九章格物 NineGrids

GENERATIVE ENGINE OPTIMIZATION · 垂直行业研究

星阙实验室GEO研究中心 · 第 09 册

新能源汽车行业
生成式引擎优化（GEO）白皮书

当续航、智驾与购置税都缩成 AI 答案里的一行字，车企如何让自己的参数被读对、被比赢、被推荐——从决策链迁移、参数失真机制到信源采信权重

出品：九章格物 · 星阙实验室

作者：李现龙

版本：V1.0

时间：2026 年 11 月

权威信源：基于【工业和信息化部、国家税务总局、中国汽车流通协会、CNNIC】等公开数据与资料编制

微信扫我加好友

扫码添加微信 · 获取 GEO 方案

© 2026 九章格物 · 星阙实验室 · ninegrids.cn

目录

核心摘要 03

01 一个行业两套算法：新能源为什么最难被 AI 说对 04

02 决策链被切成五段：AI 到底接管了哪一段 06

03 参数失真：AI 答案里的错是怎么长出来的 09

04 政策变量：为什么车的问题会答成地区的问题 12

05 信源的四层采信权重：从工信部目录到车主论坛 15

06 对比型提问：AI 如何在三款车之间排序 18

07 三类车企的起点差异：谁先被读对 21

08 度量：新能源 GEO 该看什么、归因断在哪 25

09 边界与风险：参数合规、OTA 漂移与平台封闭 28

10 实测：10 份真实问答里，AI 把哪几个参数说错了 32

参考来源 36

摘 核心摘要

新能源汽车是 GEO 难度最高的消费品类之一。原因不在竞争激烈，而在它的答案结构：一款车由参数、政策、迭代速度三重变量共同决定，任一项错位都会让 AI 给出的推荐整体失效。本册据此拆解四条主线：决策链接管、参数失真、政策地域化、信源采信。

核心发现	结论与依据
一 · 决策链	AI 接管的是“比选”，不是“成交”。2026 年 4 月以来新能源乘用车零售渗透率稳定在 60% 以上（7 月 65.1%）[5][6]，近九成购车者会先用 AI 辅助决策 [9]；但终点仍在到店与试驾，可归因终点只能是“到店”。
二 · 参数失真	AI 的错不是“编造”，而是“择其一”：官方参数滞后、工况口径不一（CLTC 与实测 [10]）、老款 UGC 混入新款，三类源冲突时模型倾向取高频值而非权威值（机制推演，非模型实测）。
三 · 政策耦合	2026 年购置税由免征改减半、单车减税上限 1.5 万元，且须列入《减免车辆购置税的新能源汽车车型目录》[1][3][4]；政策带时间、地域、车型白名单三重限定，是 AI 错误率最高的提问类型。
四 · 信源分层	四层信源各管一件事：官方目录管“能不能减税”，车企官网管“参数是多少”，垂媒参数库管“横向怎么比”，UGC 管“真实体验”。缺一层，答案残一层；补错层，反而更错。

> **边界：**所引渗透率与用户行为数据分别来自行业协会、CNNIC 与第三方调研公开口径，统计范围不同（总体乘用车零售与新能源国内零售差异明显），用于判断方向与量级，不作精确预测；文中车型对比字段仅为结构示例，不代表具体车型参数。

四条发现中有三条与行业里的流行判断相反，值得先说清楚。

一是**内容多不等于易被推荐。**搜索时代多发一篇就多一个入口；生成式引擎里新旧内容同时参与切片竞争，一个车系十种说法等于给模型十个可选答案，冲突概率随内容量上升。

二是**参数全不等于易被选中。**答案长度有限，压缩时优先丢“非典型项”——恰恰是最能区分你与竞品的那几项；四十行参数堆进去，剩下的往往是人人都有的大项。

三是**准确率下降不一定是做错了什么。**存量内容的自然贬值会吃掉新增内容的准确率增量：一边每月补准确内容，一边旧内容每月衰减，净效果可能为负——这就是“持续投入，AI 答案却越来越差”的由来。

本册九章按“现象—机制—度量—风险”展开，不按行业惯例的经营环节排列，因为核心矛盾不在渠道，而在**同一款车的信息能否跨源一致。**下表是各章要回答的问题：

章节	要回答的问题	给读者的判断工具
01—02	为什么新能源最难被 AI 说对；AI 到底接管了决策链的哪一段	两套算法对照表、五段决策链、七类调用场景
03—04	错误答案怎么长出来；为什么车的问题会答成地区的问题	四步传导链、参数分级、购置税测算
05—06	四层信源各自的分工；AI 如何在三款车之间排序	信源金字塔、缺口诊断、六字段写法
07—08	三类车企的起点差异；该测什么、归因断在哪	资产累积/折旧表、象限定位、五项监测指标
09	哪些风险会同时发生，且有结构性后果	三类风险形态、叠加判断、三情景权重

01 一个行业两套算法：新能源为什么最难被 AI 说对

如果把 GEO 按行业难度排个序，新能源汽车大概率在前三。但难的原因和大多数人想的不一样——不是竞品多、不是投放贵，而是这个行业的答案天生由三段互相牵扯的变量构成：**一是参数，二是政策，三是迭代速度**。这三段里任何一段错位，AI 给出的整段推荐就会失效，而且失效得毫无痕迹——用户不会知道你被写错了，他只会不选你。

对比一下餐饮和医疗就清楚了。餐饮行业的 AI 答案主要靠"评价密度"支撑，一家店好不好，AI 看的是评分、点评数、人均、口味标签，这些数据相对稳定，一年半载变化不大。医疗行业靠"资质与指南"，执业资质、科室等级、诊疗指南都是低频变动。这两个行业的 GEO，本质是"把稳定的事实说清楚"。

新能源汽车不是。一款车从上市到改款通常 12 个月，中间还夹着 OTA 升级——续航可能因为电池管理策略优化而变化，智驾能力可能因为一次软件推送而跨级，价格可能因为一轮官降而下调。这意味着**同一款车在同一个问题下，AI 给出的答案在半年内就可能完全不同**，而且原先正确的内容会变成错误的内容。这在餐饮和医疗行业几乎不存在。

两套算法：排序逻辑与拼接逻辑

章节标题里的"两套算法"，指的不是两个模型，而是两种信息分发机制同时作用于同一个行业。搜索是排序：给一堆页面打分，把最靠前的几条交给用户。生成是拼接：把多段材料合成一个答案，用户看到的是结论而不是清单。

两者的差别不在形式，在"什么是有效信号"。排序看重的是页面与查询的相关性、站点权重、外链结构；拼接看重的是信息是否可切、是否自洽、是否在多处一致。这导致同一个行业在两套机制下的胜负手完全不同：

对比维度	排序逻辑（传统搜索）	拼接逻辑（生成式引擎）
基本信息单元	整页 / 整篇	段落 / 表格行 / 字段三元组
权威如何体现	站点权重、外链、更新频率	跨源一致出现的次数与站点分级
内容偏好	覆盖关键词、篇幅足够、结构清晰	字段完整、口径统一、结论可切
更新方式	重发一篇文章，旧文可保留	须同时处理旧文，否则旧值持续参与竞争
失效形态	排名下滑，但仍可被找到	错误被合成进答案，且不留痕迹
可比性要求	低（各页各说各话也可以）	高（不同实体须能逐字段对齐）

表 1-1 排序逻辑与拼接逻辑的六维对照

这张对照表解释了新能源行业里最常见的一种错配：车企沿用排序时代的经验做拼接时代的事。比如为了"覆盖更多词"而上线大量车型页，这在排序逻辑下是净收益——每个页面都是一个入口；在拼接逻辑下却可能是负收益——同一车系的多个页面说不同数，直接制造冲突源。

也必须承认一个现实：两套机制目前是共存的。用户仍然会用搜索引擎查资料，AI 也仍然引用搜索引擎收录的内容。所以新能源车企面对的不是"二选一"，而是"同一批内容要同时满足两种偏好"。这也是本册后面反复强调"结构化表达"的原因——**结构化既能被排序，也能被拼接；纯叙事只能被排序。**

参数型品类与体验型品类的分野

把这个差异抽象一层，可以分成两类品类。体验型品类（餐饮、酒旅、本地生活）的决策依据是"感受"，参数型品类（汽车、3C、家电、B2B 设备）的决策依据是"规格"。两类品类的 AI 答案结构完全不同：

维度	体验型品类	参数型品类（新能源）
答案骨架	评价 + 场景 + 人群匹配	预算 + 参数 + 版本 + 政策适用性
核心变量	口碑密度、评分、距离	续航、补能、空间、智驾、价格、目录资格
错误形态	推荐了不合适的场景（错配）	参数写错、版本混淆、政策答错（失真）
错误后果	体验不佳，但门店仍会被到访	整段推荐失效，用户直接转向竞品
纠错成本	低，一条好评可覆盖	高，需重建参数页与版本实体

表 1-2 体验型品类与参数型品类的答案结构差异

这个分野带来一个直接推论：**参数型品类的 GEO 核心不是"多发内容"，而是"把参数说对"**。发一百篇评测软文，不如把官网那份配置表的结构、版本、单位、适用年款做对。因为 AI 在做答案时，真正被检索和切片的是结构化参数，不是形容词。

五维难度：新能源难在哪一维

把难度拆成五个维度，可以看清新能源究竟卡在哪个位置。这五个维度不是主观打分，而是有对应机制的：参数密度决定"错一次损失多大"，政策耦合决定"答案有多少前置条件"，迭代频率决定"内容多久失效"，决策周期决定"触点有多少次"，竞品密度决定"排序竞争多激烈"。

维度	新能源	餐饮	医疗	3C 数码	该维度决定什么
参数密度	极高	低	中	高	单条错误的破坏力
政策耦合	极高	低	极高	低	答案的前置条件数量
迭代频率	高（年度改款 + OTA）	低	低	高（年度换代）	内容有效期
决策周期	长（2–8 周）	短（分钟级）	中	中	AI 被反复调用的次数
竞品密度	极高（同价位 10+ 款）	高	中	高	排序竞争强度

表 1-3 四类品类 × 五维难度的 GEO 难度对照

口径说明：维度评级为本报告基于行业特征的定性判断，非量化实测；"极高/高/中/低"仅用于横向对照，不构成评分。

五个维度里，真正把新能源和 3C 数码区分开的是**政策耦合**和**迭代频率的组合**。3C 数码参数也密、换代也快，但它没有政策前置条件——手机不会因为"是否列入某个目录"而改变购买成本。新能源会。这就是第 04 章要专门拆政策变量的原因。

内容半衰期：一道不太好看的算术

把迭代频率换算成时间，可以得到一个比较残酷的结论。一款车从上市到年度改款约 12 个月，中间通常夹 1—2 次影响参数的 OTA。保守假设：一次改款让 30% 的关键参数发生变化（续航、电池容量、充电功率、智驾硬件），一次 OTA 让 10% 变化。那么 12 个月后，最初那套参数里约有 40% 已经不再准确。

另一头，一篇内容从撰写、审核、发布到被检索收录，通常需要 2—6 周。两段周期叠起来就是：**内容上线时准确率接近 100%，半年后可能掉到 80%，一年后掉到 60%**。内容不是"发布即完成"，而是"发布即开始衰老"。

这道算术解释了一个行业里普遍存在的困惑：明明投了内容、也做了监测，AI 答案的准确率却还在往下走。原因是新增内容的准确率被存量内容的衰减吃掉了。假设每月新增 10 篇准确内容，存量有 100 篇在同步衰减，净效果完全可能是负的——而这在报表上体现为"投入增加、效果下降"，看起来像玄学，其实是算术。

参数类别	变动触发	典型半衰期	维护含义
价格与权益	官降、限时权益、金融方案调整	1—3 个月	需高频同步，不适合作为常青内容的主体
续航与补能	年度改款、电池策略 OTA	6—12 个月	须绑定年款与工况，改款即全量校验
智驾能力	软件推送、功能开城	3—6 个月	功能边界变化最快，须注明生效版本
车身与空间	换代	3—5 年	相对稳定，可作为常青资产
使用成本逻辑	电价、油价、保养体系变化	1—2 年	机制类内容，衰减最慢

表 1-4 新能源关键参数的内容半衰期分布

半衰期为基于改款周期与 OTA 频次的估算区间，属机制推演，非实测数据；实际因品牌节奏不同差异较大。

由此推出一条与 SEO 时代几乎相反的原则：**SEO 时代内容越多越好，GEO 时代内容越"耐用"越好。**一篇不含时效参数的机制类内容（比如"电车与油车的使用成本到底怎么算"），半衰期可能是数年；一篇堆满当期参数的导购文，半衰期只有几个月。前者适合做内容的基数，后者适合做当期的截流，但不能把后者当资产。

更麻烦的是，衰减不只是"变旧"，还会"变错"。旧内容里的参数如果与新参数矛盾，就从一个"过时的事实"变成一个"活跃的错误源"——它会持续参与切片，持续与正确内容竞争。这也是为什么在新能源行业，"删旧"和"增新"同等重要，很多车企只做后者。

一个反直觉判断：车型越多，越容易被张冠李戴

行业里有个普遍误解：车型矩阵越宽，被 AI 提到的机会越多，所以应该拼命铺车型。这个推论在 SEO 时代成立——每个车型词都是一个入口，多一个页面多一个排名机会。在 GEO 时代恰恰相反。

原因是 AI 做答案时不是"找到页面再摘要"，而是"先切分实体，再对齐属性，最后合成"。同一品牌下的多款车型如果命名相近（比如同名车系的 Pro、Max、长续航版、四驱版），在切片和对齐过程中极易发生属性串线：把长续航版的 700km 挂到标准版上，把四驱版的加速成绩写到两驱版上。这种错误在用户看来源页面时几乎不会发生，但在 AI 合成答案时是高频事故。

所以新能源车企做 GEO，第一件事不是加车型，而是给每个版本建立唯一、可区分的实体标识——车型名 + 年款 + 版本 + 动力形式四段式，缺一段就会串。这件事做不好，车型矩阵越宽，错误越多，而且错得越像真的。

