

# 基于在线评论的国产生成式人工智能产品 用户满意度分析\*

杨怡 万晨硕 李瑾 胡泽文  
(南京信息工程大学管理工程学院, 南京 210044)

**摘要:** 针对现有国产生成式人工智能 (Generative Artificial Intelligence, GenAI) 产品用户满意度研究普遍依赖问卷调查, 以及难以刻画属性层面非线性影响的问题, 本研究旨在构建基于在线评论数据的用户满意度分析与优化决策框架, 系统揭示不同产品属性对整体满意度的影响机制及其改进优先级。以豆包、DeepSeek、通义千问、Kimi和文心一言等主流国产GenAI产品为研究对象, 首先通过隐含狄利克雷分布 (Latent Dirichlet Allocation, LDA) 主题模型识别用户关注的核心属性, 并对属性绩效进行量化; 随后利用SHAP (Shapley Additive Explanations) 方法估计各属性对用户满意度的影响程度; 最后结合重要性-绩效分析 (Importance-Performance Analysis, IPA) 模型对属性优化优先级进行分类。结果表明, 用户关注的GenAI产品属性主要包括输出质量、效率、图片识别、多场景支持、自主化、文本创作、学习辅助和语音交互。其中, 输出质量为优势属性, 多场景支持与学习辅助为重点改进方向, 图片识别为潜在改进属性, 效率、自主化、文本创作和语音交互属于继续维持属性。本研究为国产GenAI产品功能优化提供了数据驱动的决策依据。

**关键词:** 生成式人工智能; 用户满意度; 产品属性; 细粒度情感分析; 机器学习; 重要性-绩效分析

**中图分类号:** G350 **DOI:** 10.3772/j.issn.1673-2286.2025.11.007

**引文格式:** 杨怡, 万晨硕, 李瑾, 等. 基于在线评论的国产生成式人工智能产品用户满意度分析[J]. 数字图书馆论坛, 2025, 21(11): 62-72.

生成式人工智能 (Generative Artificial Intelligence, GenAI) 已成为人工智能技术的战略新高地, 是驱动技术范式革新、引领产业升级的核心引擎。2023年, 国家网信办等七部门印发《生成式人工智能服务管理暂行办法》<sup>[1]</sup>, 通过备案制与算法透明化要求, 构建安全发展框架, 推动GenAI应用生态优化发展。国产GenAI呈现蓬勃发展态势。截至2025年6月30日, 全国累计439款GenAI服务完成备案, 236款GenAI应用或功能完成登记, 覆盖金融、医疗、教育等领域<sup>[2]</sup>。然而, 问题与机遇

并存。墨尔本大学与毕马威会计师事务所对来自47个国家或地区的约4.8万人进行调查, 发现超过半数 (54%) 的受访者对信任人工智能持谨慎态度, 但大多数人 (72%) 接受使用人工智能<sup>[3]</sup>。

随着国产GenAI产品的快速迭代与广泛应用, 用户满意度已成为影响其技术落地与市场扩展的关键因素。陈鹏等<sup>[4]</sup>研究表明, GenAI产品的用户接受程度受多种因素共同作用。既有研究<sup>[5-7]</sup>从应用场景或使用意愿等角度对GenAI产品的用户满意度展开探讨, 为理解

收稿日期: 2025-10-09

\*本研究得到教育部人文社会科学研究规划基金项目“虚拟学术社区科研数据流转演化机制及‘共享—重用’提升策略研究” (编号: 23YJA870012)、江苏省社会科学基金项目“基于高价值专利的江苏未来产业前沿交叉技术识别与攻关机制研究” (编号: 24TQB005)、江苏高校哲学社会科学研究重大项目“江苏省未来产业高价值专利智能识别与培育机制研究” (编号: 2024SJZD065) 资助。

用户行为提供了有益参考。然而, 现有研究普遍依赖问卷调查等结构化数据, 样本规模有限且主观性较强, 难以全面反映用户体验的动态变化与真实需求。

近年来, 部分研究开始利用在线评论等非结构化数据分析用户满意度, 但总体而言, 针对国产GenAI产品的系统性满意度研究仍较为缺乏, 尤其在属性层面上, 现有方法难以刻画不同因素对用户满意度的相对重要性及其复杂作用关系。鉴于此, 本研究建立可解释机器学习模型, 深入捕捉属性情感值对满意度的非线性影响, 并结合SHAP (Shapley Additive Explanations) 方法进行属性重要性估计, 深入挖掘GenAI产品的短板与优化路径, 为产品迭代提供科学支撑。

## 1 相关研究

### 1.1 GenAI产品用户满意度研究

当前, 关于GenAI产品用户满意度的研究或采用访谈<sup>[8]</sup>、定性比较分析<sup>[9-10]</sup>等方法, 探讨学术、教育等特定环境中的伦理风险、技术缺陷等因素对用户行为的影响; 或采用描述性统计、回归分析、结构方程模型等量化统计方法对学生、旅行者等特定用户群体的GenAI使用意愿、满意度及影响因素进行实证分析<sup>[11-12]</sup>。总体来看, 上述研究多以国外主流GenAI产品为研究对象<sup>[13]</sup>, 在用户群体与产品选择上存在一定局限, 尚缺乏针对国产GenAI产品的系统性用户满意度分析研究。

从方法层面看, 问卷与访谈等传统数据采集方式易受样本偏差和社会期望效应影响, 难以充分反映用户在真实使用场景中的自然反馈<sup>[14]</sup>; 而基于回归分析或结构方程模型的量化研究通常难以刻画变量间复杂的非线性关系, 且重要性估计易受到混淆变量干扰<sup>[15]</sup>。由于在线评论数据的广泛可获得性, 学者们开始利用应用商店评论开展GenAI满意度研究。Alabduljabbar<sup>[16]</sup>基于ISO 9241标准分析了多款GenAI产品的可用性与满意度; Chhetri等<sup>[17]</sup>利用大语言模型对评论进行分类与情感分析, 识别了主要满意与不满因素; Bilal等<sup>[18]</sup>通过机器学习方法比较了ChatGPT、Copilot与Gemini的用户满意度水平。相比之下, 国产GenAI产品的相关研究仍较为匮乏。因此, 本研究基于在线评论, 对国内主流GenAI产品的用户满意度及影响满意度的关键属性进行系统分析。

