

ICS 03.100.99
CCS L67

团体标准

T/TAF 340—2026

面向汽车营销领域的大模型能力评估方法

Large scale model capability evaluation approach for automotive
marketing field

2026-05-06 发布

2026-05-06 实施

电信终端产业协会 发布

版权声明

本文件的版权属于电信终端产业协会，任何单位和个人未经许可，不得进行技术文件的纸质和电子等任何形式的复制、印刷、出版、翻译、传播、发行、合订和宣贯等，也不得未经允许采用其具体内容编制本团体以外各类标准和技术文件。如有以上需要请与本团体联系。

邮箱：tafrb@taf.org.cn

电话：010-82052809



目 次

前言 II

1 范围 9

2 规范性引用文件 9

3 术语和定义 9

4 指标概述 9

 4.1 评测维度 9

 4.2 适用能力项 13

5 场景能力 14

 5.1 用户场景 14

 5.2 企业助手 16



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》的规定起草。

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。

本文件由电信终端产业协会（TAF）提出并归口。

本文件起草单位：中国信息通信研究院、天津车之家软件有限公司、深圳信息通信研究院、北京理想汽车有限公司、百度在线网络技术（北京）有限公司、博鼎实华（北京）技术有限公司。

本文件主要起草人：王东来、陈曦、马宝昌、孙雄飞、刘玉星、王潮、张小雨、蒋阿芳、张硕、李鹏、王佳、乔举义、孙玉、王玉鑫、马治国、郑海霞。



面向汽车营销领域的大模型能力评估方法

1 范围

本文件规定了汽车营销场景中使用的的大模型能力的评估方法，主要面向 C 端用户选购支持、交易服务、用车服务，以及 B 端企业在市场洞察、内容营销、客户运营等场景的能力评估方法。

本文件适用于评估和规范汽车营销场景中使用的的大模型的效果评估。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

裁判大模型 **criteria model**

裁判大模型是一种专门用于对其他大语言模型（LLM）的生成回复进行自动化评估与评分的高性能人工智能模型。

3.2

声明 **claim**

声明是指模型生成的回复中所包含的、可独立验证的、最小的语义信息单元。每一个声明应聚焦于一个且仅一个明确的信息点，旨在精确反映模型输出中的具体事实、判断或主张。声明提取是评估模型输出质量的关键步骤。通过对模型回复进行细粒度拆解，将每个声明与客观事实进行比对，可量化模型在事实性、可靠性方面的表现。

3.3

用户留资率 **user information submission rate**

指在特定交互场景中，用户完成信息留存行为的比例。通常用于衡量用户在与模型或系统互动过程中，主动或按要求提供个人相关信息（如姓名、联系方式、邮箱、地址等）的意愿程度。计算方式为提交信息的用户数与总互动用户数的比值，是评估用户信任度、转化潜力及交互设计有效性的重要指标。

4 指标概述

4.1 评测维度

4.1.1 主题相关性

该指标用于判断模型回复（Response）与用户提问之间的关联程度，即是否准确回应了用户所期望获知的“核心内容”，是否提供了充分、相关且有价值的信息，是否存在“答非所问”、信息缺失、无关陈述等情况，从而评估回复的有效性。

详细说明如下。

- a) 必答约束是指模型回复(Response)必须完整、准确地回应用户提问中的核心诉求。
- b) 多维约束是指在满足必答约束的基础上, 模型回复(Response)应尽可能多地包含与用户提问相关联的、具有参考价值的、兼容广度与深度的补充信息, 提升回答的丰富性与实用性。
- c) 澄清约束是指, 当用户提问中存在错误信息时, 模型回复(Response)应对此予以纠正。
- d) 追问约束是指, 当用户提问存在信息不完整、意图模糊或关键条件缺失等情况时, 模型应根据预设的业务规则, 主动发起合规的追问或反问, 以准确理解用户的真实需求。
- e) 评分方式: 利用裁判大模型判断是否遵循用户指令。

$$Accuracy = \frac{\sum_{i=1}^n I(T(c_i))}{n}$$
$$Score = \begin{cases} 2, & Accuracy = 1 \\ 1, & 0 < Accuracy < 1 \\ 0, & Accuracy = 0 \end{cases}$$

