

星澜客户洞察 GEO Signal Monitor 系统监测报告

合成公开样例。品牌、指标、采样和任务均为演示数据。

1. 执行摘要与关键判断

判断	结论	证据	下步动作
监测目标	建立 AI 答案可见性、引用质量、事实准确、纠偏和谨慎归因闭环。	五平台 Prompt、指标、来源账本、证据等级和纠偏任务均已定义。	进入周期性采样、证据入库和复盘。
真实数据状态	本次为合成公开样例，不包含真实客户、真实平台答案或真实 CRM 转化数据。	sample_mode = synthetic_replay, evidence_level = E0/E1。	如需真实月报，先接入 M1-M4 的可审计答案样本。
当前风险	不能只看品牌出现率，必须同时看推荐、引用、事实、稳定性和证据等级。	出现率与推荐率、引用召回率存在差距。	建立阈值告警和 PO 纠偏机制。
报告完整性	本报告按系统性、详细度、完整性三层自检。	覆盖来源、场景、Prompt、数据接入、采样、六层分析、治理和附录。	月报沿用同一结构。

2. 任务范围、边界和采样声明

项目	说明
语言	中文简体
平台	DeepSeek、豆包、千问、Kimi、元宝
采样声明	本次为合成公开样例，不包含真实客户、真实平台答案或真实 CRM 转化数据。

项目	说明
边界	不绕过登录、验证码、限流、付费或平台条款；归因默认从观察相关开始。
真实数据要求	必须提供答案原文、Prompt、时间、平台、账号/地区/联网状态和截图、导出或接口日志。

3. 权威参考与来源账本

source_id	类型	标题/事实	URL/来源	用途	置信度
S1	合成声明	示例品牌与样本为合成	report_input.json	防止误用为真实客户	高
S2	方法来源	GEO、可验证性、RAGAS、NIST	references/research-basis.md	支撑指标和治理	高

4. 公司/品牌事实基线

事实	官方来源或核验方式	风险
星澜客户洞察是合成品牌。	报告首段	不能冒充真实客户。
样本为合成回放。	采样声明	不能写成真实平台结论。
监测主题为 B2B 客户洞察。	Prompt 和场景	场景过窄会影响推荐率。

5. 公司测试场景发现

场景	业务含义	对应 Prompt 组	风险	正确答案应覆盖
XL-01	B2B 销售团队选择客户洞察工具	推荐、比较	只看 CRM，不看客户洞察	场景、推荐理由、竞品对照
XL-02	制造业客户分层	场景问法	缺少行业证据	行业案例、适用边界
XL-03	价格和实施成本	价格、风险	旧价格被复述	当前价格事实页
XL-04	品牌真实性核验	品牌验证	资质事实缺失	官网事实卡和引用

6. 监测 Prompt 库与对照组

组别	核心 Prompt	对照 Prompt	观测重点
推荐	适合星澜客户洞察目标用户的方案有哪些？	不含品牌名的同类推荐	候选率、推荐率、排序。
比较	星澜客户洞察 与主要竞品怎么选？	调换品牌顺序	排序、优劣描述、引用源。
替代	竞品有哪些替代方案？	只问竞品	品牌是否被召回。
价格	星澜客户洞察 价格、套餐或成本如何？	只问贵不贵	价格事实、适用边界。
风险	星澜客户洞察 有什么限制或风险？	行业通用风险问法	负面表述、误解来源。
品牌验证	星澜客户洞察 是什么？有哪些产品/能力？	只问品牌是否正规	事实准确率、引用质量。
场景问法	具体业务场景如何选择工具？	不含品牌名场景问法	场景召回、推荐理由。

7. 真实数据接入模式与证据等级

模式	当前状态	进入正式指标条件	报告措辞
M0 合成回放	已用于本示例	仅用于流程验证，不进入真实月报指标	方法演示，不代表真实平台表现。
M1 用户提供真实样本	可接入	答案文本 + Prompt + 采样环境 + 截图/导出	可作为客户样本分析。
M2 人工授权采样	可接入	人工采样记录 + 频率边界 + 复核人	可作为小规模真实样本。
M3 授权 API/连接器	条件可用	API 权限、接口日志、频率、失败重试	可进入看板趋势。
M4 浏览器辅助合规采样	条件可用	人工授权登录、无绕过、截图和采样日志	可用于复核和截图证据。
M5 CRM/转化数据导入	可接入用户授权数据	脱敏、字段口径、时间窗口、拥有方授权	只能辅助归因，不能替代 AI 答案样本。