### 1.2 细粒度情感分析技术

情感分析是在线评论与用户反馈研究的重要技术手段。早期研究多为粗粒度情感分析, 通过识别文本的整体情感极性来反映用户态度。例如: Pang等<sup>[19]</sup>使用机器学习方法对电影评论进行文档级情感分类; Turney等<sup>[20]</sup>利用词汇极性与共现信息计算文本情感倾向。然而, 该类方法仅关注整体情感倾向, 难以揭示用户在具体属性层面的细微感知差异, 从而限制了其在满意度归因与产品优化研究中的应用价值。为弥补上述不足, 学界逐渐转向细粒度情感分析, 其核心在于从属性或方面层面刻画用户的情感倾向<sup>[21]</sup>。与粗粒度分析相比, 细粒度情感分析通常围绕属性识别与情感极性计算两个关键任务展开<sup>[22]</sup>, 能够建立属性与情感之间的对应关系, 从而更精准地反映用户对不同产品特征的评价。近年来, 该方法已在社交媒体分析<sup>[23]</sup>、心理健康文本分析<sup>[24]</sup>等领域得到验证。鉴于细粒度情感分析在揭示用户关注焦点和情绪差异方面的优势, 本研究将其引入GenAI产品用户满意度研究, 通过在属性层面对用户情感进行量化刻画, 为后续的重要性分析与决策支持提供基础。

### 1.3 IPA模型

重要性-绩效分析 (Importance-Performance Analysis, IPA) 模型由Martilla等<sup>[25]</sup>提出, 旨在通过比较产品或服务属性的重要性与实际表现, 识别优先改进方向。传统IPA模型主要依赖问卷数据获取用户评价, 并通过二维矩阵进行属性定位, 已广泛应用于旅游服务<sup>[26]</sup>、酒店管理<sup>[27]</sup>、高等教育<sup>[28]</sup>等领域。然而, 受限于样本规模、数据时效性及受访者主观判断, 该方法在重要性与绩效估计的客观性方面存在一定不足<sup>[29]</sup>。随着文本挖掘技术的发展, 学者们开始将在线评论数据引入IPA模型框架, 利用机器学习方法从非结构化文本中挖掘用户偏好信息<sup>[30]</sup>。此基础上, 部分研究开始构建基于在线评论文本的IPA框架。例如: Shen等<sup>[31]</sup>基于在线评论与维护记录构建集成化IPA框架, 对产品属性的重要性与绩效进行联合评估; Nam等<sup>[32]</sup>则通过文本分析从航空服务评论中提取属性并开展IPA研究, 验证了IPA模型在非结构化文本数据场景下的适用性。然而, 既有研究在重要性估计上多依赖整体特征权重或主题概率, 模型输出的可解释性有限, 难以揭示属性在个体层面的

具体贡献。为弥补上述不足, Joung等<sup>[33]</sup>引入SHAP方法对IPA模型进行扩展, 将博弈论思想融入重要性测度, 实现了从整体分析向个体解释的转变。因此, 本研究融合细粒度情感分析与基于SHAP的可解释重要性估计方法, 引入IPA模型, 对国产GenAI产品用户满意度进行系统分析, 以期对产品优化提供更具解释力的决策依据。

## 2 研究设计

本研究的研究思路如图1所示。首先, 收集国产GenAI产品的用户在线评论并进行数据预处理; 接着, 基于细粒度情感分析识别用户关注的产品属性, 据此计算其平均情感得分, 以完成绩效估计; 最后, 构建深度神经网络预测模型, 引入SHAP方法评估属性情感对

满意度的影响程度, 获得重要性估计, 并结合绩效完成IPA优先级分类。在此基础上, 依据不同属性在IPA矩阵中的定位结果, 进一步提出针对性的产品优化策略与改进建议, 实现从分析结果到决策支持的有效衔接。

### 2.1 数据收集与处理

依据产品影响力和用户基础, 本研究选取豆包、DeepSeek、通义千问、Kimi、文心一言5款具有代表性的国产GenAI产品作为研究对象, 并依托第三方数据服务平台(点点数据)获取应用商店中的用户评论和评分数据。评分采用5分制, 最低分为1分, 最高分为5分。数据时间范围为2024年7月7日—2025年7月5日, 共收集30 859条数据, 部分示例如表1所示。

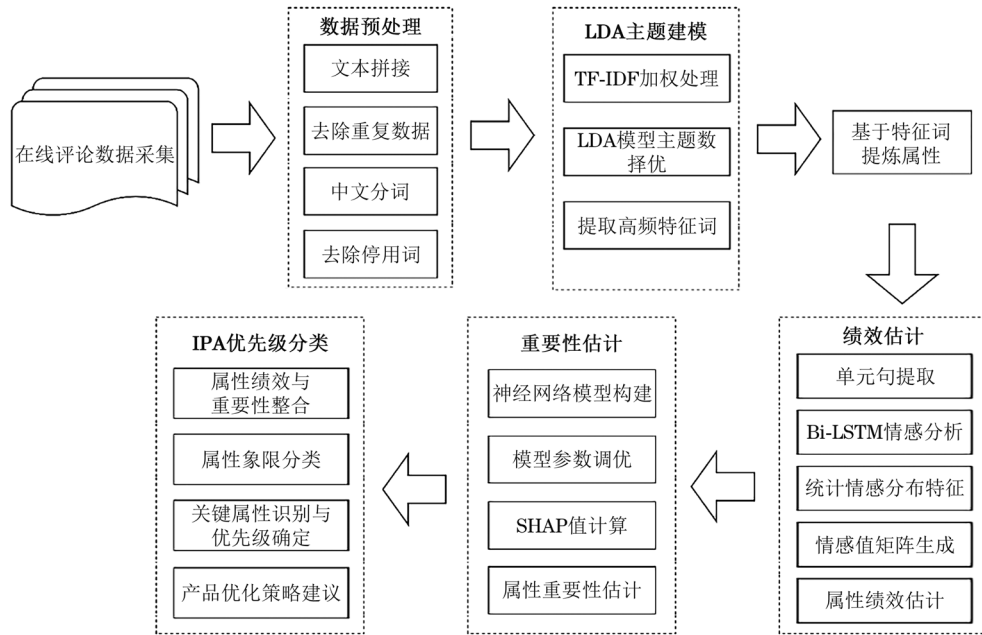


图1 研究思路

表1 应用商店部分用户评论内容及评分

标题	内容	评分
非常好	希望能扩展一下对话长度上限, 如果能增加一个新对话继承历史对话的设定就更好了	4
对话太容易达到上限	刚想问下一个问题, 结果显示对话上限了, 要开启新对话, 希望能优化, 增加对话次数	1
体验感很好	解答很详细, 之前觉得国产不给力, 一直用Copilot, 尝试了DeepSeek确实很给力! 纯国产真开心, 感谢默默付出的科技工作者, 真给国家长脸	5