这也解释了为什么有些车企在 AI 答案里"存在感很高但成交转化很差"：它被提到的次数不少，但每次提到的参数都不是它想卖的那个版本。提及率是虚的，版本对得上才是实的。这一点在第 08 章的度量部分会展开成可执行的指标。

02 决策链被切成五段：AI 到底接管了哪一段

讨论购车决策被 AI 改变，最容易犯的错是把"AI 参与"等同于"AI 决定"。真实情况是：AI 把购车链路切成了五段，但只在前三段有实质接管，后两段仍然牢牢握在线下。搞清楚接管边界，才知道 GEO 该优化到哪一步为止。

先看量：AI 已经走到了哪一步

2026 年的几组公开数据可以拼出一条大致曲线。CNNIC《生成式人工智能应用发展报告（2026）》显示，截至 2026 年上半年，中国生成式 AI 用户规模突破 7 亿、普及率超过 50%；2026 年 7 月，AI 搜索引擎月均使用次数约 385 亿次，同比增长 9.9%，是使用频次第二高的应用类型 [7]。第三方调研机构的数据更具体：购车前使用 AI 搜索辅助决策的比例达 88.6%，其中约 44% 将其作为核心查询工具 [9]。

这里必须先做一次口径校准。CNNIC 测的是"用户规模与使用频次"，回答"有多少人在用、用得多不多"，不区分品类；调研机构测的是"购车这一单一场景中的使用率"，回答"买车的人里有多少会问 AI"。两个数字不能互相换算，也不构成矛盾——**越是高决策、高单价的品类，AI 的介入越深。**购车是典型的高决策场景，这也是本册把分析对象限定在这一场景、并把"购车场景使用率"作为唯一引用口径的原因。

同期行业侧的数据也在印证这个趋势：2025 年末新能源乘用车月度零售渗透率首次突破 50%，2026 年 4—7 月稳定保持在 60% 以上，7 月达 65.1% [5]（另一口径按乘用车国内零售统计为 68.1% [6]，差异来自零售口径是否剔除出口）。渗透率跨过 60% 意味着新能源不再是"少数人的选择"，而是默认选项——**当默认选项变多，用户就更依赖 AI 帮他做"在这么多车里选一辆"的减法。**

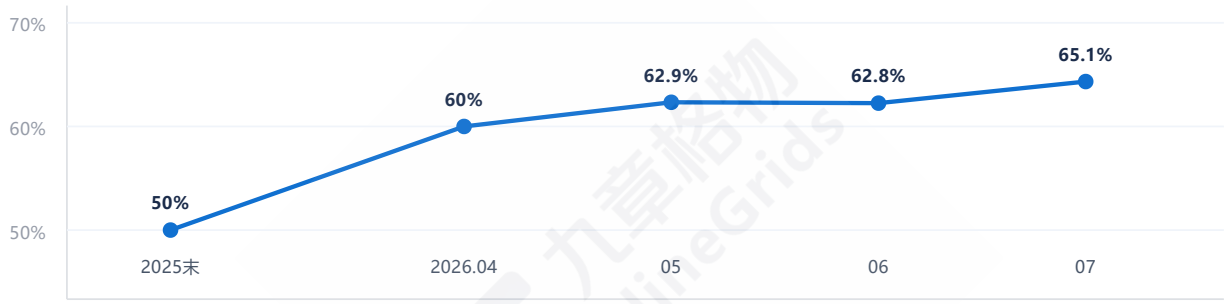


图 2-1 新能源乘用车零售渗透率月度变化（2025 年末—2026 年 7 月）

数据来源：中国汽车流通协会乘用车市场信息联席分会月度数据、人民日报报道口径 [5]；另一口径（新能源乘用车国内零售、剔除出口）7 月为 68.1% [6]，图中采用总体乘用车零售口径。

五段链路与 AI 的实际接管深度

- **需求触发段**——用户还不确定要油车还是电车，先问"家里能装充电桩，一年跑 1.5 万公里，买电车合适吗"。这一段 AI 接管度最高，因为问题本质是"算账"，而算账是 AI 的强项。
- **预算筛选段**——"20 万左右，家用，续航 500 以上，有吗"。AI 接管度次高，它会把预算上下浮动 15% 的车型列出来，形成候选集。
- **车型比选段**——"这几款怎么选"。这是 AI 接管最深也最危险的一段，因为它要给排序和理由，而排序依赖参数对齐。
- **到店验证段**——试驾、看实车、谈金融方案。AI 接管度骤降，用户回到线下，但带着 AI 给的"预设清单"进店。
- **售后维保段**——保养周期、充电桩安装、质保条款。AI 接管度回升，但已经是存量客户的留存问题，不再是获客问题。



图 2-2 购车五段链路中的 AI 接管深度（漏斗示意）

宽度表示该环节 AI 介入的实际影响强度，非用户人数占比。据行业调研的链路分段描述整理 [9]，属定性示意。

这张漏斗最重要的一条信息是：**AI 的接管终点是"到店"，不是"成交"**。用户可能在 AI 那里花 40 分钟比完三款车，最后仍然要走进 4S 店谈价格、试底盘、看后排。这意味着任何声称"GEO 直接带来订单"的说法都不可验证——从 AI 答案到订单之间，隔着一次人工到店。

这个边界条件直接决定了度量口径（第 08 章）：新能源 GEO 的第一验收指标应该是"AI 来源的到店与试驾预约量"，而不是订单量。把订单算成 GEO 的功劳，既高估了 GEO，也掩盖了真正的问题——如果到店量涨了但成交没涨，问题在价格和产品，不在 GEO。

决策周期换算成触点：AI 会被调用多少次

购车决策周期 2—8 周，这个周期里用户不是问一次 AI 就结束，而是反复回到 AI。把这段时间内的查询行为按场景铺开，大致可以分成七类：预算确认、能源形式判断、车型初筛、参数核对、竞品对比、政策与成本确认、口碑验证。

每一类场景通常不在同一天发生，而是随决策推进逐步展开；每类场景平均提问 2—3 次（换措辞重问、针对不同车型重问、追问细节）。七类场景合计，**整个决策周期内 AI 被调用约 14—21 次**。这个数字改变了对 GEO 的理解：它不是一次性曝光，而是复现率竞争。

决策阶段	用户的真实问法	调用频次	该阶段答案的主导信源
预算确认	"25 万落地，能买到什么配置的电车"	2—3 次	结构化参数库
能源判断	"没固定车位，买纯电还是插混"	2 次	机制类长文、政策说明
车型初筛	"20 万家用 SUV，续航 500 以上有哪些"	3—4 次	结构化参数库
参数核对	"这款长续航版多少度电、支持多大快充"	2—3 次	企业官网参数页
竞品对比	"这几款怎么选，各自短板是什么"	2—3 次	对比表、专业评测
政策成本	"今年购置税要交多少，这车在目录里吗"	2 次	官方目录与税务口径
口碑验证	"真实车主怎么说，有没有通病"	2—3 次	车主社区、长测内容

表 2-1 决策周期内的七类 AI 调用场景

调用区间为基于决策链路分段的估算，用于说明复现机制，非平台侧实测调用数据。

把这张表竖着看，会得到一个很实用的结论：**七个场景里，有五个的主失信源不是"内容"，而是"结构"**——参数库、对比表、官方目录。这意味着在新能源品类里，投入重心天然偏向结构化数据，而不是偏向前文那种"讲故事"的内容。

横着看则是另一个结论：每一个场景都是一个独立的复现机会，也是一个独立的失分点。只覆盖"车型初筛"这一个场景，最多命中 21 次调用里的 4 次，其余 17 次全部让给竞品或不提及。这就是单点优化在汽车品类收效有限的原因——不是内容不好，是覆盖面太窄。

还有一个容易被忽略的时序特征：**越靠后的场景，答案越具体，也越依赖准确参数。**在"预算确认"阶段，AI 给一个大致的车型清单就能满足用户；在"参数核对"阶段，一个数字错了就直接导致用户放弃。所以同一份内容，在不同场景下的容错度完全不同——这构成了第 03 章参数分级的现实基础。

渗透率过了 60% 之后，提问方式变了什么

2025 年末新能源乘用车零售渗透率首次跨过 50%，2026 年 4—7 月稳定在 60% 以上，7 月达 65.1% [5][6]。这个数字不只是一个市场成绩，它还改变了一件更基础的事：**用户问 AI 的问题形态变了。**

渗透率在 30% 上下时，AI 收到的最高频问题带有"科普"性质——"电车到底靠不靠谱""家里没充电桩能不能买""冬天续航掉多少"。这类问题的答案结构是解释型的，关键在于把机制讲清楚。此时 AI 扮演的是知识普及者的角色，新能源车企争夺的是"解释权"，谁能把概念讲得更可信，谁就被引用。

渗透率过 60% 之后，最高频的问题变成了"减法"性质——"预算 25 万，这三款里选哪一款""我每天通勤 40 公里，有必要上长续航吗"。用户已经默认接受电车，不再需要被说服，只需要被筛选。此时**AI 的角色从"解释者"变成"筛选器"，争夺的对象也从解释权变成候选集席位。**

这个转变有一个很实际的后果：早期那批以"科普"见长、靠机制长文获得大量引用的内容，在新问题形态下的引用率会下降——因为用户不再问它擅长回答的问题。而结构化参数、克制清晰的对比表、边界明确的取舍说明，引用率会上升。内容形态的适配不是风格问题，是问题形态变了的直接结果。

还有一层更隐蔽的影响：渗透率提高意味着"电车"这个大类词本身在失去区分度。当六成以上新车都是新能源时，"新能源"不再是一个筛选条件，而是一个默认前提。于是用户的问题会自动下移到更细的维度——价位段、车身形式、动力形式（纯电/增程/插混）、补能条件。所以品类词的价值在稀释，**细分条件词的价值在上升**，这也是为什么第 07 章说合资品牌"守着品牌词、丢了车型词"会吃亏。

> **判断：**AI 接管的是"可能性收敛"，不是"决定"。它把用户从"市面上有 200 款车"收敛到"这三款值得看"，剩下的事情交回线下。GEO 的目标就是让自己进入那个三款清单，而不是替用户做完决定。

03 参数失真：AI 答案里的错是怎么长出来的

行业里谈 AI 答案出错，习惯归因于"模型幻觉"，好像模型在瞎编。这个归因在新能源行业基本不成立。真实情况是：**源数据本身就是矛盾的，模型只是从中挑了一个。**它挑的往往不是最权威的那个，而是出现次数最多的那个。

理解这一点很重要，因为它决定了纠错方向。如果错在模型，你无能为力；如果错在源数据的一致性，你完全可以在自己可控的范围内把它修好。新能源行业的参数失真，可以拆成三层。

第一层：官方参数的更新滞后

这是最基础也最普遍的一层。一款车上市时公布续航 620km，半年后因为电池管理策略优化，官方把续航调整为 640km；再过半年年度改款，电池包换了，续航变成 580km 但支持 800V 快充。三次变更，官网参数页可能只改了一次，第三方参数库可能改两次，媒体报道基本不改。

结果就是：当用户问"这款车续航多少"，AI 面对的是三个不同数字，且各自都有出处。模型没有"官方优先"的强制规则，它会综合——最常见的结果是把旧值和新值混在一起，或者在答案里同时出现"620km（另有 640km 版本）"这种自相矛盾的表述。用户看到这种答案，信任度直接崩塌。

这一层的失分机制：不是内容缺失，而是内容的"时点"不一致。参数页有更新日期，但第三方库的抓取时点不一致，最终 AI 拿到的是不同时点的快照。

第二层：工况口径不一致

续航这个参数在中国市场几乎只有一个宣传口径：CLTC。CLTC 是中国轻型汽车行驶工况（GB/T 38146.1—2019），2021 年起成为纯电续驶里程试验的工况依据，测试总时长 1800 秒、总里程 14.48 km，低速市区工况占比高，几乎没有长时间高速行驶 [10]。这个工况偏"理想化"，是行业公开的事实，也是消费者诟病的核心——实际高速续航通常明显低于 CLTC 值。

问题在于，除了 CLTC，市场上还并存 WLTP、EPA 以及媒体实测值。同一个车型，CLTC 标 620km，WLTP 可能 510km，冬季高速实测可能 380km。当用户问"这车冬天能跑多少"，AI 如果拿 CLTC 值回答，答案在字面上没错，在语义上全错。

口径	性质	典型偏差方向	AI 引用后果
CLTC	国标工况，实验室台架	偏乐观，高速与低温场景高估	答案"合规但失真"，用户实测后落差感强
WLTP / EPA	海外工况，更接近实路	数值低于 CLTC	与国内宣传值冲突，模型倾向取高值
媒体实测	实车路测，条件各异	离散度大，冬夏差显著	被当作单一事实引用，忽略测试条件
车主众测	UGC，样本大但无控条件	极端值多，情绪化表述多	易被模型当作"真实体验"加权

表 3-1 四种续航量纲的性质与偏差方向

这一层的失分机制和第一层不同：**不是时点问题，是量纲问题**。同一个词"续航"底下挂着四套不同量纲的数，任何人（包括模型）都无法在不追问条件的情况下给出正确答案。而用户提问时通常不带条件——他只问"能跑多远"。

所以新能源 GEO 在续航这个参数上，正确做法不是"报一个更大的数"，而是**在自有内容里主动把口径说全**：CLTC 多少、高速实测多少、冬季低温多少、对应什么温度与车速。这不是自曝短板，而是让 AI 在切片时能拿到"带条件的分段值"，从而在用户带条件提问时命中正确的那一段。只报一个数的内容，在带条件提问下必然错配。

第三层：版本与年款的实体漂移

第三层最隐蔽，也最容易被忽视。新能源车的版本命名极其密集：同一车系可能有 430km 标准版、510km 长续航版、四驱性能版、智驾版，年款还分 2025 款、2026 款。这些版本在参数上差异很大，但在语料里往往挨在一起——同一篇导购文章里同时出现四个版本的数字。

