人工智能隐形效应：理解用户未能识别人工智能时的人机交互

Obada Kraishan媒体与传播学院，德州理工大学，美国德克萨斯州拉伯克，邮编79409
https://orcid.org/0009-0007-7180-8620

2026年1月5日

arXiv:2601.00579v1 [cs.HC] 2 Jan 2026

摘要

人工智能快速融入移动应用，彻底改变了数字生态；然而，这种变化对用户感知人工智能功能的影响仍未得到充分理解。本研究对来自 iOS App Store 和 Google Play Store 的 422 款应用程序（200 款含人工智能功能，222 款对照）中的 1,484,633 条移动应用程序评论进行了大规模分析。通过采用情感分类、主题建模和关切—收益分类，我们发现了一个显著的脱节：尽管 47.4% 的应用程序具备人工智能能力，但仅有 11.9% 的评论提及人工智能。含人工智能功能的应用程序评分显著低于传统应用程序（$d = 0.40$）；然而，层级回归揭示了一种隐蔽模式——在控制人工智能提及和评论特征后，这种负相关关系发生了逆转（$b = 0.405, p < .001$）。隐私是用户最主要的担忧（占表达担忧的评论的 34.8%），而效率则是首要收益（42.3%）。不同类别的影响差异显著：助手类应用呈正向影响（$d = 0.55$），娱乐类则呈负向影响（$d = -0.23$）。这些发现表明，人工智能功能往往在用户认知阈值之下运行；驱动负面评价的是用户明确识别出人工智能，而非人工智能仅仅存在。这一发现挑战了关于人工智能系统技术接受的基本假设。

关键词 人工智能 · 人机交互 · 移动应用 · 用户感知 · 技术接受 · 应用评论 · 用户体验

1 引言

人工智能融入日常数字体验，是过去十年最重大的技术变革之一。从能够理解自然语言的语音助手，到利用计算摄影增强图像的照片编辑器，人工智能已从研究实验室走进全球数十亿用户的智能手机 [1, 2]。这一转变在移动应用中尤为明显：人工智能功能已在各种类别的应用中日益普及，涵盖效率工具和娱乐平台等领域 [3]。然而，尽管人工智能功能已广泛应用，我们对普通用户如何感知、评价和使用这些功能仍知之甚少。

人工智能在消费类应用中的吸引力，主要在于增强个性化、提升效率，以及实现传统编程方法无法实现的功能 [4, 5]。科技公司已投入数十亿美元开发人工智能功能，并常将其定位为竞争激烈市场中的关键差异化优势 [6]。营销材料经常强调这些应用的智能水平，暗示人工智能集成能够自动改善用户体验。然而，近期研究表明，人工智能实施与用户满意度之间的关系可能比这些说法所描述的更为复杂 [7, 8]。

多个研究领域考察了人机交互的不同方面，但对于消费者应用中人工智能功能在真实场景下引发的用户反应，我们仍有许多重要问题尚未弄清。人机交互研究主要聚焦于受控的实验室实验或原型系统 [1, 9], ，而信息系统研究考察的则是组织中的人工智能采纳，而非个体消费者的体验 [10, 11]。营销研究探讨了特定情境下消费者对人工智能的态度，例如客户服务或推荐系统 [12, 13]。然而，这些研究方法通常是在受控环境中考察态度，而非收集用户在真实应用使用情境中的自然反馈。

本研究通过分析大量来自应用商店评论的真实用户反馈，弥补了这些局限，从而深入了解数百万用户在日常数字生活中如何体验和评价人工智能功能。应用商店评论为了解真实用户的情感倾向提供了独特窗口，因为它们是在现实使用情境中自愿、主动生成的反馈 [14, 15]。与实验室研究或调查问卷不同，这些评论记录了用户对已全面部署的商业应用的反应，反映了实际实施中的复杂性与限制，而非简化的测试版本。

由于大语言模型和生成式人工智能系统取得突破后，人工智能部署近期加速推进，因此本研究的时机至关重要 [16, 17]。预计2024年至2025年将见证公众对人工智能能力的高度关注，ChatGPT 和 Claude 等工具也将成为家喻户晓的名字 [18]。随着公众对人工智能能力的关注不断增加，一个重要问题随之浮现：用户是否以及如何在日常应用程序中识别人工智能功能。基于这些观察，本研究提出了三个研究问题，共同揭示当前人工智能在消费者移动应用中的集成状况：

- **研究问题1：** 含人工智能功能的应用程序与非 AI 应用在用户满意度和参与模式方面有何不同?
- **研究问题2：** 用户在讨论移动应用中的人工智能功能时，会表达哪些具体担忧和感知收益?
- **研究问题3：** 不同应用类别和移动平台对人工智能功能的接受度有何差异?

本研究为我们理解消费情境中的人机交互作出了多项重要贡献。首先，我们对真实环境中用户对人工智能驱动的移动应用的反应进行了大规模分析，考察了来自 iOS 和 Android 平台上 422 款应用程序的逾 148 万条自发评论。其次，我们考察用户是否以及如何在自然情境评价中识别并描述人工智能功能。第三，我们通过层次回归分析，揭示人工智能实施与用户满意度之间的复杂关系。第四，我们构建了一套全面的分类体系，归纳与人工智能功能相关的用户担忧和感知收益。最后，我们证明，人工智能接受度在不同应用类别和平台之间存在显著异质性。

2 文献综述

2.1 消费者应用程序中人工智能的发展

人工智能在面向消费者的应用程序中的应用经历了几个不同阶段，每个阶段都以不同的技术能力和用户期望为特征。早期应用侧重于基于规则的系统和简单的模式识别，提供预测文本和基本推荐算法等功能 [19]。这些最初的人工智能尝试往往不为用户所察觉，它们作为后台进程运行，在没有明确用户认知的情况下增强应用功能。2011年，Siri 等虚拟助手的出现，标志着人工智能功能朝着更加可见、互动的方向迈出了一大步，但技术限制和用户疑虑仍然制约着其普及 [20]。

2020 年之后出现的这一代 AI 驱动的应用程序，在能力和普及程度上都实现了质的飞跃。深度学习的发展，尤其是自然语言处理和计算机视觉领域的进步，使应用程序能够执行以往被认为需要人类智能才能完成的任务 [21, 22]。照片编辑应用程序如今可以利用计算摄影技术自动增强图像，其效果对大多数用户而言胜过手动编辑 [23]。基于大语言模型的写作助手能够以接近人类的流畅程度生成、编辑和翻译文本 [16]。这些能力改变了移动应用，使人工智能功能从非同寻常之举变成了用户的预期 [10]。

尽管取得了这些技术进步，关于用户对人工智能功能的接受度的研究却呈现出复杂且往往不一致的图景。研究记录了用户对人工智能能力的热情，也揭示了他们对人工智能实现方式的深切怀疑[24, 25]。研究表明，即使算法决策的表现明显优于人类替代方案，用户往往仍更倾向于人类判断

。另一方面，其他研究发现，在某些情况下，用户会过度依赖人工智能推荐，这可能导致技能退化或批判性思维能力下降 [26]。这些相互矛盾的发现表明，用户对人工智能功能的反应在很大程度上取决于具体情境、实施方式和个体差异。然而，尽管以往研究通过实验室实验、调查问卷或针对特定人工智能系统的孤立研究考察了人工智能接受度，但针对多个类别和平台中真实 App Store 与 Google Play 评论的大规模实证分析——即用户如何评价人工智能功能——仍然有限。鉴于 App Store 和 Google Play 评论反映的是用户对已全面部署的商业应用程序所提供的真实、自发反馈，而非原型系统或假设情境，这一研究空白尤为重要。

2.2 用户感知与人工智能悖论

越来越多的文献指出，消费者对人工智能的反应中存在一种悖论——既渴望人工智能功能，又担忧其影响。用户对人工智能所承诺的效率与便利性充满热情，同时也对隐私、个人自主权以及人工智能中介体验的真实性表示担忧 [27]。这种悖论在不同人口统计群体中表现各异；与通常更为怀疑的年长用户相比，年轻用户往往更能接受人工智能功能 [28]。

重要的是，现有关于人工智能可接受性的研究通常假定用户能够有意识地识别并区分自己正在评估的人工智能功能。然而，关于用户是否确实能在消费类应用中识别人工智能，或者人工智能功能是否可能在用户未明确察觉的情况下影响用户体验的研究，仍然有限。值得审视这种关于有意识的认知的假设，因为人工智能日益隐蔽地运行于应用程序之中，并嵌入用户可能无法识别为人工智能的功能中。

