

ICS 35.240.01

CCS L76

团 体 标 准

T/CAACCHINA 003-2026

生成式引擎优化 (GEO) 第 3 部分：计价、评估与测量规范 Generative Engine Optimization (GEO) — Part 3: Pricing, evaluation and measurement specification

202X-XX-XX 发布

202X-XX-XX 实施

中国商务广告协会 发 布

目次

目次..... II

前言.....IV

引言..... V

1 范围 1

2 规范性引用文件1

3 术语和定义2

3.1 生成式引擎优化 generative engine optimization （GEO） 2

3.2 GEO 服务商 GEO service provider2

3.3 委托方 principal 2

3.4 GEO 评估 GEO assessment 2

3.5 GEO 测量 GEO measurement2

3.6 GEO 计价 GEO pricing3

3.7 GEO 最小单元 GEO minimum unit3

3.8 GEO 问题 GEO query 3

3.9 GEO 意图词 GEO intent3

3.10 GEO 采集组 GEO collection group3

3.11 GEO 分析组 GEO analysis group 4

3.12 语料 corpus 4

3.13 信源 information source 4

3.14 可见率 visibility rate4

3.15 AI 情感中正率 AI emotional positive rate4

3.16 可见份额 share of visibility 4

3.17 平均排名 average ranking 4

3.18 指定信息正确率 information accuracy rate 5

3.19 指定信息覆盖率 information coverage rate5

3.20 信源数量引用率 information source citation rate-frequency 5

3.21 信源问题引用率 information source citation rate- query 5

3.22 语义一致性 semantic Consistency 5

4 GEO 计价方法6

4.1 计价原则6

4.2 计价的服务最小单元 6

4.3 计价的模式说明 7

4.4 报价单规范 10

5 评估与测量11

5.1 评估与测量的基本原则 11

5.2 指标计算与使用说明 11

5.3 测量要求15

5.4 数据采集方法 16

附 录 A: 采集数据量要求说明18

参考文献19

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》给出的规则起草。

本文件是《生成式引擎优化（GEO）》系列标准的第 3 部分。该系列标准包括以下部分：

第 1 部分：通论及术语与定义；

第 2 部分：服务商评估规范；

第 3 部分：计价、评估与测量规范；

第 4 部分：可信语料规范；

第 5 部分：服务流程合规与安全要求。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国商务广告协会提出并归口。

本文件起草单位：北京明略昭辉科技有限公司、重庆灵狐科技股份有限公司、上海质安华数字科技有限公司、上海曼朗市场营销策划有限公司、上海源易信息科技有限公司、北京清蓝智汇科技有限公司、上海明堂文化传播有限公司、三六零科技集团有限公司、北京搜狐新媒体信息技术有限公司、广州欢网科技有限责任公司、利欧集团数字科技有限公司、科大讯飞股份有限公司、上海市新榜信息技术股份有限公司、浙文互联集团股份有限公司、北京深演智能科技股份有限公司、钛镁时代（上海）人工智能科技有限公司、北京特比昂科技有限公司、上海喜马拉雅科技有限公司、上海臻奥数智科技有限公司、深圳智汇蓝媒科技有限公司、致维科技（北京）有限公司、复知智云（上海）科技有限公司。

本文件主要起草人：招熙琳、赵金夺、周小榆、余海鹏、张妍、余杰、王云、鲁扬、尚晶、梁亚楠、吴哲、高鹏、袁浩、邓礼帆、张浩然、王先超、方瑜骅、宋明玉、侯艺、曾杰、周日、张国鹏、李文哲、蔡晓旭、魏倩、吕虎军、庞成博、张龙飞、傅海波、牛婷婷、李铭、张晶一、邱文斌、栾一闻、刘嘉奇、谭北平、李浩宇、张慧。

本文件主要审查人：丁俊杰、陈徐彬、吴明辉、吴默、陈忠伟、任鹏、全佳欣、李丹、严禧汶、唐兴通。

引 言

随着生成式人工智能技术的大规模应用，生成式引擎优化（GEO）作为连接品牌营销与 AI 信息分发的新型服务形态，正在快速发展。GEO 服务的计价方式、效果评估与测量方法，是行业健康有序发展的基础性规范，直接影响服务商与委托方之间的合作质量与信任关系。

当前 GEO 行业在计价、评估与测量方面仍存在以下突出问题：计价方式不统一、服务单元划分不清晰，导致服务商与委托方在服务定价上缺乏共同语言；效果评估指标体系不完善，各服务商自行定义指标，难以进行横向对比；测量方法缺乏规范性，数据采集、处理和报告输出流程不一致，影响评估结果的可信度。