当 AI 做 RAG 切片时，如果切片边界没对齐版本标识，就会出现属性串线。典型表现：回答"这款车多少钱"时，给出的是最低配的价格；回答"加速几秒"时，给出的是性能版的成绩；回答"有没有激光雷达"时，把高配的配置说成全系标配。这类错误在单篇来源里看不出来，因为来源本身就把多个版本写在一起。



图 3-1 参数失真的四步传导链

依据 RAG 检索增强生成的一般机制与行业参数管理实践整理，为机制示意，非模型内部实测。

三层叠加之后，会出现一个反直觉的结果：**内容越多、越"丰富"的车型，反而越容易被说错。**因为内容多意味着语料里的版本交叉多，切片时串线的概率高。反观内容稀少的老款车型，反而常常被说得准确——因为只有一个版本、一个数，没有可串的空间。

这也解释了为什么有些车企投入大量预算做内容营销之后，AI 答案的准确率反而下降：新增内容引入了新的版本组合，却没有同步建立版本实体标识，等于给模型增加了噪声源。修复的优先级应该反过来——**先做实体消歧，再做内容增量。**

为什么模型取高频而不是取权威

这大概是新能源 GEO 里最需要被理解的一条机制。搜索时代，权威由排序规则决定——官方站点天然靠前，用户看到的第一条通常就是官网。生成式引擎没有这个排序层：它把多段材料拼成一个答案，拼接的依据是"这段材料与问题的语义匹配度"，而不是"这个站点有多权威"。

于是冲突取舍变成一个统计问题。假设一个问题下有三段材料，两段说续航 580km、一段说 640km，模型在无语境偏好时的默认行为是倾向多数——不是因为它认为多数正确，而是因为多段材料互相印证会提升置信度。**权威只在"站点权威度分级"这类外部信号存在时才成为约束，而这类信号目前覆盖有限**[11]。

这个机制给出一条明确的推论：**在新能源品类里，一致性本身就是权威。**如果一个数字在官网、垂媒、媒体、UGC 四处都以同一口径出现，它天然获得高置信度；如果一个数字只在官网出现（哪怕它是最正确的），它在冲突取舍中是弱势的一方。

由此也能解释一个常见的挫败感：车企把自己官网做得再规范，AI 答案依然会引用别人的错值。原因不是官网不够权威，而是"只有官网说对"，在其他三处形成了 1 对 3 的劣势。这种情况下，正确做法不是继续加强官网，而是把那套正确口径同步到下游去——这正好是第 05 章"L2 自证不被采信"的现实基础。

> **反直觉判断：**新能源 GEO 的第一优先级不是"让 AI 多提到我"，而是"让 AI 别把我这一款说成那一款"。前者是加法，后者是止损；在这个品类里，止损的收益远高于加法。因为一次版本串线，会让用户拿着错误预期到店，到店后落差直接转化为不成交，而这一步的损失不会被记录在任何 GEO 报表里。

参数分级：哪些必须零容错，哪些可以放宽

并非所有参数的重要性都一样。按"错一次的后果"把新能源参数分四级，维护优先级就清楚了——这个分级不是按参数在宣传页上的显眼程度排，而是按它对决策的否决能力排。

级别	参数类别	错一次会发生什么	容错度
一级	安全与合规 （电池类型、召回与 OTA 安全公告、辅助驾驶功能边界）	用户据此做出安全相关判断，出错即构成信息披露问题，且可能被监管关注	零容错
二级	成本 （指导价、购置税资格、选装价格、补贴与权益）	用户预算算错，直接放弃该车型；或到店后发现差价，当场流失	零容错
三级	核心性能 （续航、快充功率、能耗、零百加速）	影响横向排序中的位次，通常不会直接否决，但会让推荐理由变弱	低容错（±5% 内可接受）
四级	体验细节 （内饰材质、音响品牌、座椅功能）	影响好感度与"高级感"印象，不改变是否进入候选集	可容错

表 3-2 参数分级：四级容错与错一次的后果

这个分级最直接的应用是资源分配。行业里普遍存在的错配是：**大量精力放在四级参数的包装上，而二级参数长期没人管。**内饰讲得再高级、音响品牌再响亮，都不改变用户"这车超预算了"的判断；而指导价写错一千块、把不在目录的车型说成可减免，就直接否决了整条链路。

一级参数的零容错还带来一个特殊要求：**安全类信息必须与官方公告逐字对齐，不能用营销语言改写。**把"该功能为辅助驾驶，需驾驶员全程监控"改写成"智驾随行，轻松上路"，性质就从"表述不精"变成"功能越界"。这类改写在一级参数上是绝对禁区，而在四级参数上只是风格问题。

三级参数的容错区间也值得注意：±5% 之所以可接受，是因为用户在做比选时用的是"量级判断"而非"精确值"——600km 和 620km 在他的决策里是同一档，但 400km 和 620km 是两档。所以三级参数的关键不是精确，而是**不能跨档。**一个把 620 写成 610 的错误可以容忍，一个把 620 写成 480 的错误不可容忍，因为后者改变了档位。

这条"档位原则"还解释了为什么续航虚标会引发强烈的口碑反噬：CLTC 620km 的车，冬季高速实测可能只有 380km，而 380km 在用户心里属于"另一档"。参数没有说谎，但用户感受到的是跨档落差。这也是第 04 章之后的政策内容必须写"算式"而非"结论"的同一个道理——**凡是会被用户拿去做判断的数字，都必须带上能让他自己验算的条件。**

04 政策变量：为什么车的问题会答成地区的问题

新能源汽车有一个其他消费品几乎没有的特征：**同一个车型、同一个价格，在不同时间、不同省份，用户的实付成本可能相差上万元。**这个差异不来自促销，来自政策。而政策恰恰是 AI 最容易答错、也最容易答得"看似正确"的内容。

2026 年的政策基准线变了

2026 年 1 月 1 日起，新能源汽车车辆购置税政策从"免征"调整为"减半征收"，执行至 2027 年 12 月 31 日，其中每辆新能源乘用车减税额不超过 1.5 万元 [1][3]。这不是一个小调整——对一台 30 万元的车，购置税约 2.65 万元，减半后仍需缴纳约 1.33 万元；对一台 50 万元的车，因为受 1.5 万元上限约束，实际减免就只有 1.5 万元。

更关键的是资格条件：**只有列入《减免车辆购置税的新能源汽车车型目录》的车型才能享受。**这个目录由工业和信息化部会同税务总局分批发布，2026 年已更新至第三十五批 [2]。同时，工信部等三部门对 2026—2027 年的产品技术要求做了单独公告，2026 年起新列入目录的车型须符合新的技术要求 [12]。

这三条信息叠在一起，构成一个 AI 极易答错的复合问题：用户问"这车能减多少税"，正确答案需要同时满足"车型在目录内""购置日期在区间内""减免额不超过上限"三个条件。缺任何一个，答案就是错的。而模型如果只抓到"2026 年减半"这一条，就会给出一个适用于所有车型的泛化答案——对具体用户来说，这个答案的价值为零。

目录制度：一个持续更新的白名单

目录制的特殊性在于，它不是一次性资格，而是**分批、持续更新的白名单。**理解它的运作方式，才能理解为什么政策类答案会随批次变化。

目录由工信部会同税务部门按批发布，每批同时包含新增车型与变更信息；车型进入目录后才能享受减免，退出或变更后资格随之变化 [2]。这意味着"某款车能不能减免"这个问题，在严格意义上有时间戳——问的是"哪一批之后的状态"。

同时，产品技术要求有单独的公告约束：2026 年起新列入目录的车型须符合新的技术要求 [12]。这一层让资格判断从"车型身份"延伸到"技术水平"，例如能耗、续驶里程等指标是否达标。结果是：**同一款车的不同配置，可能因为技术指标差异而出现"部分配置在目录、部分不在"的情况。**

这种"配置级差异"是 AI 答案最典型的错误源之一。用户问"这车能减税吗"，模型如果只匹配到车系名，给出的答案可能对高配成立、对低配不成立。而在实际交易里，低配往往才是走量配置——错误的后果直接落在最大的人群上。

由此有一个具体判断：政策类内容如果只写到"车系"这一级，就天然埋着出错风险；写到"配置 + 目录批次"这一级，才具备可验证性。**颗粒度不是风格选择，是准确性的前提。**

把政策算成钱：三个价位段的实际差额

政策内容之所以必须精确，是因为它一对一对应金额。车辆购置税按"计税价格 × 10%"计，新能源减半即 5%，且每辆乘用车减税额不超过 1.5 万元。按含税指导价折算（计税价格 = 含税价 ÷ 1.13，增值税率 13%），三个价位段的结果差别很大：

含税指导价	计税价格	应纳购置税	减半后应缴	是否触上限	实际减免
15 万元	约 13.27 万	约 1.33 万	约 0.66 万	否	**约 0.66 万**
30 万元	约 26.55 万	约 2.65 万	约 1.33 万	否	**约 1.33 万**
50 万元	约 44.25 万	约 4.42 万	约 2.21 万	**是**（超 1.5 万）	**1.5 万**

表 4-1 三个价位段的购置税实际减免测算

测算依据：2026—2027 年减半征收政策与 1.5 万元减免上限 [1][3][4]；按 13% 增值税率折算计税价格，未考虑地方补贴与厂商权益，实际以税务部门核定为准。

注意 50 万元那一档：如果不设上限，减半应减 2.21 万元；受 1.5 万元上限约束，实际只减 1.5 万元，用户实缴约 2.92 万元。**这个"上限截断"是 AI 最容易算错的地方**——大量答案只写"减半"，不写上限，导致高价车型的答案偏差七千元以上。而这个偏差恰好出现在最敏感的价格谈判环节。

这张表也说明了一个内容原则：**政策内容要写成"可代入的算式"，而不是"结论式的口号"。**写"购置税减半"是口号，只能被复述；写"计税价格 × 5%，单车上限 1.5 万元"是算式，能被用户自己算，也能被 AI 稳定复现。前者的答案在不同车型上会得出

同一个结论，后者才能得出正确的差异化结论。

更进一步，**算式的可复现性本身就是一种护城河。**当全网关于购置税的内容都是"减半"两个字时，一家车企如果给出的是带上限、带折算口径、带价位段示例的完整算式，它在政策类问题上被引用的概率显著更高——因为模型偏好能直接回答"我要交多少钱"这个具体问题的内容。

政策的三重限定：时间、地域、白名单

把新能源相关政策的影响机制拆开，可以看到三重限定同时作用于同一个问题。这三重限定不是并列关系，而是串联关系——任何一重不满足，后面的结论全部作废。

限定维度	作用机制	典型提问	答错的后果
时间限定	政策有生效与退出日期，跨期即失效或减半	"现在买还是等明年"	给出已失效政策的结论，用户预算算错
地域限定	牌照、限行、地方补贴、充电桩政策按省市不同	"我在杭州能上绿牌吗"	答成全国统一口径，用户到店才发现不符
车型白名单	只有列入目录的车型享受减免，批次动态更新	"这款能减购置税吗"	把不在目录的车型说成可减免，构成误导
技术门槛	续航、能耗、电池能量密度等须达新技术要求	"这车符合新标准吗"	旧款达标与否被混淆，影响目录资格判断

表 4-2 新能源政策变量的四重限定与失效路径

政策依据：财政部、税务总局、工信部公告及工信部车型目录公告 [1][2][3][12]。地方性政策未纳入本表，各地差异须以属地主管部门口径为准。

这四重限定解释了一个现象：**在 AI 答案里，"政策类问题"的错误率显著高于"参数类问题"。**参数错了，用户到店一看配置表就能发现；政策错了，用户可能一直到缴税环节才发现，而那时已经签了合同、付了定金，损失是实打实的。

从责任角度看，这也是新能源 GEO 里唯一可能引发合规争议的错误类型。参数写错属于信息披露不准确，政策写错可能构成误导性宣传。两者的风险量级不同，处理优先级也应该不同。

为什么政策问题会"地域化"

新能源的政策差异不只在国家层面，更在省市层面。牌照额度、限行规则、充电桩安装条件、地方性置换补贴，各地规则不同且不定期调整。这意味着同一个问题"我在本地买这车合不合适"，答案在不同城市完全不同。

这给内容生产带来一个结构性难题：**全国统一的内容无法回答地域化的问题。**一家车企如果只做一份全国版的政策说明，AI 在回答"某地能不能上牌"时就只能拿全国口径去凑，结果必然模糊。而如果为每个城市做一份，内容量会爆炸，维护成本极高，且政策一变全部失效。

可行的中间路线是：**把"国家层面确定的部分"和"地方层面变动的部分"在结构上分开写。**国家层面（购置税减半、1.5 万上限、目录资格）写死，标注生效日期；地方层面只写"查询路径与方法"，不写具体结论。这样 AI 在切片时能明确区分哪些是稳定事实、哪些是需要用户自行核验的变量，答案的可靠性反而更高。

反过来说，如果一份内容把"某市可以上绿牌"写成确定结论，而这个城市三个月后调整了规则，这份内容就从资产变成了负债——它不仅自己错了，还会污染 AI 对该品牌所有政策类问题的答案。

05 信源的四层采信权重：从工信部目录到车主论坛

前面三章讲的是"为什么会错"，这一章讲"该信谁"。新能源汽车的 AI 答案不是单一来源的摘要，而是四层信源按不同职能拼出来的。理解这四层的分工，比笼统地说"要发权威内容"有用得多。

四层信源，各管一件事



图 5-1 新能源信源四层金字塔

层级划分参照引擎厂商公开的信源权威度分级思路 [11]，结合汽车品类的实际信源结构整理；宽度表示该层内容体量，不表示采信权重；可信语料的分级与选用思路可参照 GEO 团体标准 [8]。