信任是影响用户接受人工智能功能的核心因素。人机交互研究已识别出人工智能系统信任的多个维度，包括能力信任、善意信任和诚信信任[29, 30]。用户对人工智能系统建立信任的模式不同于对人类代理或传统软件建立信任的模式；用户往往难以形成关于人工智能能力与局限性的恰当心智模型 [31]。

人工智能系统的透明度是影响用户感知的另一个关键因素。尽管在研究和政策讨论中，对可解释人工智能的呼声日益高涨 [32, 33], ，但研究表明，用户未必总是希望了解或能从人工智能运作的详细说明中获益。Kizilcec 的研究 [34] 发现了透明度悖论：关于算法过程的信息过多，可能会降低信任和满意度。这一发现使人工智能透明度与用户接受度之间的关系不再那么简单，也表明有效的人工智能集成需要审慎考虑用户想要了解哪些信息，以及他们能够有意义地处理哪些信息。

2.3 隐私担忧与数据焦虑

隐私担忧或许是相关记录最为一致的用户忧虑，尤其涉及消费类应用中的人工智能功能。机器学习系统对数据的高度依赖，要求广泛收集和处理用户信息，由此引发了对数据安全、使用方式和对照权的疑问 [35, 36]。研究表明，用户往往不了解人工智能功能所需的数据收集范围，因而出现了 Barocas 和 Nissenbaum [37] 所称的“语境坍塌”：个人信息不同领域之间原本预期的边界遭到打破。

隐私权衡框架指出，用户在决定是否使用人工智能驱动的功能时，会进行成本效益分析，权衡感知收益与隐私风险 [38, 39]。然而，实证研究表明，这种权衡往往存在偏差，用户的偏好和行为并不一致：这就是所谓的隐私悖论，即用户表达的担忧与实际使用模式并不相符 [40]。在人工智能应用的情境中，数据处理过程不透明，以及用户难以了解自己的信息如何被用于训练和改进人工智能系统，可能会加剧这一悖论。

近期研究还强调了对推断隐私的担忧：人工智能系统能够从看似无害的数据中推断出敏感信息 [41]。用户可能愿意分享某些类型的信息，却不愿接受人工智能系统从这些数据中得出的推断。在使用人工智能进行个性化或预测的应用程序中，这种担忧尤为突出，因为系统预测用户需求的能力可能会被视为侵扰，而非助益 [42]。

2.4 平台与类别差异

移动应用生态系统在平台和应用类别方面存在显著异质性，而这些因素可能会显著影响用户对人工智能功能的接受度。关于 iOS 和 Android

用户的研究发现，他们在用户人口统计特征、技术采纳模式和隐私偏好方面存在系统性差异[43, 44]。iOS 用户的收入水平往往较高，也更愿意为应用程序付费；而 Android 用户则代表着更多元化的全球人口，其技术素养水平各不相同 [45]。这些平台差异可能会导致用户对人工智能功能抱有不同的期望，并作出不同的评价。

应用类别是影响人工智能接受度的另一个重要因素。信息系统领域的研究早已认识到，技术接受度会因任务特征和使用情境而异 [46]。在人工智能功能方面，用户对效率工具、娱乐类应用或创意平台的预期和容忍度可能各不相同。例如，用户评价效率类应用中的人工智能辅助时，主要可能考虑效率；而创意类应用中的人工智能则可能引发对真实性和人类能动性的担忧 [47, 48]。

类别层面的人工智能就绪度这一概念表明，某些应用领域可能比其他领域更适合人工智能集成。Jussupow 等人 [49] 的研究指出，任务复杂度、结果可测量性和错误容忍度等因素决定了人工智能实施能否成功。例如，娱乐类应用中的人工智能功能若损害用户愉悦感，可能面临较低的错误容忍度；而实用工具类应用则可能受益于人工智能对常规任务的自动化。理解这些类别特有的动态，对于解释用户对人工智能功能的反应差异至关重要。

2.5 理论框架

基于这些不同的文献脉络，我们采用多理论视角，结合技术接受模型、信任理论和隐私权衡框架的见解。技术接受模型（TAM）及其扩展模型为理解感知有用性和易用性如何影响用户对人工智能功能的评价提供了基础 [50, 51]。然而，传统的 TAM 构念可能不足以捕捉人工智能系统的独特特征，尤其是对自主性、真实性和算法决策的担忧。
因此，我们将信任理论中的概念纳入 TAM，认识到用户对人工智能功能的接受度不仅取决于功能评价，还取决于对系统能力、善意和正直的信任 [52]。Lee 和 See [53] 提出的自动化信任框架进一步揭示了用户如何校准对人工智能系统的依赖，并指出了导致过度信任和信任不足的因素。该框架对于理解许多消费类应用程序中观察到的人工智能能力与用户使用模式之间的脱节尤为相关。

隐私权衡理论为理解用户如何权衡人工智能驱动的功能所固有的利弊提供了一个视角 [54, 55]。我们扩展了传统的隐私权衡模型，以纳入人工智能系统的独特特征，包括数据处理过程的不透明性、发生推断性隐私侵犯的可能性，以及用户评估隐私风险时面临的困难。这个扩展框架有助于解释用户为何可能一边表达隐私担忧，一边继续使用人工智能驱动的功能，也有助于说明隐私显著性为何会因应用程序情境不同而有所差异。

3 方法

3.1 数据收集与抽样

本研究采用大规模观察性设计，分析了来自移动应用市场的公开用户评论。为确保数据覆盖全面，数据收集历时较长，并通过自动化网络爬取完成，同时设置适当的速率限制以遵守平台指南。根据下载量和用户评分，从 Apple App Store 和 Google Play Store 排名前 500 的免费应用程序中系统地筛选应用程序。筛选依据三项纳入标准：（1）连续三个月保持在排行榜中，表明用户参与度稳定；（2）同时适用于 iOS 和 Android 平台，以便进行跨平台比较；（3）至少拥有 1,000 条用户评论，以确保应用层面分析具备足够的统计功效。

我们通过对多个指标进行加权分析，识别出具有人工智能功能的应用程序：描述中明确提及人工智能相关术语、分析记录人工智能功能推出情况的发行说明，以及通过科技新闻来源和开发者公告进行核实。通过这一多来源方法，我们确认了200款具有人工智能功能的应用程序，其中包括推荐算法、自然语言处理和生成式人工智能能力等具体示例。其余222款对照应用程序没有显示出具备人工智能功能的证据；我们根据类别、用户群规模和发布时间范围，采用详细的方法将其与人工智能应用进行匹配，以尽量减少混杂变量。

3.2 样本特征

最终数据集采集于2024年1月至2025年1月，包含来自422款应用程序（200款具有人工智能功能，222款对照应用程序）的1,484,633条评论。评论在不同平台上的分布并不均衡（Android：72.2%，n = 1,072,246；iOS：27.8%，n = 412,387），这可能反映了应用市场动态。应用程序涵盖23个类别，其中效率（18.2%）、娱乐（14.5%）和创意（12.8%）占比最高，表明研究重点集中在这些领域。每款应用平均产生3,518条评论，标准差为4,892，说明评论数存在显著差异。表1详细列出了样本特征，包括各类别和平台的分布情况。

表1：样本特征与分布

特征	n (%)
评论总数	1,484,633
应用程序总数	422
含人工智能功能的应用	200（47.4%）
对照应用	222（52.6%）
提及人工智能的评论	176,658
占总数的百分比	11.9%
平台	
iOS	412,387 (27.8%)
Android	1,072,246 (72.2%)
时间段	2024年1月—2025年1月
平均评论长度（词数）	47.3（SD = 68.2）
平均评分（总体）	3.58（SD = 1.52）

3.3 数据提取流程

数据提取分为三个不同阶段，每个阶段都使用专门的 Python 库，具体内容将在后续章节中详述。应用元数据通过 iOS 平台的 App Store Scraper (v1.0.0) 和 Android 平台的 Google Play Scraper (v1.2.0) 收集，提取的标准化字段包括标识符、开发者信息、类别、版本历史、评分和描述。

评论收集针对各个平台采用了不同策略，以有效应对应用程序编程接口的局限性，例如在 iOS 平台上结合直接抓取与 iTunes RSS 订阅源。对于 iOS 评论，我们采用了多方法方案，将直接抓取、iTunes RSS 订阅源以及作为备用方案的 iTunes Lookup API 相结合。Android 评论通过支持分页的 Google Play Scraper 应用程序编程接口收集。我们设置了每个应用最多收集 1,000 条评论、评论长度至少为 10 个字符等参数，以确保数据集规模可控，同时保证数据质量。每条评论记录均包含文本内容、评分（1–5 星）、提交日期、应用版本，以及在有开发者回复时记录的回复内容。

