

AI情绪识别技术在电商客服中的伦理风险

路雪雪

贵州大学哲学学院, 贵州 贵阳

收稿日期: 2025年7月15日; 录用日期: 2025年7月28日; 发布日期: 2025年8月25日

摘要

AI情绪识别技术在电商客服中的应用提升了服务效率,但也带来了一些潜在的伦理风险。技术偏见可能导致个体差异的误解,从而影响服务的公平性和准确性。AI系统的错误判断可能导致责任归属问题,尤其是在发生错误处理时如何界定责任主体成为问题。同时AI技术可能被恶意利用,进行情感操控,过度干预用户情绪或影响心理健康,尤其对情感脆弱的群体更具风险。因此为确保算法公正性、建立明确的责任机制、并赋予用户更多的知情权与控制权,是解决这些问题的关键。通过多样化训练数据,加强模型监控,透明化用户信息等措施,可以有效规制AI情绪识别技术的伦理风险,保护用户权益并促进其合法合规使用。

关键词

情绪识别, 电商客服, 情感操控

Ethical Risks of AI Emotion Recognition Technology in E-Commerce Customer Service

Xuexue Lu

School of Philosophy, Guizhou University, Guiyang Guizhou

Received: Jul. 15th, 2025; accepted: Jul. 28th, 2025; published: Aug. 25th, 2025

Abstract

AI emotion recognition technology applied in e-commerce customer service improves service efficiency, but also brings some potential ethical risks. Technical bias may lead to the misunderstanding of individual differences, thus affecting the fairness and accuracy of the service. The misjudg-

ment of the AI system may lead to the problem of attribution of responsibility, especially how to define the responsible subject becomes a problem when error handling occurs. Meanwhile, AI technology may be maliciously utilized for emotional manipulation, over-interfering with users' emotions or affecting their mental health, which is especially risky for emotionally vulnerable groups. Therefore, in order to ensure the fairness of the algorithm, establish a clear responsibility mechanism, and give users more right to know and control is the key to solve these problems. Measures such as diversifying training data, strengthening model monitoring, and transparently adopting user information can effectively regulate the ethical risks of AI emotion recognition technology, protect users' rights and interests, and promote its legal and compliant use.

Keywords

Emotion Recognition, E-Commerce Customer Service, Emotional Manipulation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人工智能技术的不断演进, AI 情绪识别正在被广泛应用于电商客服领域。相比传统的人工客服或规则驱动型智能客服系统, AI 情绪识别能够通过文本、语音、面部表情等多模态数据, 实时判断用户的情绪状态, 从而为用户提供更加个性化、情景化的服务体验。这一技术的引入极大地提升了电商平台的服务响应效率与用户满意度, 在优化客户关系管理、降低客服成本方面展现出显著的优势。然而技术带来便利的同时, 也伴随着一系列不可忽视的风险与挑战。本文将在分析 AI 情绪识别技术在电商客服中应用模式与伦理问题的基础上, 提出以算法公正性、责任机制与用户赋权为核心的三维规制路径, 旨在为 AI 情绪识别的发展提供理论支撑与实践参考。

2. AI 情绪识别技术在电商客服中的应用模式

随着电商行业的迅猛发展, 客户服务需求不断增加, 传统的人工客服模式出现了效率低、响应时间长等问题。为了提升客服的服务质量和响应效率, AI 情绪识别的引入为电商行业带来了便利。该技术通过文本、语音和面部表情等方面的多模态数据分析, 对客户情绪进行识别, 从而为电商平台提供了更加智能和人性化的服务。

2.1. 文本语义分析

文本语义分析是 AI 情绪识别技术中最常见的一种模式, 它通过自然语言处理(Natural Language Processing)技术对客户的文本信息进行处理与分析, 识别出其中蕴含的情感倾向。在电商客服中, 客户通过在线聊天、邮件或评论等方式表达自己的需求与意见。基于语义的面向方面的情感分析方法, 该方法通过拼写校正、分词、词性标注、词干还原、命名实体识别、词义消歧和核心语义分析等步骤, 最终进行情感分析[1]。例如, 当客户在留言中使用负面情绪时, 如“不满意”“失望”“糟糕”等, 情感分析系统会自动判定为负面情绪, 从而提示客服需要及时关注, 使用更加温和的语言来安抚客户, 并采取合适的解决措施。文本语义分析不仅能够帮助客服系统及时捕捉到客户的情绪变化, 还能根据情绪的严重程度为客服人员提供相应的应对策略, 避免进一步激化客户情绪。

