策、竞品对比和风险问题；能力不足的服务商往往只会把官网标题改成问题。

作业二：样本复核

给服务商10条AI回答，让其标注目标品牌是否可见、是否Top 3、是否有效引用、是否存在事实错误、应采取何种动作。

判断标准：专业服务商会区分正向可见、错误可见、无引用可见和风险可见；能力不足的服务商只会统计有没有出现品牌名。

作业三：内容改写

给服务商一段产品介绍，让其改成更适合AI理解和引用的企业事实页。要求保留事实边界，不能夸大。

判断标准：专业服务商会增强结构、定义、适用场景、限制条件、证据和FAQ；能力不足的服务商会堆关键词、堆形容词、堆“行业领先”。

作业四：风险判断

给服务商一个敏感行业问题，例如“某医疗器械是否能替代医生诊断”，要求其说明企业内容和AI答案中应如何设置风险边界。

判断标准：专业服务商会先问合规边界、监管要求、产品适用范围和免责声明；能力不足的服务商会把问题当作普通营销文案处理。

06

第六部分

服务商评估维度与评分框架

总评分建议 62

不同企业的权重调整 63

服务商访谈评分表 64

报价评估 66

评分案例：如何避免被销售话术带偏 67

方法参考：报价拆解表 68

总评分建议

企业可以用100分制评估GEO服务商。评分不应只看方案漂亮程度，更要看方法是否可复核、AI原生能力是否真实、SEO与营销落地是否完整、合规是否稳妥、团队是否能长期协作。



GEO服务商100分评分框架

维度	建议权重	评估要点
测量与评估方法	20分	问题库、重复采样、平台记录、原始样本、指标定义、Top 1与Top 3口径、复测方法、波动处理
AI原生与平台理解	15分	模型行为、检索增强、搜索/联网模式、答案生成、采样自动化、异常样本解释、五个平台差异
SEO、内容与营销复合落地	15分	抓取索引、结构化数据、页面结构、FAQ、产品文档、品牌叙事、案例、权威证据
数据与监控能力	15分	可视化、日志保存、竞品对比、异常告警、数据导出、权限管理、长期趋势
合规与安全	15分	数据处理、个人信息、客户案例授权、敏感行业审查、内容审核、合同禁止行为
项目管理	10分	节奏、周会、月报、协同工具、变更管理、验收和交接
商业透明度	10分	报价清晰、资源关系披露、不可承诺事项、退出机制、案例真实性

不同企业的权重调整



按企业类型调整评分权重

企业类型	权重调整建议
小型企业	提高内容与实体治理、项目性价比和短期可执行动作权重；降低复杂数据看板权重
中型企业	保持测量、AI原生、SEO、内容、项目管理均衡；重点看90天内可落地结果
行业龙头	提高数据与监控、竞品对比、平台全覆盖、AI原生评测和组织协同权重
上市公司	提高合规、安全、事实准确、信息披露一致性、供应商审计和数据权属权重
高敏行业	将合规与安全提升到最高优先级，并要求专家复核流程
出海企业	提高海外平台、本地化、语言质量、目标市场权威来源和跨境数据合规权重

服务商访谈评分表



服务商访谈评分问题

问题	优秀回答特征	风险回答特征
你们如何定义GEO可见性?	能拆分目标品牌可见率、Top 1概率、Top 3概率、引用率、事实错误率、竞品共现和波动	只说排名、曝光或首页
你们团队的AI原生能力体现在哪些具体工作上?	能解释模型行为、检索增强、采样波动、评测框架、自动化监测和平台差异	只说会使用AI写稿或做提示词
如何构建问题库?	基于品牌、品类、采购、竞品、风险、售后和用户旅程构建	只用品牌词和供应商自己想的问题
如何采样?	记录平台、入口、时间、账号、联网状态、重复次数、原始样本和有效样本排除规则	只给截图或口头结论
如何处理Top 1和Top 3?	明确排序编码、首提规则、并列处理、有效样本和不可判定样本	把任何主观好结果都叫Top 1
如何处理语义一致性和情感积极性?	承认其主观性，用人工评审、典型样本和风险标签辅助判断	把主观评分包装成精确量化KPI
如何处理平台波动?	多次采样、分场景看趋势、保留置信度和异常说明	把一次好结果当作稳定效果

问题	优秀回答特征	风险回答特征
如何解释后端效果?	区分前端指标、后端观察和因果归因，说明归因限制和实验条件	承诺一定带来线索或成交
如何把SEO、内容营销和AI评测串联起来?	能从指标发现问题，再落到页面、结构化数据、内容、案例和权威来源	只会单点交付，无法形成闭环
如何处理合规内容?	提供审批流、敏感词、客户授权、数据处理和禁止行为清单	认为GEO只是内容发布，无需合规
如何做竞品分析?	比较提及、引用、来源、理由、场景和错误点	只列竞品名称或截图
如何交接数据?	交付问题库、样本、报告、看板、内容记录和技术建议	数据只在供应商系统里，客户无法导出

报价评估

GEO报价通常由诊断、监测、内容、技术、数据看板、项目管理和第三方资源等部分构成。

企业评估报价时，应把重点放在两个层面：

第一，**结果考核以核心可量化指标为主。**企业不宜把所有指标都纳入硬性考核，应优先选择最能代表项目成效的核心指标，例如目标品牌Top3概率、目标品牌可见率、引用率、事实错误率等。过程指标可以用于策略复盘和执行优化，但不宜全部变成付款或验收条件。

第二，**过程方法论要重点评估内在含金量和专业水平。**GEO服务的价值，不能简单用执行动作数量或人天投入衡量。不同服务商可能采用不同策略、不同路径和不同组织方式，真正需要评估的是其底层方法、运营思路、判断原则、团队能力、行业理解和持续迭代能力。

企业不应只比较总价，而应要求服务商说明报价背后的服务逻辑，包括核心方法论、项目推进机制、数据评估原则、内容策略框架、技术优化思路、团队配置方式、复盘机制，以及这些能力如何支撑核心指标提升。

诊断费	包含问题库设计、基线测量、竞品分析、风险诊断和策略建议。重点看诊断框架是否专业，是否能识别真实问题，是否具备行业洞察。
AI评测费	包含采样系统、重复采样、指标计算、异常样本分析、平台差异分析和看板维护。重点看评测方法是否稳定，指标口径是否清晰，数据解释是否有专业判断。
执行费	包含内容改写、页面优化、结构化数据建议、实体资料治理、平台监测和项目管理。重点看执行是否围绕核心指标展开，是否有清晰的优先级和迭代逻辑。
数据费	包含采样工具、看板、导出、告警和历史数据保存。重点看数据是否能支持周期对比、竞品对比、策略复盘和管理决策。
第三方费用	包含媒体、研究报告、认证、翻译、本地化、专家审查等。重点看是否必要、是否透明、是否能形成可信证据资产。
绩效费用	如采用绩效条款，应基于可复核的前端可量化核心指标，并明确平台范围、样本范围、统计周期、排除项和非量化指标边界。

因此，报价评估不宜简单看“便宜还是贵”，而应综合判断：**核心指标是否可考核，方法论是否有含金量，运营思路是否清晰，团队能力是否匹配，交付机制是否能支撑长期提升。**

评分案例：如何避免被销售话术带偏

虚构企业“云杉财税”准备采购GEO服务，收到三家服务商方案。

服务商	报价	主要卖点	初步判断
A公司	12万元/年	每月提供AI可见度报告，覆盖五个平台	适合基础监测，但执行能力待核验
B公司	36万元/年	提供问题库、基线、内容改造、技术建议、月度复盘	更接近复合型服务，需看方法和团队
C公司	80万元/年	承诺三个月全平台Top 1，含大量软文发布	存在高风险承诺，需谨慎

如果只看报价，A公司最便宜，C公司看起来最强。

但按评分框架拆开后，云杉财税发现B公司虽然报价居中，但能提供原始样本、指标口径、内容计划、技术改造建议和合规流程。C公司的承诺最吸引人，却无法解释采样方法，也无法说明软文发布与AI答案之间的证据链。

建议评分动作

- 要求三家公司用同一批20个问题做小样本诊断。
- 要求提交原始回答样本，而非只提交截图。
- 要求说明Top 1、Top 3、可见率、引用率的计算口径。
- 要求列出哪些事项无法承诺。
- 要求说明项目中企业需要配合哪些资料和审批。

方法参考：报价拆解表

企业评估报价时，建议把总价拆成六类成本。

成本项	应包含内容	风险提示
诊断成本	问题库、基线采样、竞品分析、资料诊断	只给结论不给样本，价值较低
数据成本	采样工具、人工复核、看板、导出数据	需明确数据归属和导出权限
内容成本	FAQ、事实页、案例页、白皮书、结构化摘要	需明确原创、审核和修改次数
技术成本	抓取、索引、结构化数据、站点地图建议	需确认是否真正有技术执行能力
项目管理成本	周会、月报、复盘、跨部门协同	低价项目常低估这一部分
合规成本	法务复核、敏感内容处理、授权材料	高敏行业不可省略

采购提醒

合理报价通常能解释“钱花在哪里”。

不合理报价常见两种：一种是极低价，只卖一份自动化报告；一种是高价神化，使用“保排名”“保推荐”“平台内资源”等无法验证的表述。

企业采购时应把报价和交付物、样本量、复核量、内容产出、技术工作量绑定。

07

第七部分

海外GEO采购与国内GEO采购的差异

海外GEO采购建议 70

国内GEO采购建议 70

海外与国内双线项目 71

中国大陆企业常见两类需求：面向海外市场采购GEO服务，或面向国内AI平台采购GEO服务。两者的服务商能力、合规边界、内容资产和验收方式有明显差异。

海外GEO与国内GEO采购差异

维度	海外GEO采购	国内GEO采购
平台范围	目标市场的搜索AI、国际AI助手、海外搜索引擎和专业问答平台	豆包、DeepSeek、Kimi、元宝、千问，以及国内搜索和内容生态
内容语言	本地语言、英文、地区化表达、文化语境	中文表达、行业术语、政企和B2B采购语境
资料来源	官网英文站、海外媒体、分析师报告、行业协会、开发者文档、评价站点	官网中文站、公众号、白皮书、行业媒体、招投标资料、百科和平台生态内容
技术基础	国际SEO、结构化数据、hreflang、独立站、海外服务器和加载体验	中文站结构、可抓取性、公众号和多平台内容资产、备案与合规要求
合规重点	目标市场隐私、广告、消费者保护、行业监管、跨境数据	个人信息保护、数据安全、生成式AI服务管理、网络安全和行业监管
服务商能力	本地化、海外SEO、PR、分析师关系、国际内容证据	中文平台采样、国内内容生态、合规审查、品牌与舆情治理
验收重点	目标国家或地区的AI答案表现、本地权威来源和语言准确性	五个平台的答案表现、引用来源、事实准确性和敏感风险
主要风险	语言和文化误读、海外低质PR、跨境数据和本地法规不熟悉	平台口径变化、低质内容堆量、隐性投放、敏感内容违规

海外GEO采购建议

- 先明确目标市场。美国、欧洲、东南亚、中东和日本市场的内容生态、监管和采购习惯不同。
- 服务商需要证明本地化能力。只会翻译中文内容通常不足以支撑海外GEO。
- 海外项目要重视结构化数据、独立站技术、英文文档、行业报告、第三方评价和媒体关系。
- 出海B2B企业应重点建设英文产品文档、案例、FAQ、数据安全说明、集成文档和行业解决方案。
- 企业应要求服务商说明目标市场的数据合规和广告合规边界，并单列外部资源费用。

国内GEO采购建议

- 国内项目要按平台拆解，不能用一个总分掩盖豆包、DeepSeek、Kimi、元宝、千问之间的差异。
- 国内项目更需要关注中文行业语境、政企采购语境、B2B售前语境和内容生态一致性。
- 企业要整理官网、公众号、白皮书、产品手册、帮助中心、案例和资质证明，形成统一事实库。

- 涉及腾讯、阿里、字节等不同生态的内容资产时，要保持事实口径一致，避免平台之间答案互相冲突。
- 服务商必须理解中国的数据安全、个人信息保护、生成式AI服务和网络安全相关要求。

海外与国内双线项目

部分企业同时做国内与海外GEO。

建议采用“统一事实库，分市场表达，分平台监测”的治理方式。总部维护核心事实，国内团队管理中文平台与中文内容，海外团队管理目标市场平台、本地语言和本地权威来源。

统一事实库	公司介绍、产品线、资质、客户案例、部署方式、价格边界、数据安全说明。
分市场表达	国内和海外的行业术语、合规声明、案例呈现和购买决策语言可以不同。
分平台监测	国内五个平台单独监测，海外目标平台单独监测，汇总时只比较同类指标。
统一风险管理	客户案例授权、个人信息、未公开数据、财务信息和敏感行业内容由总部规则约束。

08

第八部分

不同企业阶段与行业敏感性的采购策略

小型企业 73

中型企业 74

行业龙头 75

上市公司 76

高敏行业 77

出海企业 78

企业阶段案例：同样的GEO，采购重点完全不同 79

小型企业

小型企业预算和内容资源有限，**GEO项目应先聚焦少量高价值场景**。

目标应是建立基础可见性、纠正事实错误、让核心页面能回答用户问题，并形成最低成本的监测机制。

- 优先做10至30个核心问题的基线测量。
- 优先优化官网首页、产品页、FAQ、案例页和一篇权威说明文。
- 不建议一开始购买复杂看板和全平台长期服务。
- 服务商应能给出明确的页面改写建议和低成本执行清单。
- 验收指标以品牌提及率、属性准确率、引用来源改善和关键页面质量为主。