通过分析发行说明和版本历史，确定人工智能功能的推出时间，从而了解其开发时间线。iOS 发行说明来自 iTunes Lookup API，而 Android 数据则来自 recentChanges 字段。若无法获取发行说明，则检查应用程序描述以识别人工智能功能，作为功能识别的替代方法。

3.4 隐私与伦理考量

所有数据收集均遵循针对公开内容的伦理标准。为防止重新识别，同时保留分析所需的唯一性，审阅者姓名均替换为经过 MD5 哈希处理的标识符。系统自动扫描评论文本，并对其中的个人信息（包括电子邮件地址、电话号码和真实姓名）进行删减。研究遵守平台服务条款，通过速率限制控制访问，且未尝试访问受限内容或规避平台限制。

3.5 测量与操作化

AI 相关内容检测使用了一个包含 74 个术语和短语的词典。这些术语和短语源自学术文献，并经人工审核完善，用于识别相关内容。该词典涵盖三个与检测过程相关的类别：人工智能核心概念（如“人工智能”“机器学习”）、特定模型（如

“GPT”“Claude”）以及功能实现（如“AI助手”“智能回复”）。每条评论均按是否提及人工智能进行二元分类，并计算每 100 个词的关键词密度这一连续指标，以评估人工智能相关内容的普及程度和强度。

情感分析结合了基于规则的方法和机器学习方法。基于规则的部分使用了由正面和负面词语构成的加权词典。机器学习部分采用了在产品评论数据上微调的 DistilBERT 模型，可生成各情感类别的概率分布。当 Transformer 模型的置信度超过 0.7 时，它会作为主要分类器；同时，基于规则的系统用于验证 Transformer 模型的结果。

内容分析通过将文本与基于定性评论分析构建并经过验证的词典进行模式匹配，识别其中的担忧和益处。研究将八种担忧类别和八种益处类别操作化；为体现短语在具体语境中更高的特异性和相关性，短语匹配的权重设为单个关键词的两倍。

3.6 统计分析

我们使用 Python 进行统计分析，主要依赖 pandas 库进行数据处理、scipy 进行统计检验、statsmodels 进行回归分析，并使用 scikit-learn 处理机器学习部分。为尽量减少研究者自由度，分析计划已预先注册；事后开展的探索性分析则在结果部分明确标出。

在完成探索性分析后，我们使用带有 Welch校正的独立样本 t 检验，对含人工智能功能的应用与非 AI 应用进行组间比较，以处理组间方差不齐的问题。我们采用卡方检验分析类别变量的比较，包括评分分布和 AI 提及模式。对于连续结果，使用 Cohen’s d 计算效应量。对于类别变量之间的关联，我们计算 Cramér’s V，作为考虑列联表维度的效应量指标。

我们通过多元回归模型考察人工智能功能与用户评分之间的关系，同时控制潜在的混杂变量。基准模型将人工智能应用状态设为主要预测变量，并以用户评分为结果变量。后续模型进一步纳入应用类别、平台、评论长度和时间因素作为对照变量。我们检验了人工智能状态与其他预测变量之间的交互效应，尤其是与平台和类别之间的交互效应，以识别潜在的调节因素。
逻辑回归模型用于考察预测五星评论概率的因素，并计算优势比以量化效应大小。我们采用方差分析（ANOVA）检验多个应用类别之间的差异，并报告 η平方（η^2）作为效应量指标。双因素方差分析用于考察平台与人工智能状态之间的交互效应。事后成对比较采用 Bonferroni校正，以识别具体的类别差异。Pearson 相关系数用于评估具体担忧/益处与评分之间的关联。在时间趋势分析中，Cochran-Armitage检验用于评估研究期间隐私担忧是否呈单调上升趋势。所有模型均报告适当的拟合统计量，包括线性模型的 R^2 和逻辑回归模型的伪-R^2 。

除统计模型外，主题建模还采用潜在狄利克雷分配（LDA），以识别提及人工智能的评论中的主题模式。文本预处理包括使用 spaCylibrary 进行分词、将词形还原为基本形式、使用扩展列表移除停用词（其中包括领域专用词），以及提取一元词和二元词，以捕捉多词概念。为支持主题建模过程，研究使用词频-逆文档频率（TF-IDF）加权构建文档-词项矩阵，并将词汇表上限设为 1,000 个特征，以兼顾全面性与计算效率。模型选择通过网格搜索，在 k = 5 至 20 个主题范围内进行，并根据困惑度分数和主题一致性指标评估模型。最终采用的 k = 10 个主题模型，在可解释性与统计拟合度之间实现了最佳平衡。

4 结果

以下各节将详细呈现研究发现，分别回应各项研究问题，并考察满意度差异、用户担忧与益处，以及不同平台和类别之间的异质性效应。

4.1 研究问题 1：含人工智能功能的应用程序与非人工智能应用程序

具有人工智能功能的应用程序获得的平均评分显著低于对照应用程序（M = 3.28， SD = 1.54；对照应用程序：M = 3.87， SD = 1.43），$t(420) = -4.13$， $p < .001$， $d = 0.40$。这种中等程度的效应在不同分析方法中均持续存在。人工智能应用的五星评论占比更低（32.4% vs. 41.8%），一星评价占比更高（24.7% vs. 16.2%）， $\chi^2(4, N = 1,484,633) = 18,942.31$， $p < .001$，Cramér’s $V = .113$。评论长度也

有所不同，人工智能应用生成的反馈更长（M = 52.1 个词），而非 AI 应用为（M = 42.5 个词），t(1，484，631) = 47.93，p <.001，d = 0.14。表2汇总了比较统计数据。

表2：含人工智能功能的应用程序与对照应用程序的比较统计数据

测量	数值	95% 置信区间
评分，M（SD）		
含人工智能功能的应用	3.28 (1.54)	[3.26，3.30]
对照应用	3.87 (1.43)	[3.85, 3.89]
差异	-0.59***	[-0.62，-0.56]
Cohen's d	0.40	[0.37, 0.43]
五星评价，%		
含人工智能功能的应用	32.4	[32.1，32.7]
对照应用	41.8	[41.5，42.1]
一星评价，%		
含人工智能功能的应用	24.7	[24.4, 25.0]
对照应用	16.2	[15.9, 16.5]
评论长度（词数），M（SD）		
含人工智能功能的应用	52.1 (71.3)	[51.7，52.5]
对照应用	42.5（64.8）	[42.1，42.9]

尽管近一半的应用程序使用了人工智能，但只有11.9%的评论提及了人工智能——人工智能应用中占24%，而对照组中仅占0.8%（ϕ = .34，p < .001）。iOS上的提及率略高于Android（15.2%比10.8%；χ^2(1，N = 1,484,633) =6,234.45，p < .001；见图1和表3）。

表3：人工智能提及模式与分布

模式	频率	效应量
提及人工智能的评论（%）	176,658 (11.9%)	
在人工智能应用中	23.7%	ϕ = 0.34***
在非人工智能应用程序中	0.8%	
人工智能提及中的情感		
仅正面	45.2%	
仅负面	31.8%	
混合	16.3%	
中性	6.7%	
按提及类型划分的评分 (M ± SD)		
提及人工智能	3.42 ± 1.61	
未提及人工智能	3.59 ± 1.49	
差异	-0.17***	d = 0.11
热门人工智能关键词（%）		
“AI”	68.4%	
“ChatGPT”	24.7%	
“机器人”	18.3%	
“智能”	15.2%	
“智能”	12.1%	

分层回归分析考察了预测用户评分的因素（表4）。模型1显示，人工智能应用的评分较低（b = −0.59, p < .001）。加入人工智能提及后（模型2），结果显示这两个因素均独立预测较低的评分。交互效应极小（模型3：b = 0.03, p = .464），说明这些因素的影响彼此独立，而非产生协同效应。值得注意的是，在完整模型（模型4）中，纳入对照变量后，人工智能应用的系数反转为正（b = 0.405, p <.001）；模型解释了 R^2 = .084的方差，F(23, 99,971) = 397.83, p < .001。这一模式表明存在抑制效应：负向双变量关联被人工智能应用吸引更长、更具批判性的评论所掩盖。

预测五星评论的逻辑回归结果印证了这些发现（表5）。人工智能应用最初获得五星评分的几率较低（OR = 0.67，95% CI [0.65，0.69]，p < .001），但加入对照变量后，获得五星评分的几率更高（OR = 1.55，95% CI [1.51，1.59]，p < .001）。提及人工智能的评论获得五星评分的可能性仍然较低（OR = 0.75，95% CI [0.73，0.77]，p < .001）。