本文件旨在建立 GEO 服务计价、评估与测量的统一规范，明确服务单元的划分标准，建立科学的效果评估指标体系，规范测量方法与技术要求，为 GEO 服务的交易、执行和验收提供标准化依据。

本标准与 GEO 团体标准体系内其他标准的关系如下：与通论及术语与定义（第 1 部分）的术语与定义保持对齐；与服务商评估规范（第 2 部分）的服务商能力评估指标对接；与可信语料规范（第 4 部分）的语料质量标准衔接；与服务流程合规与安全要求（第 5 部分）的数据安全与合规要求对齐。

生成式引擎优化（GEO）第 3 部分： 计价、评估与测量

1 范围

本文件规定了生成式引擎优化（GEO）服务流程中的计价方式、效果评估、数据采集和指标测量的通用规范。

本文件适用于以下主体：

- a) 实施 GEO 策略的品牌方和广告主；
- b) 提供 GEO 服务的技术服务商和营销服务机构；
- c) 开展 GEO 效果评估的第三方监测机构；
- d) 从事 GEO 相关学术研究的教育和科研机构。

本文件不适用于：

- a) 传统搜索引擎优化（SEO）的技术实现细节；
- b) 生成式 AI 模型的训练与推理技术规范；
- c) 互联网广告投放的竞价与程序化交易技术规范。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 34090.1—2017 互动广告 第 1 部分 术语概述

GB/T 34090.3—2017 互动广告 第 3 部分：效果测量要求

GB/T 1.1—2020 标准化工作导则 第 1 部分：标准化文件的结构和起草规则

GB/T 42454—2023 信息技术服务 数字化营销服务 移动营销技术规范

TC260—005 人工智能应用伦理安全指引 1.0

3 术语和定义

下列术语和定义适用于本文件。

3.1 生成式引擎优化 generative engine optimization (GEO)

通过系统性的技术与内容策略，增强数字内容在生成式 AI 搜索引擎结果中的可见性、准确性和可信度的优化活动。

注 1: GEO 概念由 Aggarwal 等人于 2023 年在论文《GEO: Generative Engine Optimization》(arXiv:2311.09735) 中首次系统性提出。

注 2: GEO 的核心目标是让内容在生成式引擎响应中被发现、被引用、被推荐。

[来源: T/CAACCHINA 001-2026, 术语与定义 3.1.1]

3.2 GEO 服务商 GEO service provider

向品牌主或代理机构提供 GEO 相关服务的具有独立法人资格的企业或机构。

注: 服务范围可涵盖策略咨询、技术工具提供、内容生产与优化、效果监测与分析、语料管理等一个或多个领域。

[来源: T/CAACCHINA 001-2026, 通论及术语与定义 3.1.9]

3.3 委托方 principal

基于服务协议接受 GEO 服务商提供的生成式引擎优化服务的组织或个人。

注: 委托方通常指具有品牌推广需求的品牌主、广告主或其代理机构。

[来源: T/CAACCHINA 001-2026, 通论及术语与定义 3.1.11]

3.4 GEO 评估 GEO assessment

在 GEO 服务实施过程中，委托方或 GEO 服务商发起的，以评估委托方品牌或产品在生成式引擎中被推荐的表现情况以及服务商的服务效果为目标的工作。

注: GEO 评估可在服务实施前、中、后由需求方、服务商或第三方监测机构发起，评估维度包括可见性、位序、情感倾向、信息准确性和信源引用等。

[来源: T/CAACCHINA 001-2026, 通论及术语与定义 3.1.7]

3.5 GEO 测量 GEO measurement

GEO 评估中用于数据采集、指标计算和统计分析的技术方法与操作程序。

注：包括数据采集、数据处理、指标计算、统计分析与报告输出等环节。

[来源：T/CAACCHINA 001-2026，通论及术语与定义 3.1.8]

