

青少年对人工智能与人类行为的赔偿接受度差异

张义玲

广西师范大学教育学院心理学系雁山校区, 广西桂林, 541006;

摘要: 本研究采用 2 (赔偿主体: 人工智能 vs 人类) \times 2 (道德类型: 功利主义 vs 义务论) 的两因素被试间实验设计, 探讨不同赔偿主体与道德类型对高中生赔偿接受度的影响。被试为 204 名。方差齐性检验显示 Levene's 检验与异方差性检验均不显著, 方差齐性假设成立, 因此采用常规两因素方差分析。结果表明, 赔偿主体主效应显著, 人工智能造成赔偿的赔偿接受度高于人类; 道德类型主效应显著, 功利主义情境下的接受度高于义务论情境。此外, 赔偿主体与道德类型存在显著交互作用, 简单效应分析显示, 该交互主要由人工智能条件下功利主义与义务论接受度的显著差异驱动。本研究揭示了高中生在不同赔偿来源与道德框架下对赔偿的差异性判断模式。

关键词: 人工智能; 赔偿主体; 青少年; 功利主义; 义务论

DOI: 10.64216/3104-9702.25.05.048

引言

近年来, 人工智能在服务、教育与社会生活中的广泛应用, 使公众在评估 AI 与人类的责任归属、情感反应和补偿期望时呈现出新的心理特点。首先, 大量服务心理学研究关注了 AI 与人类在服务失败场景中的差异反应。当服务机器人在服务失败后采用幽默的恢复策略时, 消费者更容易给予宽恕。这一结果表明, AI 的表达方式能够显著调节用户对其责任感与可补偿性的判断^[1]。类似地研究进一步指出, 服务失败的类型与恢复策略会共同影响顾客对 AI 与人类客服的补偿期待, 不同主体在同类失败中并非被等量对待^[2]。此外, 与人类相比, 服务机器人在失败后更容易被视作缺乏意图性, 因此顾客对其补偿要求往往更低^[3]。这一结论表明, 主体类型会引导消费者形成差异化的责任推断, 从而对赔偿接受度产生显著影响。

与此同时, 反事实思维研究进一步揭示了人们在评价不同主体时的心理差异。反事实思维在 AI 与人类服务失败中的作用并不一致: 个体在回溯 AI 的行为时更倾向于关注系统性错误, 而在评价人类行为时则更容易产生“本可更好”的反事实, 从而增加对人类的责备。这意味着, 在相同失败条件下, 人类主体可能因更强的行为可控性而承受更高的补偿或道德责任期待^[4]。

另一方面, 关于 AI 的心智归因研究也为解释主体差异提供了理论支撑。人们普遍认为 AI 缺乏意图、情感与道德意识, 因此其在违法或不当行为中的“应受责备性”显著弱于人类。该研究说明, 低心智赋予会降低

人们对 AI 责任的归属程度^[5]。在服务失败情境中, 人们往往认为 AI 的行为更具程序性与机械性, 相应地在评价其补偿义务时会表现出更高容忍度。这与既有研究中关于功能性代责任弱化的观点一致^[6]。当 AI 在失败后呈现“温暖回应”时, 能够显著提升用户的再次使用意愿, 说明情感线索在调节补偿接受度中扮演关键角色^[7]。

因此, 本研究旨在填补这一空白, 通过实验方法系统检验青少年在不同赔偿来源与道德框架下的赔偿接受度差异, 从而进一步深化对 AI - 人类心理差异机制的理解。

1 方法

1.1 参与者

在河北省石家庄市某所高中开展, 共发放问卷 220 份 (均为高中生), 排除了 16 名未能正确回答相关验证问题、答案不完整的被试, 最终样本为 204 名被试 (51.96% 为女性, 平均年龄=16.76)。所有被试智力正常, 没有情绪相关障碍或精神疾病, 同时没有参加过类似实验。

1.2 研究设计

本研究采用 2 (赔偿主体: 人工智能 vs 人类) \times 2 (道德类型: 功利主义 vs 义务论) 的两因素被试间设计, 因变量为被试对赔偿的接受程度, 采用 7 点评分量表进行测量, 识别人们对赔偿主体的赔偿接受度是否存在差异。