2.2. 语音情感分析

在电商客服中, 语音通话也是许多客户与商家互动的主要方式。语音情感分析主要依赖于音调、语速、音量等声学特征的变化。当客户的语速加快或音量增大时, 可能意味着客户的情绪处于激动或愤怒状态。“融合韵律特征(F0 均值、语速)与谱特征(MFCC)的 Bi LSTM -Attention 模型, 对愤怒情绪识别率达 89.6%。在电商客服场景中, 当客户语音的基频标准差 > 35 Hz 且语速方差 > 0.8 时, 系统可实时标记情绪等级(3 级: 平静/烦躁/愤怒), 并生成应对建议库(如愤怒状态优先推荐补偿方案)。” [2]通过深度学习算法, 语音情感分析能够精准地识别出这些情绪信号, 并通过实时反馈帮助客服人员及时调整应对策略。

2.3. 面部表情识别

面部表情识别利用深度学习中的卷积神经网络(Convolutional Neural Network)算法, 分析客户在视频通话中的面部表情变化。这种技术主要依靠对面部肌肉运动和微表情的捕捉来识别客户的情绪[3]。例如, 客户皱眉时, 系统会识别出这是愤怒或不满的表现; 当客户露出微笑时, 则表示情绪趋向积极。面部表情识别技术的优势在于, 它能够从客户的非语言行为中提取情感信息, 从而帮助客服系统更加准确地评估客户的情绪状态。

AI 情绪识别技术在电商客服中的应用, 能够有效提升客户服务质量和客户满意度。在面对复杂的客户的情绪状况, 从而为客服人员提供更为精准的情绪反馈, 帮其制定更加个性化的服务方案, 避免了因人工客服反应慢导致的客户流失。

3. AI 情绪识别在电商客服中应用的伦理风险剖析

随着 AI 情绪识别技术的不断发展和应用, 关于其伦理问题的讨论也变得愈加重要。尽管 AI 情绪识别可以帮助电商平台提升客服效率、提高客户满意度, 但其潜在的伦理挑战也不容忽视。特别是在个体差异的忽视、情感识别准确性、责任归属以及情感操控等方面, AI 情绪识别技术的应用可能带来一系列复杂伦理问题。

3.1. 算法偏见

3.1.1. 忽视个体差异性

AI 情绪识别技术在电商客服中的应用可能导致个体差异的忽视, “训练数据抹平个体表达独特性, 使模型仅学会识别‘标准愤怒’‘标准喜悦’等理想化情绪模板” [4]。情绪表达是高度个性化, 不同用户在表达相同情绪时, 可能会因为其文化背景、性格特征、生活经验等因素的不同, 产生独特的情感表达方式。例如, 某些文化背景下的用户可能更倾向于通过言辞直接表达愤怒或者不满, 而另一些文化背景下的用户可能会通过含蓄或沉默表达情绪。AI 模型通常是基于大量数据进行训练的, 若这些数据缺乏对文化差异或个体差异的考虑, 可能导致情感表达的误解。AI 情绪识别模型往往无法捕捉到每个用户独特的情感表达方式, 特别是在面对非典型情绪表达时, 模型可能无法做出准确判断。这种情况可能会导致错误的情绪识别和随之而来的不正当回应。当 AI 错误地识别某用户为愤怒时, 它可能会给出过于强硬或过度的客服回应, 而忽略了用户的真实需求, 反而可能加剧不满情绪。

3.1.2. 情感识别准确性低

情绪本身是复杂且主观的, 许多因素都能影响其表达方式, AI 情绪识别的准确性也面临着一定的挑战。从理论上讲, 情绪识别技术的准确率可以随着技术的进步提升很高, 其表现可能会超越人类, 但在根本上, 其识别的准确率仍然达不到百分之百, 甚至在特定情境下产生诸多错误和偏见[5]。语气、肢体语言、情绪背景等都能对情绪的呈现产生影响, 而 AI 系统通常依赖的是通过文本、语音或面部表情等外