在文本分析前, 对评论数据进行清洗以剔除低质量样本, 包括空文本及无语义评论。具体步骤包括: ①合并标题和内容字段; ②使用正则表达式清除表情符号、标点等无关字符; ③过滤过短评论, 删除字符数 $\leq 5$ 的评论, 如“很好”“一般”等, 以确保评论内容具备分析价值, 获得有效评论29 094条用于后续研究; ④采用Python软件中的jieba工具对评论文本进行分词, 并保留词性为名词的词汇, 以提升主题分析的准确性与参考价值; ⑤基于常用的中文停用词表去除无属性意义的停用词。经分词与去停用词处理后, 最终保留

2 722个有效词项。

## 2.2 基于细粒度情感分析的绩效估计

细粒度情感分析通过刻画属性与情感表达之间的对应关系, 能够从在线评论中提取用户对不同产品属性的主观评价信息<sup>[34]</sup>, 在用户满意度研究中具有重要应用价值。相较于整体情感分析, 细粒度方法能够在属性层面刻画用户满意度的形成机制, 更契合真实使用体验导向的研究目标。由于在线评论文本通常未对产品属性进行显式标注, 属性识别需要借助无监督的主题分析方法。主题模型可通过词语共现关系识别评论中的潜在属性关注点。相较于规则或关键词方法, LDA模型具备无监督、可扩展和语义解释性优势<sup>[35]</sup>, 因此采用LDA模型进行属性识别。具体操作包括: ①基于词频和逆文档频率(Term Frequency-Inverse Document Frequency, TF-IDF)算法对文本加权, 降低高频通用词权重, 以提升主题分布的表征效果; ②利用主题一致性指标确定最优主题数; ③在此基础上进行主题建模, 筛选并优化属性特征词列表, 据此完成属性命名。

需要进一步说明的是, LDA模型识别的主题并不完全对应传统意义上的功能属性, 部分主题反映的是用户在特定使用情境下的体验需求。已有研究<sup>[36]</sup>指出, 用户满意度受到功能性能与情境体验的共同影响。基于此, 在主题建模结果的基础上对各主题进行语义审查, 将其划分为功能属性与情境属性两类。尽管二者在实际使用中存在一定关联, 但经语义校准后均可视为相对独立的满意度影响因素, 因此在后续分析中统一作为IPA模型中的属性维度进行处理。

在属性识别的基础上, 为刻画用户在各属性下的情感倾向, 引入百度开源情感分析工具Senta, 利用其基于双向长短期记忆(Bidirectional Long Short-Term Memory, Bi-LSTM)网络的预训练模型计算属性级情感值。考虑到单条评论中可能存在多次属性提及和情感识别噪声, 先对评论进行句级切分, 得到单元句, 再通过均值聚合的方式对情感结果进行平滑处理, 以降低个别误判对总体绩效估计的影响。最终, 各属性在所有相关评论中的平均情感值被用作其绩效水平测度, 为后续基于IPA模型的分析提供量化基础。

对每条评论的属性情感值进行离散化编码, 具体

规则如式(1)所示。

$$e' = \begin{cases} 1, & \text{若 } 0 < e \leq 0.2 \\ 2, & \text{若 } 0.2 < e \leq 0.4 \\ 3, & \text{若 } 0.4 < e \leq 0.6 \\ 4, & \text{若 } 0.6 < e \leq 0.8 \\ 5, & \text{若 } 0.8 < e \leq 1.0 \end{cases} \quad (1)$$

式中:  $e$ 表示原始情感值, 由情感分析模型对评论文本进行计算生成;  $e'$ 表示离散化后的情感得分, 1表示极度消极, 2表示较消极, 3表示中性, 4表示较积极, 5表示极度积极。若同一评论同时提及多个属性, 则分别计算并形成属性-情感矩阵, 为后续绩效估计提供基础数据。

在绩效估计方面, 将各属性在所有评论中的平均情感得分视为其绩效水平, 可在一定程度上平滑个别情感值偏差, 提高估计结果的稳定性。产品属性绩效的计算公式如式(2)所示。

$$S_j = \frac{1}{n_j} \sum_{i=1}^{n_j} s_{ij} \quad (2)$$

式中:  $S_j$ 表示第 $j$ 个属性的绩效;  $s_{ij}$ 表示第 $i$ 条评论中第 $j$ 个属性的情感得分;  $n_j$ 表示包含该属性的评论数量。绩效值越高, 说明用户对该属性的体验越积极。绩效值能够反映属性在整体满意度中的表现水平。

## 2.3 基于可解释机器学习SHAP方法的重要性估计

在重要性估计方面, 引入SHAP方法, 将评论中的属性情感值视为模型输入特征, 用户评分视为预测目标, 构建可解释机器学习模型, 以揭示国产GenAI产品属性对用户满意度的非线性影响。为避免单一模型结构可能带来的估计偏差, 同时构建深度神经网络和支持向量机(Support Vector Machine, SVM)模型进行对比, 检验属性重要性排序的一致性。平均SHAP绝对值的聚合可在一定程度上抵消噪声, 提高重要性估计稳定性。特征对预测结果的贡献可通过SHAP值计算, 其平均绝对值表示特征重要性。特征重要性计算公式如式(3)所示。

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [v(S \cup \{j\}) - v(S)] \quad (3)$$