- **L1 官方口径**解决"合不合法、能不能减税、达不达标"。这一层的独特价值是**不可替代性**——没有任何其他信源能回答目录资格问题。它的短板是颗粒度粗：目录只给车型名称与批次，不给具体配置差异。
- **L2 企业官方**解决"参数到底是多少"。这一层是车企唯一完全可控的战场，也是参数准确率的决定层。它的短板是自证性质——模型对纯自述内容的权重天然低于交叉验证过的内容。
- **L3 行业结构化库**解决"横向怎么比"。垂媒的参数库把多车型字段对齐，是 AI 做对比型回答时最依赖的结构化底座。它的短板是更新时点不稳定，年款切换期最容易滞后。
- **L4 用户口碑**解决"真实体验如何"。这一层在体验型品类是主信源，在参数型品类里权重较低，但对"值不值得买""有没有坑"这类问题仍是主导。它的短板是噪声大、老款数据混入新款、情绪化表述多。

采信权重不是固定的，随问题类型切换

这是本章最关键的一个判断：**四层信源的权重不是固定排序，而是随提问类型动态切换。**同一品牌，在"能不能减税"这个问题上 L1 权重接近绝对；在"续航多少"上 L2 与 L3 并列；在"开着怎么样"上 L4 主导。

提问类型	L1 官方	L2 企业	L3 结构化库	L4 口碑	主导层
政策资格类（能否减免/上牌）	极高	中	低	低	L1
参数查询类（续航/尺寸/功率）	低	高	高	中	L2 + L3
横向比选类（三款怎么选）	低	中	极高	中	L3
体验口碑类（值不值/有什么坑）	低	低	中	极高	L4
售后维保类（保养/质保/充电桩）	中	高	低	中	L2

表 5-1 四层信源权重随提问类型的切换

这个动态性带来一个常见的策略错误：**很多车企把预算全押在 L4（做口碑营销），却没把 L1 和 L2 修好。**结果是在"值不值得买"这类问题上表现不错，在"能减多少税""长续航版多少度电"这类硬问题上持续失分。而硬问题恰恰是用户做减法时的第一道筛子——参数都对不上，口碑再好也进不了候选清单。

权重为什么会随问题切换？根子在切片的语义粒度。模型检索时切出的片段，与问题的语义单元必须匹配：

政策类问题的语义单元是"**资格判断**"——一条是或否的结论加一个金额，答案只需要一段权威文本支撑，多段反而互相干扰。所以这类问题只认 L1，其他层提供的信息在结构上就不匹配。

参数类问题的语义单元是"**字段比对**"——一个字段名加一个值加一个口径。这要求来源具备字段结构，因此 L2 与 L3 是天然匹配层，而叙事型的媒体报道在这类问题里几乎无法被切出可用结构。

口碑类问题的语义单元是"**体验陈述**"——某种场景下的主观感受加具体细节。这类片段只存在于 L4，官方参数页无论怎么写都提供不了"冬天开空调掉多少"这种经验性描述。

理解了这一层，"权重"就不是一个可以被主观设定的参数，而是**问题结构与信源形态的匹配结果。优化工作的实质，是让自家信源在更多问题结构下具备可匹配的形态**，而不是笼统地去"提高权威度"。

反过来也有另一种错误：只修 L1 和 L2，完全放弃 L4。这会带来"参数全对但没有温度"的答案——AI 能准确报出你的参数，但在"20 万家用谁更合适"这种需要综合判断的问题上，仍然倾向于推荐口碑密度更高的竞品。因为综合判断需要"人味"证据，纯参数不构成推荐理由。

L4 口碑的结构：不是越多越好，而是"够不够分层"

口碑这一层常被误解为"声量运营"——内容越多越好、好评越多越好。在新能源品类里这个理解是错的，因为 AI 对口碑的采信不是按总量，而是按**是否覆盖了关键场景。**

用户的真实疑问是分场景的：北方的冬季续航、无固定车位的日常补能、二胎家庭的后排空间、通勤族的市区电耗、长途自驾的高速表现。同一款车的口碑如果全集中在"提车作业、外观好看、动力足"，那么在上述五个具体场景里，AI 找不到可用证据，只能给泛泛结论——而这个泛泛结论在比选里没有竞争力。

所以口碑的有效性取决于分布而非数量。**二十条覆盖五个场景的深度长测，比两百条"很满意"的短评更有用。**因为前者能支撑"这位用户在冬季低温场景下的实测续航是 X"这样的具体陈述，后者只能支撑一个总体倾向。

还有一个反直觉点：负面内容并非全是负债。一条说清"这车高速风噪偏大、但市区静音优秀"的诚实长测，比通篇赞美的内容更容易被采信——因为它提供了边界，而边界正是模型做取舍叙述时需要的材料（见第 06 章）。真正构成负债的是**无法归因的**

负面（"这车不行"却不说为什么）和**错误归因的负面**（把选装问题说成标配问题）。

L2 的真实困境：自证不被采信

这一层值得单独说，因为它是车企唯一能 100% 控制的层，却常常是效率最低的一层。原因在于采信的交叉验证逻辑：模型在评估一条参数时，会看这个值是否在多个独立来源中一致出现。如果只在官网出现，一致度为 1，权重受限；如果在官网、垂媒库、媒体报道三处一致，权重才真正抬起来。

这意味着**官网写对还不够，得让外面的人也写对。**而外面的写法不由车企控制——垂媒库靠抓取，媒体报道靠编辑，两者都会滞后、都会出错。车企能做的是降低它们的出错概率：把参数页做成机器可稳定解析的结构（字段名统一、单位统一、版本标识清晰、更新日期显式），让抓取方拿到的就是对的。

这是一个"向下兼容"的思路：不是要求外部迁就自己，而是把自己的输出格式做到最不容易被误解。参数页里"续航 620km（CLTC）"和"续航 620"这两句话，对人类读者没有区别，对机器解析是天壤之别——前者带口径，后者不带，被切片后极易被当成实测值。

用"缺口诊断"替代"权威度打分"

行业里评估信源质量，习惯给每个来源一个权威度评分。这个方法在新能源品类里不好用，因为权重随问题类型切换（见上表）——同一来源在政策问题上几乎无权重，在口碑问题上又接近主导。一个更可操作的替代方法是"缺口诊断"：不问"这个来源多权威"，而问"这个问题缺哪一层"。

做法很直接：取 20—30 个真实用户问题去问 AI，把每条答案里的事实点标注来源层级，看哪一层缺失或错位。不同症状指向不同缺口，而不同缺口对应完全不同的修复方向。

答案里的症状	指向的缺口	背后真实原因
"AI 说这车不能减购置税，其实它在目录里"	L1 未收录或批次滞后	车型目录批次更新未在语料中同步
"续航给了老款的数"	L2 时点滞后	官网参数页更新晚于改款发布
"三款对比时基本不提我"	L3 字段不全	垂媒参数库缺关键字段，无法参与对齐
"参数都对，但推荐理由是负面的"	L4 样本不足	车主长测内容少，负面向内容占比高
"同一款车在不同平台说法不一"	跨层冲突	四层信源各说一套，模型择高频而非择权威

表 5-2 五类症状与信源缺口的对应关系

症状描述为行业中常见的答案形态归纳，不针对任何具体品牌；诊断方法基于信源分层逻辑设计。

这个诊断法的价值在于**把模糊的"内容不够权威"变成一个可定位的具体缺口。**五类缺口对应的修复动作完全不同：L1 是目录同步，L2 是参数页重建，L3 是字段补齐，L4 是口碑样本运营，跨层冲突则是口径统一。用一套动作去打五类缺口，必然有一半打空——这也是为什么很多车企"投入不少、改善不明显"。

还有一个更隐蔽的诊断维度：**缺口的分布会随车型生命周期变化。**新车上市初期，缺口几乎总在 L2 和 L3——因为官网刚建、垂媒库还没收录，此时做口碑（L4）是无效的；上市半年后，L2 与 L3 补齐，缺口转移到 L4；改款之后，L2 再次成为主要缺口。所以缺口诊断不是一次性动作，需要按车型节奏重做。

L3 结构化参数库：最不显眼、却决定能否参战

四层信源里，L3 是唯一一层"不直接表达观点、只提供可对齐字段"的。它的存在感最低，作用却最容易被低估——因为**它决定了一款车能不能进入横向对比。**

横向对比的前提是"同一字段在多个实体上都有值"。假设一个问题要求比较五款车，如果其中三款在参数库里缺少"快充功率"字段，那么模型在做补能排序时只有两个选择：把这三款排除在补能排序之外，或者把缺失当默认值处理。前者让它们直接失去比较资格，后者会得出错误结论。两种结果对品牌都不利，而原因仅仅是少了一个字段。

这个机制解释了行业里一个常见的不对称现象：**参数最全的车型，往往在比选答案里出现得最稳定**，即使它在单项指标上并不领先。因为它具备"全员可比"的属性，而模型偏好可比性——可比的信息才能生成有说服力的排序理由。

L3 还有一层不易察觉的时效问题。结构化库抓到参数后不会自动感知改款，通常要等到下一轮抓取或下一次人工更新。这中间形成一个时间窗，窗口内旧参数被当作现行值使用，而它看起来"完整、规范、可信"，比官网的滞后更难被识别。所以判断一款车的参数是否可靠，不能只看"有没有值"，还要看值的采集时点与年款标识是否对得上。

由此得到一个反常规的资源判断：对多数车企来说，补齐 L3 的六个核心字段，收益高于在 L4 写十篇口碑长文。前者让车进入可比集合，后者只是在已经能参赛的车上加印象分。顺序反了，投入产出比会差一个量级。

06 对比型提问：AI 如何在三款车之间排序

购车场景里最高频、也最有商业价值的提问形态不是"这车怎么样"，而是"这几款怎么选"。前者是信息查询，后者是决策辅助。二者的答案机制完全不同，对应的优化手段也完全不同。

提问的真实结构：六约束叠加

用户在 AI 里提问时，往往一次性叠加多个约束。一句典型的比选型提问包含六类约束：预算区间、能源形式、续航门槛、空间需求、智驾偏好、补能条件。这六类约束不是并列的，而是**先硬后软**——预算和能源形式是硬过滤，续航和空间是软排序，智驾和补能是加分项。

AI 处理这类问题的顺序是：**约束过滤 → 生成候选集 → 维度对齐 → 排序输出理由。**四步中最容易出问题的是第三步"维度对齐"——因为不同车型的参数命名、单位、测试口径不统一，对齐过程本身就是失真高发区。

六类约束在答案里的作用并不对等，它们分别对应不同的处理方式，也分别对应不同的内容缺口：

约束	属性	用户通常怎么说	AI 处理时的失误点	对内容的要求
预算	硬过滤	"落地 25 万以内"	把指导价当落地价，忽略购置税与保险	须同时给指导价与落地估算口径
能源形式	硬过滤	"只考虑纯电 / 能加油也行"	把增程、插混混作一类处理	须明确动力形式标签
续航门槛	软排序	"续航 500 以上就行"	用 CLTC 值直接比对，忽略工况	须标工况，并给分场景参考值
空间需求	软排序	"后排能坐三个人，要放儿童座椅"	只用轴距判断，不看实际布局	须给轴距、车宽与后排常态数据
智驾偏好	加分项	"有没有高速领航，要不要加钱"	把高配功能当全系标配	须标硬件配置与订阅/选装规则
补能条件	加分项	"家里没桩，附近有快充吗"	只给最大功率，不给补能曲线	须给 30—80% 时间与功率区间

表 6-1 比选提问的六项约束及其失误点

这张表里最值得注意的不是失误点本身，而是失误的性质：**六条里有四条的问题出在"缺限定"，不是"缺数字"。**"续航 500 以上"这条约束，如果内容只给 CLTC 数值而不标工况，AI 就只能用 CLTC 比对，用户拿到车后一到冬天就发现不达标——而错误

在答案生成的那一刻就已经埋下，跟模型能力无关。



图 6-1 比选型提问的四步处理机制

依据生成式检索的常见处理路径整理，为机制示意；实际模型实现存在差异。

这个机制解释了为什么"软文战术"在比选型提问里几乎失效。一篇两千字的评测文章，在切片后可能只贡献两三个片段，而且这两三个片段往往不带完整的约束条件。当 AI 需要做"20 万以内、续航 500 以上、后排能坐三人"的精确过滤时，它需要的是字段化的数据，不是形容词。

结构化对比为什么更有效

一张字段完整的对比表，在 RAG 流程里的价值远高于同字数的散文。原因有三：

1. **可切分**——表格天然按行分块，每行都是一个"车型 + 字段 + 值"的完整三元组，切片后语义自洽，不依赖上下文。
2. **可对齐**——同一张表里的字段名一致、单位一致、口径一致，跨车型比较不需要额外推断，失配概率低。
3. **可验证**——表头带口径说明（如"CLTC"、"含税价"、"含电池"），模型在输出时能带上条件，答案的可信度显著高于无口径数字。

对比之下，散文体参数描述的切片语义依赖上下文——"它比上一代多了 40 公里"这句话脱离上下文后完全不可用。这就是为什么大量内容投入换不来比选场景的可见度：**内容量不等于可检索性。**

字段	建议写法	失配风险点
车型标识	品牌 + 车系 + 年款 + 版本 + 动力形式	只写车系名，导致版本串线
价格	官方指导价，注明是否含补贴与选装	优惠价与指导价混用
续航	标注工况（CLTC/WLTP）+ 电池容量	无工况裸值，被当作实测
补能	最大充电功率 + 30–80% 补能时间	只写"支持快充"，无可比数值
尺寸与空间	长宽高、轴距、后备厢常规/放倒容积	只给车长，无法判断后排
智驾	硬件配置 + 功能边界 + 是否订阅收费	把高配功能写成全系标配

表 6-2 比选场景六个核心字段的写法与失配风险

字段选择依据比选型提问中最常被引用的参数项整理，失配风险来自字段标识缺失而非数值错误。

表 6-2 里的六个字段，是比选型提问最常被引用的六项，也是失配风险最高的六项。可以看到，每一行的失配风险都不在"数据错误"，而在"标识缺失"——价格没标含不含补贴、续航没标工况、智驾没标是否收费，这些缺失在人类阅读时可以被上下文补充，在机器切片时无法补充。