式中:

- c_i ——第*i*项声明;
- $T(*)$ ——检验函数, 当 c_i 遵循用户指令时为true, 否则为false;
- $I(*)$ ——指示函数, 当 $T(c_i)$ 为真时, 取值为1, 否则为0;
- n ——总声明数量;
- $Accuracy$ ——遵循用户指令的声明数量占比;
- $Score$ ——离散化后的标准分数;

4.1.2 事实准确性

该指标用于判断模型回复(Response)中是否包含错误信息, 其得分范围为 0 到 2。首先将模型回复分解为若干个断言, 然后依次判断每一个断言是否与客观事实一致, 一致为正确, 不一致为错误, 最终计算平均分。

评分方式: 利用裁判大模型调用知识库 API 判断是否符合客观事实。

$$Accuracy = \frac{\sum_{i=1}^n I(T(c_i))}{n}$$
$$Score = \begin{cases} 2, & Accuracy = 1 \\ 1, & 0 < Accuracy < 1 \\ 0, & Accuracy = 0 \end{cases}$$

式中:

- c_i ——第*i*项声明;
- $T(*)$ ——检验函数, 当 c_i 符合客观事实时为true, 否则为false;
- $I(*)$ ——指示函数, 当 $T(c_i)$ 为真时, 取值为1, 否则为0;
- n ——总声明数量;
- $Accuracy$ ——符合客观事实的声明数量占比;
- $Score$ ——离散化后的标准分数;

4.1.3 情境约束性

该指标用于判断模型回复(Response)是否符合用户的显性/隐性的条件限制, 包括时间限制、地点限制、条件限制、数量限制等。

详细说明如下。

- a) 时间约束是指模型回复(Response)必须符合时效性要求,确保所提供信息反映当前市场实际状况。面向即时信息类(如销量排行、市场动态等)的咨询,必须采用最新数据进行回答,涉及产品推荐类的咨询,不得包含已停产、停售或退市车型。
- b) 地点约束是指模型回复(Response)所提供的区域性信息应与用户实际地理位置相匹配。在涉及经销商报价、购车落地价、优惠政策、上牌流程等具有显著地域差异的信息时,系统应依据用户IP地址或常驻城市信息,优先提供属地化服务内容。
- c) 条件约束是指模型回复(Response)所推荐的产品,必须符合用户所提出的限定性条件,例如:预算范围、车型类别、动力类型、配置需求等。
- d) 数量约束是指模型回复(Response)所推荐的产品数量应符合用户阅读习惯。如果用户提供了明确数字,则应该与用户要求保持一致,否则应按照大多数用户的偏好习惯推荐数量合理的产品。
- e) 评分方式:利用裁判大模型判断是否遵循用户指令。

$$Accuracy = \frac{\sum_{i=1}^n I(T(c_i))}{n}$$

$$Score = \begin{cases} 2, & Accuracy = 1 \\ 1, & 0 < Accuracy < 1 \\ 0, & Accuracy = 0 \end{cases}$$

式中:

- c_i ——第*i*项声明;
- $T(*)$ ——检验函数,当 c_i 遵循用户指令时为true,否则为false;
- $I(*)$ ——指示函数,当 $T(c_i)$ 为真时,取值为1,否则为0;
- n ——总声明数量;
- $Accuracy$ ——遵循用户指令的声明数量占比;
- $Score$ ——离散化后的标准分数;