3.6 GEO 计价 GEO pricing

GEO 服务交易中，由服务商与委托方协商约定的服务付费方式。

注：包括按项目计价、按服务内容和数量计价、按效果计价、按系统交付计价、混合计价等模式。

3.7 GEO 最小单元 GEO minimum unit

GEO 服务实施过程中，由服务商与委托方协商确定的用于计算服务的最小对象。

注：包括问题、服务周期、生成式引擎、报告、语料、信源等。

3.8 GEO 问题 GEO query

在 GEO 评估、测量或优化服务中，用于向目标生成式引擎发起提问的自然语言问题。

注 1：GEO 问题反映用户的具体意图、消费场景和信息需求。示例：推荐 20 万元左右的家庭 SUV 新能源车。

注 2：语义相近的一组 GEO 问题可归并为一个 GEO 意图。

3.9 GEO 意图词 GEO intent

在 GEO 服务中，用于归纳一组语义相近 GEO 问题的主题词、需求词或关键词。

注 1：GEO 意图词通常由一组指向同一品类、场景、需求的 GEO 问题构成，可用于服务范围界定、问题分组、数据汇总；实际测量时，应明确每个意图词对应的问题清单。意图词示例“防脱洗发水”。

注 2：在本标准体系中，GEO 意图词与 GEO 关键词的含义相同，GEO 时代下，GEO 意图词相比 GEO 关键词更契合业务场景的表达。

3.10 GEO 采集组 GEO collection group

在 GEO 测量中，由相同的意图、问题、生成式引擎、模式配置、端口、地域、语言、统计周期等维度共同限定的一组数据采集任务。

注：采集维度包括关键词或问题、生成式引擎、模式配置、端口、地域、语言、统计周期等。组内采集是指在同一采集组内重复采集样本，并基于该组样本计算指标。

[来源：T/CAACCHINA 001-2026，通论及术语与定义 3.1.10]

3.11 GEO 分析组 GEO analysis group

在 GEO 评估中，用于汇总、对比或分层分析的一个数据组。

3.12 语料 corpus

支撑生成式人工智能模型训练、调优、评测与推理运行的原始数据集合。

注 1：该类数据为生成式人工智能模型完成知识学习、语义理解、内容生成提供基础输入载体。

注 2：按用途可分为预训练语料、微调语料、评测语料和推理语料。

[来源：T/CAACCHINA 001-2026，通论及术语与定义 3.2.5]

3.13 信源 information source

产生或承载语料信息的原始载体。

注：信源包括但不限于品牌官方网站、权威媒体、行业数据库、官方认证文档、经审核的用户生成内容（UGC）等。

[来源：T/CAACCHINA 001-2026，通论及术语与定义 3.2.4]

3.14 可见率 visibility rate

指定品牌或产品在生成式引擎回答中被提及、推荐或进入指定位序范围的比例。

注：该比例在约定的问题集和采集条件下计算，用于衡量品牌或产品被生成式引擎展示和推荐的概率。

[来源：T/CAACCHINA 001-2026，通论及术语与定义 3.2.7]

3.15 AI 情感中正率 AI emotional positive rate

评估生成式引擎回答输出中，对指定品牌或产品的情感倾向的指标。

注：情感倾向分为正面、中性、负面，也可根据服务双方约定增加“不确定”“混合评价”等分类。

3.16 可见份额 share of visibility

指定品牌或产品在生成式引擎回答中被提及的频次占同品类全部品牌或产品被提及总频次的比例。

注：该指标反映品牌或产品在特定品类中被生成式引擎提及的相对份额。

[来源：T/CAACCHINA 001-2026，通论及术语与定义 3.2.8]