式中:  $\phi_j$ 为属性 $j$ 的SHAP值;  $N$ 为属性集合;  $|N|$ 为

属性总数;  $S$ 为不包含属性 $j$ 的任意属性子集;  $v(S)$ 为当模型仅使用属性子集 $S$ 时的预测值。SHAP值的正负表示属性对预测结果的正向或负向影响,其绝对值则反映影响强度。平均SHAP绝对值可视为该属性在整体样本中的平均贡献,能够量化其对满意度预测的重要性。从统计角度看,平均SHAP绝对值能够刻画属性重要性,并避免回归权重受多重共线性与尺度差异的影响,为IPA模型提供稳健的重要性估计。

属性重要性估计包括以下具体操作。①对用户评分进行二值化处理(用户评分 $>3$ 为积极,用户评分 $\leq 3$ 为消极)。②基于全连接神经网络构建预测模型,模型包含两个隐藏层,能够有效捕捉各属性情感值对评分的非线性影响。将该模型与SVM模型进行对比验证,以确保模型的稳健性。③通过计算各属性平均SHAP值的绝对值来确定其重要性。平均SHAP值的绝对值越大,说明该属性对满意度的影响越强。属性重要性估计值的计算公式如式(4)所示。

$$I_j = \frac{1}{n_j} \sum_{i=1}^{n_j} |\phi_{ij}| \quad (4)$$

式中:  $I_j$ 为第 $j$ 个属性的重要性估计;  $\phi_{ij}$ 为第 $i$ 条评论中该属性的SHAP值。

## 2.4 IPA优先级分类

在获得各属性的重要性与绩效估计后,参考IPA分类方法,以所有属性的重要性均值和绩效均值为分界完成四象限划分。由于IPA更关注属性在二维空间中的相对位置而非绝对数值,即便重要性或绩效存在轻微估计误差,其象限归属也基本保持稳定,因此IPA结果具有较好的稳健性。属性重要性均值与绩效均值的计算公式如式(5)~式(6)所示。

$$\bar{I} = \frac{1}{|N|} \sum_{j=1}^{|N|} I_j \quad (5)$$

$$\bar{S} = \frac{1}{|N|} \sum_{j=1}^{|N|} S_j \quad (6)$$

式中:  $\bar{I}$ 为属性重要性均值;  $\bar{S}$ 为绩效均值。具体而言:如果  $I_j \geq \bar{I}$  且  $S_j \geq \bar{S}$ , 则该属性位于优势提升区;如果  $I_j < \bar{I}$  且  $S_j \geq \bar{S}$ , 则该属性位于继续维持区;如果  $I_j < \bar{I}$  且  $S_j < \bar{S}$ , 则该属性位于后续机会区;如果  $I_j \geq \bar{I}$  且  $S_j < \bar{S}$ , 则该属性位于重点改进区。

## 2.5 基线模型对比

本研究构建了一个基于传统方法的基线模型用于对照实验,以评估所提方法的性能优势。该基线模型包括以下步骤:①采用TF-IDF与K-Means算法识别评论主题;②利用SVM回归预测用户评分;③通过SVM回归的特征重要性估计各属性的重要性;④结合IPA对属性进行优先级分类。基于该基线模型与所提方法,从主题一致性、SHAP覆盖度和IPA分类稳定性3个维度进行比较:主题一致性衡量属性识别结果在语义聚合与解释性层面的表现;SHAP覆盖度反映重要性估计对模型预测的解释能力;IPA分类稳定性通过Bootstrap重抽样检验属性象限划分对样本扰动的鲁棒性。通过上述比较,可系统评估两种方法在属性识别、重要性估计与优先级分类3个核心环节的差异,对所提方法的优越性进行验证。

## 3 结果与讨论

### 3.1 国产GenAI产品属性识别与情感分析

基于Python的gensim库构建LDA主题模型,对评论文本进行主题分析。主题一致性得分情况如图2所示。当主题数为8个时,一致性得分最高,模型表现最优。

在最优主题数条件下,基于各主题的高频特征词并结合具体评论语义,对国产GenAI产品属性进行命名,产品属性及特征词分布如表2所示。输出质量属性反映用户对生成能力与内容表现的综合评价,效率属性反映用户对性能表现与时间成本的评价,图片识别属性反映用户对图像识别准确性与呈现效果的评价,多场景支持属性体现用户对多模态与多应用场景功能的使用需求,自主化属性反映用户对产品自主研发能力与