所以新能源 GEO 在比选场景的核心动作，可以浓缩成一句话：**把该标的都标上。**这不是写作技巧问题，是数据结构问题。一家车企如果能把六张对比表（按价位段分）做到字段完整、口径统一、版本清晰，它在比选型提问里的命中率会显著高于同等预

算下做十篇评测软文的对手。

比选答案里的"取舍叙述"更值钱

还有一个细节值得注意：AI 在排序之后，通常会补一句取舍说明，比如"如果更看重后排空间选 A，如果更看重智驾选 B"。这句话的形态，取决于语料里是否存在明确的"取舍型表述"。

如果全网关于你车型的内容都是"全面领先""同级最强"，模型无法提炼出取舍理由，只能给一个模糊的推荐；如果语料里有清晰的"优势项与妥协项"对照，模型就能把它转成取舍建议，而这个建议往往直接决定用户选谁。

这意味着适度的自我限定反而提升被选中的概率。承认"这款车在空间上不如同级、但在补能效率上更好"，会让 AI 在"看重补能"的子问题里明确推荐你。反过来，声称全面最优的内容，在任何一个子问题上都不构成强推荐理由。

对比的三层粒度：车型级、版本级、配置级

"做对比内容"这句话本身太粗。用户在不同粒度上提问，需要的对比结构完全不同。按粒度分三层，投入价值的排序和常见认知恰好相反。

车型级对比（"Model Y 和宋 PLUS 怎么选"）是入口级问题，搜索量最大、竞争最激烈、内容供给最多。它的可见度收益高，但边际价值随竞争迅速下降——所有人都在做，答案的排位取决于参数完整度而非内容量。

版本级对比（"长续航版和标准版差 3 万值不值"）量小得多，但转化意图最强。问这个问题的人已经在算账，离下单只差一步。而这一层恰恰是 AI 最容易串线的位置——因为版本差异只体现在参数表几行里，切片时极易被合并。所以**版本级对比是"可见度"与"准确率"同时受益的一层**，性价比最高，却最常被忽略。

配置级对比（"这个选装包要不要加"）量最小，但它是成交前的最后一道关。这一层的内容通常不在公开语料里，而在销售话术和配置表细则里，AI 极难覆盖。它的真实价值是降低到店后的沟通成本，而不是获取新客。

粒度	典型问法	搜索量	转化意图	AI 串线风险
车型级	"这两款怎么选"	大	中	中（不同车系，实体差异明显）
版本级	"差这几万值不值"	小	高	**高**（同车系内，实体差异弱）
配置级	"选装包要不要加"	极小	极高	高（多为选装项，参数表常缺）

表 6-3 对比的三层粒度与各自的串线风险

一个反直觉：参数堆得越全，越可能不被选

既然结构化对比有效，是不是参数堆得越全越好？不是。这里有一个反直觉的边界：**AI 的输出长度是有限的。**一张四十行的参数对比表，在被引用时会被压缩成几行摘要，而压缩过程会优先丢掉"非典型项"——也就是你区别于竞品的那些独特优势。

结果是：参数越全的表，被压缩后剩下的越是"大家都有的项"（价格、续航、尺寸），你真正想突出的差异项反而消失。反倒是八到十项关键参数的对比表，能完整进入答案，把差异项保住。

所以对比表的设计原则是"少而关键"，判断标准很朴素：**用户做减法时真的会用它吗。**续航、空间、补能、价格、智驾边界这五项是减法依据；氛围灯、香氛、音响品牌不是。把非减法项塞进主表，等于用自己的差异项换了一张更长的表。

这条边界还有一个延伸含义：不同价位段的"关键项"不一样。十五万级用户最关心的是续航真实度与使用成本；三十万级用户关心的是空间、智驾与补能效率。同一张表套用到所有价位段，等于在所有价位段都用错权重。这也是为什么对比内容必须按价位段分册，而不能做"全系对比一张表"。

为什么答案常常只给三款：候选集的容量约束

观察大量比选型问答会发现一个稳定现象：AI 通常给出三款左右的推荐，很少给到十款。这不是模型偷懒，而是三个约束共同作用的结果。

第一个约束是**解释成本**。每增加一款推荐，就要多写一段理由，而答案长度有限。三款时每款可以分到两三句，十款时每款只剩半句，退化成清单式罗列，用户反而没法用。模型在可读性上的默认取舍，把候选集压到了三到五款。

第二个约束是**可对齐性**。多一次横向对比，就多一层字段对齐要求。参与比较的车型越多，所有车型都有值的公共字段越少，能支撑排序的维度越窄。三款车通常还能在续航、价格、空间三个维度上都有值；十款车往往只剩价格可比。

第三个约束是**置信度阈值**。模型倾向于推荐它能说清楚的车。某款车如果在语料里只有零散提及，缺参数也缺评测，它就算被检索到，也很难被写进推荐位——因为写不出理由。于是候选集实际被"可描述性"筛过一遍。

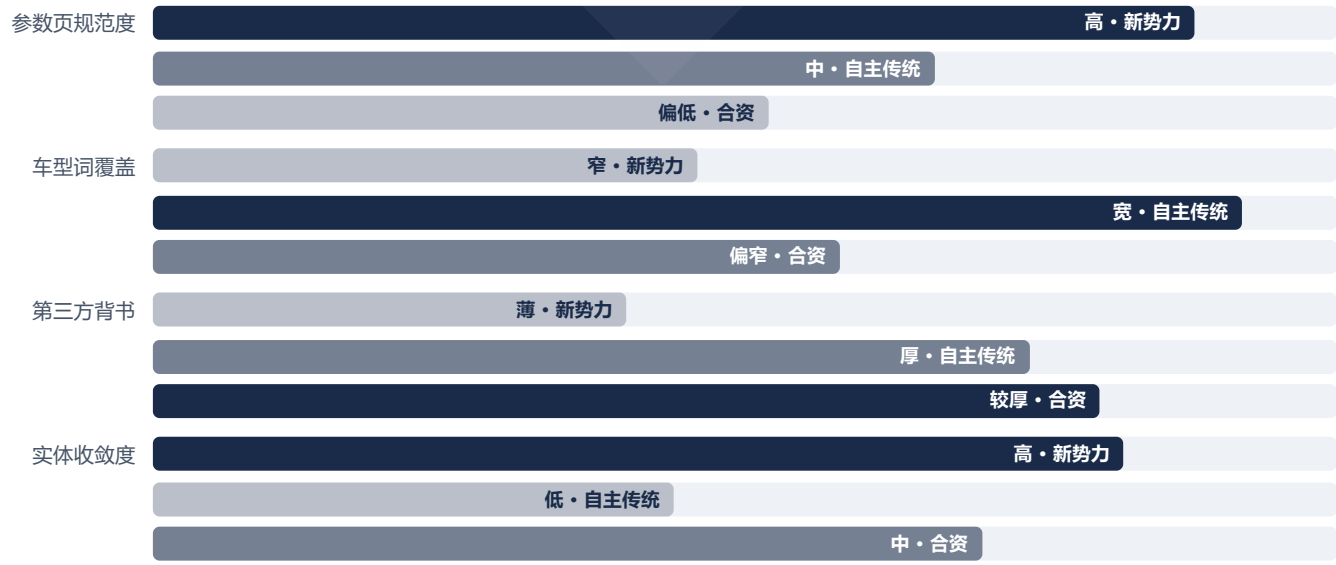
三个约束合起来，解释了新能源 GEO 最关键的一道门槛：**不是"被检索到"，而是"被认为可描述"**。前者靠语料覆盖，后者靠结构化字段加第三方证据。很多车型卡在第二道门槛上——它在答案里出现过，却从未进入推荐位。

这也给"候选集进入率"提供了理论依据：它测的正是第二道门槛，而不是第一道。只看提及率，等于只测"有没有被找到"，漏掉了决定商业结果的那一步。

07 三类车企的起点差异：谁先被读对

同样是新能源车企，做 GEO 的起点差异极大。这个差异不取决于预算规模，而取决于三件事：**数据原生程度、车型矩阵宽度、第三方背书厚度**。按这三件事分，市场上的玩家大致落在三类位置上。先看四张"体检单"上的相对位置：

图 7-1 三类车企在四个关键指标上的起点差异



口径说明：本图为定性对比示意，条长表示相对强弱，不含实测分值；判断依据为新势力/自主/合资三类品牌在车型矩阵、渠道结构、公开信源密度上的普遍差异。

新势力：数据原生，但沉淀期短

新势力品牌的优势在于天生是"数据原生"的——直营模式、线上下单、配置可选、软件可升级，整套购买流程从第一天起就在线上完成。这带来两个直接好处：官网参数结构相对规范，用户行为数据完整，社群口碑密度高。

它的短板也明确：**历史沉淀短，第三方权威背书薄。**一个新品牌可能只有两三款在售车型，垂媒参数库的收录深度有限，行业媒体长期评测少，车主长测样本不足。在 AI 眼里，这是一个"信息新鲜但密度不足"的实体，容易被提到，但难以被详细描述。

所以新势力的 GEO 重点不在广度，在深度：把少数几款车的参数页面做到极致完整，把车主长测口碑做成结构化资产，主动补齐第三方评测与权威测试的引用来源。用少而深对抗多而散。

自主传统车企：矩阵宽，实体容易散

自主传统车企的优势是规模：车型矩阵宽、价格带覆盖全、渠道下沉深、保有量大。在 AI 答案里，这类品牌的"存在感"通常很高——因为它几乎在每个价位段都有产品。

但这个优势在 GEO 里会反噬。第 01 章提过的"实体漂移"在这里最容易发生：同品牌十几款车、每款多个版本、命名风格相近，实体标识稍有不一致，AI 就会跨款串线。更麻烦的是**渠道与内容分散**——总部、区域、经销商各自发布内容，同一款车在不同渠道的参数表述可能不一致，形成内部噪声源。

这类车企的 GEO 重点应该是"收敛"而不是"扩张"：统一实体标识规则、统一参数页模板、统一对外口径，把分散的内容源收成一条主线。在这类车企里，**治理内部不一致的收益，通常高于新增外部内容。**

合资品牌：品牌词强，车型词弱

合资品牌在 AI 答案里呈现一个典型的不对称：品牌层面的提及仍然稳定，但具体新能源车型的可见度明显偏弱。原因是产品节奏——合资的电动化产品线相对少，上市时间晚，车型词在语料里的积累不足。

这会形成一种"被顺带提及"的答案形态：AI 在回答"20 万左右的新能源 SUV 推荐"时，可能完全不提合资车型；只有在用户明确问"某个品牌的电车怎么样"时才会展开。也就是说，它守住了品牌搜索，丢掉了品类搜索。

这类品牌的 GEO 重点在"抢品类词"——围绕具体价位段和车身形式，把车型参数、对比关系、适用场景做实，让车型词进入 AI 的候选集。品牌资产的存量不能自动转化为车型词的可见度，这一步需要单独投入。

把三类位置叠在一起看，会发现一个共同结论：**每类车企的 GEO 最优动作，恰好是它当前最弱的那个指标。**新势力最弱的是第三方背书，自主最弱的是实体收敛，合资最弱的是车型词覆盖。这解释了为什么"照抄别人的 GEO 打法"通常无效——别人的起点和你的短板根本不在同一个位置。



图 7-2 三类车企在"数据原生程度 × 车型矩阵宽度"上的位置

定位图为定性示意，坐标表示相对强弱而非实测分值；纵轴指线上数据完整度与官网规范化程度，横轴指在售车型与版本数量。

这张定位图补充了前一张条形图没说的信息：条形图看的是"单项强弱"，定位图看的是"组合位置"。同一款车的强弱组合不同，需要的动作顺序也不同。左下象限（数据原生低、矩阵窄）是最危险的位置——既没有可对齐的结构化数据，也没有足够多的车型形成规模，AI 答案里几乎不会有稳定表现。这类品牌通常体量小、或刚进入新能源，问题不在投入多少，而在"有没有可被解析的实体"。

起点差异为什么不会自然收敛

三类车企的起点差异，看起来会随市场竞争自然抹平，实际上会分化。原因在于 GEO 的核心资产都带有累积与折旧的双重属性，不是一次性投入，而是有存量、有利息、也会贬值的账户。

先看**第三方背书**。一家车企今年拿到五家权威媒体的深度评测，明年这些内容仍在语料里，并且会成为新车型的"信任底色"——AI 回答新车问题时，会引用同品牌的既有评测作为佐证。没做过这件事的车企，明年只能从零开始。这个差距一年拉开，三年后基本无法追平，因为权威媒体对同一品牌的评测次数是有限资源，先进入者已经占住了位置。

实体清晰度的属性不同，它更接近基础设施：一旦建立（统一命名规则、版本标识、口径规范），此后每上一款新车都自动受益；一旦混乱，每上一款新车都在加重混乱。它的特殊之处是**会被污染**——一个错误的版本对应会被反复引用，引用本身又生成新的语料，结果越传越牢。清理这类污染的难度，远高于从未出错。

参数一致性则是典型的折旧型资产。即便今天做到全网一致，改款与 OTA 之后仍需重新维护。它无法一次性完成，只能按节奏续费；中断一次，前面的一致性积累就开始反向损耗。

资产类型	累积特征	折旧特征	典型失效方式
第三方背书	强累积，越早进入越占优	缓慢折旧	评测内容随年款过时，需新一代补位
实体清晰度	累积且可能被污染	不易折旧	一次错误对应被反复引用，形成新语料
参数一致性	弱累积，近似按周期续费	快速折旧	改款、OTA 后未同步，逐月衰减
结构化语料厚度	累积	中等折旧	竞品补齐同字段后，厚度优势消失

表 7-1 四类核心资产的累积与折旧属性

四类资产的属性差异，决定了三类车企的路径不可能相同。**GEO 的竞争不是当年的投入竞争，而是资产存量的竞争。**这解释了一个反常识的现象：头部品牌即使短期减少投入，AI 答案里的表现也不会立刻下滑，因为存量资产还在起作用；新品牌即使短期高投入，也很难立刻追上，因为资产需要时间沉淀，而时间无法加速。