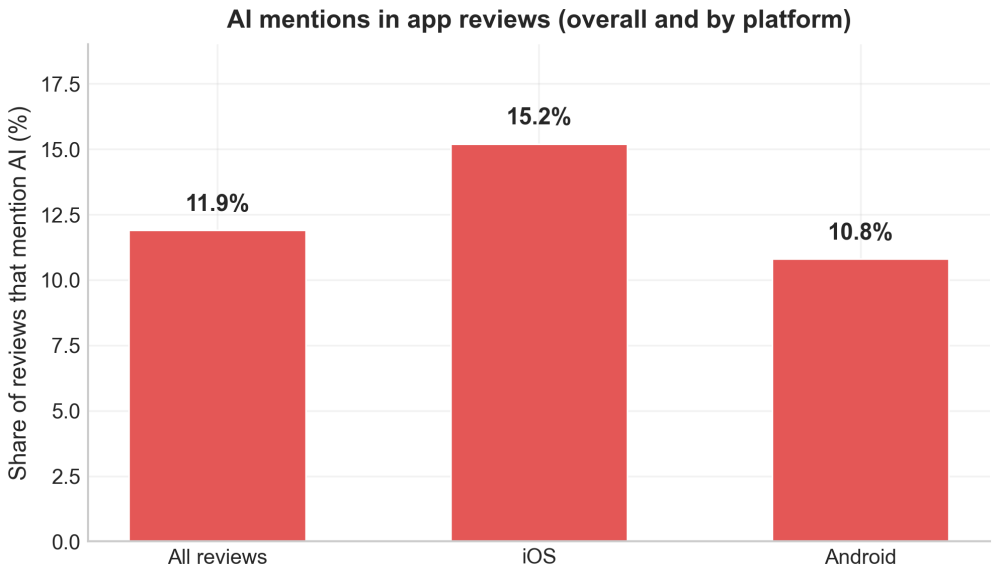


图1：应用评论中的人工智能提及

表4：预测评论评分的分层回归模型

变量	M1	M2	M3	M4
截距	3.29***	3.29***	3.29***	3.13***
	(0.01)	(0.01)	(0.01)	(0.01)
人工智能应用	−0.59***	−0.48***	−0.48***	0.41***
	(0.02)	(0.02)	(0.02)	(0.03)
AI 提及		−0.18***	−0.18***	−0.22***
		(0.02)	(0.02)	(0.02)
人工智能 × 提及			0.03	
			(0.04)	
平台				0.00
				(0.01)
评论的对数				0.06***
				(0.00)
词数				−0.02***
				(0.00)
R^2	.012	.012	.012	.084
调整后 R^2	.012	.012	.012	.084
N	100,000	100,000	100,000	100,000

注释。 M1 = 基准模型；M2 = + AI 提及；M3 = 交互项；M4 = 完整模型。非标准化 b (标准误)。因变量：星级评分（1–5）。 ***$p < .001$。

这些发现表明，具有人工智能功能的应用程序在满意度和参与模式方面与非 AI 应用存在显著差异。人工智能存在（占应用程序的47.4%）与用户明确认知（占评论的11.9%）之间的显著差距表明，用户很少意识到或明确提及人工智能功能。抑制效应表明，表面上的负面反响源于明确提及人工智能的评论者，而非人工智能本身的存在。控制人工智能提及和评论特征后，人工智能应用获得五星评分的几率更高，这表明满意度差异取决于实施质量和用户认知，而非人工智能技术本身。

4.2 研究问题 2：用户担忧与感知益处

用户既表达了对人工智能功能的担忧，也提及了其益处。提及 AI 的评论情感倾向更加两极化，评分也略低于非 AI 评论。在 176,658 条提及 AI 的评论中，32.1% 表达了担忧，45.3% 提及了益处，29.0% 同时表达了两者。与非 AI 评论相比，这些评论的情感倾向更加两极化（45.2% 强烈正面；31.8% 强烈负面，而非 AI 评论中分别为 33.6% 和 24.1%）， $\chi^2(2, N = 50{,}000) = 157.59$, $p < .001$, Cramér's $V = .056$。提及

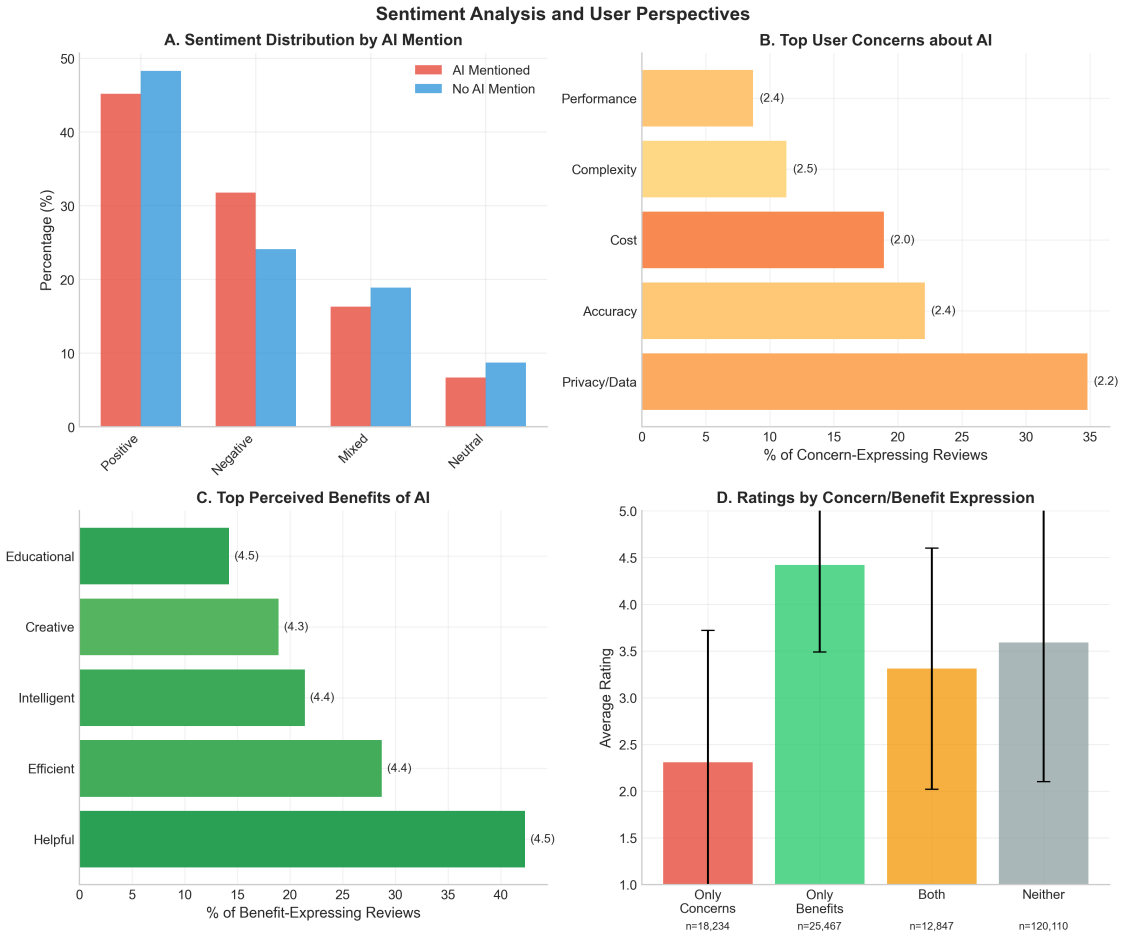


图2：情感分析与用户观点

表5：预测五星评价的逻辑回归模型

变量	B	SE	OR	95% 置信区间
截距	-0.90***	0.01	0.41	[0.41, 0.41]
人工智能应用（0/1）	+0.44***	0.02	1.55	[1.51, 1.59]
AI 提及（0/1）	-0.29***	0.02	0.75	[0.73, 0.77]
平台（iOS）	+0.08*	0.04	1.08	[1.00, 1.16]
Log（评论总数）	+0.06***	0.00	1.06	[1.06, 1.06]
伪 R^2	0.007			
N	100,000			

人工智能的评论评分略低（M = 3.42, SD = 1.61），未提及人工智能的评论评分则为（M = 3.59, SD = 1.49），t(1,484,631) =−43.52, p < .001, d = 0.11，且大写比例更高。图2展示了包含四个面板的分析，其中 (A) 比较提及人工智能与非 AI 评论的情感分布；(B) 用户对人工智能功能最主要的担忧；(C) 用户感知到的人工智能功能主要益处；(D) 按担忧/益处表达模式划分的平均评分，误差线表示标准差。

最主要的担忧是隐私/数据（34.8%）、准确性/错误（22.1%）和费用/订阅（18.9%）；最主要的益处是实用性/效用（42.3%）以及效率/节省时间（28.7%）（表6）。