3.17 平均排名 average ranking

评估指定品牌或产品在本品类范围下，生成式引擎的回答中首次出现的排名均值指标。

注：该指标反映品牌或产品被 AI 提及和引用的优先程度。

3.18 指定信息正确率 information accuracy rate

评估生成式引擎回答输出中，对委托方指定的信息是否正确表达的指标。

注：指定信息可包括品牌名称、企业名称、产品型号、功能参数、价格区间、核心卖点、资质认证等。反映生成式引擎引用指定信息的正确程度

3.19 指定信息覆盖率 information coverage rate

生成式引擎多轮回答中，包含委托方指定信息的回答数量占总回答数量的比例。

注 1：指定信息可包括品牌名称、企业名称、产品型号、功能参数、价格区间、核心卖点、资质认证等。

注 2：该指标反映生成式引擎引用指定信息的覆盖广度。

3.20 信源数量引用率 information source citation rate-frequency

在约定采集组内，统计指定信源在生成式引擎回答中被引用的次数占全部引用信源总次数比例的指标。

注：该指标以“引用次数”为计数基础，适用于评估指定信源在全部引用信源中的数量占比和重要程度。

3.21 信源问题引用率 information source citation rate- query

在约定采集组内，统计指定信源在生成式引擎回答中被至少引用一次的提问次数占全部有效提问次数的比例的指标。

注：该指标以“问题的提问次数”为计数基础，适用于评估指定信源对问题集合的覆盖程度

3.22 语义一致性 semantic Consistency

评估 GEO 的服务商生成或维护的 GEO 语料，在不同页面、不同信源和不同渠道中的表达是否一致的指标。

注：重点检查同一实体（品牌、产品、人物、参数、功效、价格等）的名称、属性、数值和描述是否无矛盾、无歧义、无冲突。

4 GEO 计价方法

4.1 计价原则

GEO 服务的计价建议遵循以下原则：

——对象明确：计价模式和最小单元清晰，如意图词/问题、生成式引擎、服务周期、报告等的界定清晰、无歧义；

——可量化：价格基于明确的最小单元及其数量统计，总价格的计算方式公开透明；

——符合规范：报价文档或报价单符合规范，包括清晰的 GEO 服务、数量、价格计算公式，并明确每项计价最小单元所涉及的服务内容边界、交付物和验收方式。

4.2 计价的服务最小单元

GEO 服务中，可基于最小单元进行服务的组合和量化计价，最小单位有七类：内容单元、生成式引擎单元、时间单元、数据单元、交付单元、信源单元、语料单元。

4.2.1 内容单元：问题数量

GEO 问题对计价的影响主要体现在数量和内容类型两个方面。问题数量越多，服务范围越大，计价通常越高。

问题的内容类型不同，GEO 服务难度也不同，因此会影响单价设定。例如问题归属于品类、场景、品牌、地区排名、全国排名的不同类型的意图，可根据竞争强度、优化难度和采集成本设置不同单价。通常情况下，问题归属于大品类、高竞争场景、全国排名等意图，则单价较高；问题归属于品牌、低竞争场景、地区排名等意图，则单价较低。

4.2.2 生成式引擎单元：数量及模式设置

生成式引擎及其模式设置对计价的影响主要体现在两个方面：数量和模式设置的参数选择。

生成式引擎和模式设置的数量直接影响 GEO 服务的计价，数量越多，计价越高。受生成式引擎模型设置影响，不同的生成式引擎及不同模型设置优化难度不一，难度越高，计价越高。当服务覆盖多个生成式引擎时，宜根据服务的数量、模式设置难度评估计价，不建议简单按固定倍数计价。

4.2.3 时间单元：服务周期

服务周期的时间要求对计价有直接影响，周期越长，计价越高。当前 GEO 服务通常以“月”或“季度”为最小周期单位。

4.2.4 数据单元：数据要求

GEO 服务过程中，委托方的数据要求对计价的影响主要体现在两个方面：数据数量和数据维度。

委托方对于数据数量（例如：监测采集的提问次数，也被称为样本量）的要求对计价有直接影响，次数越高，计价越高。

委托方要求采集和分析的数据维度越多，计价越高，数据维度的价格差异一般体现在交付报告的单价差异上。

4.2.5 交付单元：交付物要求

GEO 服务过程中，委托方的交付物要求对计价的影响主要体现在两个方面：交付数量和内容要求。

委托方对交付数量的要求越高，计价越高。例如：要求提供周报+月报的计价高于只要求提供月报的计价。

委托方对交付内容的要求决定了交付物的单价。分析内容越复杂，包含的数据维度越多，单价越高。例如：结案服务和月报的单价通常高于周报。

4.2.6 信源单元：信源数量

围绕提升品牌信息可采集、可验证、可引用能力，对委托方的官方信源（例如官方网站、官方电商）进行建设或优化的服务。

委托方要求优化的信源数量决定计价，数量越多，计价越高。

4.2.7 语料单元：语料数量

语料建设是绕品牌、产品、服务及行业主题建立标准化内容体系，包括品牌语料、产品语料、FAQ 语料、应用场景语料、案例语料、参数语料及行业知识语料等，为生成式引擎理解、关联、引用品牌信息提供基础内容支撑。

委托方要求提供的语料数量决定计价，数量越多，计价越高。

4.3 计价的模式说明

4.3.1 按项目计价

委托方以项目制为计价单位，与服务商协商，综合确定单个项目的范围、服务内容、服务周期、生成式引擎范围、意图词/问题范围、交付物及数量、效果目标及承诺等，并进行整体打包定价。