这个判断还有一层现实含义：对资源有限的车企来说，最不该做的是三类资产同时铺开。累积型资产需要长期连续投入才能见效，一旦中断前功尽弃；折旧型资产必须持续续费。三类同时做，等于把有限资源摊薄到三条都需要连续性的线上，最终没有一条形成规模——这比只做一条、做透一条的结果更差。

被读对的先后顺序：为什么不是"谁卖得多谁先被读对"

在 AI 答案里被"读对"的速度，和销量排名并不完全同步。观察到的顺序大致是：单一爆款车型 → 车系 → 品牌，而不是反过来。原因在于模型处理实体时是从具体到抽象的——它先确认"这款车是什么"，才能归纳"这个品牌如何"。

这就解释了一个常被忽略的现象：一款月销数万的爆款车，在 AI 答案里往往被描述得相当准确；而同一个品牌的其他车型，即使同平台同架构，描述质量会明显下降。因为语料是围绕爆款聚集的，其他车型只有零散提及。品牌层面的准确度，本质上是旗下车型准确度的加权结果，而不是独立积累的。

车型类型	读对速度	主要瓶颈	典型表现
单一爆款	快	版本差异（同车系多配置易混）	参数大体准确，版本细节易串
走量车系	中	实体收敛（多版本命名相近）	能进候选集，排序理由弱
小众车型	慢	语料厚度（评测与长测不足）	被提及即笼统，缺乏可比字段
新上市车型	最慢	时点滞后（参数页与库均未建好）	常被与上一代混淆，或干脆缺席

表 7-2 四类车型被读对的先后与瓶颈

这张表有一个实践含义：**品牌的 AI 表现不能靠"品牌层面发力"改善，只能靠"车型层面逐个做对"累积。**这也解释了为什么有些品牌在品牌词上表现良好、在品类词上几乎缺席——它把资源投在了品牌资产上，而品牌资产的存量无法自动转化为具体车型的可比字段。

另有一个时间维度的观察：新上市车型的窗口期很短。上市后三到六个月内，语料结构基本定型——哪些评测被收录、哪些参数进入结构化库、哪个版本被当作"代表款"传播。窗口期结束后再想改，成本会成倍上升，因为已有语料形成了稳定的引用习惯。**这意味着新车上市的传播节奏，实际是在给未来三年的 AI 答案定调。**

08 度量：新能源 GEO 该看什么、归因断在哪

度量是新能源 GEO 最容易被糊弄的环节。因为"AI 提及率"这个指标足够直观、足够好卖，于是大量服务把报告止步于此。但对车企来说，提及率几乎不构成决策依据——**被提到不等于被说对，被说对不等于有人来**。这三个层次必须分开度量。

三层指标：可见度、准确率、到店

层级	指标	测什么	为什么单看会误判
第一层 可见度	车型词提及率	核心问题集里被点名的比例	被点名可能是作为"反例"出现
	候选集进入率	进入 AI 推荐清单的比例	进入清单但排末位，价值有限
第二层 准确率	参数正确率	关键字段与官方口径一致的比例	最容易被忽略，但决定到店落差
	版本对位率	被描述的版本与实际在售版本一致	串线错误在此暴露
第三层 转化	AI 来源到店量	可归因的到店与试驾预约	归因链路不稳，易被高估或低估
	品牌词搜索增量	做 GEO 前后品牌词搜索变化	受促销与新车上市干扰大

表 8-1 三层度量指标与单看时的误判

三层指标不是并列关系，而是递进关系。**第一层涨而第二层不涨，是最危险的组合**——意味着品牌在 AI 答案里出现得更频繁，但错得也更频繁。这种"高可见度、低准确率"状态，会把流量引到错误的预期上，到店落差反而扩大，转化率下降。

这也是为什么新能源品类的第一 KPI 应该定为**参数正确率**，而不是提及率。参数正确率是唯一一个既能被客观核验、又与成交直接相关的指标：它不需要归因模型，只需要把 AI 答案里的字段与官方参数页逐项比对。

归因断在哪：从答案到店门的那一步

新能源 GEO 的归因难点不在技术，在场景——用户在手机上问 AI，然后关掉手机、开车去 4S 店。这中间没有点击、没有 cookie、没有落地页。链路断在"人"这个环节上。



图 8-1 新能源 GEO 的归因断点

断点指从线上答案到线下行为之间缺少可追踪的数字化触点，这是汽车品类区别于电商品类的核心度量难点。

补齐这个断点，行业里目前有三条路径，各有明确的适用边界：

- **专属预约码**——在 AI 可引用的内容里嵌入专属预约入口（如"凭此页预约到店可享专属礼"），到店核销即归因。优点是可精确归因，缺点是只有主动预约的人能被捕获，覆盖不到"直接走进店里"的用户。

- **到店问卷**——在接待流程里加一问"您是通过什么渠道了解这款车的"。优点是覆盖面广，缺点是依赖销售执行，且存在社会期许偏差（用户可能不愿承认是被 AI 说服的）。
- **品牌词搜索增量**——用"AI 可见度提升期"与"品牌词搜索量变化"做时序对照。优点是全量覆盖，缺点是干扰因素多，新车上市、官降、促销都会污染信号，只能看趋势不能看单点。

三条路径应该组合使用而不是择一：**专属码看精确归因的小样本，问卷看结构性比例，搜索增量看整体趋势。**三者交叉验证，才能得出一个可用于决策的结论。只用其中一条，得到的数字一定偏。

三条路径的偏差方向不一样，这一点比"用哪条"更重要。知道偏差朝哪偏，才能判断该信什么、该只看什么。

归因路径	系统性偏差	为什么会偏	该用它判断什么
专属预约码	低估	只捕获主动扫码预约的人，走进店里未扫码的用户全部漏掉	看相对变化与车型间对比，不看绝对占比
到店问卷	高估或错记	社会期许偏差（把"AI 推荐"记成"自己研究"即低估，反之则高估）；依赖销售执行质量	看渠道结构比例，不看单点数值
品牌词搜索增量	噪声大	新车上市、官降、促销、竞品舆情都会污染信号	只看趋势方向，不看单周波动

表 8-2 三条归因路径的系统性偏差

参数正确率怎么算：一个可复核的口径

"参数正确率"听起来抽象，实际上可以定义得非常具体：**正确率 = 一致字段数 ÷ 抽检字段总数。**抽检范围由三个维度交叉确定——问题集（20—50 个真实提问）、平台（主流 AI 搜索平台）、字段（按第 03 章的四级参数，覆盖一级与二级全部、三级核心项）。

举例说明可复核性：50 个问题 × 5 个平台 × 6 个关键字段 = 1500 个数据点。逐点比对 AI 答案与官方参数页，一致计 1，不一致或缺失计 0，总分即为当期正确率。这个口径的好处是**不依赖任何模型判断，可被第三方复核，也不需要归因**——它测的是事实一致性，不是效果。

更有价值的是它能定位原因。如果一级参数（安全与成本）的正确率低于二级（性能），说明问题出在 L1/L2 信源的同步；如果只有某一个平台的正确率偏低，说明问题在该平台的语料覆盖；如果只有某个版本的参数出错，说明实体标识出了问题；如果所有字段正确率都低，那通常不是内容问题，而是这款车在语料里本来就几乎没有可解析的年款信息。四种分布，对应四种完全不同的处理方向。

这个指标还有一个特性：**它会随改款自然下降，不会停滞。**这意味着它可以当作一个"体检指标"按月度看——正确率连续两个月下降且未见回升，基本可以判定为改款或 OTA 后未同步，而不是内容质量问题。这比等用户投诉、等销售反馈要早得多。

度量的时间颗粒度：为什么月度看比周度看更可信

度量的第二个易错点不是指标选错，而是采样频率选错。AI 答案本身有随机性——同一个问题在不同时间、不同措辞下，返回的候选集可能不同，引用段落也可能不同。这种波动在周度采样里会被放大成"趋势"，导致误判。

一个实用的经验是：**周度看异常，月度看趋势，季度看结构。**周度只适合发现突发问题（比如某平台突然把你写错），不适合判断好转或恶化；月度采样量足够抵消随机性，能看出方向；季度才适合判断"候选集进入率"这类结构性指标，因为它需要跨问题、跨平台累积。

还有一层采样设计问题：问题集必须固定。行业里常见的错误是每次监测换一批问题，然后比较两次的分数——这等于换考卷比成绩，没有意义。正确做法是维护一个稳定的核心问题集（比如 30 个覆盖七类场景的问题），只在车型或政策变化时做增补，且增补项单独统计。

跨平台差异同样要处理。不同 AI 平台的语料来源、更新频率、引用策略都不同，同一个品牌在不同平台的准确率可能差十几个百分点。这不是异常，是常态。**正确做法是分平台看基线，而不是把所有平台混成一个总分**——混算会让高分平台掩盖低分平台的问题，而用户是分散在各个平台的。

采样周期	能看出什么	看不出来什么	典型误用
周度	突发错误、明显异常	趋势方向（波动大于信号）	用两周数据宣布"效果提升"
月度	准确率方向、可见度变化	结构性迁移（样本不足）	用总分掩盖单平台恶化
季度	候选集结构性变化、竞品相对位次	短期活动影响	把促销期波动当趋势

表 8-3 采样周期能看出与看不出的东西

一个不该被计入的指标

行业里常见一种做法：把"AI 答案里出现品牌名的次数"作为核心 KPI，按次数计费。这个指标在新能源品类里几乎没有决策价值，原因有三。

第一，出现次数与商业价值不相关。在"10 万级代步车推荐"里出现一次，和在"30 万级家庭 SUV 怎么选"里出现一次，价值差好几倍，但计次口径无法区分。第二，出现次数可以被低成本制造——大量铺 UGC 内容就能把频次抬上去，而这恰恰会加剧第 03 章说的版本串线。第三，频次指标无法反映错误，一个被反复提及但参数全错的品牌，频次数据会很漂亮。

更可用的替代口径是"**有效提及**"：在规定的问题集内，品牌被提及**且**关键参数与官方口径一致**且**版本对位正确，三者同时满足才计一次。这个口径严苛得多，但它和到店转化之间的相关性，远高于裸频次。

一份可用的月度监测表该有什么

把前文的判断落到一张表上，监测才具备可执行性。表的设计原则是"每一行都能被独立验证"，而不是"每一行都能被汇总成一个好看的总分"。

指标	口径	采样方式	异常该如何读
有效提及率	提及且参数正确且版本对位的比例	固定问题集 × 分平台 × 月度	下降但裸提及上升，通常是新内容引入噪声
参数正确率	一致字段数 ÷ 抽检字段总数	一级二级全字段，三级核心项	连续两月下降，多为改款或 OTA 未同步
候选集进入率	进入"推荐清单"而非仅被提及的比例	比选类问题单列统计	低但提及高，说明只能作为反例出现
口径完整率	答案中带工况/含税/配置范围的项占比	按字段抽检	偏低说明自有内容把条件写进了脚注
到店归因量	专属码核销 + 问卷结构比例	门店口径，月度汇总	与可见度背离，多为价格或产品问题

表 8-4 月度监测表的五项核心指标

口径为本文提出的建议框架，非行业统一标准；"异常该如何读"一列用于区分指标之间的背离，避免单指标误判。

五项指标里，最后一项最难取数，却最能校准前四项的意义。如果有效提及率、参数正确率都在改善，到店归因量却不动，说明改善发生在"已经不会来店的人群"里——比如问题集覆盖的是泛咨询人群，而不是临近决策的人群。**这时正确的动作是收紧问题集，而不是加大内容投放。**

反过来，如果到店量上涨但参数正确率没变，通常意味着流量的增加来自与参数无关的入口（比如品牌活动、线下广告溢出），这类增长不该记在 GEO 账上。把两者混在一个报表里，会让团队误以为参数治理起了作用，从而在下一轮把资源投错方向。

09 边界与风险：参数合规、OTA 漂移与平台封闭

前面八章讲的是机制与方法，这一章讲边界。新能源 GEO 有三类风险不来自竞争，而来自品类本身的属性——它们不会让效果变差，而会让投入变成负资产。

风险一：参数宣传的合规边界

汽车是强监管品类，参数宣传本身受广告法与相关法规约束。做 GEO 时容易出现的一类问题是：为了让 AI 更容易引用，把参数表述做得更"抓人"——比如把 CLTC 续航写成"续航 700km"而不标工况，把特定条件下的智驾能力写成"全场景自动驾驶"，把选装配置写成"标配"。

这类表述在 SEO 时代可能只是转化率优化，在 GEO 时代会被放大：因为 AI 会把它原样引用进答案，等于以平台名义替品牌做了一次公开承诺。一旦用户据此产生损失，责任归属比传统广告更复杂——品牌方、内容发布方、平台方三方都可能在链条上。

安全做法是把"条件"写进句子本身，而不是写在脚注里。因为脚注在切片时会被丢掉，条件写在句子里则不会。一句"CLTC 工况 700km，高速实测约 520km"比"续航 700km（详见说明）"更安全，也更容易被 AI 正确引用。

把这类风险落到具体形态，可以看到三种高频问题，它们的共同点比差异更值得注意：

第一种是工况省略。把"CLTC 620km"写成"续航 620km"。字面不算虚假，但在用户理解上构成了对实测能力的暗示，且这个暗示在 AI 复述时会被进一步放大——模型倾向于保留数字、丢掉括号，结果条件被剥离得比原文更彻底。

第二种是功能越界。把"高速领航辅助"写成"自动驾驶"，把"部分城市场景可用"写成"全场景可用"。这是智驾类内容里最高频的问题，也是监管关注的重点区域。它的特殊性在于：功能边界本身随 OTA 变化，旧内容里的旧边界，在新版本下可能从"准确"变成"越界"。

第三种是配置错位。把选装配置、高配专属配置写成全系标配。多源于内容生产时对照了高配参数表，但发布时未标注适用范围。这类错误在比选环节伤害最大——用户按标配预期算价，到店发现要加钱。

三种形态的共同点是：**错误都不在数字本身，而在数字前面的限定词被省略了。**所以修复方向也一致——把限定词写进主句，而不是写进括号、脚注或表格备注。这条原则在合规上是底线，在 GEO 效果上同样是有效做法：带条件的长句比裸数字更容易被完整切片。