独立样本t检验显示，仅表达担忧的评论评分远低于仅表达益处的评论（Δ =2.11； M = 2.31， SD = 1.41 对比 M = 4.42， SD = 0.93），t(43,701) = 89.23， p < .001， d = 1.77。混合评论

表6：提及 AI 的评论中的用户担忧与益处

类别	AI 占比评论	Avg 评分
担忧（占表达担忧的评论）	32.1	2.31
隐私/数据	34.8	2.18
准确性/错误	22.1	2.42
费用/订阅	18.9	1.97
复杂性/可用性	11.3	2.54
性能/技术	8.7	2.38
伦理/偏见	3.1	2.89
依赖/过度依赖	0.8	3.12
真实性/虚假性	0.3	2.76
益处（占比其中表达益处的评论)	45.3	4.42
有帮助/有用	42.3	4.51
高效/节省时间	28.7	4.38
智能/聪明	21.4	4.42
有创意/创新	18.9	4.29
教育/学习	14.2	4.48
易于使用/简单	13.6	4.31
个性化/自适应	7.8	4.27
娱乐性/趣味性	6.9	4.19

处于中间水平（M = 3.31， SD = 1.29）。皮尔逊相关显示，隐私（r = –.34）和效率（r = .41）与评分的关联最强，两者均 p < .001。

表7：LDA分析的主题模型结果

主题	标签	% 文档	平均评分	主题描述
0	Google助理/Gemini 故障	6.7	2.43	平台集成问题
1	ChatGPT 的帮助性	9.4	4.22	问题、回答、帮助
2	AI聊天与角色	10.4	3.25	Chat, bots, 角色人工智能
3	Copilot与智能功能	5.9	3.20	搜索、语音、Alexa
4	费用与变现	11.1	2.27	免费、付费、订阅
5	最佳 AI 热情度	12.9	4.59	最佳，惊艳，喜爱
6	更新与问题	12.9	2.43	修复，更新，故障
7	AI 比较	5.0	3.80	更好，用过的应用程序
8	对话体验	16.9	3.73	交谈、乐趣、喜欢
9	技术卓越性	8.8	4.26	Tool, 准确 ，模型

主题建模（LDA； k = 10）凸显了三种模式：一个占比较大的“对话体验”主题（16.9%； M= 3.73， SD = 1.45），分布广泛；一个评价高度正面的“最佳 AI 热情度”主题（12.9%； M = 4.59， SD =0.78）；以及两个评价较低的主题簇——“成本／变现”（11.1%； M = 2.27）和“更新／问题”（12.9%； M =2.43）。此外，一个规模较小的“平台集成问题”主题（6.7%）评价也较低（M = 2.43， SD = 1.38）；其占比在2024年11月前后增加了约23%。图3展示了对176,658条提及 AI 的评论进行潜在狄利克雷分配（LDA）主题建模（k=10）生成的词云，其中词语大小表示每个主题中词语的词频及重要性，这些主题代表了用户关于人工智能功能的讨论中的不同主题簇。完整分布请参见表7。

使用 Cochran–Armitage 检验进行时间分析后发现，隐私担忧有所增加，从 2024年第一季度的 28.3% 上升至 2025年第一季度的 39.2%（z = 18.92， p < .001）。研究期间，每月人工智能提及率也从 8.3% 上升至 14.7%。效率效益保持稳定（27.1%–29.8%），而在 2024年10月推出订阅模式后，与成本相关的投诉增加了 47%。主题波动性分析表明，技术问题的变化幅度较大（CV > 0.30），而核心 AI 能力则保持稳定（CV < 0.10）。

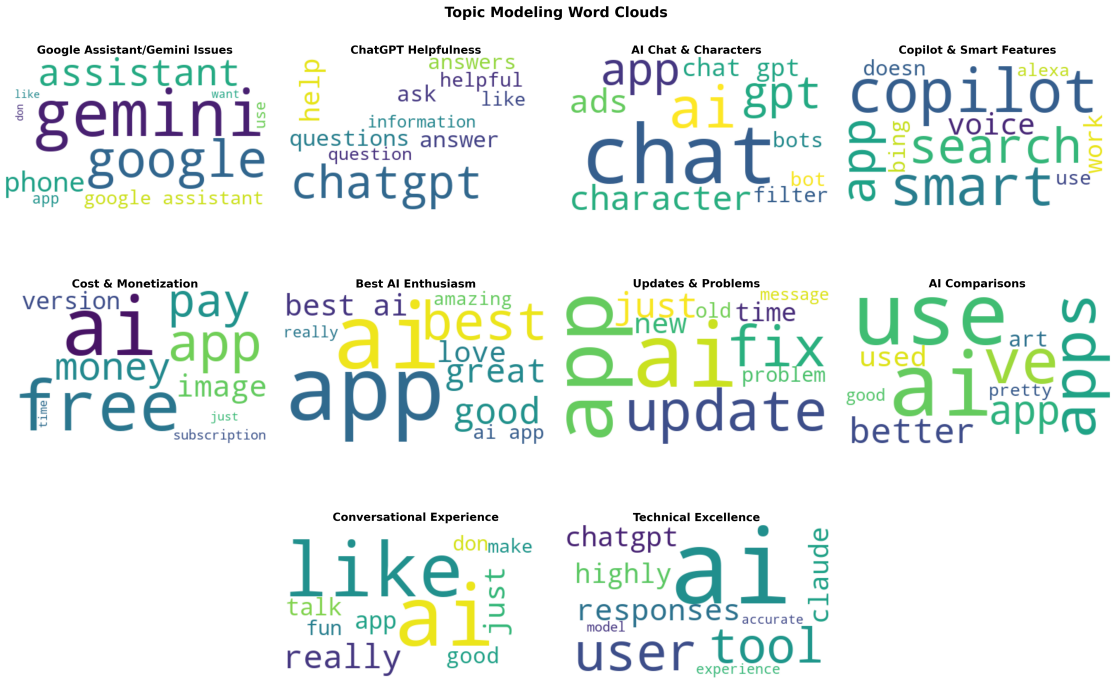


图3：主题建模词云

这些发现通过明确指出用户对人工智能功能的具体担忧与益处，直接回应了研究问题2。隐私/数据安全是最主要的担忧（34.8%），而实用性/效用则是用户感知到的主要益处（42.3%）。仅表达担忧的评论与仅表达益处的评论之间效应量较大（d = 1.77），这表明用户评价在很大程度上受其感知到的成本效益权衡影响。情感分布呈现两极化（Cramér's V = .056），而主题建模得出的主题模式多样，这说明人工智能功能会引发强烈且各不相同的情绪反应，而非一致的反应。成功似乎取决于能否解决隐私担忧，同时带来切实的效率提升；不同用例的实施质量也有所差异，这一点从对话式人工智能体验呈现的双峰分布中可见一斑。

4.3 研究问题3：平台与类别异质性

为考察不同平台和类别中用户对人工智能功能接受度的差异，我们使用独立样本t检验比较了 iOS 和 Android，采用双因素方差分析检验平台与人工智能状态的交互作用，并通过单因素方差分析及后续事后比较评估类别差异。

平台分析显示，用户对人工智能功能的接受度存在系统性差异。与 Android 用户相比（M = 3.04， SD = 1.56），iOS 用户对含人工智能功能的应用程序给出了显著更高的评分（M = 3.52， SD = 1.51），t(298,651)= 84.32， p < .001， d = 0.31。人工智能应用与对照应用之间的平台差异幅度明显更大（分别为 Δ = 0.48 和 Δ = 0.31）。检验平台与人工智能状态交互作用的双因素方差分析显示，交互效应显著， F(1, 1,484,629) = 234.78， p < .001， η^2 = .012，证实平台会调节人工智能功能与用户满意度之间的关系。在控制应用类别、评论日期和评论者特征后，这种平台效应仍然显著。

使用单因素方差分析进行的类别层面分析显示，不同应用类型的人工智能接受度存在显著异质性，F(5, 99,994) = 412.67, p < .001, η^2 = .020。助手类应用主要包括人工智能驱动的聊天机器人和虚拟助手，其人工智能效应最为积极（d = 0.55）；含人工智能功能的助手所获评分显著高于非人工智能替代品（M = 3.61, SD = 1.50 vs. M = 2.68, SD = 1.50）， t(24,290) = 47.82, p < .001。创意类应用表现出第二强的正向效应（d = 0.47），用户尤其看重人工智能功能在图像编辑、音乐生成和艺术创作方面的作用。含人工智能功能的创意应用（M = 3.90, SD = 1.50）表现显著优于非人工智能创意应用（M = 3.13, SD = 1.50）， t(79,954) = 33.91, p < .001。