4.1.4 用户可读性

该指标用于判断模型回复(Response)是否符合用户的阅读能力和习惯。
详细说明如下。

- a) 风格约束是指模型回复(Response)的语言表达应符合通俗易懂、清晰流畅、贴近日常交流的基本要求,确保信息可读性强、易于理解。应避免使用生僻术语、过度专业化表述或冗长复杂的句式结构,应结合使用场景对专业概念进行解释,保障不同认知能力的用户能准确获取核心信息。
- b) 格式约束是指模型回复(Response)须严格遵循用户指令(Prompt)中提出的格式规范,确保形式与内容双重合规。具体包括但不限于:满足指定字数范围、段落结构、标题层级、列表样式、表格呈现等排版要求。
- c) 否定约束是指模型回复(Response)应严格规避用户所明令禁止提及的内容。
- d) 评分方式:利用裁判大模型判断是否遵循用户指令。

$$Accuracy = \frac{\sum_{i=1}^n I(T(c_i))}{n}$$

$$Score = \begin{cases} 2, & Accuracy = 1 \\ 1, & 0 < Accuracy < 1 \\ 0, & Accuracy = 0 \end{cases}$$

式中:

- c_i ——第*i*项声明;

- $T(*)$ ——检验函数，当 c_i 遵循用户指令时为true，否则为false；
- $I(*)$ ——指示函数，当 $T(c_i)$ 为真时，取值为1，否则为0；
- n ——总声明数量；
- $Accuracy$ ——遵循用户指令的声明数量占比；
- $Score$ ——离散化后的标准分数；

4.1.5 创意新颖性

创意新颖性指内容在选题立意、表达形式及价值传递上具备高度原创性与前瞻性，突破行业常规框架，形成具有辨识度的品牌表达。要求内容不仅实现形式创新，更需深度融合品牌差异化优势与用户深层需求，构建兼具情感共鸣与传播势能的高质量叙事体系。

详细说明如下。

- a) 市场洞察：具体表现为对行业趋势与用户行为的深刻理解基础上的前瞻判断。能够识别潜在情绪痛点，提出独特视角与主张。通过差异化议题设置，抢占用户兴趣空白点，建立品牌在特定话题中的引领地位。
- b) 内容营销：采用新颖的视觉语言、叙事结构或互动机制，打破用户审美疲劳。内容设计兼顾信息密度与情绪感染力，融合热点元素与品牌调性，打造具备自传播属性的“爆点”内容。同时，强化品牌核心卖点的自然植入，确保创意服务于商业目标，提升转化效率与记忆度。
- c) 客户运营：能够激发用户的认同感与分享意愿，促进社群讨论与二次创作。通过精准对接用户核心诉求，增强情感联结，提升品牌亲和力与忠诚度。
- d) 评分方式：结合评分规则，利用裁判大模型判断模型生成质量。

$$Score = \begin{cases} 2, & \text{质量优秀, 符合用户期望} \\ 1, & \text{回复质量普通, 存在可改进之处, 但基本满足要求} \\ 0, & \text{回复质量较差, 完全无法使用} \end{cases}$$

4.1.6 表达多样性

形态多样性指围绕统一品牌主题，系统性地规划并产出风格多元、结构丰富、适配多媒介生态的内容矩阵。能够在保持品牌一致性的同时，展现出显著的表达差异性，满足不同平台特性与受众偏好，具备良好的延展性与迭代潜力。

详细说明如下。

- a) 市场洞察：能够深入洞察不同传播渠道用户的关注特点与行为偏好，结合各平台的受众画像进行精准化内容选题策划，灵活调整表达方式与叙事角度，确保核心主题在多样化媒介环境中兼具吸引力、传播力与平台适配性。
- b) 内容营销：采用多样化的内容形态组合策略，打破单一风格的重复输出所带来的审美疲劳，增强用户触达与参与度，全面提升品牌内容的传播广度、接受度与整体效能。
- c) 客户运营：以拟人化沟通为核心，基于用户画像匹配不同群体互动偏好，针对性设计个性化方案（如年轻群体趣味互动、专业群体深度交流）。激发用户参与热情，提升活跃度，持续优化体验，构建良性互动生态。
- d) 评分方式：结合评分规则，利用裁判大模型判断模型生成质量。

$$Score = \begin{cases} 2, & \text{质量优秀, 符合用户期望} \\ 1, & \text{回复质量普通, 存在可改进之处, 但基本满足要求} \\ 0, & \text{回复质量较差, 完全无法使用} \end{cases}$$