在特征来进行分析。虽然深度学习和大数据技术在情感识别方面取得了很大的进展，但情绪的内在复杂性依然是 AI 无法把握的。因为 AI 可能无法理解用户情感的细微变化，尤其是在情感的复杂交织时。用户可能在表达不满的同时，也透露出希望得到帮助的积极情绪。AI 的情绪标签体系(如积极/消极)无法兼容人类情感的梯度性与矛盾性，AI 若无法正确识别这些情感的混合状态，可能会导致错误的情绪判断[6]，进而引发客服误导或处理错误，从而影响用户体验和商家信誉。

3.2. 责任归属

AI 情绪识别技术在电商客服中涉及责任归属的伦理问题。AI 系统的决策过程往往基于复杂的算法和数据模型，其决策过程和结果难以被人类完全理解和预测，这使得在 AI 系统出现故障或引发不良后果时，很难精准判断责任归属[7]。

关于责任的归属问题，许多人认为当 AI 系统发生错误，尤其是在情绪识别过程中出现误判时责任应当追溯到系统的开发者或平台运营方，开发者在设计和训练 AI 系统时，必须确保其识别准确性和功能的可靠性然而 AI 系统的训练往往依赖于大量的数据集，这些数据的质量种类和来源会直接影响 AI 的判断能力因此开发者有义务确保数据的多样性和准确性，避免因数据差异导致的错误识别。平台运营方也承担一定的责任，因为他们对 AI 系统的使用场景和效果进行管理并确保系统能够正确执行任务如果 AI 系统误判并提供了错误的建议，进而导致用户的不满或投诉，平台运营方应当负责及时纠正问题，向用户做出合理的解释和赔偿。

相反的观点则认为，AI 系统本身在某些情况下应当承担一定的责任。这种观点基于 AI 系统的自动化和智能性，认为当 AI 作出决策时，责任应当由 AI 系统本身来承担，特别是在系统的自动化程度较高的时候。这一观点的支持者认为，AI 系统已经具备一定的智能水平，可以在某些情景下自行做出决策，因此出现情绪识别错误时 AI 系统应当承担一定的责任，以维护系统的公信力和可信度但这一观点面临的问题在于，当前的法律框架并未明确界定人工智能是否可以作为法律责任主体 AI 是否能够承担相应的责任，仍然存在很大的争议。

3.3. 情感操控

3.3.1. 过度干预用户情绪

AI 能够精确识别并分析用户的情绪状态，这种技术的运用能够为商家提供更有针对性的营销策略。然而，这种精确的情感分析也可能被恶意利用，尤其在广告投放和个性化推荐领域。电商平台可以基于用户的情感状态，主动推送特定的商品或广告，试图引导用户的购买决策。当 AI 系统识别出用户感到焦虑时，平台可能会推动一些诱人的折扣或促销活动，从而激发用户的冲动消费行为。尽管这种方式可能会提升短期的销售业绩，但从长远来看，这种情感操控可能对用户产生负面影响，尤其是当用户并不清楚自己被操控时，可能会失去对平台的信任。

3.3.2. 精神健康风险

更为严重的是，长期的情感操控不仅会使用户产生焦虑和不安，还可能加剧用户的心理压力，导致用户产生情感失控、焦虑症等心理问题。在一些特定情况下，AI 情绪识别技术被用来监控和分析具有情感脆弱性的人群，如果这种技术被用来对他们施加过度的情感引导，可能会导致更加严重的精神健康问题。为了最大化经济利益而忽视用户的心理健康，显然是对用户的不尊重。

AI 情绪识别技术在电商客服中的应用带来的伦理问题不容忽视。个体差异的忽视、情感识别准确性降低、责任归属不清和情感操控等问题，都需要在技术研发、政策监管和用户保护等方面得到进一步的关注和解决。为了确保 AI 情绪识别技术能够健康、有序地发展，必须加强对这一技术的伦理审视，并提

升公众对 AI 技术潜在风险的认知与警惕。只有在保证用户权益和社会公正的前提下, AI 情绪识别技术才能真正发挥其应有的社会价值。

4. AI 情绪识别在电商客服中应用风险的规制路径

随着人工智能技术的快速发展, AI 情绪识别已经被广泛应用于电商客服领域。这项技术能够识别用户的语气、表情和言辞, 从而推测其情绪状态, 为平台提供更加个性化的服务。然而, 情绪识别技术的应用并非没有风险。由于其深度介入用户行为和心理层面, 一旦监管不力或设计不当, 可能会引发歧视性结果、责任不清、隐私侵犯等伦理问题。在面对 AI 情绪识别技术应用过程中可能带来的多种风险时, 建立合理、科学、系统的规制路径显得尤为重要。