相反，娱乐类应用呈现出显著的负向人工智能效应（d = −0.23），人工智能功能降低了用户的愉悦感。含人工智能功能的娱乐应用（M = 3.42, SD = 1.50）获得的评分显著低于

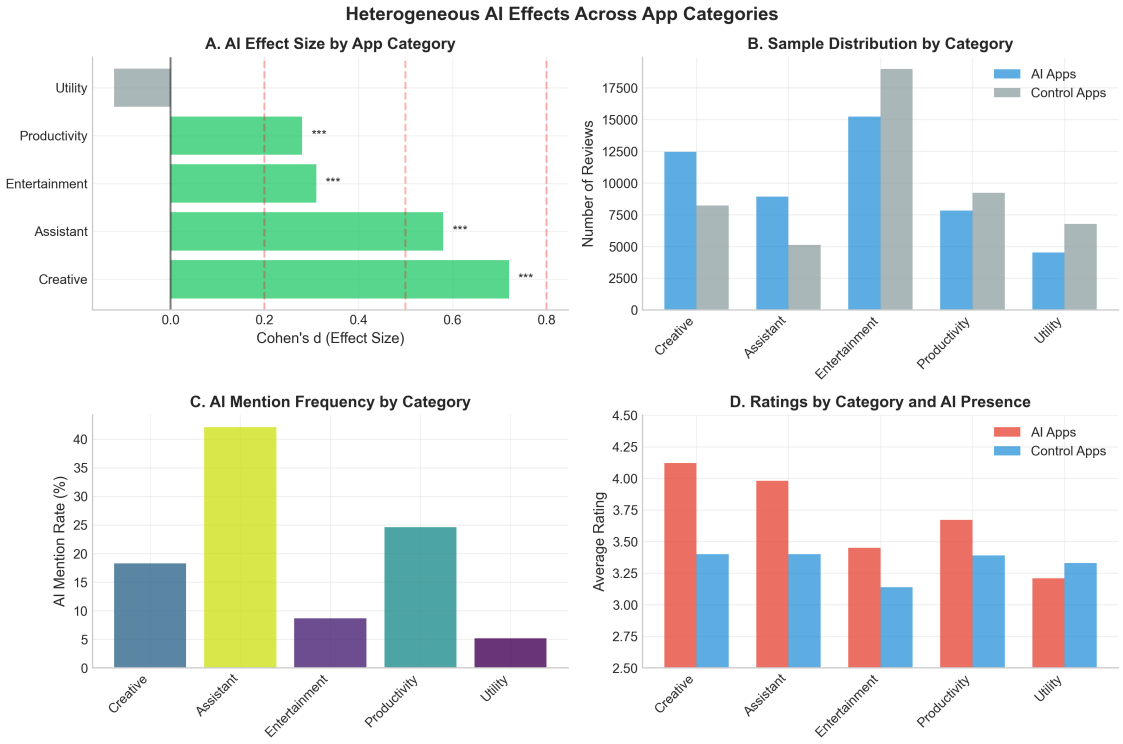


图 4：不同应用类别的异质性人工智能效应

表 8：不同应用类别的异质性人工智能效应

类别	数量（人工智能）	数量（非人工智能）	人工智能 M (SD)	非人工智能 M (SD)	d	p
助手	15,207	9,085	3.61 (1.50)	2.68 (1.50)	0.55	<.001
创意	46,473	33,483	3.90 (1.50)	3.13 (1.50)	0.47	<.001
其他	8,829	19,553	3.60 (1.50)	3.24（1.50）	0.22	<.001
效率	6,566	18,618	3.74（1.50）	3.57 (1.50)	0.10	<.001
效用	267	11,179	3.59 (1.50)	3.64 (1.50)	−0.03	.630
娱乐	22,658	8,082	3.42 (1.50)	3.79 (1.50)	−0.23	<.001

注释。 M = 均值；SD = 标准差；d =Cohen’s d 效应量。

非 AI 娱乐应用（$M = 3.79$, $SD = 1.50$）， $t(30{,}738) = -11.24$, $p < .001$。实用工具类应用几乎没有人工智能效应（$d = -0.03$），人工智能实用工具（$M = 3.59$, $SD = 1.50$）与非人工智能实用工具（$M = 3.64$, $SD = 1.50$）之间没有显著差异， $t(11{,}444) = -0.48$, $p = .630$，这表明用户使用实用工具时更看重基本功能，而非智能化。表 8 展示了完整的类别比较。图 4 通过四个面板分析展示了这些异质性效应：（A）各应用类别的 Cohen’s d 效应量；（B）样本分布，按类别和人工智能状态展示评论数量；（C）各类别中提及人工智能的评论比例；以及（D）各类别中含人工智能功能的应用与对照应用的平均评分比较，并标示 95% 置信区间。

效率类应用呈现出复杂且取决于实施方式的模式。任务自动化和智能日程安排功能获得了积极评价，而笔记应用中的 AI 写作建议则引发了褒贬不一的反应。“其他”类别呈现出较小的正向效应（$d = 0.22$， $p < .001$），代表了人工智能集成方式和成效差异较大的各种应用程序。

进一步分析平台、类别和 AI 状态之间的三重交互发现，平台差异在创意类应用中最为显著（iOS 优势 $\Delta = 0.72$），在实用工具类应用中最不显著（iOS 优势 $\Delta = 0.14$）。这表明，平台特定的实现质量因应用类型而异， $F(5, 1{,}484{,}617) = 89.34$, $p < .001$, $\eta^2 = .003$。

这些发现直接回答了研究问题3，表明不同平台和应用类别的人工智能接受度存在显著异质性。平台分析显示，iOS 用户对人工智能功能的接受度显著更高（$d = 0.31$），且人工智能存在会进一步增强这一效应（$\eta^2 = .012$）。类别分析显示，人工智能有效性呈现出一个连续谱：在助手类应用中（$d = 0.55$）显著为正，因为人工智能与其核心功能相契合；而在娱乐类应用中（$d = -0.23$）则显著为负，因为人工智能可能会干扰享乐体验。实用工具类应用中的无显著效应（$p = .630$）表明，对于为特定实际用途设计的工具，用户更看重可靠性，而非智能程度。观察到的异质性（方差分析 $\eta^2 = .020$）进一步表明，人工智能功能是一种情境依赖型创新，其成功取决于是否恰当地应用，而不仅仅是是否存在。

5 讨论

5.1 理论启示

本研究通过指出，大多数用户在评估应用程序时，并不会明确注意到其人工智能功能，从而推进了消费者人机交互理论。尽管人工智能已得到广泛部署，但只有少数评论提及人工智能，这揭示了实施与表述之间存在认知差距。当人工智能以不可见的方式运行时，假设用户会进行有意识评估的接受理论无法解释用户的认知；要解释用户评价，就必须考虑低于意识层面的运行。

经典接受模型（以感知有用性和易用性为核心）假设用户会有意识地评估功能[51, 46]。我们的证据表明，对于嵌入式人工智能，许多用户可能永远不会进行明确评估，因为他们没有意识到某项功能属于人工智能。因此，我们提出将无意识采纳作为一个可检验的构念：用户可以在没有明确认知或有意采纳的情况下从人工智能中受益（或受到影响）。按照这一修订，认知与有用性和易用性并列，成为接受的近端前因，有助于解释为什么人工智能即使未被提及，也能影响满意度。

其次，我们解释抑制效应。双变量分析显示，人工智能存在与较低的满意度相关；在控制明确的人工智能提及和评论协变量后，这种关联出现反转。这种抑制模式意味着，不满意主要集中在注意到人工智能并对其提出批评的用户中，而不显眼的人工智能则可能带来中性或积极的体验。这一结果拓展了对算法厌恶的解释，表明驱动负面评价的并非“算法”本身，而是显著性与期望违背共同作用 [7]。

信任理论解释了为何隐私担忧主导用户反应。隐私和数据处理对应基于诚信的信任，是用户接受服务的前提 [56], ；而效率提升则对应基于能力的信任和工具性价值。对话式人工智能中的主题模式表明，交互质量存在差异：有些体验让人感觉自然且有帮助，另一些则显得不合拍或令人沮丧。需要通过实验操纵接近人类的措辞与明显机器化的措辞，以检验发生错误时，接近人类的文本是否会加剧负面反应。同样，特定更新时段前后平台集成投诉的增加，应被视为时间上的描述性现象，而非因果效应，除非采用事件研究设计。

尽管解释的方差较小（这在多因素结果中很常见），但理论信号依然清晰：在各项分析中，效应方向和效应大小均与这样一种解释相吻合，即认知、信任（正直与能力）和类别契合度共同影响人们的接受程度。