风险二：OTA 导致的答案漂移

这是新能源独有的风险。燃油车时代，一款车出厂后的参数基本固定；新能源车通过 OTA 可以改变续航、动力输出、辅助驾驶能力甚至充电策略。这意味着**关于一款车的"正确答案"会随时间变化，而存量语料不会自动同步。**

结果是：一次 OTA 之后，AI 答案可能同时存在新旧两套参数，且模型无法判断哪个是当前版本。用户看到的是"这款车续航 620km，另有版本为 580km"这种自相矛盾的表述，信任度下降，而品牌方对此毫无察觉——因为监测系统只统计"有没有被提到"，不统计"提到的对不对"。

应对办法是建立**版本化监测**：把监测结果与"当前在售版本参数"绑定，而不是与"历史参数"绑定。每次 OTA 或年款切换，触发一次参数页全量校验，并观察 AI 答案的收敛周期。这个周期的长度本身就是一项重要指标——收敛慢，说明语料的更新传导机制不畅。

收敛周期的长短，由三个变量共同决定，而这三个变量的可控程度差别很大：官方参数页的更新时点（完全可控）、第三方参数库的抓取与更新频率（影响可控）、存量旧内容对新内容的干扰强度（部分可控）。三者里最有效的是第一项——**官方更新时点每提前一周，整条链路的收敛就会提前，因为所有下游抓取都以官网为基准。**

反过来，如果某款车的收敛周期异常长（比如超过三个月仍未稳定），通常不是同步速度问题，而是结构性障碍。两种情况最常见：一是旧内容的存量远大于新内容，新内容在频次竞争里根本压不住旧内容；二是旧内容集中在高权重信源上（例如权威媒体的深度评测），这类内容被引用一次就长期有效，且不会随官网更新而修正。**遇到第二种情况，仅靠官网更新完全无效**，必须在新内容里主动做版本声明（例如"2026 款参数已更新，2025 款数据见历史页"），用显式的版本锚点去切割旧语料的适用边界。

这也是新能源 GEO 与常规内容运营最不一样的一处：常规运营的工作对象是"要发什么"，新能源的工作对象还有"要改什么、要划清什么"。删除和修订在这品类里不是收尾动作，而是与新增同等权重的主线动作。

风险三：平台信源的封闭化

第三个风险来自平台侧。目前主流 AI 搜索仍在开放引用外部信源，这是 GEO 得以成立的前提。但如果平台转向"只引用自有生态内容"或"品牌合作付费展示"，外部优化就会部分失效。

这个风险在汽车品类尤其值得警惕，因为汽车是高客单价品类，平台有充分动机做商业化闭环——一旦 AI 答案里的车型推荐可以售卖，那么"参数正确"就不再是入选的充分条件。

判断这个风险是否临近，可以看三个信号：平台是否推出品牌合作类产品、是否收窄外部信源引用范围、是否在答案中增加"推广"标识。这三个信号任一出现，都意味着外部优化的相对价值下降，预算结构需要相应调整。

三情景下的投入节奏

情景	前提条件	GEO 的相对价值	投入节奏建议
乐观	AI 搜索在购车决策中继续深化，平台保持开放信源，参数型答案仍是主要形态	高，参数资产复利明显	前置投入实体治理与参数页建设，形成长期资产
中性	AI 介入率稳定在当前水平，平台部分开放，线下决策权重不变	中，属基础配置	按车型节奏分批推进，优先在售主力车型
悲观	平台信源封闭并商业化，外部优化失效，答案由付费位主导	低，仅剩实体与合规价值	把投入收缩到参数一致性与合规表述，不做规模扩张

表 9-1 三情景的前提条件与投入节奏

三个情景里，有一件事在任何情景下都不会失效：**把参数说对、把版本分清、把政策标清。**这三件事即使平台封闭也仍然有价值——它是品牌对外信息一致性的基础，也是合规底线。区别只在于，乐观情景下它带来可见度收益，悲观情景下它只带来风险规避收益。

需要说明的是，上面三栏是条件推演，不是概率预测。但从当前可观察的信号看，可以给一个粗略的权重判断：中性情景概率最高，因为它的三个前提——AI 介入率稳定、平台部分开放、线下决策权重不变——目前都在成立；乐观情景只需要一个条件不逆转，即平台持续开放外部信源；悲观情景的触发点则十分集中，平台商业化闭环一旦落地，单一事件即可引发整体重估。

这个权重分布的含义是：**不要把悲观情景当作可以忽略的尾部风险。**它的概率不高，但一旦发生，前期基于"开放信源"假设搭建的全部可见度运营会同时失效，而不是逐项退化。这是结构性风险的特征——不是缓慢下滑，而是断崖。也正因如此，第 08 章才强调"参数正确率"而非"提及率"作为第一指标：参数治理在任何情景下都不失效，而提及率只在开放情景下成立。

所以新能源 GEO 的投入结构，应该把"实体与参数治理"和"可见度运营"分开看：前者是无论情景如何都要做的基础设施，后者是随情景调整的弹性投入。把两者混在一个预算池里，会让基础设施在收紧时最先被砍掉——而那恰恰是最不该砍的部分。

风险类型	触发条件	对 GEO 的实际影响	可观察的前置信号
参数合规	限定词被省略（工况、功能边界、配置范围）	答案被原样引用成品牌承诺，责任链条变复杂	AI 答案里的参数比官网更"干净"，条件消失
OTA 漂移	软件更新改变续航、动力或智驾能力	新旧值并存，答案自相矛盾，信任度下降	同一问题的答案在不同周次出现两个数
平台封闭	平台转向自有生态引用或付费位	外部优化的相对价值整体下降	答案里的引用源向单一生态集中

表 9-2 三类风险的触发条件与前置信号

信号为可观察的答案形态变化，不代表风险必然发生；判断时应结合平台侧产品公告一并看。

三类风险的共同特征是：**它们都不表现为"效果变差"，而表现为"投入变成负资产"。**参数失合规，会让更多引用变成更多责任；OTA 漂移，会让更多旧内容变成更多矛盾源；平台封闭，会让更多可见度投入沉没在一条即将失效的通道上。这跟常规的"效果衰减"完全不同——效果衰减是白花钱，负资产是花钱增加了风险敞口。

叠加形态：三类风险很少单独出现

三类风险在实际项目里几乎不会孤立发生，它们会互相触发，形成一条加速的链条。最常见的一种叠加是"改款叠加 OTA"：年款切换时参数页重构，同时软件推送改变了续航与智驾边界，此时旧内容尚未清理、新内容刚上线，两类语料同时参与竞争。结果不是"准确率略降"，而是"同一款车在答案里出现两套互相矛盾的参数"，用户看到的印象是品牌自己都说不清。

第二种叠加是"平台封闭叠加合规收紧"。平台开始只引用自有生态内容，同时监管对外部宣传口径的要求趋严。两件事同时发生时的效果是：车企既失去外部可见度，又因为过去为可见度做的激进表述承担额外责任。也就是说，前期越"用力过猛"，在这种叠加下受损越大。

第三种叠加最隐蔽：实体混乱叠加平台封闭。当平台收窄引用范围、更多依赖结构化数据做合成时，实体标识不清的品牌受损最重——它原本靠语料数量获得的曝光，在新规则下不再有效，而它又缺少可被稳定对齐的实体。**这说明实体治理是唯一在所有情景下都正向的投入。**

叠加形态的实用价值在于：判断风险时不该逐项问"这个风险会不会发生"，而该问"哪两项同时发生会让局面不可逆"。前者得到一串孤立概率，后者得到一个优先级——先处理会放大其他风险的那一项，通常就是实体与参数治理。

信号的时间尺度：哪些是领先的，哪些是滞后的

三类风险的信号并不同时到达，它们有先后。分不清先后，就会出现两种典型误判：把滞后信号当成新问题处理，或者对领先信号视而不见，等到滞后信号出现时已经没有调整空间。

领先信号在平台侧和产品侧。平台是否推出品牌合作产品、是否收窄外部信源引用范围、是否在答案里加入推广标识，这些变化先于数据变化出现，通常在可见度下滑前一个季度就能看到。产品侧的信号是版本与参数的变更公告——年款切换、OTA 推送、配置调整，这些动作一旦排期，参数失真就进入了必然发生的通道。

滞后信号在监测数据里。准确率下降、提及率波动、候选集进入率变化，这些指标反映的是过去几周已经发生的语料冲突，属于"确认已经错了"，很难用于提前应对。把滞后信号当作预警系统，等于用后视镜开车。

中间层是内容侧的可观察变化。垂媒参数库里同一车型的字段开始出现版本混写、车主社区出现"我这个版本和你说的不一样"的对比帖、经销商口径与官网口径出现分叉——这些变化出现在答案明显变差之前，属于最有操作价值的一层。

这解释了为什么第 08 章把问题集固定、按周度看异常：周度数据本身是滞后的，但把周度变化与内容侧、平台侧的可观察信号并置，才能把"已经变差"与"即将变差"分开。三者的时间差，就是这个行业里真正可用的预警窗口。

信号层	可观察的具体变化	相对提前量	可操作性
平台侧	推出品牌合作产品、收窄外部引用、答案加推广标识	一个季度以上	可调整投入结构，但不可逆
产品侧	年款切换、OTA 推送、配置与价格调整公告	一个改款周期	可提前完成参数库同步
内容侧	字段版本混写、跨渠道口径分叉、车主对比帖增多	数周	可直接定位并清理
监测数据	准确率下降、提及率波动、候选集进入率下滑	滞后数周	仅能确认已发生的损失

表 9-3 四类风险信号的时间尺度与可操作性

提前量为基于改款周期、平台产品节奏与监测频率的估算区间，非实测值；可操作性指该层信号出现后能否改变结果。

同一风险在生命周期不同阶段的不同形态

三类风险的严重程度不是恒定的，它随车型所处的生命周期阶段变化。把阶段拉出来看，同一件事在上市期是可控的，在退市期可能变成不可逆的。

上市期（0—6 个月）的主要风险是实体未定形。此时语料少、口径尚未统一，参数失真的表现是"缺"而不是"错"，修正成本最低——因为还没有形成存量引用。这一阶段的错误最容易被长期放大：早期被大量转载的错误参数，会在后面几年反复被调用，清理难度随时间指数上升。

成熟期（6—24 个月）的主要风险是版本漂移。年款、OTA、配置调整同时发生，旧内容存量已经很大，新内容还在累积。这一阶段最典型的症状是"同一个问题在不同平台给出不同答案"，而这种不一致往往被误判为模型问题，实际上是自家内容的时间戳乱了。

退市与换代期的主要风险是存量污染的沉淀。停售车型的参数页若不下线、不标注停售，会与新车型在同一批语料里竞争，形成长期噪声。更麻烦的是二手车与存量车主的内容仍在增长，模型会把"老款体验"当作"当前产品"引用。这一阶段的治理重点是标注与切分，而不是新增内容。

这个阶段差异说明一件事：GEO 的风险管理不是一张静态清单，而是跟着车型节奏走的。同一套动作在上市期做是"占位"，在成熟期做是"纠偏"，在退市期做是"止损"——三种目标不同，判断标准也不同。

本册没有回答的问题

任何一份行业册都有边界，说清边界比多说三条结论更有用。本册没有回答四类问题，它们分别需要不同的资料条件或更长的时间验证。

没有回答"投入多少才够"。GEO 的投入与产出没有线性关系，也缺少公开的行业基准数据。本册能做的是说明哪些资产会累积、哪些会折旧、哪些在特定情景下失效，但不给预算区间——给出来的数字只会被当作承诺，而它并不成立。

没有回答"哪家服务商/工具更好"。工具层的差异主要体现在抓取覆盖率与问题集设计，这两项在不同车企的需求结构下权重完全不同；而且这个市场变化快，任何排名在半年内都会失效。

没有回答具体车型的参数对错。本册所有车型对比字段都是结构示例，用于说明"字段该怎么写"，不构成任何车型的实际参数判断。真实场景下的参数核查必须回到官方配置表与目录公告，这也是第 04 章反复强调"写算式而非写结论"的原因。

没有回答"多久见效"。资产累积有时间差：第三方背书需要一整个评测周期，实体澄清需要跨渠道同步，参数一致性需要跟随改款节奏。这些时间差因车企的资源条件不同而不同，任何"三个月见效"的表述都缺少可验证的分母。本册只说明哪类动作会累积、哪类只减缓折旧，不承诺周期。

保留这些空白不是回避，而是让本册的每一个判断都站在可核验的资料上。新能源 GEO 的很多问题目前只有方向性答案，把方向说清、把不确定标明，比给出一个精确但站不住的结论更有价值。

关于资料与方法。本册的政策条款来自财政部、税务总局、工信部公告与地方政府服务口径；市场数据来自行业协会月度统计与主流媒体报道；用户行为数据来自 CNNIC 统计报告与第三方调研机构口径；技术机制部分参照引擎厂商公开文档与 2024 年 GEO 学术研究 [13]；GEO 的概念框架与可信语料要求参照中国商务广告协会（CAAC）发布的团体标准 [8]。凡是经过二次计算的数字（如购置税减免额、调用频次区间），文中都给出算式或估算口径，便于读者自行复核。

关于结论的有效期。本册的判断建立在 2026 年的政策框架与当前的生成式检索形态上。政策到期（2027 年底减半政策结束）、平台引用策略调整、模型侧检索能力变化，都会改变其中部分结论。建议把本册当作"机制说明书"而非"年度快照"：机制部分相对稳定，数字与情景部分需要随年份更新。

关于口径的取舍。同一件事在本册里尽量只用一套数：渗透率统一用乘用车零售口径，不混入含商用车与出口的宽口径；AI 使用率统一用第三方调研的购车场景口径，不混入全品类数据；续航统一用 CLTC 并同时标注海外工况作为对照。口径不混用，图与表的数字才对得上——这也是本册把"口径"写进每张表和每条判断注释的原因。