按项目计价是一种面向特定委托方、定制化程度较高、灵活性较强的计价方式，一般适用于稳定、长期或综合性合作。

报价单应明确项目名称、项目范围、项目服务内容、项目服务周期、项目交付物（包含要求）、项目价格。

表 1 按项目计价报价单示例

项目名称	项目范围	服务内容	服务周期	交付物	项目价格

4.3.2 按服务内容和数量计价

将 GEO 服务内容进行拆解，按交付服务的最小单元、单价和数量核算整体服务费用的模式。该模式适用于大部分 GEO 服务，是当前 GEO 服务中较常见的计价方法。

按不同的交付物类型，服务内容与数量计价方法如下：

a) 交付物为“数据”时

计价的最小单元以“数据单价 × 数据量要求 × 问题数量 × 模型数量 × 服务周期”为基准。

报价单应明确数据采集处理单价、数据采集处理的意图词或问题、数据量要求、数据采集的生成式引擎及模式、服务周期、数据交付形式和数据服务总价。

表 2 按服务内容和数量计价报价单示例-交付数据

数据单价	数据量要求	问题数量	模型数量	服务周期	交付要求	数据总价
	如：提问次数		含设置要求			

b) 交付物为“报告”时

计价的最小单元以“报告单价 × 报告数量”为基准。

报价单应明确报告名称、报告内容要求和交付形式、报告单价、报告数量、报告服务总价。

表 3 按服务内容和数量计价报价单示例-交付报告

报告名称	报告单价	报告数量	报告要求	报告总价
如：策略报告				

周报				
月报				
结案报告				

c) 交付物为“提供优化服务”时

计价的最小单元以“服务单价 × 问题数量 × 模型数量 × 服务周期”为基准。

报价单应明确服务单价、优化的意图词或问题、优化模型、服务周期、服务交付要求（如内容数量、知识库、效果承诺等）和服务总价。

表 4 按服务内容和数量计价报价单示例-优化服务

服务单价	问题数量	模型数量	服务内容	服务周期	交付要求	服务总价

d) 交付物为“GEO 语料”时

计价的最小单元通常以“1 条语料单价 × 语料数量”为基准。

报价单应明确语料服务单价，语料形式，交付数量，交付要求，服务总价。

表 5 按服务内容和数量计价报价单示例-交付语料

语料单价	语料形式	交付数量	交付要求	服务总价

4.3.3 按效果计价

当委托方有明确效果承诺要求，并将是否达成效果或达成程度作为计价因素时，可采用按效果计价模式。效果计价通常作为 GEO 服务的重要辅助计价方法。

效果计价通常按考核 KPI 达成程度分档计费。

报价单应明确效果指标的定义、计算方法、数据来源、达成程度分档和各档支付要求。

表 6 按效果报价单示例

效果考核指标	考核指标说明	达成程度分档	分档支付比例	支付总价
当前值	计算方法			
目标值	数据要求			
	数据来源			

4.3.4 按系统交付计价

通过系统进行交付的计价模式，在 GEO 服务中常见于评估或测量系统平台交付。该模式可分为“系统定制开发”和“系统使用”两种计价方式。

表 7 按系统报价单示例-系统使用

账号类型	账号服务范围与 权限	账号单价 年/月	账号数量	服务周期 年/月	总价

表 8 按系统报价单示例-系统定制开发

系统要求	开发工作内容	工时单价 人/日	工时数量	开发总价

4.3.5 混合式计价

上述多种计价模式的组合使用。

注：如采用“（问题数量×生成式引擎数量×优化月数）+效果达标绩效”的组合报价模式。

4.4 报价单规范

GEO 服务商向委托方提供的报价单应遵循以下规范：

- a) 计价模式声明：明确所采用的计价模式，例如按服务内容及数量计价、按效果计价、按项目计价；

- b) 计算过程透明：报价中应明确价格计算过程和计算逻辑。当委托方需要调整服务内容、效果要求或交付内容时，能够根据变化计算出新的价格；
- c) 计价涉及的 GEO 最小单元表述明确：不同计价模式中，对总体价格产生影响的最小单元，有详细说明。例如：当以意图词为计价最小单元时，说明每个意图词包含的问题和问题数量；当以生成式引擎为计价最小单元时，说明每个生成式引擎的模式设置要求。