1.3 研究材料

1.3.1 电车困境、天桥困境实验材料

使用改编的电车困境(foot, 1967)、天桥困境(Thomson, 1976), 数据选取电车与天桥困境的均值。其中对赔偿的接受程度采用 7 点 Likert 量表评分(1=完全不接受 2=很不接受 3=有些不接受 4=不太接受 5=比较接受 6=很接受 7=完全接受)。为了更好的筛选数据, 因为对人工智能不熟悉的被试, 从而影响本研究的结果。

1.3.2 过程

首先, 被试随机分配到 4 个实验组, 被试阅读指导语后, 阅读经典的电车困境、天桥困境, 分别填写接受人工智能、人类赔偿的接受程度, 得出两个困境的接受度均值。

其次, 人口统计学包括性别、年级、年龄, 对人工智能的熟悉程度(采用 7 点 Likert 评分(1=完全不熟悉, 7=完全熟悉))为了更好的筛选数据, 排除对人工智能熟悉度为“完全不熟悉、很不熟悉、有些不熟悉、不太熟悉”的被试, 以免影响本研究的结果。

最后, 并告知学生本次实验是虚拟的情境。

1.4 结果

1.4.1 方差齐性与分析策略

Levene's 检验结果显示, 基于均值的莱文统计值为 $F(3, 200)=1.32, p=0.27$; 基于中位数的结果亦不显著($F(3, 200)=1.05, p=0.37$), 基于中位数并调整自由度的检验同样不显著($F(3, 183.742)=1.05, p=0.37$)。此外, 基于修剪后均值的检验结果亦不显著($F(3, 200)=1.35, p=0.26$)。

为进一步检验误差方差是否依赖自变量水平, 进行了异方差性检验。结果显示, 异方差 F 值为 $F(1, 202)=0.14, p=0.71$, 同样不显著。

1.4.2 主效应与交互作用

两因素方差分析结果显示, 两个主效应及一个交互效应均显著。赔偿主体的主效应显著, $F(1, 202)=23.21, p<0.001$, 偏 $\eta^2=0.10$ 。具体来看, 人工智能造成的赔偿接受度($M=3.97, SD=1.47$)显著高于人类造成的赔偿接受度($M=3.10, SD=1.16$)。道德类型的主效应亦显著, $F(1, 202)=9.40, p<0.01$, 偏 $\eta^2=0.05$, 表明功利主义情境下的接受度($M=3.81, SD=1.38$)显著高于义务论情境($M=3.26, SD=1.35$)。此外, 赔偿主体与道德类型

的交互效应显著, $F(1, 202)=6.62, p<0.05$, 偏 $\eta^2=0.03$, 说明道德类型对接受度的影响在不同赔偿主体下存在差异。

为进一步解释交互效应, 按赔偿主体分别检验道德类型的简单效应: 在人类组中, 功利主义与义务论的接受度差异不明显($M=3.15$ vs 3.06), 而在人工智能组中, 功利主义接受度显著高于义务论($M=4.47$ vs 3.46), 提示交互效应主要由人工智能组的道德类型差异驱动。

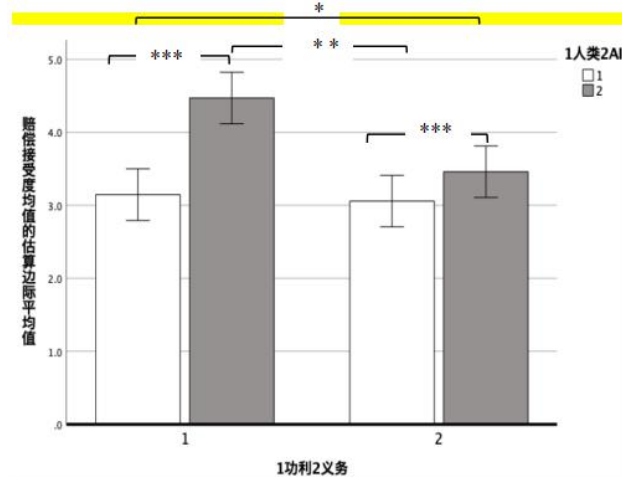


图 1. 赔偿主体(人类 vs 人工智能)与道德类型(功利主义 vs 义务论)对赔偿接受度的交互作用, 误差线表示 95% 置信区间, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