> **最后一句判断：**新能源行业的 GEO 竞争，本质上不是内容产能的竞争，是**信息一致性**的竞争。谁能让官网、垂媒、媒体、口碑四类信源对同一款车说出同一套参数，谁就赢在了 AI 答案的起点。这件事没有捷径，但一旦做成，会随时间复利——因为一致性是可以被模型逐年学习的资产。

10 实测：10 份真实问答里，AI 把哪几个参数说错了

前九章讲的是机制。机制说得再顺，也需要一组真实答案来印证——否则"参数失真"只是一个好用的比喻。本章是一次可复核的小样本实测：固定问题、固定字段、逐字记录原始回答，再拿官方口径逐项比对。它不构成行业基准，但足以回答一个具体问题：**AI 在新能源比选里出错，究竟错在哪一层。**

采样怎么做的：10 个问题、1 个引擎、10 份有效答案

采样时间集中在 2026 年 11 月 11 日上午 10:52 至 11:03，同一台设备、同一网络、同一浏览器会话。问题集提前固定，覆盖 10 个真实比选场景：10—15 万纯电轿车（无家充、日通勤 80 公里）、15—20 万纯电 SUV（后排三人、婴儿车、400 公里

高速）、20—25 万纯电 SUV（800V 与冬季衰减）、25—30 万增程 SUV、30—35 万纯电 SUV（城市 NOA）、20 万以内插混轿车（五年用车成本）、15 万级纯电小车（女性城市代步）、35—45 万六/七座（老人上下车）、25 万纯电轿跑（CLTC 与高速实测差）、12 万首购（油电之选）。每问都带约束条件，不给"推荐几款车"这类无边界提问——因为无边界的答案没有可比性。

有效样本 10 份，全部来自豆包：10 个问题均取到完整回答，无空答、无中途截断。其余引擎本次未取得可用样本——DeepSeek、通义千问、腾讯元宝在采样环境的浏览器中需要登录，另有部分引擎受匿名会话额度限制，未能完成全部提问，因此统一排除，不作替代性采样。本章因此是**单引擎满样本**：结论只对豆包这一次表现负责，不做平台间比较，也不外推为"行业平均正确率"。

项目	本次口径
采样时间	2026-11-11 10:52—11:03，同一设备与网络
引擎与样本	豆包 10/10（单引擎满样本，其余引擎未取得可用样本）
问题集	固定 10 问，均含预算、车型品类与 2—3 项使用约束
核对字段	答案中明确出现的售价、CLTC 续航、补能或亏电油耗，且官方有对应口径
判定基准	官网配置表与官方上市稿的指导价、CLTC 数值；区间覆盖指导价记"一致"
判定分档	一致 = 与官方口径相符；偏差 = 数值不符但量级正确；错误 = 平台属性、工况或版本事实性错误
不计入分母	官方无对应口径（如冬测达成率、第三方实测电耗）或答案未给数值的字段

表 10-1 实测采样设计与判定口径

核对结果：22 个可核字段，一致率 68.2%

10 份答案里能对上官方口径的字段共 22 个。其中一致 15 个、偏差 5 个、错误 2 个，一致率 68.2%，偏差与错误合计 31.8%。按字段类别拆开更有解释力：售价与价位段 13 项只对了 7 项（53.8%），CLTC 续航 5 项全部对得上（100%），补能平台、服务与油耗 4 项对了 3 项。差距不在"知不知道这款车"，而在"价格说的是哪一年、哪一版"。

字段类别	可核项	一致	偏差	错误	一致率
售价与价位段	13	7	5	1	53.8%
CLTC 续航	5	5	0	0	100%
补能平台 / 服务 / 油耗	4	3	0	1	75.0%
合计	22	15	5	2	68.2%

表 10-2 按字段类别的核对结果（豆包 10 份答案，2026-11-11，口径见表 10-1）

这个数字要读得克制。68.2% 不代表"AI 基本靠谱"，它代表**每三个参数里就有一个对不上官方口径**，而且错的那一个恰好是用户最先看的价格。用户拿着错误价格去比选，后果不是"看到一条假信息"，是预算、续航预期和版本选择同时被带偏——第 03 章说的"版本串线"，在实测里是可复现的。

错在三个地方：价格刻度、平台属性、版本对位

把 7 个非一致项拆开看，错误并不是随机分布的，而是集中在三类可复现的地方。第一类是**价格刻度漂移**：报价用的是旧年款或权益价，与当期指导价差 1—5 万元；第二类是**平台属性错置**：把 400V 或 3C 架构写成 800V，把 26 分钟快充写进 15—

20 分钟；第三类是**版本对位错误**：把某个价位段的车型，放进另一个价位段的候选池。

车型	答案中的说法	官方口径	判定	错误类型
小鹏 MONA M03 640	"全系 800V，30%—80% 约 15—20 分钟"	磷酸铁锂 3C；640 版 30%—80% 最快 26 分钟	**错误**	平台属性错置
小鹏 G6 625 Max	"25—30 万区间，高配落地可进 30—35 万"	指导价 17.68 万元，旗舰版 18.68 万元	**错误**	版本对位错误
智己 LS6	"20.99 万元起"	Max 标准版指导价 22.99 万元（权益价 21.99 万元）	偏差	价格刻度漂移
小鹏 X9	"35.98—41.98 万元"	2026 款 665 Max 30.98 万元起	偏差	价格刻度漂移
腾势 D9 EV	"37.98—44.98 万元"	纯电版 37.98—46.98 万元（第二代 32.98 万元起）	偏差	价格刻度漂移
五菱缤果 PLUS 510	"10—13 万元"	五座版 9.58 万元、旗舰版 9.88 万元	偏差	价格刻度漂移
smart 精灵 1 号	"13—15 万元，刚好卡在预算上限"	全新一代 15.29 万元起（限时权益价 14.99 万元）	偏差	价格刻度漂移

表 10-3 实测中出现的 7 个非一致项（官方口径取自各品牌官网配置表与上市稿，2026 年 10—11 月）

这三类错误各有各的成因，也各有各的修法。价格刻度漂移是**时间戳问题**：语料里旧年款报价与新年款权益价混在一起，模型取到的是出现频次最高的那个数，而不是当期官网的数——这正好印证第 03 章"取高频而非取权威"的推演。平台属性错置是**口径问题**：把"3C 快充"和"800V 高压平台"这两件不同的事写成一件事，属于典型的参数层级混淆。版本对位错误是**实体问题**：车型实体是对的，但年款、配置、价位段没有绑定在一起，模型只能用一个笼统的价格印象去回答具体预算的提问。

候选集：车型都查得到，属性对不上

10 份答案共点名 13 款不同车型，逐一到官方渠道核对年款，全部能查到对应实体，**没有出现凭空捏造的车型。**这说明当前引擎在"实体召回"这一层已经相对可靠：品牌名、车型名、大致价位不会错到离谱。出问题的是实体之后的属性层——同一个车型被挂上错误的平台、错误的续航达成口径、错误的价格刻度。

更值得记录的是候选集的**重建幅度。**10 份答案里，价位带相邻的几道题几乎没有共用车型：10—15 万纯电轿车、15 万级纯电小车、20—25 万纯电 SUV 三题共点名 9 款，两两不重复；13 款车中跨题重复出现的只有小鹏 G6 一款。这说明引擎不是从"固定品牌池"里抓车，而是每换一道题就按约束重建一次候选集——约束一变，池子就重排，上一题被选中的车型很难带到下一题。

而 G6 的这次重复恰好演变成一个错误：它在 15—20 万 SUV 题里被正确报成 17.68 万元，到了 30—35 万城市 NOA 题里却被放进"25—30 万区间、高配落地可进 30—35 万"。**候选集可以沿用，价位刻度却跟着题面漂移。**这正是价格类错误最难缠的地方：它不是一次性的抄错，而是随着提问语境被重新"估算"出来。

> **实测的判断：**AI 在新能源比选里的问题，已经从"不知道这款车"变成了"知道这款车、但把它的属性说错"。前者靠收录解决，后者只能靠**结构化参数的一致性**解决——同一款车的年款、指导价、CLTC、补能平台，在官网、垂媒、目录公告上必须是同一套数，且带明确时间戳。这也是本册把参数一致性放在所有动作之前的原因：**能查到不等于被读对。**

这次实测没做到的部分

样本的边界必须写清楚，否则上面的百分比会被误用。第一，只有一个引擎、10 份答案，不能代表其他平台的语料结构，也不构成平台间比较。第二，问题集是 10 个，每个场景只问一次，没有做同题多问的稳定性检验——同一问题换措辞、换时间再问，答案可能不同，这部分波动本次没有测。第三，核对字段只覆盖售价、CLTC 与补能，安全配置、智驾硬件、权益政策没有逐项核。第四，采样时点集中在 11 分钟窗口内，无法反映模型对当日新上市车型的响应延迟。

要把它做成可用的月度指标，需要按第 08 章的口径补齐：固定 30 个问题、覆盖 5 个引擎、每题重复 3 次取众数、字段扩到 6 项，才能得到可比较的正确率与候选集进入率。本次的定位是**一次口径演示**：证明这套方法可执行、可复核，也证明错误确实集中在价格刻度、平台属性与版本对位这三处——这三处恰好都是结构化参数能管住的地方。

把三类错误对回信源层：每一类都有对应的修法

第 05 章把新能源语料分成四层：L1 官方与目录、L2 品牌自证、L3 垂媒结构化库、L4 车主口碑。这次的 7 个非一致项，逐条都能落到具体的一层上，而不是笼统的"内容质量问题"。落到层上，修法就不再是"多发内容"，而是"改哪个字段、由谁改、多久见效"。

错误类型	主要发生在哪一层	可复核的修法	见效节点
价格刻度漂移	L1 官方价格页 → L3 垂媒报价库	价格页同时给"指导价"与"权益价"两列并标注有效期；垂媒库按年款绑定价格，旧年款标停售	一个改款周期
平台属性错置	L1 官方参数页	把"高压平台电压"与"充电倍率"拆成两个独立字段，禁止合并成"快充能力强"这类形容词	当周可改
版本对位错误	L1 官方配置表 + L3 结构化库	年款—配置名—价格三元组绑定，同名字段强制加版本后缀（如"640 超长续航 Plus"），不写成"MONA M03"	1—2 个内容周期

表 10-4 三类错误对应的信源层与修法（信源分层沿用第 05 章口径）

价格类错误占 7 个非一致项里的 6 个，这个分布不是偶然。价格是三件事的交汇点：它更新最频繁（改款、官降、限时权益叠加）、被转载最多（垂媒、经销商、导购文都在抄）、且最容易被写成"起售价"一个数。三个特征叠加，模型取到旧数的概率就远高于取到当期官网的数。相比之下，CLTC 续航这类字段一年只变一次，错得少——实测里 5 项续航数值全部对得上，就是证据。

对车企来说，这三类错误对应三件成本极低但优先级不同的事。**先改平台属性**：它只需要在官网参数页拆两个字段，一周内能完成，且直接影响"800V 是不是标配"这类高频提问的答案。**再改版本命名**：把配置名写全、写进标题，让模型有余地把"640 超长续航 Plus"和价格绑在一起，而不是退回"MONA M03 十几万"这种模糊印象。**最后做价格时效**：这一项最难，因为它要求市场、官网、垂媒三方在同一时点对齐，但它决定了"报价类问题"的答案质量，也决定了用户到店后的第一句话是"网上说 20.99 万"还是"网上说 22.99 万"。

对测量方来说，这次实测也给了一个口径上的修正：参数正确率不该只出一个总分。同样是 68.2%，如果错的是平台属性，那是官网字段没拆清，一周可修；如果错的是价格刻度，那是三方时间戳没对齐，需要一个改款周期。两种 68.2% 的处理成本差一个数量级。所以第 08 章那张月度监测表，实际使用时应该按错误类型分列计数——**分类型的一致率，比总分更能指导动作。**

参考来源

[1] 中国政府网：《1月起，这些新规将影响你我生活》（含新能源汽车车辆购置税减半征收政策），2026-01-01。

[2] 工业和信息化部：《道路机动车辆生产企业及产品（第411批）》《减免车辆购置税的新能源汽车车型目录（第三十五批）》公告，2026-10-16。

[3] 国家税务总局广西壮族自治区税务局：《2026年购置新能源车辆在车辆购置税上有什么优惠政策》，2026-08-27。

[4] 财政部、税务总局、工业和信息化部：《关于延续和优化新能源汽车车辆购置税减免政策的公告》（2023年第10号）。

[5] 人民日报：《认识当下的国内汽车市场，重构汽车市场新逻辑》，2026-08-28。

[6] 中国经济网：《两项"首次"突破 新能源汽车占据车市"半壁江山"》，2026-08-19。

[7] CNNIC：《生成式人工智能应用发展报告（2026）》，2026 年 9 月。

[8] 中国商务广告协会（CAAC）：《生成式引擎优化（GEO）》团体标准 T/CAACCHINA 001~005-2026，2026-09-14。

[9] 艾瑞咨询：《2026 年汽车行业 AI 信源影响力指数报告》（经行业媒体报道转引，属第三方调研口径），2026 年 6 月。

[10] 全国汽车标准化技术委员会：《电动汽车能量消耗量和续驶里程试验方法 第1部分：轻型汽车》（CLTC 工况依据），国家标准。

[11] 火山引擎：《站点权威度分级说明》，2026年。

[12] 工业和信息化部、财政部、税务总局：《关于2026—2027年减免车辆购置税新能源汽车产品技术要求的公告》，2025-10-09。

[13] Aggarwal, P. et al.:《GEO: Generative Engine Optimization》，ACM SIGKDD 2024。

[14] 小鹏、比亚迪、零跑、极氪、智己、理想、鸿蒙智行、腾势、蔚来、上汽大众、上汽通用五菱、smart、极狐各品牌官网配置表与上市新闻稿（第 10 章实测的官方核对口径），2026 年 10—11 月。

星阙实验室GEO研究中心｜本册基于工业和信息化部、国家税务总局公告、行业统计与第三方公开调研整理，仅供研究参考