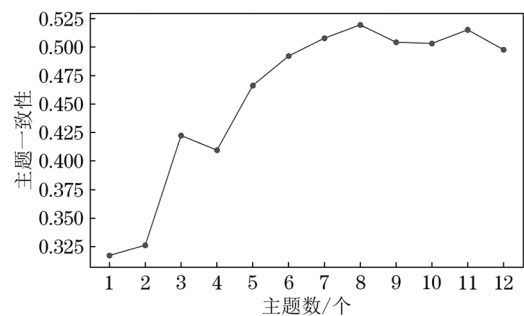


图2 主题一致性得分情况

表2 产品属性及特征词分布

产品属性	高频特征词	评论示例
输出质量	智能、体验、价值、专业性、理解能力、高质量、逻辑性、观点、瑕疵、太烂	更新后理解能力直线下降, 我真不理解为什么更新后DeepSeek变得听不懂话了, 我就让他给我总结个表格, 都说了要分类, 我提醒了好多遍它都一直不分类, 我真的好难受, 甚至想把版本调回去! 希望优化
		算是相对早使用的一批人吧, 查询一些关于考试的内容, 真的很高质量且非常好地解决了我的问题。分析思考能力非常令人震撼
效率	工具、解决问题、工作效率、效率、实用性、性能、反应速度、压力、规范、节省时间	确实是一个提升效率的好工具
		智能助手首选! DeepSeek彻底改变我的效率体验。这款AI工具简直让人惊艳! 通过深度学习的智能推荐系统, 它能精准预测我的需求, 无论是工作文件检索还是知识查询都秒出结果。界面设计简约大气, 交互体验流畅自然, 特别点赞“智能联想”功能, 输入关键词就能生成完整解决思路。更支持多终端同步, 通勤路上用手机也能处理复杂任务。自从使用后工作效率提升200%, 堪称职场人和学生的必备神器! 强烈推荐给所有追求高效生活的朋友, 快来体验未来科技的魅力吧
图片识别	图片、精准、效果、指令、格式、画面、准确率、错误率、尺寸、版权	图片识别不了文字就无法使用, 所以你也无法提问图片是什么
		要么就是图片上传失败, 要么就是繁忙请稍后再试。这体验感特别不好特别让人生气哦! 真的哪怕你收费呢, 每个月收多少钱, 但是别这么折磨用户啊啊啊啊啊啊啊什么时候能没有毛病啊
多场景支持	功能、视频、新闻、历史记录、编程、生图、音频、艺术、核心、历史数据	有很多功能, 也挺好用, 能满足客户不同需求
		集成了多种功能, 包括智能问答、文案创作、图片生成、编程助手、起名神器、英语学习助手等, 能够满足多个场景下的需求, 非常好
自主化	模型、能力、国产、科技、水平、用心、思想、国际、趋势、技术人员	R1模型的确很好用, 尤其是思考分析的能力
		国产大模型来看, 通义还是蛮稳定和强大的, 进步挺大的。生成文件可能会略久, 估计渲染能力的缘故, 再接再厉
文本创作	专业、文章、写文章、法律、英语单词、干货、撰写论文、语言学、字幕、证明	夸豆包, 豆包特别好用, 还是免费, 原来买的撰写文章的都不用了, 豆包是大千世界无所不能
		很好用! 写文章作文文案很好! 思考很全面
学习辅助	答案、记忆、题目、模式、风格、记忆力、仔细检查、参考值、误导性、做作业	我只是在问学术问题, 一直被和谐答案我真服了
		用了好久的豆包, 起初觉得非常好, 但用到后来, 当我问了一道数学题, 就是关于柯西不等式的, 我们老师说豆包这答案是错的; 然后还有一次就是生物问题, 好像是那个什么胃蛋白酶原, 他竟然说胃蛋白酶原的合成, 其始于附着在内质网上的核糖体, 可是正确答案是游离, 真的对我们学生特别地有误导性。反正多改正吧, 希望豆包变好, 我的作文还靠它呢, 所以两颗星哈
语音交互	交流、语音输入、信任、心扉、识别率、结构化、文化背景、音效、编辑、语音沟通	我很喜欢豆包, 跟他交流得很愉快
		那个英语可以陪练, 真的太有趣了, 虽然他真的是有时候你发音不太标准, 但是他就能很好地帮你圆过去, 还是一本正经的那种, 挺好玩的哈, 他那种语气是起起伏伏的, 就是有种喘不过气的感觉, 但是感觉还是比较良好的, 而且他说这些英语的话你还是能听得懂的, 而且你也是能说中文和他交流的, 就是我英语可能比较差, 但是还是很好玩

技术水平的关注, 文本创作属性反映用户在文本生成与语言表达方面的应用需求, 学习辅助属性反映用户在学习与作业辅助场景中的使用行为, 语音交互属性反映用户对自然语言交互与语音沟通体验的期待。其中, 学习辅助和多场景支持主要反映使用场景与行为特征, 属于情境属性, 其余主题为功能属性, 此区分有助于更清楚地识别不同属性对用户的具体影响。

将已标注产品属性的单元句按属性类别进行汇总, 统计不同属性下正向与负向情感单元句的数量及其占比, 结果如图3所示。输出质量、效率和自主化的正向

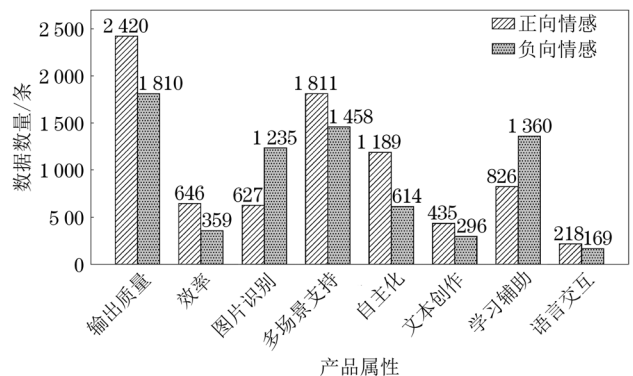


图3 产品属性的情感倾向分布

情感倾向明显高于负向情感，表明用户普遍认可国产 GenAI 在生成质量、效率提升和自主化能力方面的表现。相比之下，图片识别与学习辅助的负向情感占比较高，说明用户对产品的图像处理 and 教学辅助能力存在较多不满。多场景支持与语音交互的情感分布较为均衡，

反映了用户在多样化使用环境与交互体验上的分歧。产品属性层面的情感差异为后续产品优化提供依据。

基于情感分析结果得到原始属性情感值，进一步构建了评论文本的情感矩阵。按照式 (1) 的情感值编码规则，计算各属性的情感得分，部分结果如表3所示。

表3 属性原始情感值和归一化的情感得分 (部分)

评论数据	用户评分	属性原始情感值 (属性情感得分)							
		输出质量	效率	图片识别	多场景支持	自主化	文本创作	学习辅助	语音交互
功能少, 希望以后可以分析视频和图片	4			0.74 (4)	0.13 (1)				
很好, 直到最近出现了“推荐”智能体, 没有任何价值的“广告”吧? 会莫名其妙出现了“推荐”置顶, 每天都要手动删除“推荐”, 很莫名其妙的功能	3	0.30 (2)			0.04 (1)				
好用, 非常非常好用, 支持国产品牌, 还有深度解感和日常模式, 真的是让生活变得更便捷了	5					0.98 (5)		0.49 (3)	

### 3.2 产品满意度分析及属性分类

基于细粒度情感分析得到的情感矩阵与用户评分数据, 按照8:2的比例划分训练集和测试集。经过参数调优, 构建一个具有两个隐藏层的深度神经网络预测模型, 隐藏层神经元数量逐层递减, 激活函数为 ReLU, 并引入早停机制以防止过拟合。实验结果 (见表4) 表明, 深度神经网络模型在准确率和F1分数上均优于SVM模型。

表4 深度神经网络预测模型和SVM的预测结果

训练模型	准确率	F1分数
深度神经网络预测模型	0.79	0.85
SVM	0.76	0.83

在深度神经网络预测模型的基础上, 进一步计算各属性的平均SHAP绝对值, 并由式 (4) 计算得出属性重要性。同时, 依据式 (2) 的绩效估计方法, 量化用户对各属性的评价, 结果如表5所示。