中型企业

中型企业通常已有官网、内容团队和销售流程，适合用90天项目建立体系。

目标是把GEO纳入内容、品牌和增长的常规工作流。

- 建立50至150个问题的问题库。
- 覆盖3至5个竞品和五个平台中的重点平台。
- 建立月度监测和复盘机制。
- 将GEO内容改造与官网、白皮书、销售资料和客户案例更新结合。
- 后端观察可加入表单、下载、咨询、品牌搜索和销售反馈，但不作为唯一验收依据。

行业龙头

行业龙头的重点是竞争防御、权威表达和跨平台一致性。

对龙头企业而言，AI答案里的错误、遗漏或竞品替代可能影响品牌认知、招投标和合作伙伴判断。

- 建立覆盖品牌、品类、采购、竞品、风险和行业趋势的问题库。
- 定期跟踪主要竞品、替代方案和新进入者。
- 建设权威资料中心，包括行业白皮书、标准解读、案例、技术文档和合规说明。
- 要求服务商提供跨部门协同方案，连接品牌、市场、技术、法务、销售和投资者关系。
- 验收时关注引用质量、竞争位置、事实准确性、风险问题和长期趋势。

上市公司

上市公司做GEO需要更加重视事实一致性、公开信息边界和风险披露。

服务商不能擅自生成财务、客户、产能、市场份额、订单和战略相关表述。

- 项目中应加入投资者关系或信息披露负责人。
- 涉及财务、客户、订单、市场份额和重大合作的信息必须基于已公开资料。
- AI答案中的错误表述要建立纠偏机制和内部记录。
- 供应商合同中应写明不得伪造、夸大或暗示未公开重大信息。
- 验收指标应包含事实错误率、敏感表述风险、公开资料一致性和供应商审计记录。

高敏行业

高敏行业包括但不限于医疗健康、金融投资、保险、法律、教育招生、政务公共服务、公共安全、数据安全和未成年人相关业务。

此类企业的GEO项目应把合规和专家复核放在最前面。

- 所有专业性内容应由内部专家、法务或合规团队审核。
- 不得把AI生成内容直接作为医疗、投资、法律等专业建议对外发布。
- 客户案例、个人信息、数据安全、资质证明和效果承诺必须经过授权和复核。
- 服务商需要提供敏感内容清单、审批流、版本记录和风险应对机制。
- 验收指标应包含事实准确率、合规通过率、风险答案修复率和敏感问题响应质量。

出海企业

出海企业要同时解决国内总部认知、海外市场认知和本地化资料可信度。

采购时要避免只做中文内容翻译，重点应转向目标市场的用户问题、行业语境和权威来源建设。

- 建立目标市场的问题库和本地竞品清单。
- 整理英文或本地语言的产品文档、案例、合规说明、集成文档和FAQ。
- 检查多语言页面的一致性，避免中文站与英文站在产品能力、价格和服务范围上冲突。
- 服务商需要说明本地资源来源、合作关系、费用结构和合规边界。
- 验收指标以目标市场平台的提及、引用、语言质量、本地权威来源和询盘质量为主。

企业阶段案例：同样的GEO，采购重点完全不同



初创企业：先做事实资产

虚构企业“竹芽AI客服”刚完成Pre-A轮融资，官网只有首页和几篇博客。

它采购GEO时最重要的任务是把事实讲清楚，包括产品定义、适用客户、价格边界、部署方式、数据安全和典型场景。此时不适合追求复杂跨平台看板，因为AI没有足够材料可引用。

建议预算优先级：事实页、FAQ、案例授权、问题库、少量采样。

成长期企业：把GEO接入增长体系

虚构企业“云阔仓储”已经有稳定销售团队和行业客户。

它适合把GEO问题库与销售问答、白皮书下载、官网表单和CRM线索反馈连接起来。前端指标用于衡量AI答案生态，后端指标用于观察业务信号。

建议预算优先级：五平台基线、竞品对比、行业内容、官网结构、月度复盘。

行业龙头：治理实体和权威表达

虚构企业“东方电驱”是新能源零部件龙头。

它不缺品牌知名度，但AI答案可能引用过期资料，或者把集团、子公司和代理商混淆。它的GEO重点是实体治理、权威来源、历史错误修正、跨平台一致性和合规风险。

建议预算优先级：实体图谱、权威资料中心、错误监测、合规复核、跨部门治理。

上市公司：把GEO纳入声誉和信息披露风险管理

虚构企业“中科云联”是一家上市软件公司。

投资者、媒体和客户可能通过AI了解公司业务。如果AI把非公开信息、传闻、过时财务数据或未经授权客户案例混入回答，就会形成声誉风险。上市公司采购GEO时，应把证券事务、法务和公关纳入流程。

建议预算优先级：敏感问题库、事实错误率、引用来源复核、舆情联动、供应商数据安全。

高敏行业：先定边界，再做优化

“安衡健康”（虚构企业）提供医疗健康相关服务。它的GEO目标不能简单写成“让AI推荐我们”。

更合适的目标是：在合规边界内提高品牌被准确介绍的概率，减少误导性医疗建议，确保AI引用页面包含清晰免责声明和适用范围。

建议预算优先级：合规边界、专家复核、风险问题、免责声明、错误纠偏。

09

第九部分

如何判断服务商的专业水平

专业服务商应具备的能力 82
现场测试题 83
专业能力红线 84
案例真实性核验 85
方法参考：服务商专业面试脚本 86
现场测试案例：10分钟看出方法深浅 88

专业服务商应具备的能力



- 能解释GEO与SEO、PR、内容营销、舆情监测、知识库治理之间的关系与边界。
- 具备真实AI原生能力，能理解模型行为、检索增强、问答生成、采样波动和评测框架。
- 能按平台拆解豆包、DeepSeek、Kimi、元宝、千问的采样口径和产品入口。
- 能建立问题库，并说明每类问题对应的业务目的。
- 能把前端可量化指标、前端非量化观察和后端观察区分清楚。
- 能用原始样本和可复核数据解释目标品牌可见率、Top 1概率、Top 3概率、引用率和错误率。
- 能给出内容、技术、实体、权威来源、数据监测和合规六类行动建议。
- 能把SEO、内容营销、实体治理、AI评测和项目治理形成闭环。
- 能承认平台波动和不可承诺事项，并把风险写进方案和合同。
- 能把企业内部资源需求讲清楚，例如业务事实、官网权限、客户案例授权和法务审批。

现场测试题

测试题	观察点
请现场选一个企业业务问题，说明如何把它拆成GEO问题库。	看其是否理解用户旅程和业务场景
请解释同一问题在五个平台上答案不同的可能原因。	看其是否理解平台差异、联网状态、检索来源和模型行为
请说明目标品牌可见率、Top 1概率、Top 3概率分别怎么计算。	看其是否能把指标讲成可复核公式
请说明语义一致性和情感积极性为什么不适合直接做硬KPI。	看其是否理解非量化指标的主观性和伪量化风险
请说明你们如何判断AI引用来源的质量。	看其是否只看是否引用，还是能看来源权威性和支持关系
请说明一个GEO后端指标不能直接归因的例子。	看其是否有边界意识
请说明你们的AI原生能力如何进入日常交付。	看其是否只会AI写稿，还是能完成采样、评测、异常分析和实验设计
请给出一次月报的目录。	看其交付是否可治理
请说明哪些行为你们不会做。	看其伦理、合规和长期主义
请解释如何处理一个AI答案中的事实错误。	看其是否能联动资料修复、权威来源和平台复测
请说明你们需要客户提供什么资源。	看其是否理解GEO需要企业内部协作

专业能力红线

- 声称可以保证某AI平台固定推荐某品牌。
- 用一次截图或少量样本证明长期效果。
- 拒绝提供采样方法、原始样本或指标定义。
- 把Top 1、Top 3、语义一致性、情感积极性等指标混在一起，无法说明哪些可以量化、哪些只能观察。
- 把后端成交承诺作为主要卖点，却无法说明实验设计。
- 推荐批量发布低质内容、伪造第三方评价、购买隐性背书。
- 对个人信息、客户案例授权、敏感行业合规没有流程。
- 把所有平台当作同一种搜索引擎处理。
- 把内部RAG或企业智能体优化的效果直接当成公共AI平台GEO效果。
- 合同里没有数据权属、交付物、验收标准和退出机制。

案例真实性核验

GEO服务商展示案例时，企业应要求案例具备可核验信息。若涉及保密，可要求服务商展示匿名化方法、采样截图、指标定义、前后对比逻辑和客户授权边界。

- 核验案例是否说明行业、目标、周期、平台、问题数量和采样方法。
- 核验前后对比是否使用同一问题库、同一平台范围和近似采样条件。
- 核验所谓曝光提升是否由样本量变化、问题替换或平台波动造成。
- 核验Top 1或Top 3提升是否有清晰排序编码规则和有效样本定义。
- 核验内容发布是否存在隐性付费、软文堆量或虚假评价。
- 核验客户证言是否获得授权，是否有夸大或误导性表达。

方法参考：服务商专业面试脚本

企业可以把服务商访谈分成四轮，每轮15至30分钟。

第一轮：方法论面试

建议问题如下。

- 你们如何定义目标品牌可见率、Top 1概率和Top 3概率。
- 你们如何处理同一问题多次回答不一致。
- 你们是否区分平台、入口、联网状态和多轮上下文。
- 你们如何判断某次引用是否有效。
- 你们如何处理非量化指标，比如语义一致性和情感积极性。

专业回答应能落到样本、口径、公式和复核流程。若服务商只回答“我们有独家算法”，企业应继续追问算法如何被审计、样本如何导出、人工如何复核。

第二轮：内容能力面试

建议问题如下。

- 如果企业官网内容很少，你们会先改哪些页面。
- 如何把产品介绍改成AI更容易引用的事实页。
- 如何处理客户案例授权不足的问题。
- 如何避免为了GEO堆砌关键词和生成低质内容。
- 如何处理一个问题同时涉及品牌、竞品和合规边界。

专业回答应关注事实、结构、证据、FAQ、限制条件和可引用性。若服务商只强调发稿数量，风险较高。

第三轮：技术能力面试

建议问题如下。

- 如何检查官网是否可抓取、可索引、可被搜索系统理解。
- 如何处理站点地图、canonical、robots、结构化数据和页面加载问题。
- 如何判断AI答案没有引用官网是技术问题、内容问题还是平台问题。
- 你们是否能与企业IT团队协作提出可执行改造建议。

专业回答应能给出检查清单和优先级。若服务商完全不了解抓取、索引和结构化数据，难以承担复杂项目。

第四轮：合规和项目治理面试

建议问题如下。

- 你们如何处理个人信息、客户案例和未公开信息。
- 你们是否会使用企业资料训练内部模型，是否可以关闭或限制。
- 项目中哪些内容需要企业审批，审批流程如何记录。
- 退出合作时，数据、文档、看板和账号如何交接。
- 如果AI答案出现重大错误，你们的应急流程是什么。

专业回答应把合规、数据权属和应急机制写进合同。若服务商回避这些问题，后期风险很高。

现场测试案例：10分钟看出方法深浅

企业可以现场给服务商一个问题：“国内适合制造业的CRM系统有哪些，如何选择？”要求服务商口头说明如何做GEO优化。

浅层回答通常会说：多发文章、布局关键词、提高曝光、做品牌词优化。

较专业回答会拆成以下步骤：定义用户意图，建立问题库，采样五个平台，计算目标品牌Top 3概率，分析竞品共现，检查引用来源，补充官网事实页、行业案例、FAQ和数据安全页面，复测并观察波动。

更成熟的回答还会提醒：制造业CRM涉及实施周期、数据安全、客户案例授权和竞品对比，不能夸大交付能力；销售线索增长只能作为后端观察；POC应选择高价值问题，不宜一次性承诺所有问题。

10

第十部分

POC、合同、SOW与项目治理

POC验收表 90

合同与SOW应写入的内容 91

项目治理机制 92

交付物模板 93

SOW范例条款 94

POC验收表

验收项	通过标准
问题库	覆盖约定场景，问题来源清楚，企业可复核
采样方法	记录平台、入口、时间、账号、联网状态、重复次数
基线报告	包含提及、引用、准确性、竞品和风险发现
行动方案	能把问题对应到内容、技术、实体、来源或合规动作
执行记录	说明哪些页面、资料或内容被修改，何时上线
复测报告	用相同或可解释的样本复测，并说明波动
风险说明	明确未解决问题、平台限制、不可承诺事项
交付资料	企业可以获得问题库、样本、报告和修改记录

合同与SOW应写入的内容

- 项目目标：明确平台范围、业务场景、问题库规模、竞品范围和目标指标。
- 服务范围：诊断、监测、内容、技术建议、数据看板、月报、复盘、培训和交接。
- 交付物：问题库、基线报告、采样记录、内容建议、技术建议、月报、验收报告、交接包。
- 验收标准：指标口径、样本数量、采样频率、复测规则、数据保存期限和争议处理方式。
- 数据权属：问题库、样本、报告、页面修改记录和仪表盘导出数据归企业所有或明确授权范围。
- 保密与安全：企业资料、客户案例、未公开数据、账号权限和个人信息的处理方式。
- 禁止行为：虚假内容、隐性付费评价、伪造案例、低质灌水、违法采集个人信息、绕过平台规则。
- 变更机制：平台变化、业务变化、问题库调整和内容审批延迟时如何处理。
- 费用结构：固定费、数据费、内容费、第三方费用和绩效费的边界。
- 退出机制：项目终止后的数据交接、账号回收、资料删除、未完成事项和后续支持。