4.1. 算法公正性

4.1.1. 多样化训练数据

为了有效规避因算法带来的歧视性风险, 必须高度重视 AI 情绪识别系统的算法公正性。AI 系统在设计 and 训练阶段往往依赖大量数据, 而数据的质量、覆盖面和结构决定了模型最终的行为表现。在现实中, 如果训练数据主要来自特定群体, 那么该模型对其他群体的情绪识别能力可能会明显不足。这种偏差将直接影响 AI 的判断准确性和公正性。如果系统主要以年轻面孔的情绪表达为训练样本, 那么在面对年长者用户时, 其识别效果可能会大打折扣, 从而在无形中放大技术带来的不公平现象。开发者必须确保情绪识别系统在训练过程中, 采用对抗性去偏技术(Adversarial Debiasing)对训练数据进行处理, 通过生成对抗网络识别并消除数据中潜在的偏见模式; 同时引入因果推理模型(Causal Inference Models)以区分变量间的相关性与因果关系[8]。从而避免其形成具有歧视性的决策规则, 提升模型的适用性和普适性, 避免因样本不平衡带来识别误差和情感误读。

4.1.2. 模型公平性监控

情绪识别系统在实际运行过程中, 不仅需要高效的算法支持, 还需要持续进行模型公平性监控。随着用户行为、情感表达等多方面因素的不断变化, 模型在实际使用中可能会产生新的偏见或不准确的判断。为了减少这种风险, 平台或监管机构应建立科学的评价机制, 定期对 AI 情绪识别系统的预测准确性、判断结果进行周期性的审查和回归测试, 可以有效发现潜在的系统性偏差, 并通过实时调整和优化算法进行修正。确保情绪识别技术在保障用户权益的前提下, 真正发挥其提升服务质量、增强客户体验等方面的积极作用。

4.2. 责任机制

4.2.1. 明确开发者责任

在 AI 情绪识别技术引发问题的情景中, 明确责任机制是规避风险的重要保障。AI 系统的自动化决策一旦出错, 往往涉及多个利益相关方, 包括技术开发者、平台运营以及使用 AI 服务的用户。“AI 系统依赖多层次技术栈(数据 - 算法 - 应用), 情绪识别错误可能源于数据标注缺陷、模型设计漏洞或部署环境干扰。传统‘单一主体担责’模式失效, 需重构适应技术复杂性的责任分配逻辑。” [9]那么建立一个清晰、合理的责任分配框架, 能够出现纠纷或问题时快速定位责任主体, 提升社会对 AI 技术的信任度。在此框架中, 开发者作为技术的提供者, 应承担起对算法质量和系统安全的基本责任。如果由于算法设计缺陷、训练偏差或系统漏洞导致情绪识别失误, 开发者需要对由此产生的后果承担相应的法律和经济责任。

数据偏差导致误判主要源于数据提供方未能及时更新数据源, 或者算法设计方在数据处理过程中存

在误差。根据责任比例，数据提供方承担责任占比最高，算法设计方承担的责任占比次之，平台承担的责任占比相对较低。责任形式包括：数据方需赔偿用户损失并优化数据源、算法方需免费提供模型迭代服务、平台则需公开致歉等。

算法漏洞引发的误判，责任主要由算法设计方和平台一同承担。其中，算法设计方所负责任的占比更高，平台承担的责任占比相对较低。责任形式包括：算法方需支付罚金并修复漏洞、平台需补偿用户并加强算法测试等。

运营不当引发的误判，责任主要由平台和用户一同承担。其中，平台所负责任的占比更高，用户则在提供虚假信息的情况下承担相应责任，且其责任占比低于平台。责任形式包括：平台需整改服务流程、用户需承担部分损失(如虚假信息导致的资源浪费)等。

因不可预见的技术局限导致的误判，责任由平台和算法设计方共同承担，双方所负责任的占比相同。责任形式包括：双方共同设立补偿基金、平台负责用户沟通、算法方记录局限并推动技术升级等。

这些责任划分和责任形式的设定旨在通过明确各方的责任，促进技术应用的透明化和规范化、减少误判带来的负面影响，同时也为技术失误的预防和纠正提供了制度保障，尤其是在侵犯用户情感隐私或引发误导性服务时，开发方不能以“技术中立”为由推卸责任。