从重要性分布来看, 输出质量的重要性显著高于其他属性, 是决定用户满意度的核心因素; 多场景支持和学习辅助的重要性均在0.040以上, 属于关键影响因素。从绩效分布来看, 效率和自主化得分较高, 说明用

表5 属性重要性与绩效估计

产品属性	重要性	绩效
输出质量	0.070	3.45
效率	0.014	3.74
图片识别	0.026	2.65
多场景支持	0.040	3.36
自主化	0.026	3.74
文本创作	0.013	3.64
学习辅助	0.055	2.90
语音交互	0.007	3.56

户对其评价积极; 图片识别的绩效最低, 显示用户对该属性还存在较多不满。

基于IPA模型, 根据重要性与绩效水平, 将8个属性划分为优势提升区、继续维持区、后续机会区和重点改进区 (见图4)。

用户满意度由多属性共同作用形成, 不同属性的影响存在显著差异<sup>[37]</sup>。影响国产GenAI用户满意度的属性可划分为优势属性与劣势属性。①输出质量位于优势提升区, 属于高重要性-高绩效区域, 是当前国产GenAI产品的优势属性, 取得了大部分用户的关注和高度评价, 应继续保持并加以强化。②效率、自主化、文本创作和语音交互位于继续维持区, 属于低重要性-高绩效

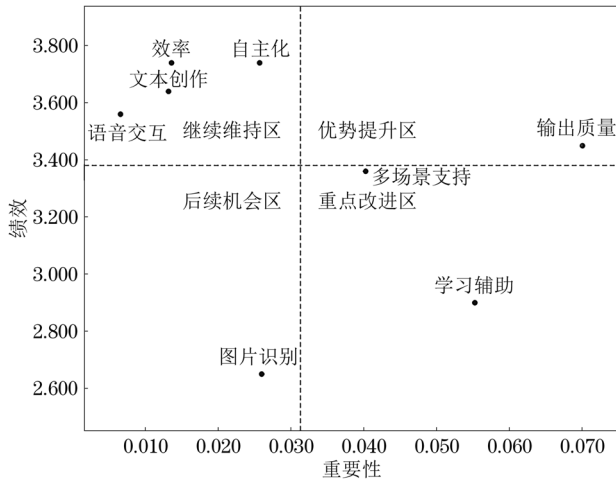


图4 IPA分类结果

区域,说明尽管这些属性对用户满意度的影响相对有限,但当前的使用体验已能较好地满足用户需求,可在现有水平的基础上稳步提升。③图片识别位于后续机会区,属于低重要性-低绩效区域,是当前国产GenAI产品较为劣势的属性,虽对整体满意度影响有限,但其绩效偏低,反映用户认可度不足。应持续跟踪其用户需求变化,避免其因外部环境或使用情境变化转变为影响满意度的关键因素,并在技术和资源充足的情况下,优先改进该属性。④多场景支持和学习辅助位于重点改进区,属于高重要性-低绩效区域,是当前国产GenAI产品的主要劣势属性,用户对该类属性具有较高的使用预期,但目前产品表现尚未有效满足用户需求。因此,国产GenAI产品开发应优先针对该类属性进行优化,切实提升整体用户满意度。需要注意的是,学习辅助与多场景支持虽属于情境属性,但其重要性得分较高,表明用户在特定使用场景下的体验质量对整体满意度具有显著影响,因此学习辅助与多场景支持同样是产品优化的重要方向。

### 3.3 模型有效性验证

对照实验结果(见表6)表明,在主题一致性、SHAP覆盖度以及IPA分类稳定性等评价指标上,本研究所提方法均表现更优。具体而言,在主题一致性方面,所提方法能够生成语义更为集中、边界更清晰的属性主题,其得分明显高于传统的TF-IDF+K-Means方法;在SHAP覆盖度方面,所提方法在该指标上的显著提升表明,预测结果中可被解释的贡献占比更高,属性重要性排序也更加可信;在IPA分类稳定性方面,通过

多轮Bootstrap重抽样验证,所提方法的属性象限划分在不同抽样条件下保持更高的一致性,显示出更好的稳健性。

表6 对照实验结果

模型	主题一致性	SHAP覆盖度	IPA分类稳定性
所提方法	0.523 9	0.759 3	0.703 2
基线方法	0.503 1	0.692 2	0.659 0

## 4 结论与展望

### 4.1 研究结论

当前,GenAI已成为全球科技竞争的核心,其技术突破度、功能完善度与应用生态成熟性直接关系到一个国家在数字经济时代的竞争力<sup>[38]</sup>。本研究融合可解释机器学习SHAP方法与IPA模型,对国产GenAI产品的用户满意度进行了系统分析,主要结论如下。

(1) 基于LDA主题建模与细粒度情感分析,从用户在线评论中识别出核心属性,并划分为功能属性(输出质量、效率、图片识别、自主化、文本创作、语音交互)和情境属性(学习辅助、多场景支持)。功能属性是满意度的基础来源,情境属性影响使用体验与用户黏性。研究表明,用户普遍对输出质量、效率、自主化、文本创作及语音交互等属性持积极态度,说明国产GenAI在生成能力、运行流畅性与人机交互等方面获得较高认可;用户对图片识别与学习辅助属性存在明显负向情感,主要源于识别稳定性不足和答题逻辑失误;用户对多场景支持属性的情感整体偏正向,但体验差异明显。

(2) 基于可解释机器学习方法量化各属性对用户满意度的非线性贡献,并结合IPA确定各属性的优化优先级,结果显示:输出质量是核心优势;效率、自主化、文本创作与语音交互属于支撑型优势,应保持稳定迭代;图片识别是潜在的薄弱环节;多场景支持与学习辅助为关键优化方向。整体来看,功能属性与情境属性在满意度形成中存在互补关系:前者决定产品的可用性与性能基准,后者则塑造产品的差异化体验。

(3) 国产GenAI产品在文化贴合度与语言理解力方面用户满意度更高,反映其在本土化语义适配与本地场景应用上的优势。相比而言,美国和欧洲国家的GenAI产品在生态系统建设、算力资源与产业协同上

具有明显优势<sup>[39]</sup>。未来,国产GenAI产品可进一步发挥这一优势,通过优化本地化体验和持续改进跨模态能力,提升用户满意度,实现与国际主流产品的差异化竞争。