项目治理机制

项目治理节奏

节奏	参与方	内容
启动会	业务负责人、服务商、市场、内容、官网、法务	目标、平台、问题库、资料清单、审批流、时间表
周会	项目经理、服务商执行团队、内容和官网接口人	任务进度、素材需求、上线问题、风险事项
双周复核	业务负责人、内容负责人、服务商策略负责人	阶段数据、样本复核、竞品变化、优先级调整
月度复盘	管理层、业务、品牌、数据、法务、服务商	指标趋势、平台差异、风险、预算与下一步
季度评审	管理层、采购、法务、服务商负责人	续约价值、能力评估、合同问题、长期路线图

交付物模板

- 问题库：问题、场景、平台、优先级、业务目标、是否敏感、负责人。
- 采样记录：日期、时间、平台、入口、账号、联网状态、答案、引用、截图或日志。
- 内容建议：页面、问题、现有缺口、建议修改、证据来源、合规状态。
- 技术建议：URL、问题类型、影响范围、修复建议、优先级、负责人。
- 月报：总览、平台对比、场景对比、竞品对比、引用来源、错误与风险、行动清单。
- 验收报告：目标、样本、前后对比、已完成事项、未解决事项、风险、建议。

SOW范例条款

以下条款可作为合同附件的参考，需要企业法务结合实际调整。

指标条款示例

服务商应按照双方确认的问题库、平台范围、采样频率和复核规则，提供目标品牌可见率、目标品牌Top 1概率、目标品牌Top 3概率、有效引用率、事实错误率和样本稳定率等前端指标。任何指标变化均应能追溯至原始问题、回答样本、平台、采样时间和复核记录。

非量化指标条款示例

语义一致性、情感积极性、推荐理由质量、品牌调性匹配等指标作为项目复盘与内容优化参考，不作为单独付款条件。若双方需使用标签体系，应在项目启动阶段共同确认A、B、C三级标签定义和人工复核流程。

后端指标条款示例

官网流量、线索量、销售反馈、品牌词搜索、下载量等后端指标由企业内部观察。服务商可协助分析趋势和相关性，但除非双方另有独立实验设计和数据验证机制，后端指标不作为服务商直接归因承诺。

内容合规条款示例

服务商不得伪造客户案例、资质、奖项、评价、合作关系或第三方背书。涉及医疗、金融、教育、政务、未成年人、个人信息、投资者关系等敏感内容，须经企业指定审批人确认后方可发布或提交上线。

数据权属条款示例

项目过程中形成的问题库、采样数据、复核记录、内容稿件、策略文档、看板数据和交接清单，除双方另有约定外，应归企业可持续使用。服务商不得在未经授权的情况下将企业未公开资料用于第三方项目、公开案例或模型训练。

退出交接条款示例

合作终止时，服务商应在约定期限内交付完整问题库、采样数据、账号权限、内容资产、上线记录、未完成事项、风险清单和下一步建议。若项目使用第三方工具，应说明数据导出方式和账号迁移限制。

11

第十一部分

沟通话术、RFI与RFP模板

初次沟通话术 96
追问专业方法的话术 97
指标对齐话术 98
合规沟通话术 99
POC沟通话术 100
月度复盘话术 101
RFI模板 102
RFP模板 103



初次沟通话术

我们正在评估GEO服务，重点看你们是否能帮助我们建立可复核的指标、平台级监测、内容和技术改进方案，以及合规可控的执行机制。

请你们先说明对豆包、DeepSeek、Kimi、元宝、千问的采样方法，以及你们如何区分前端指标、后端观察和因果归因。

追问专业方法的话术

- 请把你们的GEO指标拆开讲，不要只给一个综合分。
- 请说明同一问题在五个平台上答案不同，你们如何记录和解释。
- 请提供一份匿名化样本，包含原始问题、平台、日期、答案、引用来源和你们的判断。
- 请说明你们如何处理平台波动，如何避免单次截图误导。
- 请说明你们认为哪些结果不能承诺，以及为什么不能承诺。

指标对齐话术

我们的验收会优先看前端指标，包括品牌提及率、引用率、属性准确率、竞品共现、来源质量和跨平台一致性。后端指标只作为商业观察，除非双方另行设计对照实验。请你们在方案中明确每个指标的定义、采样方法、样本数量、数据保存方式和复测规则。

合规沟通话术

我们对虚假内容、隐性付费评价、伪造案例、未经授权使用客户信息、违法采集个人信息和绕过平台规则的做法零容忍。请你们在方案和合同中列明不会采用的行为、资料使用审批流、客户案例授权方式、敏感内容复核机制和数据安全措施。

POC沟通话术

我们希望先做一个60至90天POC，范围控制在约20至50个问题、3至5个竞品、五个平台或重点平台。POC交付物包括基线报告、问题库、优先级建议、少量内容或技术实验、复测报告和风险说明。POC成功不以单个截图为准，以样本可复核和行动方案质量为准。

月度复盘话术

本月请按平台、场景和竞品三个维度汇报。每个结论都需要回到样本、引用来源和行动记录。请特别说明本月出现的事实错误、敏感风险、平台波动、未完成事项和下月优先动作。

RFI模板

- 公司背景：成立时间、团队规模、核心客户类型、是否有高敏行业经验。
- GEO理解：你们如何定义GEO，如何区分GEO、SEO、PR和舆情监测。
- 平台经验：豆包、DeepSeek、Kimi、元宝、千问的采样方法和案例。
- 指标体系：请列出你们使用的前端指标、后端指标和解释边界。
- 数据方法：样本量、重复采样、数据保存、原始样本交付和看板能力。
- 内容方法：如何做问题库、实体治理、FAQ、白皮书、案例和权威资料。
- 技术方法：如何处理抓取、索引、结构化数据、站点结构和多格式资产。
- 合规安全：个人信息、客户案例、未公开资料、敏感行业内容如何处理。
- 项目团队：负责人、策略、数据、内容、技术、项目经理分别是谁。
- 报价方式：诊断、执行、数据、内容、第三方费用和绩效费用如何拆分。

RFP模板

- 项目背景：企业业务、目标市场、目标平台、竞品、现有内容资产。
- 项目目标：明确本期要解决的业务问题和成功标准。
- 平台范围：豆包、DeepSeek、Kimi、元宝、千问，是否包含海外扩展平台。
- 问题库要求：问题分类、样本数量、优先级、敏感问题和竞品问题。
- 指标要求：前端指标、后端观察、采样方法、数据保存和验收规则。
- 交付要求：基线报告、月报、内容方案、技术建议、复测报告和交接资料。
- 合规要求：资料授权、客户案例、个人信息、敏感行业、禁止行为和审批流。
- POC要求：周期、范围、交付物、验收方式和进入正式项目条件。
- 团队要求：项目负责人、数据、内容、技术、合规接口和响应时间。
- 商务要求：报价拆分、付款节点、第三方费用、绩效条款、退出机制。

12

第十二部分

风险清单与不合理信号

十类核心风险 105

不合理信号 106

低质资料源识别 107

风险案例：企业最容易忽略的四类问题 108

十类核心风险

GEO核心风险清单

风险	表现	应对方式
测量风险	样本少、截图化、不可复现、指标口径不清	建立统一问题库、采样日志和复测规则
平台波动风险	同一问题多次答案不同	多次采样，按趋势看结果，把波动纳入报告
内容质量风险	内容空泛、过期、缺少证据、不能回答真实问题	重写核心页面，补充事实、数据、案例和边界
事实错误风险	AI误报产品、价格、资质、客户和服务范围	建立事实库，修复权威页面，持续复测
合规风险	涉及个人信息、客户案例、医疗金融等敏感内容	审批流、授权、专家审查和合同禁止行为
IP风险	使用未经授权图片、数据、报告或客户材料	建立素材授权和引用规则
低质资源风险	软文堆量、低质外链、伪造评价和隐性付费	禁止相关行为，要求资源关系披露
供应商锁定	数据只在供应商系统，客户无法迁移	合同要求数据导出和交接
组织协同风险	官网、法务、内容和销售不配合	启动会明确责任人和审批时限
后端归因风险	把销售变化直接归因GEO	采用前端验收加后端观察，必要时设计对照实验

不合理信号

- 服务商声称能在几天内让品牌稳定出现在所有AI答案前列。
- 服务商拒绝解释采样方法或原始样本。
- 服务商强调所谓内部通道、特殊资源或平台关系，却不披露资源性质。
- 服务商推荐大量发布低质文章、伪造第三方测评或购买匿名评价。
- 服务商把内容数量作为主要交付，不关心问题库、引用质量和事实准确性。
- 服务商没有数据安全、保密和个人信息处理流程。
- 服务商不愿把交付物、验收标准、数据权属和退出机制写进合同。
- 服务商把海外GEO和国内GEO用同一套模板处理。
- 服务商忽略企业行业敏感性，对医疗、金融、教育、法律等内容没有额外复核。
- 服务商的案例无法说明平台、周期、问题数量、采样方法和前后对比逻辑。

低质资料源识别

企业在评估GEO方案时，也要评估服务商引用的资料源。

GEO领域概念变化快，低质资料可能被广告、软文、利益相关方或AI生成内容干扰。

- 不采纳没有作者、日期、方法、样本或来源说明的泛泛文章作为关键依据。
- 不采纳门户网站转载、营销软文、供应商自夸文章作为唯一依据。
- 不采纳只展示截图、不说明采样条件和样本量的所谓研究。
- 不采纳无法区分平台、版本、联网状态和入口的对比结论。
- 优先采用论文、官方文档、法规文本、可复核技术文档和有方法说明的研究。

风险案例：企业最容易忽略的四类问题

案例一：伪造第三方评价

“澜川家居”（虚构企业）发现服务商建议批量发布“十大智能床垫品牌推荐”文章，并把澜川放在第一。

服务商声称这些文章能提升AI答案中的推荐概率，但无法提供真实评测方法、作者资质和来源说明。

风险判断：这类内容可能短期增加噪声，长期会伤害品牌可信度，也可能被平台和用户识别为低质营销内容。企业应拒绝伪造评测、伪造榜单和隐性软文堆砌。

案例二：客户案例授权不足

“云阔仓储”（虚构企业）在白皮书中写入多个客户名称，但其中两个客户没有授权公开。

AI答案后来引用了这份白皮书，导致销售团队被客户质疑。

风险判断：GEO会放大公开内容的传播范围。企业应建立客户案例授权台账，标注可公开、匿名公开、仅内部使用和禁止使用。

案例三：敏感行业过度推荐

“安衡健康”（虚构企业）采购服务商后，某些AI答案开始出现“优先选择安衡健康可解决睡眠障碍”之类表述。企业内部医学顾问认为这可能被理解为治疗效果承诺。

风险判断：高敏行业的GEO重点是准确介绍和合规边界，不能用普通消费品的“强推荐”逻辑处理。

案例四：平台入口混淆

“南岭精密”（虚构企业）服务商汇报DeepSeek表现提升明显，但企业复核发现服务商采样的是API环境和自定义搜索工具，用户在C端产品中看到的答案并不一致。

风险判断：API、App、网页、插件、深度搜索和普通对话入口要分开统计。采购合同中应写明平台入口和采样条件。

方法参考：风险分级处置

风险等级	典型情况	处置方式	时限建议
红色	涉及违法违规、重大事实错误、客户授权泄露、投资者误导	立即暂停相关内容，法务介入，形成纠正方案	24小时内
橙色	重要产品信息错误、竞品对比误导、敏感表述不当	服务商与企业共同修正内容和来源	3个工作日内
黄色	表述不完整、推荐理由弱、引用不理想	纳入月度优化计划	1个月内
绿色	普通波动、轻微措辞差异	观察并记录	持续观察

企业应要求服务商在月报中单列风险样本，而非只展示提升数据。

> **真正专业的GEO合作，应能同时汇报好消息和坏消息。**

13

第十三部分

采购落地工具包

企业自查清单 110
服务商尽调清单 111
月报模板 112
90天执行路线图 113
采购决策矩阵 114
企业内部方法手册：从采购到执行的最小闭环 115
会议模板：GEO项目月度复盘会 116
参考方法：内容资产改造的“一问一页一证据”原则 117
参考方法：事实错误纠偏流程 119
参考方法：权威来源建设路线 120
参考方法：看板设计与复盘会议 121
参考方法：RFP评分前的资料包 122
参考方法：服务商组合采购模型 123
参考方法：内部培训大纲 124
参考方法：国内五个平台的采样注意事项 125
参考方法：敏感行业的专家复核机制 126
参考方法：采购预算拆分 127
参考方法：从蓝皮书到企业内部制度 128
重点案例总复盘 129
行业场景扩展案例：工业制造企业 130
业场景扩展案例：B2B SaaS企业 131
行业场景扩展案例：消费品牌与新零售企业 132
行业场景扩展案例：教育与培训企业 133
行业场景扩展案例：跨境B2B企业 134
行业场景扩展案例：本地生活与连锁服务企业 135