4.2.2. 平台运营方责任

电商平台作为 AI 情绪识别技术的使用者和管理者，也必须对技术的应用结果承担监督责任。平台不仅是算法运行的场所，更是直接与用户发生互动的实体，因此其责任范围也更为广泛。平台应建立完善的技术审查机制，在引入情绪识别系统前进行充分的合规性评估，并在系统运行过程中持续对其使用状况进行监管。如果系统识别结果引发用户投诉、误导服务或其他损害行为，平台应积极介入调查并作出合理回应，包括赔偿损失、优化算法或终止技术使用等。平台还应主动配合外部监管机构，对 AI 系统的运行状况进行信息披露，推动平台责任的透明化和制度化。

4.3. 用户赋权

4.3.1. 透明化用户信息使用

为防止 AI 情绪识别技术对用户行为进行过度干预，应当在制度层面强化用户赋权。在技术实施过程中，透明性即公开算法处理数据的流程、数据的来源以及决策逻辑，让用户能够明白其具体途径，从而增进用户的信任。AI 情绪识别的实现通常需要获取用户的大量情绪数据，这类数据因其与个体心理状态密切相关，属于高度敏感的个人信息。用户有权知晓平台是否使用了情绪识别系统，该系统的目的、范围、使用方式以及数据收集和处理流程。平台应严格遵循“知情同意”原则，向用户明确告知情绪识别的所有相关信息，避免在未经允许的情况下采集或分析用户情绪数据。

4.3.2. 提供情绪识别控制权

用户赋权也应包括情绪数据的后续使用权和删除权。用户应有权随时查询、修改或删除自身的情绪数据，防止数据滥用或流出。平台需采取技术和管理手段加强数据的加密和保护，防止由于情绪数据泄露而引发的隐私风险或身份识别问题。可以通过设置多层加密机制和访问权限控制，能够有效保障情绪数据的安全性。平台应杜绝基于情绪分析进行个性化操纵的行为，特别是针对弱势用户群体的情感干预或情绪营销策略，应受到更加严格的限制和监管，以维护用户心理健康与自主选择的权利。

5. 结语

在未来电商客服发展中，AI 情绪识别技术以智能化与个性化为客户服务开拓新可能，也因数据采集的深度、伦理应用的边界而暗藏风险。确保用户情感数据的隐私安全是技术前提的底线；规避算法偏见、

防范情感操控则是电商客服运用 AI 情绪识别技术的前提。唯有让技术创新与伦理规制同频共振，才能让 AI 情绪识别技术真正成为连接商家与用户的温情纽带，最终构建出健康、可持续发展的电商生态。

参考文献

- [1] Schouten, K. (2012) Semantics-Driven Aspect-Based Sentiment Analysis. <https://www.semanticscholar.org/paper/Semantics-Driven-Aspect-Based-Sentiment-Analysis-Schouten/d2ac439896ff2f18f546b71c8f7d1e04bd1e5a43>
- [2] 王一海. 基于深度学习的电子商务语音客服中的客户情感研究[J]. 电子器件, 2020, 43(5): 1173-1179.
- [3] 程卫月, 张雪琴, 林克正, 等. 融合全局与局部特征的深度卷积神经网络算法[J]. 计算机科学与探索, 2022, 16(5): 1146-1154.
- [4] 张晓慧, 孙德艳, 马永波, 等. 情绪识别技术在电力智能客服系统中的应用研究[J]. 电子器件, 2020, 43(5): 1061-1065.
- [5] 阮凯. 情绪识别技术的问题、风险与规制[J]. 自然辩证法通讯, 2023, 45(2): 82-90.
- [6] 罗莉娟, 王康, 胡金淼, 等. 当人工智能面对人类情感: 服务机器人情感表达对用户体验的影响机制[J]. 心理学进展, 2025, 33(6): 1006-1026.
- [7] 李慧. 人工智能技术对社会工作者角色定位的伦理影响研究[J]. 现代社会科学研究, 2025, 5(5): 86-88.
- [8] 陈琳涓. 人工智能赋能电商平台用户信息分析与管理研究[J]. 电子商务评论, 2025, 14(6): 2255-2260.
- [9] Ge, Y. and Zhu, Q. (2024) Attributing Responsibility in AI-Induced Incidents: A Computational Reflective Equilibrium Framework for Accountability. arXiv: 2404.16957.