5 评估与测量

5.1 评估与测量的基本原则

GEO 的评估与测量应符合以下基本原则：

- a) 指标完整：评估所采用的指标体系完整，能够反映生成式引擎中委托方的品牌和产品的当前表现及 GEO 服务商的服务成果。指标定义清晰，计算方法明确，并说明适用场景和使用限制；
- b) 方法透明：评估所采用的数据采集方法、指标计算公式、数据统计方法公开、明确、可复现；
- c) 方法科学：评估方法应依托可靠的统计学或数据分析原理，具备合理的数据样本量要求和明确的对比基准，避免因样本量不足、分组口径混淆或统计口径不一致导致误判；
- d) 流程透明：从数据采集、指数计算到报告输出的全过程应可追溯，关键步骤和决策应有记录；
- e) 分组一致：评估与测量应在服务商与委托方约定的最小分析颗粒度下开展。采集组中应保证意图词/问题、生成式引擎、模式、端口、地域/语言和统计周期等条件一致；分析组应分别列示各组结果。如需跨多个采集组的数据汇总，应披露权重设置、合并规则和适用限制。

5.2 指标计算与使用说明

5.2.1 可见率

指标计算公式：可见率/提及率 = 命中指定品牌或产品的提问次数 / 提问总次数 × 100%。

注：按位序统计时，仅统计品牌或产品首次出现位置满足约定位序要求的提问；同一回答中同一品牌或产品多次出现，按 1 次命中计。

在采集组口径明确、样本量满足要求的前提下，品牌/产品可见率可作为项目考核 KPI 指标：

——可以基于品牌、产品分别计算，按品牌统计时，为品牌可见率，按产品统计时，为产品可见率；

——可以按 TOP1、TOP3 等位序要求分别计算；

——可按日期、生成式引擎、模式、意图词或问题、本品与竞品等维度进行分析。

计算范围只限定为行业/品类问题（如“新能源汽车推荐”），不包含品牌问题（如“某品牌怎么样”）。

5.2.2 AI 情感中正率率

指标计算公式：

AI 情感中正率：指定品牌或产品的情感为正面或中性的提问次数 / 出现指定品牌或产品的总提问次数 × 100%。

AI 情感负面率：指定品牌或产品的情感为明确负面的提问次数 / 出现指定品牌或产品的总提问次数 × 100%。

注：“负面”的判定规则应由服务商和委托方预先约定，例如出现“不推荐”“问题多”“投诉”等词，或表达明确贬义。若采用 AI 进行语义判断，应说明使用的模型、提示词和复核规则，并保留人工复核机制，以降低 AI 判定口径与品牌定义口径不一致造成的误差。

该指标可作为项目考核 KPI 指标。可基于品牌或产品分别计算，并可按日期、生成式引擎、模式、意图词或问题等维度分析。

5.2.3 可见份额

指标计算公式：可见份额 = 指定品牌或产品被提及次数 / 同一统计范围内全部品牌或产品提及总次数 × 100%。

注：单次回答中同一品牌或产品多次出现时，按 1 次计；按品牌维度或产品维度统计应预先约定。

可见份额受品类边界、竞争品牌范围和问题集设置影响较大，是重要的策略分析指标，但不宜直接作为项目考核 KPI 指标。

该指标可基于品牌或产品分别计算，并可按日期、生成式引擎、模式、意图词或问题等维度分析。也可基于特定场景统计，例如在“高端家庭 SUV 推荐”中仅统计“新能源车”的份额。