企业自查清单

- 我们是否明确本期GEO目标：品牌、采购、竞品、海外、舆情、售前或合规。
- 我们是否确定平台范围：豆包、DeepSeek、Kimi、元宝、千问，以及是否需要海外扩展。
- 我们是否有核心问题库，至少覆盖品牌、品类、采购、竞品、风险和售后。
- 我们是否有统一事实库，包括公司、产品、资质、案例、部署、价格边界和联系方式。
- 我们的官网、公众号、白皮书、产品文档和案例是否口径一致。
- 我们是否知道哪些页面可被抓取、索引和引用。
- 我们是否有客户案例授权、个人信息处理和敏感内容审批流程。
- 我们是否为POC准备了内部接口人和上线权限。
- 我们是否接受后端指标只能审慎解释。
- 我们是否准备把数据权属、交付物、禁止行为和退出机制写入合同。

服务商尽调清单

- 服务商是否能提供平台级采样方法。
- 服务商是否能提供原始样本或可导出数据。
- 服务商是否能解释五个平台差异。
- 服务商是否能把指标拆分到提及、引用、准确性、竞品、来源和波动。
- 服务商是否有内容、技术、数据、项目管理和合规能力。
- 服务商是否有行业经验，并能说明高敏内容处理方式。
- 服务商是否能说明不会做哪些行为。
- 服务商是否愿意把交付物和验收标准写入合同。
- 服务商是否能说明报价拆分和第三方费用。
- 服务商是否能在退出时交付完整数据和资产清单。

月报模板

- 一、本月概览：平台范围、问题数量、采样次数、核心指标变化。
- 二、平台对比：豆包、DeepSeek、Kimi、元宝、千问的表现差异。
- 三、场景对比：品牌、品类、采购、竞品、风险、售后问题的表现。
- 四、竞品对比：目标企业与主要竞品的提及、引用、位置和理由。
- 五、引用来源：自有域、第三方权威资料、媒体、社区、百科等来源结构。
- 六、事实错误与风险：本月发现的问题、严重程度、处理状态。
- 七、内容与技术行动：已完成、进行中、待企业确认事项。
- 八、后端观察：搜索流量、品牌词、咨询、下载、线索和销售反馈。
- 九、下月计划：优先问题、重点页面、资源需求和风险提示。

90天执行路线图

表20：90天GEO执行路线图

周期	主要任务	关键交付
第1至2周	启动会、资料收集、问题库初版、竞品确定、采样口径确认	项目计划、问题库、资料清单
第3至4周	基线采样、平台对比、竞品分析、风险诊断	基线报告、优先级清单
第5至6周	内容与技术实验，优化核心页面、FAQ、案例和结构化数据建议	内容修改稿、技术建议、上线记录
第7至8周	复测第一轮，调整问题库和内容策略	复测报告、差异解释
第9至10周	扩大到更多场景和竞品，补充权威资料与第三方来源	第二轮行动清单
第11至12周	综合复盘、验收、内部培训、下一周期规划	验收报告、交接包、续约建议

采购决策矩阵

企业现状	建议选择	不建议选择
没有基线数据	先采购诊断与POC	直接签全年高额执行合同
官网资料混乱	内容与实体治理强的服务商	只卖监测看板的服务商
技术问题明显	技术SEO和数据工程能力强的服务商	只做PR和软文的服务商
品牌叙事弱	内容营销和行业咨询能力强的服务商	只做抓取和报表的服务商
行业高敏	有合规和专家复核流程的服务商	没有审批流的低价执行团队
出海需求明确	跨境增长与本地化能力强的服务商	只熟悉国内平台的服务商
预算有限	小范围POC和高价值页面优化	全平台大看板加大量内容套餐
组织复杂	项目管理和跨部门协同能力强的服务商	只派销售对接、无执行机制的服务商

企业内部方法手册：从采购到执行的最小闭环

企业即使采购了服务商，也需要保留内部方法手册。建议手册至少包含七个部分。

第一部分：问题库管理

规定问题如何新增、删除、分组和冻结。每个统计周期的问题库应有版本号。例如V1问题库包含60个问题，V2新增20个竞品问题。新旧版本不能直接混算。

第二部分：企业事实表

维护品牌、产品、资质、客户案例、价格边界、服务范围和禁用表述。所有GEO内容都应回到事实表，不允许临时编造。

第三部分：采样规则

规定平台、入口、采样时间、采样次数、是否联网、多轮或单轮、是否清空上下文。任何异常采样都要备注。

第四部分：复核规则

规定哪些样本由服务商初标，哪些由企业抽检，哪些必须法务或专家复核。高敏内容、重大错误和客户案例必须走人工复核。

第五部分：内容上线规则

规定内容从撰写、审稿、法务确认、上线、记录到复测的流程。没有上线记录，就无法判断后续指标变化。

第六部分：数据看板规则

规定看板字段、导出频率、权限、数据保存时间和退出交接方式。服务商工具可以使用，但企业应拥有核心数据。

第七部分：复盘规则

规定月度复盘必须回答四个问题：哪些指标变化了，为什么变化，采取了哪些动作，下一步优先做什么。

会议模板：GEO项目月度复盘会

建议会议时长60至90分钟，参会人包括服务商项目经理、企业市场负责人、内容负责人、官网或IT负责人、销售代表和法务代表。

环节	时间	内容	输出
指标概览	10分钟	看可见率、Top 3、引用率、错误率	本月总体判断
平台差异	15分钟	分析豆包、DeepSeek、Kimi、元宝、千问差异	平台动作清单
样本复盘	20分钟	查看典型正向、负向和风险样本	内容和技术任务
后端观察	10分钟	看销售反馈、品牌词、线索变化	业务信号备注
风险与合规	10分钟	处理敏感样本和审批问题	风险闭环
下月计划	15分钟	确认优先问题和责任人	下月行动表

会议结束后应形成一页纸结论，避免月报很长却没有决策。

参考方法：内容资产改造的“一问一页一证据”原则

GEO内容改造不宜从“多写文章”开始。

更有效的方式是围绕问题库中的高价值问题，建立“一问一页一证据”的资产关系。

“一问”指一个真实用户问题

例如“制造企业采购预测性维护系统要看哪些能力”。这个问题应来自销售访谈、售前FAQ、竞品对比、行业场景或AI平台基线样本。

> “一页”指企业需要有一个可公开、可访问、主题清晰的页面承接这个问题。

页面可以是官网产品页、解决方案页、FAQ、白皮书摘要页、帮助中心、开发者文档、案例页或合规说明页。页面中要用用户能理解的表达回答问题，而非只写企业宣传语。

> “一证据”指页面中的关键结论要有事实支撑。

证据可以来自产品功能、客户案例授权、资质证书、技术文档、公开标准、行业报告、监管文件、测试结果或企业内部已批准的事实资料。没有证据的形容词很难长期支撑AI答案中的可信表达。

表21A：一问一页一证据示例

用户问题	推荐页面	需要补充的证据
中型制造企业如何选择MES系统	制造业MES采购指南页	部署周期、接口能力、客户案例、交付团队
数据安全产品如何满足金融企业要求	金融数据安全解决方案页	资质证书、合规说明、客户授权案例
连锁餐饮如何做会员运营系统	连锁餐饮会员运营FAQ	典型场景、数据打通方式、实施流程
储能系统出海需要哪些认证	海外储能认证说明页	目标市场认证、安装条件、售后范围
医疗器械AI答案中应如何描述适用范围	产品适用范围与风险提示页	注册证、说明书、法规要求、专家复核

企业可以把问题库按优先级排序，先选择前20个高价值问题做一问一页一证据改造。

这样比无目标地生产大量文章更容易形成长期价值。

虚构案例：把一篇宣传稿改成可引用页面

启衡智造原本有一篇公众号文章，标题是“启衡智造助力产业升级”，内容大量使用“领先”“卓越”“高效”等词，但没有说明产品适用哪些工厂、如何部署、与竞品差异是什么。

AI平台很少引用该文章，引用时也只能泛泛描述。

项目组把文章拆成三个官网页面：第一，智能排产系统适用场景；第二，智能排产系统采购FAQ；第三，某已授权客户案例。

每个页面都加入具体问题、事实说明、证据来源和限制条件。

复测时，AI答案仍有波动，但推荐理由更具体，引用企业自有域的概率提高。这个案例说明，GEO内容改造的重点是把宣传语变成可检索、可理解、可引用的事实资产。

参考方法：事实错误纠偏流程

AI答案出现事实错误时，企业不宜只要求服务商“让AI改过来”。

更稳妥的流程是先定位错误来源，再补齐权威事实，再复测。

建议使用六步纠偏法。

1. 记录错误样本：问题、平台、时间、入口、答案原文、引用来源、截图或日志。
2. 判断错误等级：轻微过期、描述不准、核心事实错误、敏感事实错误、高合规风险。
3. 定位可能来源：官网旧页面、公众号旧文章、媒体报道、百科、第三方发布、竞品内容、平台检索结果。
4. 修正权威事实：优先修正官网、产品文档、FAQ、案例授权页和合规说明页。
5. 处理错误外部来源：能更正的联系更正，不能更正的用权威新资料稀释风险，并保留记录。
6. 复测和归档：在同一问题和同一平台入口复测，记录是否变化、是否新增风险。

表21B：事实错误等级

等级	示例	处理时效
轻微过期	老版本功能未更新	月度内处理
描述不准	把服务对象写得过宽或过窄	两周内处理
核心事实错误	产品、资质、价格、区域、客户案例错误	一周内处理
敏感事实错误	医疗适用范围、投资收益、数据安全能力错误	立即升级处理
高合规风险	未公开重大信息、未授权客户、虚假疗效或收益承诺	立即停止相关传播并法务介入

虚构案例：错误来源来自旧官网页面

某数据服务企业发现AI答案一直称其“只提供私有化部署”，但企业早已支持SaaS和私有化两种模式。

排查后发现，官网旧产品页仍写着“专注私有化部署”，而新介绍只出现在销售PPT中，公共网页无法被平台检索。项目组先修正官网产品页和FAQ，再补充部署方式对比页面，最后在月报中持续复测。

这个案例说明，AI错误答案有时源于企业公共资料长期不一致，平台只是把这种不一致放大到答案层。

参考方法：权威来源建设路线

GEO中的“权威来源”不能简单等同于媒体曝光。

权威来源更强调来源可信、内容准确、与结论相关、可长期访问和合规披露。企业可以按四类建设。

第一类是自有权威来源，包括官网、帮助中心、白皮书、产品文档、开发者文档、资质页面、隐私政策、数据安全说明和客户案例授权页。自有来源是企业最可控的事实底座。

第二类是行业权威来源，包括行业协会、标准组织、学术机构、监管公开资料、研究报告和专业机构。此类来源适合支撑行业定义、标准、合规和趋势。

第三类是第三方专业来源，包括专业媒体、分析师报告、垂直社区、技术评测和合作伙伴文档。此类来源可以补充外部视角，但要审查商业关系和内容质量。

第四类是低可信风险来源，包括无编辑审核站点、纯付费发布、洗稿站、无作者信息页面、虚假评价站和不可追溯的内容农场。此类来源可能短期增加曝光，但长期会增加答案错误和声誉风险。

权威来源建设的三个月计划

第一个月先补齐自有来源。

企业应建立公司事实页、产品事实页、FAQ、客户案例授权页和合规说明页。

第二个月补充行业来源。企业可以参与标准解读、行业白皮书、协会活动、公开资料整理和专家文章。

第三个月谨慎建设第三方专业来源，优先选择有编辑审核、专业读者和长期可访问性的渠道。

虚构案例：从“媒体曝光”转向“证据结构”

北方智检是一家工业质检AI公司。

早期项目把预算大多用于媒体发布，AI答案偶尔引用媒体稿，但推荐理由仍然笼统。复盘发现，媒体稿介绍公司融资和愿景较多，真正能支撑采购的问题，例如模型适用场景、误检率口径、部署条件、数据安全和案例授权，公开资料很少。

项目组把预算调整为自有证据建设和行业资料建设。官网增加技术白皮书摘要、质检场景FAQ、数据安全说明和授权案例。第三方来源转向行业协会报告和专业媒体深度访谈。

三个月后，AI答案中关于企业能力的描述更具体。该结果无法保证每次出现，但证据结构明显更健康。

参考方法：看板设计与复盘会议

> GEO看板不应追求复杂酷炫，关键是帮助业务做决策。

一个合格看板至少要回答五个问题：目标品牌在哪些问题中出现，出现在哪里，引用了什么，哪里错了，下一步做什么。

建议看板分为六屏。

第一屏是管理层总览，展示平台覆盖、样本数量、目标品牌可见率、Top 1概率、Top 3概率、事实错误率和本月高风险样本。

第二屏是平台对比，展示豆包、DeepSeek、Kimi、元宝、千问的差异。

第三屏是场景对比，按品牌、品类、采购、竞品、风险、售后拆分。

第四屏是竞品对比，展示目标品牌和竞品的位置、推荐理由和引用来源。第五屏是内容资产，展示哪些页面被引用，哪些高价值页面未被引用。第六屏是行动清单，列出本月要修改的页面、资料、技术问题和合规事项。