## 4.2 实践启示

本研究揭示了国产GenAI产品不同属性的优化优先级差异,表明用户满意度的提升不仅依赖单一技术指标的优化,更需要考虑属性绩效与实际使用体验。

(1) 改进国产GenAI产品的关键在于优先优化用户期望较高但体验不足的核心属性。研究发现,多场景支持与学习辅助是制约总体满意度提升的主要瓶颈,因此应从高频任务与真实使用情境入手,梳理用户在任务理解、信息获取与结果生成过程中的关键痛点,并通过场景化能力建设提升国产GenAI产品在不同语境下的稳定性与可用性。相关研究<sup>[40]</sup>表明,面向具体应用场景的能力优化与持续反馈有助于提升用户体验。

(2) 输出质量是国产GenAI产品的相对优势属性,其持续优化对于巩固核心竞争力具有重要意义。通过强化生成内容的一致性校验、事实约束与用户反馈驱动的迭代机制,可降低幻觉风险并提升内容可信度。因此,在保持生成能力优势的基础上,应进一步推动生成内容向专业性与适配性方向深化,使输出更好地服务于真实应用需求。

(3) 对于效率、自主化及语音交互相关属性,尽管其重要性相对有限,但其绩效水平较高,为用户体验提供了稳定支撑。系统响应速度、交互流畅度与本土化语言适配是影响用户持续使用意愿的重要基础因素,应通过渐进式优化保持其稳定表现。相比之下,当前处于低重要性与低绩效区间的图像识别属性虽尚未成为用户决策的关键影响因素,但其基础性能不足仍可能产生负向体验外溢效应。已有研究<sup>[41]</sup>指出,良好的错误反馈与可恢复路径是降低用户挫败感的重要手段。在新兴功能尚未成熟阶段,此类设计有助于缓冲用户负面体验,从而降低负面评价风险。

总体而言,本研究从实践层面表明,国产GenAI产品的满意度提升应遵循聚焦核心短板、巩固优势能力、维持基础体验、前瞻布局潜力功能的渐进式优化逻辑,为相关产品的功能迭代与资源配置提供了有针对性的决策参考。

## 参考文献

- [1] 国家互联网信息办公室. 生成式人工智能服务管理暂行办法[EB/OL]. [2025-09-12]. <http://www.zgys.gov.cn/ysgaj/c121083/202504/7d2a0fde37174ba3b90ab228ae5bb5be.shtml>.
- [2] 国家互联网信息办公室. 关于发布生成式人工智能服务已备案信息的公告(2025年4月至6月)[EB/OL]. [2025-09-12]. [https://www.cac.gov.cn/2025-07/11/c\\_1753948489002783.htm](https://www.cac.gov.cn/2025-07/11/c_1753948489002783.htm).
- [3] GILLESPIE N, LOCKEY S, WARD T, et al. Trust, attitudes and use of artificial intelligence: a global study 2025[EB/OL]. [2025-09-12]. <https://kpmg.com/xx/en/our-insights/ai-and-technology/trust-attitudes-and-use-of-ai.html>.
- [4] 陈鹏,王蓉,白佳鹭. 哪些因素影响学习者的Gen AI采纳行为?: 基于模糊集质性比较分析的发现[J]. 现代教育技术, 2025, 35(6): 36-45.
- [5] YU C C, YAN J Z, CAI N. ChatGPT in higher education: factors influencing ChatGPT user satisfaction and continued use intention[J]. *Frontiers in Education*, 2024, 9: 1354929.
- [6] CHENG Y, JIANG H. How do AI-driven chatbots impact user experience? Examining gratifications, perceived privacy risk, satisfaction, loyalty, and continued use[J]. *Journal of Broadcasting & Electronic Media*, 2020, 64(4): 592-614.
- [7] LIN Z H, NG Y L. Unraveling gratifications, concerns, and acceptance of generative artificial intelligence[J]. *International Journal of Human-Computer Interaction*, 2025, 41(17): 10725-10742.
- [8] IMRAN M, SHAHID A R, HOU M, et al. From early adoption to ethical adoption: a diffusion of innovation perspective on ChatGPT and large language models in the classroom[EB/OL]. [2025-09-12]. <https://doi.org/10.36227/techrxiv.170630660.06963201/v1>.
- [9] JO H, BANG Y. Analyzing ChatGPT adoption drivers with the TOEK framework[J]. *Scientific Reports*, 2023, 13: 22606.
- [10] CHAN C K Y, HU W J. Students' voices on generative AI: perceptions, benefits, and challenges in higher education[J]. *International Journal of Educational Technology in Higher Education*, 2023, 20(1): 43.
- [11] AL-MUGHAIIRI H, BHASKAR P. Exploring the factors affecting the adoption AI techniques in higher education: insights from teachers' perspectives on ChatGPT[J]. *Journal of Research in Innovative Teaching & Learning*, 2025, 18(2):