5.2 实践意义

对于应用开发者和科技公司而言，我们的发现提供了若干切实可行的见解，有助于改善人工智能集成和用户满意度。人工智能实施与用户认知之间的脱节表明，目前传达人工智能能力的方式大多收效甚微。此外，我们记录到的日益加剧的两极分化和不断上升的隐私担忧表明，用户对人工智能的意见正变得更加分歧，而非趋于共识。这一时间变化模式说明，建立积极人工智能认知的窗口期可能正在缩短，因此，及早、透明地传达人工智能功能变得愈发重要。

人工智能在不同应用类别中的接受度存在异质性，这为制定人工智能实施策略提供了明确指引。人工智能功能在助手和创意类应用中似乎最有益，用户看重能够提升效率和创意表达的智能能力。相反，娱乐类应用对人工智能集成表现出负面反应，而实用工具类应用则未显示出明显影响，这表明在实用工具中，用户更看重基本功能，而非智能能力。这些模式表明，人工智能实施应针对不同类别的用户需求进行审慎定制，而不应一概而论。观察到的效应量较小，这说明人工智能功能是在补充而非彻底改变移动应用的用户体验。

明确具体的用户担忧，为消除人工智能接受度的障碍提供了行动指南。隐私是影响用户信任的首要障碍，这表明开发者应优先采用透明的数据处理实践，在可能的情况下实施设备端处理等隐私保护技术，并向用户清楚说明这些措施。准确性方面的顾虑也十分突出，说明管理用户对人工智能能力及其局限性的预期至关重要。开发者与其夸大人工智能的能力，不如清楚说明人工智能功能能做什么、不能做什么，从而有望减少用户的失望和沮丧。

我们观察到的平台差异表明，优化策略应因平台而异。iOS 应用程序或许能从更醒目的人工智能功能中受益，而 Android 应用程序则需要谨慎引入人工智能功能，并强调其实用益处。这些差异可能反映了人口统计差异、技术能力和文化规范，开发者在设计特定平台的实现方案时应予以考虑。

5.3 局限性与未来研究

应当承认本研究存在若干局限。首先，横断面设计使我们无法对人工智能功能与用户满意度之间的关系作出因果推断。我们无法确定是人工智能功能降低了满意度，还是质量较低的应用程序更有可能将人工智能作为差异化手段。尽管我们的研究发现对替代设定和聚类标准误具有稳健性，但数据的观察性限制了解释空间。未来若采用纵向设计或自然实验，考察引入人工智能功能前后的用户反应，将能提供更有力的因果证据。

其次，我们的模型仅能解释用户评分中有限的方差（$R^2 = .084$）。尽管这对于多因素结果而言较为常见，但这表明人工智能功能只是影响满意度的一个组成部分，界面设计、功能、定价和个人偏好也同样起作用。这进一步印证了我们的观点：人工智能是对用户体验的补充，而非主导因素。

第三，我们依赖数字市场上公开的用户评论，这可能会引入选择偏差。撰写评论的用户可能更倾向于表达极度满意或不满， [57], 从而可能放大对人工智能功能的反应。尽管我们的基于关键词的人工智能检测方法较为全面，但仍可能遗漏细微提及，或在人工智能是在无关语境下被提及时产生假阳性。未来研究可以采用先进的自然语言处理技术，更好地捕捉隐性人工智能提及和用户认知模式。
第四，我们的二元人工智能分类掩盖了技术复杂程度和实施质量方面的显著差异。尽管用户体验可能有所不同，我们仍将采用简单推荐算法的应用与由先进语言模型驱动的应用同等对待。未来研究应构建更细致的分类体系，区分人工智能类型、突出程度和实施质量，以确定用户重视或排斥哪些具体功能。

最后，我们的分析仅限于来自美国市场的英语评论。鉴于已有研究记录了技术接受和隐私态度方面的文化差异， [58], 这些模式可能无法推广至其他国家和地区。跨文化研究将有助于阐明我们发现的担忧和益处究竟是普遍存在的，还是具有文化特殊性。

5.4 未来研究方向

我们的研究结果揭示了三个研究方向。首先，应通过定性研究探讨我们记录到的认知差距。访谈或日记研究可以确定，用户是真的没有意识到人工智能功能，选择不将其视为人工智能，还是认为人工智能显著性与他们的评价无关。理解这些认知与社会过程，将有助于推进技术认知的理论模型，并为实践中的传播策略提供参考。

其次，应通过实验研究检验人工智能框架如何影响用户评价。比较用户对标注为“智能”“人工智能驱动”或未加标注的相同功能所作出的反应，将有助于厘清负面反应究竟源于人工智能技术本身，还是源于对它的框定。纵向研究可以追踪我们观察到的两极分化会随着时间推移而趋于稳定、逆转还是加剧，并识别导致态度转变的关键事件。此类研究将检验我们提出的“无意识采纳”这一构念是否适用于不同情境和时间跨度。
第三，未来研究应在更广泛的技术生态系统中考察人工智能。随着应用程序日益将人工智能与增强现实、物联网或区块链技术相结合，理解用户如何评估这些技术组合变得至关重要。这将使我们的框架超越孤立的人工智能功能，延伸至复杂的社会技术系统。

6 结论

对422款移动应用中148万条评论的分析揭示了人工智能实施与用户认知之间存在根本性脱节：尽管近一半的应用程序具备人工智能功能，但提及人工智能的评论仅占11.9%。这一认知差距对假设用户会有意识地进行功能评估的传统技术接受模型提出了挑战，转而表明人工智能往往在识别阈值之下运行。

我们的研究发现了一种抑制效应：具有人工智能功能的应用程序最初评分较低，但在控制明确的人工智能提及和评论特征后，这种关系发生了逆转。这一模式表明，负面评价主要来自那些识别并批评人工智能的用户，而非人工智能本身的存在。隐私担忧是负面回应的主要原因，而效率提升则推动了正面评价；不同应用类别之间存在显著异质性，从助手类应用中的正面效应到娱乐类应用中的负面效应不等。

这些发现通过引入“无意识采纳”推进了人机交互理论：用户在没有明确认知或评估的情况下使用人工智能功能。这一概念要求理论框架将人工智能往往不可见的运行方式纳入考量，从而超越假设用户会有意识地评估技术的传统模型。对于从业者而言，我们的研究结果表明，成功的人工智能集成取决于因情境而异的实施方式，而非普遍适用的方案；尤其需要在提供切实效率效益的同时，重视解决隐私担忧。

随着人工智能在数字平台上日益普及，理解技术能力与用户感知之间的差距变得至关重要。实现有效的人机交互，不仅需要先进的技术，还需要密切关注用户认知、信任建立和情境需求。证据表明，人工智能在消费类应用中的未来，与其说取决于算法进步本身，不如说更取决于弥合开发者所构建的内容与用户在日常数字体验中所感知、理解和重视的内容之间的鸿沟。

致谢

本文已提交至 Computational Communication Research，目前正在接受同行评审。经合理请求后，数据和分析代码可向通讯作者索取。

参考文献

[1] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). 人机交互指南。*2019*年*CHI*计算系统中的人为因素会议论文集, 1–13.[2] Burrell, J., Kahn, Z., Jonas, A., & Griffin, D. (2023). 用户如何掌控算法：Twitter实践中体现的价值观。*ACM*人机交互论文集, 7(CSCW1), 1–20.[3] Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.-W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., Roepman, R., Dietmann, S., Virta, M., Kengne, A. P., Venditti, C., Wang, H., & Bilkey, D. (2021). 人工智能：科学研究的强大范式。*The Innovation*, 2(4), 100179.[4] Cai, C. J., Jongejan, J., & Holbrook, J. (2019). 基于示例的解释对机器学习界面的影响。第*24*届智能用户界面国际会议论文集, 258–262.[5] Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). 重新审视人机交互在设计上是否、为何以及如何具有独特的难度。*2020*年*CHI*计算系统中的人为因素会议论文集, 1–13.[6] Brynjolfsson, E., & McAfee, A. (2017). 人工智能商业。《哈佛商业评论》, 95(7), 3–11.[7] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). 算法厌恶：人们看到算法出错后，会错误地避而不用。《实验心理学杂志：综合》, 144(1), 114–126.[8] Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). 对医疗人工智能的抵制。《消费者研究杂志》, 46(4), 629–650.[9] Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). 你会接受不完美的人工智能吗？探索调整终端用户对人工智能系统预期的设计方案。*2019*年*CHI*计算系统中的人为因素会议论文集, 1–14.