计算范围只限定为行业/品类问题（如“新能源汽车推荐”），不包含品牌问题（如“某品牌怎么样”）。

5.2.4 平均排名

指标计算公式：平均位序 = 指定品牌或产品首次出现的位序数字总和 / 指定品牌或产品出现的提问次数。

注：单次回答中同一品牌或产品多次出现时，只统计第一次出现的位序；指定场景统计时，只统计该场景内的首次出现位序。

平均位序受品类、竞争品牌和问题集影响较大，是重要的策略分析指标，但不宜单独作为项目考核 KPI 指标。

该指标可基于品牌或产品分别计算，并可按日期、生成式引擎、模式、意图词或问题等维度分析。

计算范围只限定为行业/品类问题（如“新能源汽车推荐”），不包含品牌问题（如“某品牌怎么样”）。

5.2.5 指定信息正确率

指标计算公式：指定信息正确率 = 指定信息被正确表达的提问次数 / 被评估且涉及该指定信息的有效提问次数 × 100%。

指标使用说明：

- 可以基于品牌、产品、企业或具体场景的信息进行计算；
- 可按日期、生成式引擎、模式、意图词或问题等维度分析；

5.2.6 指定信息覆盖率

指标计算公式：指定信息覆盖率 = 出现指定信息的提问次数 / 包含评估品牌或产品的有效提问次数 × 100%。

指标使用说明：

- 可以基于品牌、产品、企业或具体场景的信息进行计算。
- 可按日期、生成式引擎、模式、意图词或问题等维度分析。

5.2.7 信源数量引用率

指标计算公式：信源数量引用率 = 指定信源被引用的次数 / 全部信源引用次数 × 100%。

注：同一回答中同一信源出现多个引用位置时，可按实际引用条数计；如需按回答或 URL 去重，可在测量方案中预先约定。

该指标可用于品牌当前表现评估、监测服务、阶段报告和结案报告，作为辅助指标。

信源数量引用率反映指定信源在全部引用信源中的数量占比。占比越高，说明该信源在对应采集组内的引用重要性越高，但不直接代表覆盖了更多问题。

与信源问题引用率相比，信源数量引用率关注引用次数，可能受单个回答引用多条链接或多个位置的影响；

可指定的信源包括：媒体类型（如新闻资讯、官网）、媒体（如新华网）、URL、文章（如“2026 年 6 月拍照手机测评榜”）、账号（如微信公众号）。

由于不同大模型产品的引用信源场景和位置不同，信源引用率分析应根据“模型-模式-端口”明确待分析的信源位置和场景。

当前主流大模型的引用信源位置主要包括：答案正文前的参考资料、答案正文内的引用标记、答案正文内的商品卡、答案后出现的图文和视频链接等。

表 9 主流大模型引用信源位置-以豆包和元宝为例

模型	端口	生成式引擎配置选项	信源位置说明
豆包	APP	快速	参考资料-答案正文前
豆包	APP	快速	商品卡-答案正文中
元宝	Web	HY3 模型-深度思考	参考资料-答案正文前
元宝	Web	HY3 模型-深度思考	相关视频-答案正文后

5.2.8 信源问题引用率

指标计算公式：信源问题引用率 = 指定信源被引用的有效提问次数 / 有效提问总次数 × 100%。

注：同一回答中同一信源多次出现，按 1 个问题命中计。

该指标可用于品牌当前表现评估、项目考核 KPI、监测服务、阶段报告和结案报告，作为辅助指标。

信源问题引用率反映指定信源对委托方问题集的语义匹配和覆盖程度。数值越高，说明该信源对评估问题的覆盖越全面。

信源引用率分析可指定信源：媒体类型（如新闻资讯、官网）、媒体（如新华网）、URL、文章（如“2026 年 6 月拍照手机测评榜”）、账号（如某企业微信公众号）。

该指标优先在同一 GEO 采集组内计算。跨模型、跨模式、跨意图词或跨周期进行组间比较时，应分别列示各组结果；如需汇总，应说明权重设置和合并规则。

5.2.9 语义一致性

指标计算方法：可采用“通过一致性校验的内容条数 / 参与校验的内容总条数 × 100%”计算；

注：校验规则应以委托方确认的标准知识库、产品资料或品牌事实库为基准。

用于品牌语料内容服务、可信语料建设和内容质量复核。

该指标不宜作为通用 GEO 效果指标强制适用；当合同范围包含语料生产、语料治理或知识库维护时，可作为质量验收指标。

5.3 测量要求

5.3.1 数据系统的性能要求

数据采集系统应具备高可用性、稳定性和一定的并发处理能力。对关键 API 接口（例如：用于数据采集的接口）应有明确的性能承诺，例如响应时间、成功率等。

数据采集系统应能支持 7 × 24 小时持续运行，核心任务的失败率或中断率宜不超过 2%，并应具备故障告警和补采机制。

满足透明化要求：系统中的数据应真实、可溯源，不得伪造、篡改或选择性呈现数据结果。

5.3.3 多平台统一测量机制要求

在跨多个生成式引擎采集数据时，不同生成式引擎采用统一的问题集，保持测试日期、问题表述、采集方式的一致性。

5.3.4 采集地域要求

GEO 数据采集设备应能覆盖多个市场地域。中国市场内建议不低于 3 个城市。

5.3.5 采集生成式引擎覆盖数量要求

采集设备应覆盖市场主流的生成式引擎 AI 模型产品。中国市场内建议不低于 5 个。当委托方要求采集的生成式引擎数量 ≥ 2 时，单一生成式引擎模型的样本占比不应超过全样本的 80%。