- 232-247.
- [12] KIM J H, KIM J, KIM S S, et al. Effects of AI ChatGPT on travelers' travel decision-making[J]. *Tourism Review*, 2024, 79 (5): 1038-1057.
- [13] 代宝, 杨利英, 郑怡晴. 生成式AI用户采纳行为研究的现状与展望[J]. *情报探索*, 2025 (3): 46-52.
- [14] 江立华, 陈文超. 试论问卷调查中的结构性张力及其制约: 以2010年CGSS湖北区域调查实践为例[J]. *河南社会科学*, 2011, 19 (2): 139-143.
- [15] 刘智锋, 吴亚平, 王继民. 科学数据集学术影响力归因研究: 基于回归分析与可解释机器学习的双重证据[J]. *科学学研究*, 2025, 43 (5): 976-987.
- [16] ALABDULJABBAR R. User-centric AI: evaluating the usability of generative AI applications through user reviews on app stores[J]. *PeerJ Computer Science*, 2024, 10: e2421.
- [17] CHHETRI V, UPADHYAY K, SIDDIQUE A B, et al. What users value and critique: large-scale analysis of user feedback on AI-powered mobile apps[EB/OL]. [2025-09-12]. <https://arxiv.org/abs/2506.10785>.
- [18] BILAL A, MIRZA H T, KHAN A S, et al. Who dominates generative AI? Analyzing user feedback to identify common use cases and areas for improvement in ChatGPT, Copilot and Gemini[J]. *Knowledge and Information Systems*, 2025, 67 (11): 10797-10831.
- [19] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques[EB/OL]. [2025-09-12]. <https://aclanthology.org/W02-1011.pdf>.
- [20] TURNEY P D, LITTMAN M L. Unsupervised learning of semantic orientation from a hundred-billion-word corpus[EB/OL]. [2025-11-12]. <https://arxiv.org/abs/cs/0212012>.
- [21] ZHANG W X, LI X, DENG Y, et al. A survey on aspect-based sentiment analysis: tasks, methods, and challenges[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35 (11): 11019-11038.
- [22] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2015 task 12: aspect based sentiment analysis[C]// *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Stroudsburg: ACL, 2015: 486-495.
- [23] YAN J L S, TURTLE H R, LIDDY E D. EmoTweet-28: a fine-grained emotion corpus for sentiment analysis[C]// *International Conference on Language Resources and Evaluation*. Stroudsburg: ACL, 2016.
- [24] WANG W B, CHEN L, TAN M, et al. Discovering fine-grained sentiment in suicide notes[J]. *Biomedical Informatics Insights*, 2012, 5 (Suppl. 1): 137-145.
- [25] MARTILLA J A, JAMES J C. Importance-performance analysis[J]. *Journal of Marketing*, 1977, 41 (1): 77-79.
- [26] TAPLIN R H. Competitive importance-performance analysis of an Australian wildlife park[J]. *Tourism Management*, 2012, 33 (1): 29-37.
- [27] WILKINS H. Using importance-performance analysis to appreciate satisfaction in hotels[J]. *Journal of Hospitality Marketing & Management*, 2010, 19 (8): 866-888.
- [28] O'NEILL M A, PALMER A. Importance-performance analysis: a useful tool for directing continuous quality improvement in higher education[J]. *Quality Assurance in Education*, 2004, 12 (1): 39-52.
- [29] AZZOPARDI E, NASH R. A critical evaluation of importance-performance analysis[J]. *Tourism Management*, 2013, 35: 222-233.
- [30] 吴江, 李秋贝, 胡忠义, 等. 基于IPA模型的乡村旅游景区游客满意度分析[J]. *数据分析与知识发现*, 2023, 7 (7): 89-99.
- [31] SHEN M Y, CHENG A X, BI Y Y. An integrated framework for importance-performance analysis of product attributes and validation from online reviews and maintenance records[J]. *Design Science*, 2024, 10: e19.
- [32] NAM S, LEE H C. A text analytics-based importance performance analysis and its application to airline service[J]. *Sustainability*, 2019, 11 (21): 6153.
- [33] JOUNG J, KIM H M. Approach for importance-performance analysis of product attributes from online reviews[J]. *Journal of Mechanical Design*, 2021, 143 (8): 081705.
- [34] 唐晓波, 刘广超. 细粒度情感分析研究综述[J]. *图书情报工作*, 2017, 61 (5): 132-140.
- [35] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation[J]. *Journal of Machine Learning Research*, 2003, 3: 993-1022.
- [36] HASSENZAHL M. *Experience design: technology for all the right reasons*[M]. New York: Morgan and Claypool Publishers, 2010.
- [37] KALANKESH L R, NASIRY Z, FEIN R A, et al. Factors influencing user satisfaction with information systems: a systematic review[J]. *Galen Medical Journal*, 2020, 9: e1686.
- [38] 方兴东. 从人工智能大模型看大国科技博弈态势[J]. *人民论*

- 坛·学术前沿, 2024 (13): 44-56.
- [39] CHOI W C, CHANG C I, CHOI I C, et al. A review of large language models (LLMs) development: a cross-country comparison of the US, China, Europe, UK, India, Japan, South Korea, and Canada[EB/OL]. [2025-09-12]. <https://www.preprints.org/manuscript/202504.2136>.
- [40] DZVONYAR D, KRUSCHE S, ALKADHI R, et al. Context-aware user feedback in continuous software evolution[C]// Proceedings of the International Workshop on Continuous Software Evolution and Delivery. 2016: 12-18.
- [41] NORMAN D. The design of everyday things[M]. Cambridge: MIT Press, 2013.

## 作者简介

杨怡, 女, 博士, 高级工程师, 研究方向: 网络信息管理、科学数据共享与重用。

万晨硕, 男, 本科生, 研究方向: 文本挖掘。

李瑾, 女, 本科生, 研究方向: 自然语言处理。

胡译文, 男, 博士, 教授, 博士生导师, 通信作者, 研究方向: 数据智能与情报分析、知识挖掘与知识服务, E-mail: huzewen915@163.com。

### Analysis of User Satisfaction for Domestic GenAI Products Based on Online Reviews

YANG Yi WAN ChenShuo LI Jin HU ZeWen

(School of Management Science and Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, P. R. China)

**Abstract:** Existing studies on user satisfaction of domestic generative AI (GenAI) products largely rely on survey-based methods and struggle to capture nonlinear effects at the attribute level. This study aims to develop a user satisfaction analysis and optimization framework based on online review data, systematically revealing the impact mechanisms of different product attributes on overall satisfaction and their improvement priorities. Taking mainstream domestic GenAI products—including Doubao, DeepSeek, Tongyi Qianwen, Kimi, and Wenxin Yiyan—as research objects, we first identify core attributes of user concern through the LDA topic model and quantify their performance. Then, the SHAP method is employed to estimate the influence of each attribute on user satisfaction. Finally, the IPA model is used to classify attributes according to their optimization priorities. Results show that users primarily focus on output quality, efficiency, image recognition, multi-scenario support, autonomy, text generation, learning assistance, and voice interaction. Among these, output quality is a strength attribute; multi-scenario support and learning assistance are key areas for improvement; image recognition is a potential improvement attribute; and efficiency, autonomy, text generation, and voice interaction are maintenance attributes. This study provides data-driven decision support for the functional optimization of domestic GenAI products.

**Keywords:** GenAI; User Satisfaction; Product Attribute; Fine-Grained Sentiment Analysis; Machine Learning; IPA

(责任编辑: 王玮)