[10] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., Le Meunier-FitzHugh, K., Le Meunier-FitzHugh, L. C., Misra, S., Mogaji, E., Sharma, S. K., Singh, J. B., Raghavan, V., Raman, R., Rana, N. P., Samothrakis, S., Spencer, J., Tamilmani, K., Tubadji, A., Walton, P., & Williams, M. D. (2021). 人工智能（AI）：对新兴挑战、机遇以及研究、实践和政策议程的多学科视角。《国际信息管理杂志》, 57, 101994.[11] Rai, A., Constantinides, P., & Sarker, S. (2019). 下一代数字平台：迈向人类—人工智能混合体。*MISQuarterly*, 43(1), iii–ix.[12] Huang, M.-H., & Rust, R. T. (2021). 营销中的人工智能战略框架。《营销科学学会期刊》, 49(1), 30–50.[13] Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). 服务机器人崛起：人形机器人如何影响服务体验并引发补偿性消费者反应。《营销研究杂志》, 56(4), 535–556.[14] Harman, M., Jia, Y., & Zhang, Y. (2012). 应用商店挖掘与分析：面向应用商店的软件仓库挖掘。*2012*年第*9*届*IEEE*软件仓库挖掘工作会议, 108–111.[15] Pagano, D., & Maalej, W. (2013). 应用商店中的用户反馈：一项实证研究。*2013*年第*21*届*IEEE*需求工程国际会议, 125–134.[16] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). 语言模型是少样本学习器。《神经信息处理系统进展》, 33, 1877–1901.[17] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). 通用人工智能的曙光：GPT-4的早期实验。*arXiv*预印本 *arXiv:2303.12712*.[18] Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). 我认为这是最具颠覆性的技术：利用Twitter数据探索ChatGPT早期采用者的情感倾向。*arXiv*预印本 *arXiv:2212.05856*.[19] Kaplan, A., & Haenlein, M. (2019). Siri，Siri，我手中的Siri：谁是世上最美的？论人工智能的解读、呈现与影响。商业视野, 62(1), 15–25.[20] Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., & Bandeira, N. (2017). 我能为你做些什么？智能个人助理的低频用户体验。第*19*届移动设备与服务人机交互国际会议论文集, 1–12.[21] LeCun, Y., Bengio, Y., & Hinton, G. (2015). 深度学习。《自然》, 521(7553), 436–444.[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). 注意力就是你所需要的一切。《神经信息处理系统进展》, 30.[23] Gharbi, M., Chen, J., Barron, J. T., Hasinoff, S. W., & Durand, F. (2017). 用于实时图像增强的深度双边学习。《*ACM*图形学汇刊》, 36(4), 1–12.[24] Cave, S., Craig, C., Dihal, K., Dillon, S., Montgomery, J., Singler, B., & Taylor, L. (2019). 人工智能的呈现与认知及其重要性。英国皇家学会。[25] Glikson, E., & Woolley, A. W. (2020). 人类对人工智能的信任：实证研究综述。《管理学会年鉴》, 14(2), 627–660.[26] Parasuraman, R., & Manzey, D. H. (2010). 人类使用自动化时的自满与偏见：注意整合。*Human Factors*, 52(3), 381–410.[27] Sundar, S. S., Kim, J., Beth-Oliver, M., & Molina, M. D. (2020). 可预测信息披露的在线隐私启发式。*2020*年*CHI*计算系统中的人为因素会议论文集, 1–12.[28] Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). 消费者对在服务交付中使用人工智能（AI）设备的接受度。《国际信息管理杂志》, 49, 157–169.[29] Hoff, K. A., & Bashir, M. (2015). 自动化信任：整合影响信任因素的实证证据。*HumanFactors*, 57(3), 407–434.

[30] Jacovi, A., Marasovi´c, A., Miller, T., & Goldberg, Y.（2021年）。人工智能信任的形式化：人类信任人工智能的前提、原因与目标。《2021年ACM公平、问责与透明度会议论文集》，624–635。[31] Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E.（2019年）。超越准确性：心智模型在人机团队绩效中的作用。《AAAI人类计算与众包会议论文集》，7，2–11。[32] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F.（2020 年）。可解释人工智能（XAI）：迈向负责任人工智能的概念、分类体系、机遇与挑战。《信息融合》，58，82–115。[33] Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z.（2019年）。XAI——可解释人工智能。《科学机器人》，4(37)，eaay7120。[34] Kizilcec, R. F.（2016年）。信息量多大才合适？透明度对算法界面信任的影响。《2016年CHI计算系统中的人为因素会议论文集》，2390–2395。[35] Acquisti, A., Brandimarte, L., & Loewenstein, G.（2015年）。信息时代的隐私与人类行为。《科学》，347(6221)，509–514。[36] Martin, K., & Murphy, P.（2017年）。数据隐私在营销中的作用。《营销科学学会期刊》，45(2)，135–155。[37] Barocas, S., & Nissenbaum, H.（2014年）。大数据绕过程序性隐私保护的“捷径”。《美国计算机学会通讯》，57(11)，31–33。[38] Dinev, T., & Hart, P.（2006年）。电子商务交易的扩展隐私计算模型。《信息系统研究》，17(1)，61–80。[39] Smith, H. J., Dinev, T., & Xu, H.（2011年）。信息隐私研究：跨学科综述。MIS Quarterly，35(4)，989–1015。[40] Kokolakis, S.（2017年）。隐私态度与隐私行为：关于隐私悖论现象当前研究的综述。《计算机与安全》，64，122–134。[41] Wachter, S., & Mittelstadt, B.（2019年）。合理推断权：大数据与人工智能时代对数据保护法的再思考。《哥伦比亚商业法律评论》，2019(2)，494–620。[42] Tene, O., & Polonetsky, J.（2013年）。全民大数据：分析时代的隐私与用户控制。《西北大学技术与知识产权期刊》，11(5)，239–273。[43] Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G.（2011年）。伴着《愤怒的小鸟》、Facebook和Kindle入睡：关于移动应用使用的大规模研究。《第13届国际人机交互与移动设备及服务会议论文集》，47–56。[44] Petsas, T., Papadogiannakis, A., Polychronakis, M., Markatos, E. P., & Karagiannis, T.（2013年）。应用程序星球的崛起：移动应用生态系统的系统性研究。《2013年互联网测量会议论文集》，277–290。[45] Statista。（2025年）。2009年1月至2025年1月全球移动操作系统市场份额。取自 https://www.statista.com/statistics/272698/[46] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D.（2003年）。信息技术的用户接受度：迈向统一观点。MIS Quarterly，27(3)，425–478。[47] Chamberlain, A., Crabtree, A., Rodden, T., Jones, M., & Rogers, Y.（2023年）。实地研究：理解设计与开发中的“实地”方法。《交互系统设计》，795–796。[48] Epstein, Z., Hertzmann, A., Herman, L., Mahari, R., Frank, M. R., Groh, M., Schroeder, H., Smith, A., Akten, M., Fjeld, J., Farid, H., Leach, N., Pentland, A., & Russakovsky, O.（2023年）。生成式人工智能的艺术与科学：深入探究。《科学》，380(6650)，1110–1111。[49] Jussupow, E., Spohrer, K., & Heinzl, A.（2021年）。增强医疗诊断决策？人工智能参与医生决策过程的研究。《信息系统研究》，32(3)，713–735。[50] Davis, F. D.（1989年）。感知有用性、感知易用性与信息技术的用户接受度。MIS Quarterly，13(3)，319–340。[51] Venkatesh, V., & Davis, F. D.（2000年）。技术接受模型的理论扩展：四项纵向实地研究。《管理科学》，46(2)，186–204。[52] McKnight, D. H., Choudhury, V., & Kacmar, C.（2002年）。为电子商务开发并验证信任测量：一种整合型分类。《信息系统研究》，13(3)，334–359。

[53] Lee, J. D., & See, K. A.（2004年）。自动化信任：为适当依赖进行设计。*Human Factors*，46(1)，50–80。[54] Culnan, M. J., & Armstrong, P. K.（1999年）。信息隐私顾虑、程序公平与非人际信任：一项实证研究。*Organization Science*，10(1)，104–115。[55] Xu, H., Luo, X. R., Carroll, J. M., & Rosson, M. B.（2011年）。个性化隐私悖论：基于位置的营销决策过程探索性研究。*Decision Support Systems*，51(1)，42–52。[56] McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F.（2011年）。对特定技术的信任：其组成部分与测量方法研究。*ACM管理信息系统汇刊*，2(2)，1–25。[57] Hu, N., Pavlou, P. A., & Zhang, J.（2017年）。在线产品评论中的自我选择偏差。*MIS Quarterly*，41(2)，449–471。[58] Hofstede, G., Hofstede, G. J., & Minkov, M.（2010年）。文化与组织：心智软件 （第3版）。麦格劳-希尔。