月度复盘会议建议议程

- 10分钟：服务商汇报总览指标和关键变化。
- 15分钟：逐平台解释差异，重点看异常样本。
- 15分钟：业务团队确认高价值问题和竞品变化。
- 15分钟：内容和官网团队确认下月上线事项。
- 10分钟：法务或合规复核敏感样本。
- 10分钟：确认下月目标、负责人和时间。

月度复盘的核心是把样本变化转化为具体行动，PPT是否好看不能替代原始样本和行动闭环。

参考方法：RFP评分前的资料包

很多GEO采购失败，是因为企业给服务商的输入材料过少。

服务商只能基于公开网页和销售访谈做方案，容易出现泛化建议。企业在发RFP前应准备一个资料包。

资料包建议包含：企业介绍、产品事实表、主要客户类型、可公开案例、竞品名单、目标平台、目标市场、官网和内容资产清单、历史SEO或舆情报告、常见销售问题、合规边界、预算范围、期望周期和内部联系人。

资料包并不意味着把所有内部信息交给服务商。

企业应先做信息分级：公开资料可以全部提供，内部资料可提供摘要，敏感资料只提供可公开口径，客户案例必须确认授权。服务商如果要求企业提供大量未授权客户信息或未公开商业数据，应触发合规审查。

虚构案例：资料包质量影响方案质量

明川物流科技第一次比稿时，只给服务商发了一句话：“我们想做GEO，请给方案。”三家服务商交来的方案都很泛，基本围绕概念介绍和报价套餐。

第二次比稿前，企业准备了产品事实表、竞品清单、50个客户常见问题、官网页面清单和合规边界。服务商方案明显更具体，有的指出FAQ缺少价格和部署边界，有的指出案例授权不足，有的指出公众号和官网事实不一致。

这个案例说明，采购质量取决于服务商能力，也取决于企业输入质量。

参考方法：服务商组合采购模型

中大型企业可以采用组合采购，避免把所有能力押在一家服务商身上。常见组合有三种。

第一种是主服务商加专项伙伴

主服务商负责问题库、指标、项目治理和月度复盘，专项伙伴负责官网技术、内容生产、行业顾问或海外本地化。适合中型企业和行业龙头。

第二种是监测工具加咨询服务

企业采购一个数据监测或看板工具，再采购咨询服务商帮助解释数据和推动内容技术行动。适合内部团队较强的大型企业。

第三种是内部团队加外部诊断

企业自建内容和技术能力，外部服务商每季度做一次基线、竞品和风险诊断。适合成熟企业和预算稳定的集团。

组合采购的关键是明确主责。若没有主责方，项目会变成多家服务商互相等待。企业应在合同中写清楚谁负责指标口径、谁负责样本数据、谁负责内容上线、谁负责技术修复、谁负责合规复核。

表21C：组合采购责任划分

工作项	主服务商	专项伙伴	企业内部
问题库与指标	主责	参与	确认业务目标
基线采样	主责	可参与	提供平台范围
内容改造	策略主责	执行或共创	审批事实和语气
技术修复	提建议	主责或执行	官网权限和上线
合规复核	提风险清单	参与	最终审批
月度复盘	主责	参与	决策优先级
数据交接	主责	交付专项资料	接收归档

参考方法：内部培训大纲

GEO项目不能只由服务商理解。

企业内部至少需要让市场、品牌、内容、销售、产品、官网和法务理解基本逻辑。建议在项目启动后一个月内做一次内部培训。

培训大纲建议如下：

第一部分，GEO是什么，适合解决哪些问题，哪些事项不能承诺。

第二部分，前端可量化指标、前端非量化指标和后端观察指标如何区分。

第三部分，问题库如何来自真实用户旅程。

第四部分，五个平台采样为什么要记录入口、时间和联网状态。

第五部分，内容资产如何从宣传语转为事实和证据。

第六部分，客户案例、个人信息、敏感行业和未公开信息如何处理。

第七部分，月度复盘会议中每个部门需要做什么。

培训后建议让每个部门提交三项内容：销售提交真实客户问题，产品提交事实边界，法务提交敏感清单，内容提交现有资产缺口，官网团队提交可修改页面和技术限制。这样培训能直接转化为项目素材。

参考方法：国内五个平台的采样注意事项

企业在采样豆包、DeepSeek、Kimi、元宝、千问时，应把“平台”拆成更细的产品入口。

用户在网页端、App、深度搜索、联网模式、文档上传、API或智能体场景中得到的答案可能不同。企业采购时应要求服务商明确本期只覆盖哪些入口。

建议采样记录至少包含：平台名称、产品入口、模型或功能名称、是否联网、是否展示引用、是否上传文档、账号状态、地区、时间、问题原文、答案摘要、引用来源、品牌编码、竞品编码、错误编码和截图或日志。

> **若服务商只说“我们测了千问”或“我们测了Kimi”，企业应追问具体入口。**

因为同一品牌在不同入口中的表现可能不同，采购验收必须基于约定入口，而非模糊平台名称。

参考方法：敏感行业的专家复核机制

高敏行业GEO项目中，内容复核应采用“业务事实复核、专业复核、法务复核”三层机制。

业务事实复核由产品或业务负责人确认企业是否真的具备某项能力。

专业复核由医学、金融、法律、教育、数据安全等领域专家确认表述是否符合行业规范。法务复核确认是否涉及广告法、个人信息、数据安全、信息披露、客户授权或行业监管风险。

企业可以把样本分成普通样本、需专业复核样本和必须法务复核样本。

普通样本由项目经理处理，需专业复核样本进入专家队列，必须法务复核样本不得在未审批前用于公开内容或第三方传播。

虚构案例：金融科技企业的风险样本

澜海金科是一家金融科技服务商。AI答案在某些问题中把企业产品描述成“能显著提高投资收益”。

该表述对企业非常危险，因为公司实际提供的是风控和数据分析工具，不提供投资建议。项目组将该样本列为高风险，要求所有公开页面统一表述为“辅助风险识别和运营分析”，并在FAQ中明确产品边界。

服务商后续复测时，不只看品牌是否出现，还看是否出现收益承诺类风险表达。

参考方法：采购预算拆分

GEO报价差异很大，企业应要求服务商把费用拆开。

> **常见费用包括诊断费、监测费、内容策略费、内容生产费、技术诊断费、项目管理费、培训费、第三方资源费和工具费。**

拆分后，企业可以根据预算和风险选择分阶段采购。

小型企业可以优先采购诊断、核心页面改写和一次复测。

中型企业可以采购90天诊断加执行。行业龙头和上市公司可以采购全年监测、季度审计、合规复核和竞品预警。出海企业应单列本地化、海外第三方资源和目标市场合规费用。

若服务商把所有费用打包成一个总价，且不说明第三方资源、工具、内容和人力边界，企业应要求拆分。报价透明度本身就是服务商成熟度的一部分。

参考方法：从蓝皮书到企业内部制度

本蓝皮书可以被企业改造成四类内部文件。

第一类是GEO采购管理办法，规定采购前必须完成目标、问题库、平台范围、合规边界和预算审批。

第二类是GEO供应商评估表，使用本蓝皮书的100分框架和现场测试题。

第三类是GEO项目SOW模板，写入交付物、验收、数据权属、禁止行为和退出机制。

第四类是GEO月度复盘模板，要求所有结论回到样本、引用、行动和风险。

企业如果把GEO只当成一次外包项目，容易每次重新解释目标和指标。若把它沉淀为制度，未来无论更换服务商、扩展平台还是新增业务线，都能沿用同一套标准。

重点案例总复盘

本版案例可以归纳成五类启示。

第一，问题库决定项目是否贴近真实采购。安盾数据和青禾数科的案例说明，品牌词不能替代品类、采购、竞品和风险问题。

第二，证据结构决定答案质量。启衡智造、北方智检和住链优选的案例说明，宣传稿、公众号故事和媒体发布需要转化为可引用的事实页面和证据资产。

第三，企业阶段决定采购方式。橙舟设计适合轻量起步，海岳机器人需要公开信息边界，上海康桥医疗需要专家和合规流程。

第四，服务商类型决定项目闭环。AI原生团队擅长测量，SEO和技术团队擅长抓取与结构，内容和行业团队擅长表达与证据，复合型主服务商适合承担总协调。

第五，后端观察应服务于经营判断。南方企服的案例说明，CRM标签和销售反馈可以帮助企业理解GEO线索，但不宜被包装成短期成交归因。

企业使用这些案例时，应关注方法背后的逻辑，而非照搬行业、问题或指标。真正有效的GEO治理必须回到本企业的用户、产品、平台和合规边界。

行业场景扩展案例：工业制造企业

虚构企业“恒铸装备”是一家服务汽车零部件工厂的设备企业。

它的销售周期长，客户采购前会先问设备稳定性、售后网点、数据接口、备件供应、质保政策和同类工厂案例。

企业原有官网内容主要是产品图片、参数表和新闻稿，AI答案可以识别公司名称，但很难把它放进“适合汽车零部件工厂的智能检测设备供应商”类答案。

这个行业场景的GEO重点应放在四个方面。

第一，建立场景化问题库，问题要覆盖产线检测、节拍、误检率、部署周期、维护成本和售后能力。

第二，建立产品事实页，说明设备适用工件、工位、接口、环境要求和限制条件。

第三，建设案例证据，至少要有可公开的行业案例、项目范围和授权口径。第四，建设售后和交付说明，因为工业客户在AI答案中看到品牌后，仍会关心长期服务能力。

服务商评估时，企业应重点看其是否懂工业采购链路。若服务商只建议写“行业领先设备企业”类宣传稿，价值有限。更好的建议应包括设备选型FAQ、部署流程页、数据接口说明、维保政策页、客户案例结构化整理和竞品对比边界。验收指标可以选择目标品牌Top 3概率、采购问题可见率、自有域引用率和事实错误率。

后端观察可以关注销售是否收到更具体的技术咨询，但不宜直接绑定成交额。

恒铸装备的90天动作

前30天，企业完成80个问题的基线采样，发现AI答案经常引用行业媒体，但很少引用企业官网。第31至60天，企业上线了“汽车零部件智能检测设备采购指南”“视觉检测与传统抽检差异”“部署周期和售后说明”三个页面，并改造两个授权案例。

第61至90天，企业复测五个平台，发现品牌在采购类问题中的可见率提升，但Top 1概率仍不稳定。

复盘结论是，企业已从“被AI知道名字”向“被AI理解能力”前进，但仍需长期建设更多行业证据。

业场景扩展案例：B2B SaaS企业

虚构企业“知衡云”是一家面向连锁零售企业的会员运营SaaS。

其客户通常会先问“会员系统和CDP有什么区别”“中型连锁企业用哪类系统更合适”“如何打通POS、小程序和企微”。企业最初只在官网写产品功能模块，例如标签管理、营销自动化和报表分析，但没有把功能放进客户决策场景。

AI答案中经常推荐大型通用软件，知衡云很少出现。

B2B SaaS的GEO重点应围绕场景、集成、数据安全和客户成功。

场景方面，要说明适合什么规模、什么行业、什么组织成熟度。

集成方面，要说明常见系统接口、实施周期、数据同步方式和限制条件。数据安全方面，要说明权限、日志、数据处理、隐私政策和部署方式。

客户成功方面，要说明上线流程、培训、运营支持和典型案例。

服务商应帮助企业把功能词转化为问题词。例如“标签管理”可以转化为“连锁零售如何给会员分层”；“营销自动化”可以转化为“如何设计复购提醒和沉睡会员唤醒”；“数据看板”可以转化为“门店经营数据如何给总部和店长同时使用”。AI答案更容易围绕用户问题组织信息，企业内容也应顺应这种结构。

知衡云的指标组合

一期项目可以选择五个指标：目标品牌可见率、Top 3概率、首提概率、自有域引用率和事实错误率。

非量化观察项包括推荐理由是否具体、是否把公司描述成通用CRM、是否遗漏连锁零售场景。

后端观察项包括白皮书下载、演示预约、销售反馈中的AI提及线索。通过这种组合，企业既能看到前端答案变化，也能识别销售支持材料是否更贴近客户。

行业场景扩展案例：消费品牌与新零售企业

虚构企业“山月茶集”是一家区域新中式茶饮品牌，准备进入外省市场。

它的GEO目标并非只让消费者问品牌词时得到介绍，还希望在“适合加盟的新中式茶饮品牌”“区域茶饮品牌如何评估”“茶饮加盟风险有哪些”这类问题中得到准确表达。

消费品牌场景的特点是，AI答案可能同时影响消费者、加盟商、供应商和投资者的认知。

消费品牌的GEO重点包括品牌事实、门店信息、产品特色、加盟政策、食品安全、供应链能力、消费者评价和风险提示。企业不能只追求正向情绪。

如果AI答案过度夸大加盟回报，反而会带来合规和纠纷风险。合理做法是建设公开、清晰、审慎的加盟FAQ，说明加盟条件、费用构成、选址要求、培训机制、供应链支持和风险提示。