2 讨论

本研究旨在考察青少年在不同赔偿主体(人工智能 vs 人类)与不同道德类型(功利主义 vs 义务论)情境下, 对赔偿的接受度是否存在显著差异。与人类相比, 人工智能赔偿时, 高中生表现出更高的赔偿接受度。这一现象与已有研究中“AI 责任弱化效应”相吻合。先前研究指出, 人们通常认为人工智能缺乏意图性、情感性与完整的心智能力, 因此不太倾向将其视为能够承担完全道德责任的主体(Stuart & Kneer, 2021)。此外, 在服务失败情境中, AI 常被解释为一种执行规则的系统, 其错误被视为“技术性偏差”而非“道德性过失”(Huang & Lo, 2025)。

本研究中, 功利主义情境的赔偿接受度显著高于义务论情境, 说明青少年在面对包含“最大化整体利益”价值框架的道德信息时, 更倾向于给予较高接受度。这一结果与道德心理学的经典发现一致: 功利主义框架会促使人们从“结果”而非“规则”评估行为合理性, 从而降低对赔偿行为的质疑。先前研究发现, 在服务失败

或道德判断中,只要某一行为能够提升总体利益,人们对补偿方式的要求与严格度会下降(Yu et al., 2025)。

本研究发现,道德类型×赔偿主体的交互作用显著,且主要由人工智能条件下功利主义和义务论差异显著所驱动。已有研究显示,人们对 AI 的评估高度依赖情境线索,而非其固有属性(Wang et al., 2025)。若处于功利主义框架中,更容易被视为“理性而正确的执行者”;但在义务论框架下,其行为可能显得“冷漠”“缺乏人性”,从而降低赔偿接受度。同样,服务恢复研究指出,AI 的道德评价比人类更依赖框架信息(Fürst et al., 2025)。当情境强调规则与义务,AI 会因缺乏人类式道德直觉而受到负面评价;反之,在强调结果最大化的框架下,人们反而认为 AI 更具优势。因此,本研究中只有人工智能条件呈现显著差异,说明青少年在面对 AI 时更可能“随情境波动”,表现出更强的道德框架敏感性。

参考文献

- [1] Yu, Zhaoji, et al. “The Influence of AI Service Robots’ Humorous Response Strategies on Consumer Forgiveness Following Service Failure.” *Scientific Reports*, vol. 15, no. 1, 18 Nov. 2025, pp. 40481 – 40481, [www.nature.com/articles/s41598-025-24368-y](https://doi.org/10.1038/s41598-025-24368-y), <https://doi.org/10.1038/s41598-025-24368-y>. Accessed 30 Nov. 2025.
- [2] Fürst, Andreas, et al. “Service Recovery by AI or Human Agents: Do Failure and Strategy Context Matter?” *Journal of Service Management*, 13 Apr. 2025, www.researchgate.net/publication/390743050_Service_Recovery_by_AI_or_Human_Agents_Do_Failure_and_Strategy_Context_Matter?utm_source=chatgpt.com, <https://doi.org/10.1108/JOSM-04-2024-0190>.
- [3] Song, Jinzhu, et al. “Influencing Customer Compensation Expectations in Service Failures: Comparing the Roles of Service Robots and Human Employees.” *International Journal of Hospitality Management*, vol. 126, 29 Nov. 2024, pp. 104031 – 104031, <https://doi.org/10.1016/j.ijhm.2024.104031>.
- [4] Xie, Yibo, et al. “Artificial Intelligence or Human Service, Which Customer Service Failure Is More Unforgivable? A Counterfactual Thinking Perspective.” *Psychology & Marketing*, 31 Mar. 2025, <https://doi.org/10.1002/mar.22215>.
- [5] Stuart, Michael T., and Markus Kneer. “Guilty Artificial Minds: Folk Attributions of Mens Rea and Culpability to Artificially Intelligent Agents.” *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW2, 13 Oct. 2021, pp. 1 – 27, <https://doi.org/10.1145/3479507>.
- [6] Huang, Zuwen, and Ada Lo. “Human vs. Robot Service Provider Agents in Service Failures: Comparing Customer Dissatisfaction and the Mediating Role of Forgiveness and Service Recovery Expectation.” *Information Technology & Tourism*, 18 Feb. 2025, <https://doi.org/10.1007/s40558-025-00314-6>.
- [7] Wang, Cuicui, et al. “Exploring the Effect of AI Warm Response on Consumer Reuse Intention in Service Failure.” *Computers in Human Behavior*, 12 Feb. 2025, p. 108606, www.sciencedirect.com/science/article/pii/S0747563225000536, <https://doi.org/10.1016/j.chb.2025.108606>.