5.3.6 采集频次要求

意图词或问题集的设定，应为真实的、多样化的提问场景，覆盖不同消费群体、不同提问场景、不同需求类别的意图词和问题。

由于生成式引擎的回答具有一定随机性，相同问题、相同模型和模式的不同次数采集的答案可能不完全相同，为保证统计指标的准确性，每次评估测量的单个意图词（或问题）的提问次数，应满足统计学上的最小样本量要求，以确保结果的代表性。

最低提问次数（样本量）的设定可参考中国商务广告协会 2025 年 10 月发布的《2025 年 GEO 行业发展报告》相关分析。建议依据委托方所要求的“最小颗粒度的分析维度”，按照总样本量不低于 30 的原则进行采集。例如：若委托方要求最小颗粒度为“意图词-模型-周”，则每个意图词在每个模型、每周的有效提问样本量应不低于 30 次（见附录 A）。

5.3.7 采集参数设置要求

采集数据时，应明确以下维度的参数设置：

——生成式引擎的名称；

——采集方式：包括每个模型使用的采集方式（对接生成式引擎 API、通过 PC 用户界面模拟用户提问、通过 APP 用户界面模拟用户提问）；

——生成式引擎的配置选项：包括每个模型是否联网，是否开启深度思考，使用哪个模式（例如豆包使用快速模式、或专家模式），使用哪个子模型（例如元宝使用混元或 DeepSeek）；

——国家归属与语言：确认模型采集的数据所属国家或市场，以及采集所使用的语言（例如：ChatGPT，采集新加坡市场数据，使用英语）。

5.3.8 信息采集日期要求

采集数据时应明确设置采集日期范围，例如：2026 年 7 月 1 日-7 月 30 日。

5.3.9 采集数据储存要求

原始采集数据应至少保存至项目结案后一个完整年度（12 个月），以备查询与复核。

应采用安全、可靠的方式存储，确保数据不丢失、不被篡改。敏感数据应加密存储。

5.3.10 异常检测与复核

当核心指标在连续出现异常波动且无对应业务归因时，服务商应在启动异常检测与复核流程，核查采集环境、回溯采样日志并交叉比对同期数据，复核结论应在 3 个工作日内出具。

5.4 数据采集方法

当前 GEO 数据主要采集方式包括以下两种：

——对接生成式引擎 API：直接调用生成式引擎平台开放的 API 进行测试与数据采集，该方式仅适用于开放 API 接口的 AI 模型产品。

——通过用户界面模拟用户提问并监测生成式引擎的回答：通过自动化脚本模拟不同用户，向目标生成式引擎提出预设问题，并采集其输出结果。该方式适用于未开放 API 接口或需模拟真实用户端体验的生成式引擎。

附录 A：采集数据量要求说明

生成式引擎的回答具有随机性。即使是完全相同的问题、设备和模型，多次提问得到的回答内容也可能不同。因此，GEO 评估与测量的数据采集应保证足够的提问次数（样本量），才能客观反映品牌或产品表现。

根据中国商务广告协会 AI 营销应用工作委员会与明略科技《2025 年 GEO 行业发展报告》的研究，可得到以下结论：

当样本量不低于 30 时，品牌可见率指标的统计结果标准差不超过 10%，达到较稳定要求。

因此，建议 GEO 服务商依据委托方要求的“最小颗粒度的分析维度”，按照样本量不低于 30 的原则进行数据采集和统计。

品牌可见度

提问次数 - 样本量	标准差	可见度=20%	可见度=50%	可见度=80%
	N=5	±18%	±24%	±15%
	N=10	±13%	±16%	±10%
	N=30	±7%	±8%	±6%
	N=60	±4%	±5%	±4%
	N=100	±2%	±2%	±2%

参考文献

- [1] GB/T 34090.1—2017 互动广告 第 1 部分 术语概述
- [2] GB/T 34090.3—2017 互动广告 第 3 部分：效果测量要求
- [3] GB/T 1.1—2020 标准化工作导则 第 1 部分：标准化文件的结构和起草规则
- [4] GB/T 42454—2023 信息技术服务 数字化营销服务 移动营销技术规范
- [5] TC260—005 人工智能应用伦理安全指引 1.0
- [6] 《中华人民共和国网络安全法》
- [7] Aggarwal P, Murahari V, Rajpurohit T, 等. GEO: Generative Engine Optimization [EB/OL]. (2024-06-28)
- [8] 中国商务广告协会 AI 营销应用工作委员会、明略科技：《2025 中国 GEO 行业发展报告》