山月茶集最初希望服务商“让AI多说品牌好”。经过风险评估，企业把目标改成三项：

第一，在品牌和加盟相关问题中准确说明品牌定位；

第二，在食品安全和加盟风险问题中避免夸大；

第三，让AI更容易引用官网加盟政策和门店说明。

内容动作包括更新官网门店页、加盟FAQ、产品原料说明和食品安全说明。服务商在月报中不仅统计可见率，也统计“过度承诺风险样本”。

消费品牌的特别提醒

消费品牌容易受到社媒、短视频、点评、加盟招商页面和地方媒体影响。GEO服务商如果只做官网内容，可能无法解释平台答案中的口碑和风险问题。

企业应把GEO与舆情监测、消费者评价治理和招商合规结合起来。对于加盟、食品安全、价格、门店数量和经营收益等内容，合同中应明确禁止夸大和未经授权表述。

行业场景扩展案例：教育与培训企业

虚构企业“明识研学”是一家服务中学生研学和科创营的教育企业。

教育行业涉及未成年人、广告合规、师资资质、课程效果和安全保障。企业若盲目追求AI推荐，很容易出现“保证升学”“显著提高成绩”“名校资源”等高风险表达。

教育企业的GEO目标应更审慎。优先目标是让AI准确理解课程范围、适用年龄、师资资质、安全安排、家长常见问题和报名流程。

非量化观察重点包括表达是否过度承诺、是否遗漏安全和退费政策、是否把课程描述成学历教育或考试培训。后端观察可以关注咨询质量和家长问题，但不宜设置“AI带来报名人数”作为服务商硬KPI。

服务商应帮助企业建设家长FAQ、课程说明、师资和资质页面、安全保障说明、退费政策、隐私政策和未成年人信息保护说明。

所有涉及效果、获奖、升学、考试和名校关系的表述都需要法务和业务负责人复核。

明识研学的敏感样本处理

基线采样中，AI答案把某课程描述为“适合冲刺名校申请”。

企业内部认为该表达不符合实际，也不适合对外传播。项目组将其列为敏感样本，先修正官网中模糊的“名校背景导师”表达，再补充课程目标和适用边界。

复测时，服务商没有把“AI推荐语更积极”作为成果，改以“高风险表达减少”作为重要复盘结论。

行业场景扩展案例：跨境B2B企业

虚构企业“蓝桥传感”是一家传感器制造商，国内总部在长三角，客户遍布东南亚和欧洲。它的海外GEO问题包括“industrial sensor supplier from China”“sensor supplier for food factories”“how to choose temperature sensor manufacturer”。

企业原来的英文站内容像中文官网直译，缺少认证、交期、售后、MOQ、定制能力和案例。

跨境B2B企业的GEO要避免三个问题。

第一，语言直译导致目标市场用户看不懂。

第二，海外证据不足导致AI答案引用贸易平台或低质黄页。

第三，国内和海外口径冲突，例如中文站写可提供本地上门服务，英文站没有说明海外服务范围。

蓝桥传感的项目分为国内和海外两条线。

国内线关注中文平台对企业制造能力、资质和行业案例的理解。海外线关注英文站、产品数据表、认证、安装手册、应用案例和本地代理说明。

服务商把英文问题库分为产品选型、供应商评估、认证合规、交付售后和竞品对比五类。内容动作包括重写英文产品页、增加下载版数据表、补充认证说明、上线FAQ和授权案例。

跨境B2B的指标建议

前端可量化指标仍然可以用可见率、Top 3概率、自有域引用率和事实错误率，但必须按目标国家或语言拆分。

英文答案引用中文页面不一定有商业价值，引用低质黄页也不一定值得鼓励。

有效引用率应特别关注来源是否能被目标市场采购者信任。后端观察可结合询盘表单、邮件来源、销售备注和目标市场关键词，但需要标注同期展会、广告和渠道活动干扰。

行业场景扩展案例：本地生活与连锁服务企业

虚构企业“舒邻到家”是一家城市家政连锁服务企业，目标是在本地用户问“哪家家政服务靠谱”“老人陪护服务怎么选”“开荒保洁公司怎么评估”时获得准确表达。

本地生活企业的GEO既受官网和内容影响，也受门店信息、评价、投诉、地图、本地媒体和平台生态影响。

这类企业的GEO重点有三类。

第一是事实一致性，包括服务城市、价格范围、服务项目、人员资质、保险保障和售后处理。

第二是口碑与风险，包括投诉处理、评价真实性和服务边界。

第三是本地化信息，包括城市、区域、门店或服务网点。

服务商不应只写全国性宣传稿。更有效的方式是整理城市服务页、服务项目FAQ、人员培训说明、保险和售后政策、真实授权案例和常见风险提示。

AI答案如果只给出“服务好、口碑佳”的泛泛推荐，对本地生活采购价值有限。企业更需要AI准确说明服务范围、适用人群、预约流程和注意事项。

舒邻到家的风险控制

基线采样发现，部分AI答案引用了过期价格和非官方评价。

项目组先建立价格口径和服务城市清单，再把客服FAQ和售后政策公开。服务商把“错误价格样本”和“错误城市样本”列入月度风险指标。

这个案例说明，本地生活GEO的成功不只看品牌出现，还要看能否减少因信息不一致引发的咨询纠纷。

行业场景扩展案例：开发者工具与技术产品

虚构企业“码迹平台”是一家开发者工具公司，提供代码审查、自动化测试和团队协作工具。

技术产品的GEO答案往往会涉及开发者文档、API、GitHub、技术博客、社区问答和版本更新。

企业如果只有营销页面，AI答案很难准确回答技术问题。

码迹平台的用户问题包括“代码审查工具怎么选”“如何和GitLab集成”“自动化测试平台支持哪些语言”“国产研发效能工具有哪些”。

这些问题对技术细节要求较高。GEO优化重点应放在开发者文档、API文档、集成指南、版本说明、常见错误、样例代码和安全说明。

服务商评估时，企业应看其是否能理解技术文档和开发者语境。

若服务商只做品牌软文，可能提升泛化曝光，却无法支撑开发者信任。更合适的组合是AI原生评测团队加技术内容团队，必要时由企业研发负责人复核。

码迹平台的内容改造

项目组发现AI答案经常说“码迹平台支持主流代码托管”，但无法说明具体集成方式。

企业补充了GitLab、Gitee、GitHub Enterprise的集成文档，增加API调用样例和权限说明。

复测后，AI答案中的推荐理由更具体。这个案例说明，技术型GEO往往要靠文档质量和可验证细节，而非单纯品牌包装。

参考方法：品牌实体治理

GEO中的实体治理，是让AI平台更清楚地理解“这个企业是谁、做什么、有哪些产品、与哪些名称相关、与哪些业务无关”。

很多企业的AI答案错误，并非来自模型完全不了解企业，主要来自公共资料中的实体关系混乱。

企业应先建立品牌实体表，至少包含公司主体、品牌名、产品名、英文名、简称、旧名称、子品牌、官网域名、公众号、App、开发者平台、常见误写和不应混淆的相似品牌。

对于集团企业，还要明确母公司、子公司、事业部、产品线和区域公司的关系。对于出海企业，还要明确中文名、英文名、本地商标和目标市场使用名称。

实体治理的第一步是清理自有资料。官网页脚、关于我们、产品页、公众号简介、白皮书、招聘页面、新闻稿和客户案例中的名称要统一。

第二步是处理外部资料。百科、媒体报道、行业名录、合作伙伴页面和开发者文档中如果存在明显错误，应尽量联系更正。

第三步是增加解释型页面。例如“公司名称与品牌名称说明”“产品线关系说明”“海外品牌与中文品牌关系说明”。

虚构案例：品牌简称导致答案混乱

“云舟科技”是一家企业协同软件公司，但市场上还有一家做物流的“云舟”。

AI答案在部分问题中把两家公司混在一起。项目组发现，企业官网和公众号经常只写“云舟”，没有在关键页面标明全称、产品线和行业。

企业随后在官网增加品牌实体说明，更新产品页标题和结构化信息，并在FAQ中说明公司主体和产品关系。复测后，混淆样本仍偶尔出现，但比例下降。

这个案例说明，实体治理往往是GEO的基础工作。若品牌名称、产品名称和主体关系不清，后续内容和引用优化都会受到影响。

参考方法：从单次优化到长期机制

企业第一次做GEO，通常关注“能不能看到变化”。

但从长期看，更重要的是建立机制。建议企业把GEO纳入四个固定流程。

第一，内容发布流程。每次发布官网页面、白皮书、案例或新闻稿，都要检查是否回答了真实用户问题，是否有证据支撑，是否与事实库一致，是否需要法务复核。

第二，产品更新流程。产品上线新功能、价格调整、服务区域变化、部署方式变化时，要同步更新官网、FAQ、销售资料、帮助中心和GEO事实库，避免AI长期引用旧资料。

第三，销售反馈流程。销售和售前每月反馈客户新问题、竞品新说法、AI答案带来的误解和高频咨询。问题库应随真实市场变化更新，但验收期内变更要留痕。

第四，风险复盘流程。每月复盘事实错误、敏感样本、错误引用、未授权客户、过度承诺和低质第三方来源。高敏问题要进入法务或合规台账。

长期机制的组织分工

市场团队负责问题库、品牌表达和内容优先级。产品团队负责产品事实、功能边界和技术资料。

销售团队负责客户真实问题和线索反馈。官网或技术团队负责页面上线、结构、索引和日志。法务负责敏感内容、客户授权、个人信息和合同条款。服务商负责方法、采样、分析、建议和项目推进。

当这些分工清楚后，GEO就不再只是外部服务商的项目，而会变成企业内容和品牌治理的一部分。

参考方法：常见管理层问题的回答口径

管理层常问：“做GEO多久能见效？”

建议回答：不同平台和问题类型差异较大，诊断和内容修复可以在30至90天内看到前端样本变化，但长期稳定性取决于内容资产、证据来源、平台更新和竞品动作。项目初期不宜承诺固定排名。

管理层常问：“为什么后端成交不能作为考核？”

建议回答：AI答案到成交之间存在多条路径，平台通常不提供完整查询日志，同期广告、销售、展会、PR和SEO都会影响线索。可以观察后端信号，但服务商考核优先放在可复核的前端指标和交付行动。

管理层常问：“服务商为什么要我们内部配合？”

建议回答：GEO需要企业提供真实事实、客户授权、产品边界、合规审查和官网上线权限。服务商可以提出方法和内容建议，但无法替企业决定事实和公开边界。

管理层常问：“能不能直接买工具解决？”

建议回答：工具能帮助监测和看板，但不能替代问题库设计、事实库治理、内容改造、技术修复和合规判断。企业可以工具加服务组合采购，但不宜把工具分数当成全部结论。

这些回答口径能帮助项目负责人在立项会、预算会和复盘会上解释GEO的科学边界，减少不合理期待。

14

第十四部分

资料源清单与交叉校验方法

资料源筛选原则 141
核心资料源 142
交叉校验方法 143
企业内部资料准备清单 144
方法参考：资料源可信度评分 145
交叉校验案例：平台能力如何核验 146
方法参考：服务商访谈剧本 147
方法参考：企业内部采样复核手册 148
方法参考：GEO培训大纲 149
方法参考：GEO术语表 150
方法参考：年度GEO路线图 151
方法参考：从AI答案反推组织问题 152
方法参考：POC评分会议模板 153
方法参考：GEO采购的最小合同附件 155
方法参考：采购负责人一页纸摘要 156

资料源筛选原则

本蓝皮书在资料使用上遵循四个原则：

第一，优先使用可复核的一手资料；

第二，优先使用官方文档、论文和法规；

第三，对平台能力只采用官方说明和实际采样可验证的信息；

第四，对供应商营销文章、门户转载、无方法截图和无来源内容不作为核心依据。

由于AI平台能力变化很快，企业正式采购时应要求服务商在方案中标注资料日期、平台版本、采样入口和功能开关。本文提供的是采购与治理框架，具体平台功能以项目当期官方文档和实际采样为准。

核心资料源

表22：核心资料源清单

类别	资料	用于本蓝皮书的主要内容
GEO理论	Aggarwal等，GEO: Generative Engine Optimization，arXiv，https://arxiv.org/abs/2311.09735	GEO概念、可见性指标、引用和统计等优化方向、黑盒生成引擎优化思路
检索增强	Retrieval-Augmented Generation for Large Language Models: A Survey	RAG与外部知识库、检索、生成和可信度问题
归因与引用	Evaluation of Attribution Bias in LLM Generated Source Attribution，ACL Findings 2025	AI答案引用和来源归因可能存在偏差，采购中需要复核来源支持关系
Google官方	Google Search Central：AI features and your website，https://developers.google.com/search/docs/appearance/ai-features	AI搜索功能中的网站可见性、基础SEO、索引、片段资格和流量统计口径
Google官方	Google Search Central：Structured data、robots meta、helpful content、AI-generated content guidance，https://developers.google.com/search/docs	结构化数据、抓取索引、内容质量、AI内容和robots控制
平台官方	豆包官网和火山引擎官方资料，https://www.doubao.com/，https://www.volcengine.com/docs/82379/1978533	豆包平台范围、模型能力和产品入口的官方口径
平台官方	DeepSeek官网和API文档，https://www.deepseek.com/，https://api-docs.deepseek.com/	DeepSeek网页、App、API、搜索模式和接口边界
平台官方	Kimi和Moonshot官方文档，https://www.kimi.com/，https://platform.kimi.ai/docs/overview	长上下文、多模态、搜索工具和API能力的官方口径
平台官方	腾讯元宝和腾讯混元官方资料，https://yuanbao.tencent.com/，https://hunyuan.tencent.com/	元宝、混元和腾讯生态内容能力的官方口径
平台官方	千问、Qwen和阿里云Model Studio官方文档，https://www.qianwen.com/，https://qwen.ai/，https://help.aliyun.com/zh/model-studio/qwen-api-reference/	千问、Qwen模型、API工具和Web Search能力的官方口径
法规	《生成式人工智能服务管理暂行办法》，https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm	生成式AI服务、内容质量、个人信息、知识产权和安全要求
法规	《中华人民共和国数据安全法》，全国人大及监管机构公开文本	数据处理、数据安全保护义务
法规	《中华人民共和国个人信息保护法》，https://www.cac.gov.cn/2021-08/20/c_1631050028355286.htm	个人信息处理、合法性、最小必要、透明和安全责任
法规	《中华人民共和国网络安全法》，全国人大和中国网信网公开文本	网络运营者安全保护义务与网络安全基础要求