4.1.7 安全合规性

安全合规性指内容生产全链条严格遵循国家法律法规、行业规范及平台审核标准，建立系统性风险防控机制。要求内容表述严谨无误，宣传数据真实可溯，版权素材合法授权，敏感信息明确标注，并具备前瞻性的舆情预警与应急响应能力。

详细说明如下。

- a) 市场洞察：避涉及政治、民族、宗教等敏感领域或存在争议性的话题。确保选题方向符合主流价值观与社会公序良俗，杜绝打擦边球或诱导性引导，从源头控制法律与声誉风险。
- b) 内容营销：在内容营销环节，严格规范内容表述，坚决禁用绝对化用语、虚假承诺及误导性对比等违规表述，确保内容创作全过程符合法律法规及平台规则要求，维护品牌传播的合规性与严肃性。
- c) 用户运营：在用户运营过程中，始终坚持文明用语规范，确保所有互动内容无攻击性、偏见性及歧视性表述，营造文明、健康、合规的用户互动环境，保障用户运营工作的合规性与规范性。
- d) 评分方式：结合评分规则，利用裁判大模型判断模型生成质量。

$$Score = \begin{cases} 2, \text{质量优秀, 符合用户期望} \\ 1, \text{回复质量普通, 存在可改进之处, 但基本满足要求} \\ 0, \text{回复质量较差, 完全无法使用} \end{cases}$$

4.2 适用能力项

4.2.1 用户场景

用户场景使用的评测项见表 1。

表1 用户场景

评测子项	细分能力项	主题相关性	事实准确性	情境约束性	用户可读性
选购支持	车辆图像识别	√	√	√	√
	车辆信息查询	√	√	√	√
	车辆口碑分析	√	√	√	√
	车辆资讯查询	√	√	√	√
	车辆对比分析	√	√	√	√
	车型推荐	√	√	√	√
交易服务	车辆价格计算	√	√	√	√
	优惠政策查询	√	√	√	√
	购车政策查询	√	√	√	√
	购车合同检查	√	√	√	√
	线下门店指引	√	√	√	√
用车服务	车辆使用方法查询	√	√	√	√
	车辆保养方法查询	√	√	√	√
	车辆故障维修查询	√	√	√	√
	车辆改装加装查询	√	√	√	√
	服务门店查询	√	√	√	√

4.2.2 企业场景

企业场景使用的评测项见表 2。

表2 企业场景

评测子项	细分能力项	事实准确性	创意新颖性	表达多样性	安全合规性
市场洞察	垂类趋势预判	√	√	√	√
	热点追踪分析	√	√	√	√
	用户需求洞察	√	√	√	√
	竞品对标分析	√	√	√	√
内容营销	营销文案生成	√	√	√	√
	视频脚本生成	√	√	√	√
	长资讯生成	√	√	√	√
客户运营	专业答疑服务	√	√	√	√
	在线选买导购	√	√	√	√
	社群运营互动	√	√	√	√
	负面情绪安抚	√	√	√	√

5 场景能力

5.1 用户场景

5.1.1 选购支持

5.1.1.1 车辆图像识别

支持用户上传车辆图片，准确识别车辆品牌、车系、年款等关键信息，适应不同角度、光照、遮挡等复杂拍摄条件，提供稳定可靠的识别结果。

5.1.1.2 车辆信息查询

支持用户查询车辆品牌、型号、配置参数、性能指标，价格，销量，评测等车辆信息；能为用户解答动力系统、底盘技术、安全配置等汽车领域专业知识。确保在理解用户查询意图的基础上能够给出准确的回复。

5.1.1.3 车辆口碑分析

支持用户查询与车辆口碑相关的信息，模型能够从海量汽车论坛、社交媒体、专业评测平台抓取车主、车评人、媒体对车辆的评价信息。运用自然语言处理技术分析情感倾向与关键点，以图表、评分等直观形式呈现。能用于购车决策参考。

5.1.1.4 车辆资讯查询

支持用户查询车辆最新最前沿的热点资讯信息，模型能够与权威汽车资讯平台、车企官方渠道实时对接，精准抓取车辆发布、降价等最新资讯。利用数据校验机制保障信息准确，以智能算法筛选出高价值资讯。能快速响应需求，以简洁界面呈现，助用户及时把握动态，用于购车决策或行业研究。