证据等级	条件	当前样例状态
E0	无原始答案和环境字段	不作为真实平台数据。
E1	有答案文本但缺少截图或完整环境	可作为线索。
E2	有答案文本、Prompt、平台、时间、地区、联网状态	可作为单次真实样本。
E3	E2 + 截图、导出文件、引用链接或接口日志	可审计真实样本。
E4	E3 + 多轮复采、对照 Prompt、复核人和去重记录	可进入月报统计。

8. 五平台采样口径

平台	重点	样本量建议	必填环境字段	质检重点
DeepSeek	结论稳定性、证据链、联网状态	40+	时间、设备、地区、联网、sample_mode	多次答案是否一致。
豆包	口语问答、图文输出、短答案	40+	设备、账号、地区、联网、截图	是否省略来源或过度简化。
千问	引用源、追问路径、生态信源	40+	轮次、追问、联网、引用链接	引用是否支持说法。
Kimi	深度研究、长文引用、文档站	40+	长文模式、联网、引用段落	引用召回和事实更新。
元宝	微信生态、公众号、视频号	40+	账号、地区、生态来源、可访问路径	二手中文来源是否可靠。

9. 核心指标总览

指标	合成结果	解释
品牌出现率	62%	召回改善，但不等于被推荐。
候选率	46%	替代类 Prompt 改善。
推荐率	38%	推荐理由仍不足。
描述准确率	71%	价格事实仍需纠偏。
引用召回率	44%	Kimi 深度研究引用提升。
引用准确率	68%	二手媒体支持度有限。
负面表述率	14%	旧版本误解下降。

10. 平台差异分析

平台	差异	动作
DeepSeek	结构化比较好，但引用链弱	强化官方来源账本。
豆包	短答案容易简化品牌定位	增加短事实卡。
千问	引用表现较好，追问后竞品增多	保留追问链路和 turn_index。
Kimi	长文能覆盖研究和文档站	检查旧数字和长引用支持度。
元宝	中文生态召回强	防止公众号二手内容替代官方来源。

11. 引用源追踪与证据质量

来源类型	支持等级	判断规则	纠偏动作
官方网站/文档	A	直接支持答案事实	优先作为事实卡和纠偏锚点。
官方中文资料	A/B	中文可读且支持说法	用于国内平台引用优化。
投资者/公告/标准	A	支持动态数字或治理要求	写绝对日期，避免旧数据。
媒体/评测/社区	B/C	可辅助比较但不一定支持事实	仅作参考，不作主事实。
竞品页面	C	用于对照，不验证本品牌事实	标注为竞品来源。

12. 答案事实性与描述准确率

事实类型	典型错误	核验方法	优先级
产品/能力	用旧名称、少列产品、夸大 AI 能力	对照官网和官方文档	P0

事实类型	典型错误	核验方法	优先级
价格/套餐	把付费能力写成免费	对照定价、知识库和公告	P0
客户/案例	使用旧数字或未授权案例	对照官方公告和案例页	P1
市场/适配	过度绝对化国内或海外适用性	对照场景和竞品	P1

13. 竞品、替代和负面表述分析

维度	监测点	风险	输出
竞品出现	竞品频率、排序、推荐理由	品牌被替代或弱推荐	竞品矩阵。
替代关系	国产替代、海外替代、传统方案	不同场景混在一起	场景化替代表。
负面表述	价格、本地化、数据、实施成本	被二手内容放大	风险澄清页。

14. 稳定性、波动和置信度

置信度	条件	报告措辞
高	多平台、多轮次、一致引用、E3+ 证据和对照支持	可作为稳定判断。
中	有样本和引用，但平台间有差异或证据等级不足	作为方向判断。
低	只有单次样本、合成回放或缺少引用	仅作为观察线索。