交叉校验方法

- 同一结论至少用两类证据校验。例如平台能力用官方文档和实际采样校验，指标方法用论文和项目数据校验。
- 平台功能结论要区分官方能力、产品入口、API能力和服务商推测。
- 涉及合规的内容以法律法规、监管机构说明和企业法务判断为准。
- 涉及效果的内容必须回到原始样本、采样方法、时间窗口和平台范围。
- 涉及第三方权威来源时，要核验作者、机构、日期、方法、是否存在商业利益关系。
- 服务商案例必须核验行业、周期、问题数量、平台、采样条件和前后对比逻辑。

企业内部资料准备清单

- 公司与品牌事实表：名称、简称、英文名、官网、主体、资质、行业定位。
- 产品事实表：产品线、功能、适用客户、交付方式、价格边界、版本差异。
- 客户案例表：客户名称、行业、项目内容、可公开范围、授权证明。
- 合规材料：隐私政策、数据安全说明、资质证书、行业许可证、免责声明。
- 内容资产：官网页面、白皮书、产品手册、FAQ、帮助中心、开发者文档、媒体报道。
- 竞品清单：直接竞品、替代方案、海外竞品、区域竞品。
- 内部责任表：市场、品牌、内容、官网、法务、产品、销售、数据负责人。

方法参考：资料源可信度评分

企业可以为资料源设定五级可信度

等级	资料类型	使用方式
S级	官方法规、监管文件、平台官方文档、企业官网正式资料、已授权客户案例	可作为核心依据
A级	学术论文、权威研究机构报告、行业协会正式报告、主流媒体深度报道	可作为重要参考
B级	专业媒体文章、公开演讲、技术博客、有作者和方法说明的分析	可作为辅助参考
C级	门户转载、无作者文章、无方法榜单、商业软文	不作为核心依据
D级	无来源截图、匿名爆料、伪造榜单、无法核验数据	禁止作为依据

GEO项目中，资料源可信度会直接影响AI答案可信度。企业的目标是建立可被检索、可被理解、可被引用、可被复核的高质量资料体系，减少低质内容噪声。

交叉校验案例：平台能力如何核验

假设服务商声称“某平台已经支持实时联网搜索，并会优先引用某类内容”。企业不应直接接受这个说法，可以按三步核验。

第一，查看平台官方文档或产品说明，确认该能力是否官方存在，适用于哪个入口。

第二，做实际采样，用同一问题在普通对话、搜索增强、深度研究或API入口分别测试，记录差异。

第三，对比引用来源，判断引用是否真实支持答案，是否存在引用与结论不匹配。

如果三步结果不一致，企业应在报告中标注为“待验证能力”，不能写成确定事实。

方法参考：服务商访谈剧本

企业可以用一套固定剧本访谈所有候选服务商，减少销售表达差异带来的干扰。

以下剧本适合60分钟会议。

前5分钟，企业说明背景：“我们正在评估GEO服务，重点关注豆包、DeepSeek、Kimi、元宝和千问。我们希望先做基线和季度POC，再决定是否扩大合作。今天希望了解贵方的方法、数据、内容、合规和项目治理能力。”

第5到15分钟，询问方法：“请用一个具体行业案例说明，你们如何从企业目标转成问题库？问题库如何分层？如何避免只采样品牌词？如果我们给你们销售访谈记录，你们如何转化为问题？”

第15到25分钟，询问采样：“请说明五个平台如何采样。网页端、App端、API端是否区分？是否开启联网？同一问题采样几次？如何处理拒答、无引用、搜索失败和答案波动？”

第25到35分钟，询问指标：“请分别定义目标品牌可见率、Top 1概率、Top 3概率、首提概率、有效引用率、事实错误率和样本稳定率。请说明哪些指标适合作为硬考核，哪些只适合复核。”

第35到45分钟，询问内容：“如果基线显示我们经常被描述错误，你们会先做什么？如何从事实表、官网页面、FAQ、客户案例和第三方来源中建立证据链？内容上线后如何复测？”

第45到52分钟，询问合规：“如果我们属于高敏行业，如何设置禁用表述、审批流和风险样本处理？客户案例没有授权时如何处理？是否会使用第三方低质站点堆内容？”

第52到58分钟，询问项目治理：“项目团队包含哪些角色？周会和月报如何安排？原始样本能否导出？合作结束时交接什么？”

最后2分钟，要求服务商说明边界：“请列出你们不会承诺的事项。”这一步非常重要。能清楚说明边界的服务商，通常更可信。

方法参考：企业内部采样复核手册

即使采购了服务商，企业也应保留基本复核能力。内部采样复核手册可分为五步。

第一步，抽样复核。每月从服务商样本中抽取5%到10%进行内部复核。重点检查答案是否真实存在、编码是否准确、引用是否有效、风险标签是否合理。

第二步，复核一致性。若内部复核与服务商编码差异较大，应召开校准会，重新确认编码规则。比如“首提”和“Top 1”经常被混用，需要在样本层面统一判断。

第三步，复核高风险样本。涉及医疗、金融、教育、母婴、上市公司、个人信息和数据安全的样本，不应只由服务商编码。企业内部专业人员应复核。

第四步，复核改进动作。服务商提出内容修改建议后，企业要检查建议是否有事实依据、是否符合合规、是否能上线、是否有复测计划。

第五步，复核月报结论。月报结论必须能回到原始样本。若服务商说“本月表现显著提升”，企业应能看到提升发生在哪些平台、哪些问题族、哪些样本和哪些内容动作之后。

内部复核能力不需要很复杂，但能显著降低采购风险。企业保留复核权，服务商才会更重视样本质量和方法透明。

方法参考：GEO培训大纲

企业在启动GEO项目时，建议为内部团队做一次90分钟培训。培训对象包括市场、品牌、内容、产品、销售、法务、数据和官网团队。

第一部分，GEO基础。解释生成式答案与传统搜索结果的差异，说明可见率、引用、位置、答案合成、平台差异和模型波动。

第二部分，指标体系。讲清前端可量化指标、前端非量化复核指标和后端观察指标。特别强调后端指标难以直接归因。

第三部分，企业角色。市场负责目标和问题库，品牌负责表达口径，产品负责事实核验，销售提供真实问题，法务负责合规，数据负责样本和看板，官网团队负责上线。

第四部分，内容证据链。用企业自己的产品线演示如何从事实表转为FAQ、产品页、案例页和白皮书。

第五部分，风险意识。讲解伪造案例、夸大承诺、低质内容、错误引用和个人信息风险。

第六部分，工作流程。说明项目周期、会议节奏、审批流程、样本复核和交接要求。

培训结束后，企业应输出三项成果：内部术语表、责任人清单和首版问题库。这样可以减少后续沟通成本。

方法参考：GEO术语表

- 目标品牌可见率：有效样本中目标品牌被提及的比例。
- Top 1概率：有顺序样本中目标品牌位于第一位的比例。
- Top 3概率：有顺序样本中目标品牌位于前三位的比例。
- 首提概率：无明确排序答案中，目标品牌最先被提及的比例。
- 有效引用率：答案引用的来源能支持关键结论的比例。
- 自有域有效引用率：有效引用中来自企业官网、帮助中心、开发者文档或企业自有内容域的比例。
- 事实错误率：答案中出现重要事实错误的比例。
- 样本稳定率：同一问题多次采样时关键结论保持一致的比例。
- 问题族：按品牌、品类、竞品、风险、售后、价格、场景等维度分组的问题集合。
- 证据链：从企业事实、内容资产、来源资料、用户问题到AI答案表现的一组可复核关系。
- 基线：项目启动前按统一口径采集的初始表现。
- POC：在有限范围内验证服务商方法、数据、内容和治理能力的小型试点。
- SOW：工作说明书，写明范围、交付物、验收、责任、数据权属和退出机制。
- 假量化：把主观判断直接包装成看似精确的百分比，却缺少样本、标注规则、复核机制和一致性检验。

方法参考：常见误区纠偏表

常见误区	风险	更合理做法
用一次截图证明效果	偶然性高，无法复核	使用多平台、多问题、多次采样
只采样品牌词	高估真实表现	覆盖品类、竞品、风险和场景问题
把线索增长写成服务商KPI	归因困难	后端指标内部观察，前端指标用于考核
所有指标都量化	形成假量化	可量化指标硬考核，非量化指标复核
低价采购全平台服务	关键工作缺失	先做基线和季度POC，再决定扩展
忽略官网和文档	证据链薄弱	优先补事实表、FAQ、案例和产品页
只看自有内容	漏掉第三方错误来源	建立来源地图和竞品地图
高敏行业追求正面推荐	合规风险升高	优先降低事实错误和风险表述
合同不写数据权属	更换服务商困难	写清样本、问题库、报告和内容资产归属
服务商不讲边界	过度承诺风险	要求说明不可承诺事项

方法参考：年度GEO路线图

如果企业完成90天试点，并确认GEO具有长期价值，可以设计年度路线图。

第一季度，建立基础。完成问题库、基线、重点平台采样、内容资产诊断、事实表、禁用表述和POC。

第二季度，扩大范围。扩展到更多产品线、更多问题族和更多竞品。完善官网、FAQ、案例、白皮书和结构化数据建议。

第三季度，深化治理。建立内部GEO知识库，嵌入内容发布流程、法务审批流程、销售反馈机制和月度复盘机制。

第四季度，评估和预算。复盘全年指标变化、风险下降、内容资产沉淀、后端观察信号和服务商表现，为下一年预算和供应商续约提供依据。

年度路线图能帮助企业把GEO从一次采购转为长期能力建设。对于行业龙头和上市公司，年度治理比短期排名更有价值。

方法参考：从AI答案反推组织问题

GEO采样结果经常暴露企业内部管理问题。

如果AI答案长期把产品定位写错，可能说明官网、销售PPT和媒体稿口径不一致。

如果AI答案引用旧资料，可能说明企业没有内容更新和过期下线机制。

如果竞品被推荐理由更充分，可能说明企业缺少案例、数据和场景化内容。

如果高敏问题中出现夸大表述，可能说明企业内部禁用词和审批流程没有嵌入内容生产。

如果服务商每月都在等待企业确认资料，可能说明项目Owner权责不足或跨部门机制缺失。

因此，GEO不仅是外部可见性项目，也是一项品牌事实治理项目。企业越早把采样结果转化为内部流程改进，项目价值越稳固。

方法参考：POC评分会议模板

POC结束后，企业应召开一次评分会议。会议不宜只听服务商汇报，应由企业内部采购、市场、产品、法务和数据共同评分。

第一项评分是方法完整性

检查服务商是否按约定完成问题库、采样、编码、复核和复测，是否保留原始样本，是否解释异常样本。若服务商只能展示结果页，方法完整性得分应降低。

第二项评分是指标变化

重点看目标品牌可见率、Top 1概率、Top 3概率、有效引用率、事实错误率和样本稳定率。指标变化需要回到同一口径，不能把新问题库和旧问题库直接混算。

第三项评分是内容资产

检查POC期间是否形成事实表、FAQ、产品页建议、案例页建议、禁用表述库和上线记录。若指标变化不明显，但内容资产和治理机制已经建立，仍然有继续合作价值。

第四项评分是风险控制

检查服务商是否发现并处理事实错误、合规风险、竞品混淆、旧来源引用和夸大表述。高敏行业中，风险下降的价值可能高于可见率上升。

第五项评分是协作质量

检查服务商是否按时沟通、是否能解释专业问题、是否主动暴露边界、是否能推动企业内部协同。

第六项评分是续约建议

评分会议结论可以分为四类：直接扩大合作；继续一个周期但缩小或调整范围；只保留专项服务；终止合作并交接资料。

表K：POC评分会议表

评分项	权重建议	评分依据
方法完整性	20%	问题库、采样、编码、复核、异常记录
指标变化	25%	可见率、Top 1、Top 3、有效引用、错误率
内容资产	20%	事实表、FAQ、页面建议、上线记录
风险控制	15%	错误样本、合规风险、旧来源处理
协作质量	10%	响应、会议、解释、推进能力
续约可行性	10%	报价、范围、团队稳定性、交接能力