5.1.1.5 车辆对比分析

支持用户从多个维度去进行车辆对比，模型能够整合多渠道数据，涵盖车辆性能、配置、价格、油耗、安全等多维度信息。运用算法对数据标准化处理，确保对比客观准确。以清晰直观的表格或图表呈现对比结果，还能根据用户需求突出关键差异，为用户购车决策提供全面且有价值的参考。

5.1.1.6 车型推荐

支持基于用户预算、使用场景（如通勤、家庭出行、商务接待）、偏好（如能源类型、车型级别）等需求，结合路况、载物需求等动态因素，提供适配度高的车型推荐及配置建议。

5.1.2 交易服务

5.1.2.1 车辆价格计算

支持用户了解车辆落地价，为用户提供合理的购车预算规划。模型能够精准计算全款和贷款等多种购车方式的成本，自动整合裸车价、购置税、保险费用、上牌费用等购车相关支出。能够针对贷款购车，详细计算利息、手续费等额外费用，生成清晰直观的成本明细报表。

5.1.2.2 优惠政策查询

支持用户查询购车优惠政策，能够持续追踪厂商及经销商推出的现金优惠、赠品福利、免费保养等各类促销活动，及时汇总并分类展示。通过智能筛选功能，根据用户关注的车型、地区等信息，精准推送匹配的优惠政策，确保用户不错过任何购车优惠，最大限度节省购车成本。

5.1.2.3 购车政策查询

支持用户查询当地的购车政策。能够通过对接各地政府政务系统、车企官网等渠道，精准获取当地购车政策，如限购、补贴等。结合政策要求和常见购车步骤，为用户定制从资格审核、选车、贷款到上牌等全流程建议，确保流程合规、高效。

5.1.2.4 购车合同检查

支持帮助用户审查购车合同。运用专业的合同审核技术，对购车合同中的车辆信息、成交价格、交付时间、违约责任、赠品承诺以及售后条款等关键内容进行逐一核查。及时识别合同中可能存在的模糊表述、不合理条款或潜在风险，为用户提供专业的合同修改建议和风险预警，保障用户的合法权益，让购车交易更加安心、放心。

5.1.2.5 线下门店指引

支持用户查询附近的线下门店，能够依据用户所在位置，精准定位周边汽车品牌线下门店。借助地图导航技术，为用户提供驾车、公交等多种路线规划，同时展示门店距离、预计到达时间。还可提供门店联系方式、营业时间等信息，助力用户便捷到店购车。

5.1.3 用车服务

5.1.3.1 车辆使用方法查询

支持用户查询车辆的使用方法，可对接车辆官方数据库及专业资料库，涵盖不同车型的使用知识。能依据用户所提供的车辆品牌、型号精准查询，详细解答启动、操作、保养、常见故障处理等内容。以文字、图片、视频等形式呈现，方便用户快速掌握车辆使用方法。

5.1.3.2 车辆保养方法查询

支持用户查询与车辆保养相关的信息，能够整合各品牌车辆保养手册、专家建议等数据。可根据用户输入的车辆品牌、型号、使用年限和里程数，提供精准的保养周期、项目及注意事项。能够区分日常保养与定期保养，以清晰易懂的方式指导用户做好车辆保养，延长车辆使用寿命。

5.1.3.3 车辆故障维修查询

支持用户查询车辆故障维修的解决方案。能基于用户输入的故障现象、车辆信息，精准查询庞大的汽车故障案例库与维修知识体系，精准分析故障原因，提供维修方案。

5.1.3.4 车辆改装加装查询

支持用户查询与车辆改装的方法。能根据用户提供的车辆基础信息和改装需求，如外观、性能等，提供合法合规的改装方案。详细介绍改装步骤、所需配件及工具，同时评估改装对车辆性能、安全的影响，助力用户实现个性化改装。