15. 答案差异与谨慎归因

干预	基线窗口	观察窗口	对照	归因规则
内容发布	T-14 至 T0	T+7/T+14/T+30	不相关 Prompt / 竞品 Prompt	默认观察相关。
页面修复	T-14 至 T0	T+7/T+14	未修复页面组	有对照改善才升置信。
外部信源	T-30 至 T0	T+14/T+30	未发布主题	检查索引延迟和外部事件。
CRM/转化	T-30 至 T0	T+30/T+60	非 GEO 入口或未曝光组	只能辅助解释，不能替代 AI 答案采样。

16. 纠偏任务与路线图

优先级	问题	映射资产	验收指标
P0	旧价格套餐被复述	官网价格事实页	价格错误率低于 5%
P0	官网引用不足	FAQ、案例页、文档站入口	官网引用提升 30%
P1	制造业案例缺证据	行业案例页	场景推荐率提升 8pp

17. 告警规则和复盘节奏

告警	阈值	处理
事实错误	P0 错误连续两轮出现或描述准确率低于 80%	建 P0 纠偏，14 天内复采。
引用不足	引用召回率低于 50%	补官方证据页和中文承接页。
证据等级不足	正式样本低于 E2 或截图/导出缺失	降级为待复核，不进入正式指标。

告警	阈值	处理
推荐下降	推荐率环比下降超过 10pp	检查平台更新、竞品动作和 Prompt 分布。
负面上升	负面表述率高于 18%	建风险澄清页和销售口径。

18. 仪表盘字段、数据库表和 API 草案

表/接口	字段或路径	用途
monitor_prompts	scenario_id、group、query_text、control_flag、prompt_version	管理 Prompt 和对照组。
answer_samples	sample_mode、evidence_level、platform、sampled_at、region、network_enabled、answer_text	保存采样答案。
sample_evidence	raw_answer_path、screenshot_path、api_log_id、collector、permission_basis	保存真实数据证据。
citations	source_type、source_url、claim_text、support_level	追踪引用质量。
correction_tasks	priority、mapped_asset、owner、acceptance_metric	管理纠偏闭环。
API	GET /api/geo-monitor/monthly-report	拉取月报聚合。

19. 治理、合规、数据质量和风险控制

风险	控制
平台条款	不绕过登录、验证码、付费和限流；批量采样需授权。
数据隐私	CRM、转化、账号、截图和接口日志脱敏；示例只用合成数据。
数据质量	记录采样环境、来源账本、复核人、证据等级和置信度。
生成式 AI 风险	标注幻觉、过时事实、引用不支持和过度归因。
真实数据误用	没有可审计样本时，报告必须标注为合成或待复核。

20. 自 review 结果

检查项	结果	说明
系统性	通过	覆盖来源、场景、Prompt、数据接入、采样、指标、引用、归因、纠偏、治理和附录。
详细度	通过	每个模块有字段、阈值、动作、证据等级或验收标准。
完整性	通过	结论可回到 Prompt、样本、来源、证据和纠偏任务。
HTML 菜单	通过	浏览器检查确认桌面端 fixed 菜单、移动端 sticky 菜单。
横向溢出	通过	桌面 1440px 与移动 390px 视口均无横向溢出。
Pandoc 默认 CSS	通过	HTML 生成时禁用默认文档 CSS，并显式覆盖 body 窄栏约束。

检查项	结果	说明
kami UI	通过	白底优先，采用油墨蓝、暖灰、紧凑层级、稳定表格边框。

21. 附录：Prompt 全表、指标字典、来源账本、采样字段

附录	内容
Prompt 全表	七组 Prompt、对照 Prompt、场景 ID、版本。
指标字典	出现率、候选率、推荐率、排序、描述准确、引用召回、引用准确、稳定性。
来源账本	source_id、source_type、url、fact_supported、freshness_risk、confidence。
采样字段	sample_mode、evidence_level、platform、sampled_at、device、account_state、region、network_enabled、turn_index。
真实数据证据	raw_answer_path、screenshot_path、api_log_id、collector、permission_basis、review_status。