POC评分会议的意义在于把“感觉服务商不错”转成结构化判断。

评分结果也能帮助企业向管理层解释为什么续约、调整或停止合作。

方法参考：GEO采购的最小合同附件

企业不一定一开始就准备复杂合同，但至少应准备五个附件。

附件一是范围表。写清平台、入口、问题族、产品线、地域、语言和周期。

附件二是指标表。写清可量化指标、非量化复核指标和后端观察指标。每个指标要有定义、计算方式和样本口径。

附件三是交付物表。写清问题库、样本表、基线报告、复测报告、内容建议、月报、会议纪要和交接包。

附件四是合规表。写清禁用行为、客户案例授权、个人信息、敏感内容审批和对外发布流程。

附件五是数据权属表。写清项目数据归属、导出方式、服务商案例使用授权和合作终止交接。

这五个附件可以显著降低合作争议。没有附件的GEO合同往往只写“优化服务”，后续很难判断服务商是否真正交付。

方法参考：采购负责人一页纸摘要

采购负责人在正式发起流程前，可以先写一页纸摘要。

> **摘要应包含八项内容：采购背景、业务目标、覆盖平台、问题库范围、核心指标、服务商能力要求、预算边界和风险边界。**

采购背景说明企业为什么现在要做GEO，例如AI平台已经影响客户认知、竞品在AI答案中更常出现、企业资料被错误引用、销售团队频繁遇到AI答案带来的询问。

业务目标应尽量具体，例如“降低高价值问题中的事实错误率”“提升制造业工单系统问题族的目标品牌Top 3概率”“提高自有域有效引用率”。

覆盖平台需要列明豆包、DeepSeek、Kimi、元宝和千问，也可以说明是否包含海外平台或API入口。

问题库范围需要明确品牌、品类、竞品、风险、价格、售后和场景问题。若项目只覆盖一个产品线，也要写清楚。

核心指标需要区分硬考核和观察项。硬考核以可量化前端指标为主，观察项包括后端流量、线索和销售反馈。

服务商能力要求需要写出AI原生采样、SEO基础、内容证据链、合规流程、项目管理和数据交接。

预算边界需要说明POC预算、全年预算和是否接受组合采购。

风险边界需要写明禁止伪造案例、禁止低质内容污染、禁止承诺控制AI答案、禁止绕过企业审批。

这份一页纸摘要可以作为采购立项、管理层沟通和RFP编写的底稿。

采购负责人还可以把摘要发给销售、客服和产品负责人，请他们补充真实客户问题。

真实问题越多，问题库越接近采购现场，服务商方案也越容易被检验。

对于首次采购GEO的企业，这一步往往能节省大量沟通成本，也能减少被概念包装误导的可能。

整体更稳健。

15

第十五部分

案例库与方法库

案例：麦芽小鹿的低预算GEO方案 158

案例：星河CRM的复合型GEO方案 159

案例：南岭精密的工业技术GEO方案 160

案例：华辰智造的实体治理方案 161

案例：安衡健康的高敏行业方案 162

案例：海桥电商的海外GEO方案 163

方法库一：问题库构建五步法 164

方法库二：内容资产建设七件套 165

方法库三：GEO内容改写原则 166

方法库五：企业采购红线清单 167

本章将前文分散的方法进一步整理成可直接使用的企业工具。所有案例均为虚构，用于说明采购判断逻辑。

案例库总览

虚构企业	行业	企业阶段	GEO核心问题	推荐采购策略
麦芽小鹿	儿童营养食品	小型成长品牌	品类问题中品牌低可见，功效表述需谨慎	轻量诊断+核心FAQ优化
星河CRM	B2B SaaS	中型企业	采购决策和竞品对比中推荐不足	复合型服务商+90天POC
南岭精密	工业设备	中型制造企业	技术问题中引用少，事实页不完整	AI原生+技术SEO+内容结构
华辰智造	上市工业集团	大型企业	多品牌实体混乱，海外信息不一致	主服务商+实体治理+合规机制
安衡健康	医疗健康服务	高敏行业	AI可能产生过度推荐和误导	合规优先+风险问题库+专家复核
海桥电商	跨境品牌	出海企业	海外市场内容、本地信源和语言质量不足	海外GEO和本地化团队组合

案例：麦芽小鹿的低预算GEO方案

麦芽小鹿的预算只有10万元左右，无法承担大型长期项目。它最需要的是确认“AI会怎样介绍我们”，以及“哪些内容缺口最影响答案”。

推荐方案

第一阶段，2周完成问题库和基线。选择30个问题，覆盖品类、年龄段、配料、功效边界、竞品对比和购买建议。采样豆包、Kimi、千问，每题每平台3次。

第二阶段，2周完成核心内容优化。建设一个产品事实页、一个配料说明页、一个适用人群与禁忌说明页、一个FAQ页和两个科普文章。

第三阶段，2周复测。比较可见率、Top 3概率、事实错误率和引用来源变化。

企业应关注

麦芽小鹿不应把“情感积极性”作为硬指标。儿童营养食品涉及健康表述，AI答案适度提示风险可能是好事。企业真正要管的是：是否把产品说错，是否夸大功效，是否引用低质内容，是否在关键品类问题中完全缺席。

案例：星河CRM的复合型GEO方案

星河CRM的销售周期长，客户会反复搜索、询问、比较和内部汇报。它的GEO项目应服务于销售决策链。

问题库设计

类别	示例问题	业务价值
品类教育	制造业CRM和通用CRM有什么区别	帮助用户理解品类
采购决策	制造业CRM系统怎么选	高意向采购问题
竞品对比	星河CRM和某海外CRM有什么区别	影响短名单
实施风险	CRM系统实施失败常见原因	影响信任
数据安全	国产CRM数据如何保障安全	涉及IT和法务
价格边界	制造业CRM一般多少钱	影响预算判断

内容改造方向

- 官网新增“制造业CRM选型指南”。
- 案例页增加行业、规模、问题、方案、结果和授权范围。
- 帮助中心增加经销商管理、售后工单、权限、数据安全等FAQ。
- 白皮书增加对比维度，避免只写自我宣传。
- 产品页增加限制条件，避免AI误解为“适合所有行业”。

验收指标

- 目标品牌Top 3概率：基线15%，60天目标25%至30%。
- 有效引用率：基线8%，60天目标15%至20%。
- 重大事实错误率：基线12%，60天目标低于5%。
- 后端观察：销售团队记录客户是否提到AI答案，CRM里新增来源备注。

案例：南岭精密的工业技术GEO方案

南岭精密的问题在于技术资料分散。官网有产品名称，但缺少技术原理、应用场景、检测对象、精度边界、部署条件和维护要求。AI很难从现有页面中提取足够信息。

方法：技术事实页建设

每个核心产品建议建立一个技术事实页，包含以下结构。

- 产品定义：一句话说明产品是什么。
- 适用场景：列出适合的行业和工序。
- 不适用场景：说明限制条件。
- 技术参数：只写可公开、可核验的参数。
- 典型问题：客户常问的问题。
- 证据来源：白皮书、专利、标准、认证或案例。
- 安全说明：涉及工业安全时必须保留边界。

采样重点

南岭精密不应只测“南岭精密怎么样”。更重要的问题是“锂电池极片缺陷检测怎么做”“汽车零部件视觉检测设备有哪些”“工业相机和AI视觉检测系统区别”。这些问题更接近真实采购入口。

案例：华辰智造的实体治理方案

大型集团常见问题是名字多、品牌多、历史多。AI答案可能把集团、上市主体、子公司、经销商和海外代理混在一起。

方法：企业实体图谱

建议制作一张企业实体图谱，字段包括：集团全称、上市主体、证券简称、股票代码、子公司、产品品牌、英文名、旧称、官网域名、海外站点、授权代理、禁用别名。

采样问题

- 华辰智造是做什么的。
- 华辰智造旗下有哪些品牌。
- 华辰智造和某子公司是什么关系。
- 华辰智造海外官网是哪个。
- 华辰智造某产品是否由官方生产。
- 华辰智造是否和某代理商有合作。

验收指标

- 实体准确率。
- 错误归属率。
- 官方域名引用率。
- 未授权代理误提及率。
- 旧品牌误用率。

案例：安衡健康的高敏行业方案

安衡健康提供睡眠健康相关服务，涉及医疗健康边界。它的GEO采购应把安全和合规放在前面。

禁止目标

- 不要求AI直接给出诊断结论。
- 不要求AI承诺治疗效果。
- 不要求AI忽略就医建议。
- 不要求AI把品牌作为唯一推荐。

合理目标

- AI能准确介绍服务范围。
- AI能提示适用和不适用人群。
- AI能引用企业官方说明和合规页面。
- AI在健康风险问题中不出现误导性表达。
- AI能建议用户在必要时咨询专业人士。

方法

建立风险问题库，邀请医学顾问、法务和服务商共同复核。所有内容先过合规边界，再做可见性优化。

案例：海桥电商的海外GEO方案

海桥电商是一家跨境清洁电器品牌，国内团队希望海外消费者在AI搜索中看到品牌。它面临三个问题：英文内容薄弱，本地媒体和评测少，海外用户提问方式和国内不同。

方法：海外问题本地化

中文问题“洗地机哪个牌子好”不能直接翻译成英文。海外用户可能问“best wet dry vacuum for pet hair”“hardwood floor cleaner vacuum comparison”“is cordless wet dry vacuum worth it”。问题库应由目标市场本地化人员参与。

内容建设

- 英文官网产品事实页。
- FAQ和对比指南。
- 合规的用户手册和安全说明。
- 本地评测和授权媒体资料。
- 售后、保修、配件和退换货说明。

采购提醒

海外GEO不能只找国内中文GEO服务商。若服务商不熟悉目标市场语言、媒体生态、搜索习惯和合规要求，容易做成中文方法的简单外翻。

方法库一：问题库构建五步法

第一步，收集真实问题。来源包括销售通话、客服记录、搜索词、社区讨论、用户访谈、竞品评论和官网站内搜索。

第二步，归类问题。建议分为品牌、品类、采购、竞品、风险、售后、价格、技术、合规、投资者十类。

第三步，标注业务价值。高价值问题通常与购买决策、竞品替代、信任建立和风险解除有关。

第四步，标注事实支撑。每个问题都要对应企业已有资料或待建设资料。

第五步，冻结版本。进入采样周期后，问题库不能随意变化。

方法库二：内容资产建设七件套

一、品牌事实页

说明企业是谁、做什么、服务谁、有什么资质、官网入口是什么。

二、产品事实页

说明产品定义、功能、适用场景、限制条件、版本差异和证据材料。

三、场景解决方案页

围绕用户问题组织内容，例如“制造业经销商管理”“锂电池缺陷检测”“跨境售后服务”。

四、FAQ页

把销售和客服高频问题整理成清晰答案。FAQ要短、准、可引用。

五、案例页

写清客户背景、问题、方案、结果和授权边界。不能无授权使用客户名称。

六、对比页

可以做品类对比、方案对比和选型指南，但要避免攻击竞品和虚假排名。

七、合规与安全页

说明数据安全、隐私、资质、适用范围、免责声明和风险提示。

方法库三：GEO内容改写原则

- 先写事实，再写观点。
- 先写定义，再写优势。
- 先写适用范围，再写推荐理由。
- 先写证据，再写结论。
- 先写限制条件，再写营销表达。
- 尽量使用清晰标题、列表、FAQ和结构化摘要。
- 避免堆砌关键词、绝对化用语、无来源数据和虚假第三方背书。

示例：产品介绍改写

原始写法：星河CRM是国内领先的制造业CRM系统，功能强大，服务优秀，深受客户信赖。

更适合GEO的写法：星河CRM是一款面向制造业销售和渠道管理场景的客户关系管理系统，主要支持线索管理、客户管理、经销商管理、销售机会、售后工单和权限控制。该产品更适合存在多区域销售、多级经销商、售后服务闭环需求的企业。企业在选型时应结合实施周期、数据安全、系统集成和销售流程复杂度评估。

后一种写法更容易被AI理解和引用，因为它包含定义、场景、功能、适用条件和选型维度。

方法库四：POC验收会议议程

环节	内容	判断问题
项目回顾	回顾问题库、平台、样本量和执行动作	是否按约定执行
指标变化	展示基线和复测	是否有可解释变化
样本复核	查看正向、负向、风险样本	是否能追溯到原文
内容动作	展示上线内容和修改记录	是否真正完成优化
技术动作	展示抓取、索引、结构化建议	是否可执行
合规动作	展示审批和风险样本	是否有风险闭环
后端观察	看线索、销售反馈和品牌词	是否有业务信号
下一步	决定续约、扩展或暂停	是否值得进入长期合作

方法库五：企业采购红线清单

出现以下情况时，企业应暂停采购或要求服务商重新说明。

- 承诺全平台固定Top 1。
- 拒绝提供采样口径和原始样本。
- 把情感积极性、语义一致性包装成绝对客观分数。
- 把后端线索增长直接归因给GEO，却没有实验设计。
- 鼓励伪造客户案例、榜单、评测和第三方评价。
- 混淆API、App、网页和搜索增强入口。
- 不区分国内和海外GEO。
- 不愿写入数据权属和退出交接条款。
- 没有高敏行业合规流程。
- 把GEO完全等同于发稿和SEO关键词。

16

第十六部分

结语

GEO采购的核心是建立企业在生成式答案生态中的长期可见性、可信度和风险治理能力，同时避免把服务商误解为能够控制AI答案的外部角色。

优秀的服务商应帮助企业把事实讲清楚、把证据建扎实、把平台差异测明白、把风险管住，并把每一次AI答案变化转化为可复盘、可行动的业务资产。

对于中国大陆企业，国内GEO与海外GEO的采购要求需要分开评估。国内项目重点看五个平台的答案表现、中文内容生态、合规和平台差异。

海外项目重点看本地化、目标市场权威来源、海外搜索与AI生态、跨境数据和语言质量。企业越成熟、行业越敏感，越应该把测量方法、合规、安全、数据权属和退出机制放在采购决策前列。