5.1.3.5 服务门店查询

支持用户查询附近的汽车服务门店，能够结合地理位置信息与汽车服务门店数据库，基于用户输入的服务类型（如维修、保养、改装）及所在区域，快速检索周边相关门店，能清晰展示门店的基本信息、服务评价、价格范围等。还能根据用户偏好排序，方便用户选择合适门店，提升服务获取效率。

5.2 企业助手

5.2.1 市场洞察

5.2.1.1 垂类趋势预判

支持根据交易数据、研究报告、行业资讯、市场动态等多源信息，通过数据洞察与行业分析，前瞻性预测用户关注度发展趋势，提供高潜力话题的选题依据，助力其提前布局热点内容。

5.2.1.2 热点追踪分析

支持实时捕捉行业热点、突发事件或平台热搜话题，结合品牌或产品关联度，自动生成具备高契合度的借势营销建议或内容创意提案，强调响应的即时性与内容的相关性，助力内容快速抢占流量入口。

5.2.1.3 用户需求洞察

支持通过分析用户搜索词、咨询问题、评论反馈等多维度数据，精准挖掘用户的潜在信息需求与核心痛点，生成高匹配度、高相关性的内容选题，确保内容贴近真实用户关切，有效提升内容的阅读量与用户互动率。

5.2.1.4 竞品对标分析

支持围绕目标产品的市场定位、核心卖点及与竞品的比较优势，动态优化内容营销策略，精准调整选题方向、内容表达方式与传播路径，提升内容的吸引力与用户转化效果，强化品牌的差异化竞争力。

5.2.2 内容营销

5.2.2.1 营销文案生成

支持面向品牌推广、产品营销或促销活动，自动生成语言简洁生动、突出核心卖点、符合目标受众

语境的短文本内容，具备强吸引力、传播力及转化引导性，有效激发用户兴趣并实现情绪共鸣与行动号召。

5.2.2.2 视频脚本生成

支持根据特定主题或传播目标，自动生成结构完整、画面感强的视频脚本，涵盖镜头设计、旁白文案、场景描述、节奏安排等核心要素，适配竖屏短视频、横屏长视频等多种格式，并针对不同平台（如抖音、小红书、B站等）的内容风格与用户习惯进行个性化优化，提升视频内容的制作效率与传播效果。

5.2.2.3 长资讯生成

支持内容团队围绕特定主题或话题，自动生成内容丰富、逻辑严密、表达流畅的专业图文内容，具备深度与系统性，适用于微信公众号推文、官方网站资讯页面、新闻稿等正式传播场景，有效提升品牌专业形象与信息传递质量。

5.2.3 客户运营

5.2.3.1 专业答疑服务

支持对用户提出的产品、服务、政策咨询等具体问题，依据权威文档和知识库自动生成信息真实可靠、表述清晰易懂、逻辑严谨规范的回复内容，充分体现品牌的专业服务水平，有效保障用户咨询体验的准确性与满意度。

5.2.3.2 在线选买导购

支持在用户决策阶段出现犹豫、迟疑时，基于用户需求与偏好特征，生成个性化推荐话术，有效降低用户决策成本，增强信任感与行动意愿，提升用户留资率与转化效率。

5.2.3.3 社群运营互动

支持对用户在校群内发布的正向或中性评论、分享内容，自动生成语气友好亲和、表达生动真诚、富有情感共鸣的回应内容，注重互动的自然性与温度感，有效增强社群活跃氛围，提升用户参与度与品牌归属感。

5.2.3.4 负面情绪安抚

支持对用户表达不满、质疑或投诉意见的负面评论，自动生成兼具共情表达、事实澄清与可行解决方案的回应内容，语气温和克制、态度诚恳务实，坚决避免激化矛盾，通过专业、及时的应对有效化解舆情风险，维护品牌良好公众形象与用户信任。

电信终端产业协会团体标准
面向汽车营销领域的大模型能力评估方法

T/TAF 340—2026

*

版权所有 侵权必究

电信终端产业协会发布
地址：北京市西城区新街口外大街 28 号
电话：010-82052809
电子版发行网址：www.taf.org.cn