

• 专题:双清论坛“大模型时代数智商务的理论与方法” •

DOI: 10.3724/BNSFC-2025.06.20.0002

商务人工智能的伦理挑战: 一个基于利益相关者视角的理论框架*

张紫琼¹ 苏欣¹ 王雪妍¹ 邓朝华^{2**}

1. 哈尔滨工业大学 经济与管理学院, 哈尔滨 150001

2. 华中科技大学 管理学院, 武汉 430074

[摘要] 人工智能(Artificial Intelligence, AI)在推动商业模式创新与优化消费者体验的同时,也引发了深刻的伦理挑战,其复杂性与系统性亟需理论突破。现有研究多局限于单一场景,缺乏针对商务AI伦理问题的系统性理论框架。本文基于Web of Science核心合集与CNKI数据库,剖析了161篇英文文献和156篇中文文献。首先,界定了商务AI伦理的基本概念,并提炼出五项核心议题。其次,从消费者、组织与社会三重利益相关者视角,分析了AI伦理问题的影响因素、内在机制、影响结果和边界条件,构建了商务AI伦理的理论研究框架。最后,提出了该领域未来研究方向。本文所构建的框架不仅为深入理解商务AI伦理提供了系统的理论支撑,也为企业合规部署AI与政策制定者完善伦理治理体系提供了实践启示。

[关键词] 商务AI; AI伦理; 利益相关者视角; 理论框架; 技术包容性与可控性

人工智能(Artificial Intelligence, AI)的发展正引领人类生产生活方式的变革。在商务领域, AI技术被广泛应用于多个业务环节, 以提升管理效率、创新商业模式、改善消费者体验。生成式AI、大语言模型等新型技术的出现, 不仅推动了AI能力的跃升, 也极大拓宽了AI在商务背景下的应用场景, 促使企业在数字化转型中不断挖掘AI的潜在价值。斯坦福大学以人为本人工智能研究所(Stanford HAI)发布的AI指数报告显示, 在2024年, 全球有78%的组织使用了AI技术, 且对生成式AI的总投资达到了339亿美元^[1]。2024年一项对全球近1300家企业的调查进一步显示, 已有59%的企业职能部门正在进行或已经完成了对AI项目的部署, 有34%的部门正计划实施AI解决方案, 并有51%的企业计划创建专属于自己企业的大语言模型^[2]。由此可见, 在商务活动中进行AI部署和应用已经成为一种必然的发展趋势。

然而, AI技术的广泛使用也伴随着数据隐私泄露、

技术透明度缺失、算法偏见、虚假信息传播、技术信任危机与道德责任缺位等潜在的伦理问题, 引发社会广泛关注。这些问题不仅损害了企业声誉和公众信任, 也对用户权益与社会福祉构成威胁。例如, 在电商购物平台和在线叫车服务等场景中, 存在“大数据杀熟”现象, AI系统根据建立的用户画像实行动态定价, 对老用户反而报价更高; 亚马逊一款基于AI的面部识别软件存在性别歧视问题, 该软件在识别肤色较深的女性时, 有一定概率会将其错误地标记为男性。随着此类伦理风险的不断显现, 企业对构建AI伦理框架的重视程度不断提升。全球调查显示, 81%的受访企业领导者支持制定更加“全球化”的AI使用标准, 以指导业界在伦理道德范畴内负责任地使用AI技术^[2]。

我国在积极推动AI发展的同时, 也高度重视对AI的规范、监管与治理。2021年和2023年, 国家先后发布了《新一代人工智能伦理规范》和《生成式人工智能服务

收稿日期: 2025-06-20; 修回日期: 2025-10-17

* 本文根据国家自然科学基金委员会第406期“双清论坛”讨论的内容整理。

** 通信作者, Email: zh-deng@hust.edu.cn

本文受到国家自然科学基金项目(72131005, 72271102)的资助。

引用格式: 张紫琼, 苏欣, 王雪妍, 等. 商务人工智能的伦理挑战: 一个基于利益相关者视角的理论框架. 中国科学基金, 2025, 39(5): 748-760.

Zhang ZQ, Su X, Wang XY, et al. Ethical challenges of business artificial intelligence: A theoretical framework based on stakeholder perspective. Bulletin of National Natural Science Foundation of China, 2025, 39(5): 748-760. (in Chinese)

管理暂行办法》,提出了对AI尤其是生成式AI全生命周期活动中的伦理规范要求,覆盖公平公正、隐私安全、偏见歧视、用户权益、可控可信、公开透明、责任担当等方面。规范文件明确了AI研发与应用中的伦理底线,有利于提高全社会的AI伦理意识,推动AI技术的健康发展。

尽管AI伦理问题已逐渐引起各界关注,但现有综述多集中于特定行业或学科。例如,在招聘与选拔领域,已有研究系统梳理了AI应用的伦理机会、风险及缓解方法^[3];在医疗教育与医疗实践中,也存在针对AI伦理教学或医疗AI应用的范围综述,分析了伦理风险、信任问题、法律挑战及未来研究方向^[4,5]。此外,部分研究通过内容分析或知识图谱的方法,对图书情报学科及整体AI伦理领域的文献进行了系统分析,总结了研究主题、知识结构与发展趋势^[6-8]。总体来看,这些综述主要关注行业应用或学科视角,分析框架侧重于风险、规范、方法或主题聚类。而商务环境作为AI技术深度应用的重要场景,AI伦理问题虽已有研究,但相关文献仍较为分散,缺乏系统化的理论框架。

为填补这一空白,本文通过检索Web of Science核心合集和中国知网数据库,梳理并剖析了161篇英文文献和156篇中文文献,从消费者、组织与社会三个层面的利益相关者视角出发,分析AI伦理相关问题的影响因素、内在机制、影响结果和边界条件,构建面向商务环境的多层理论分析框架,并提出未来可能的研究方向。本文可以为在商务领域进行AI伦理研究的学者提供相应的理论支撑,并为企业部署道德、安全、合规的AI应用以及政策制定者进一步完善AI伦理规范和治理提供实践启示。

1 研究设计

1.1 文献检索

为识别与商务环境下的AI伦理问题相关的学术研究,本文进行了系统性的文献检索,结合关键词逻辑分类与人工筛查,对中英文文献进行了检索与分析。在关键词设置上,本文从三个维度进行构建:第一,针对AI技术主体,涵盖“artificial intelligence”“machine learning”“deep learning”等常规术语^[9,10],以及近年来兴起的“Generative AI(GAI)”“Explainable AI(XAI)”“Large Language Model(LLM)”等新型技术^[11-13];同时考虑到AI在实际应用中的表现形式,纳入了“robot”“chatbot”“intelligent agent”“digital human”等具体实体^[14-16]。第二,围绕AI的伦理问题,选取了“ethic”“moral”“awareness”“privacy”“transparency”“bias”“fairness”“trust”“disinformation”等高频概念,以覆盖当前学术界对AI伦理风险的核心关注点^[17-20]。第三,结

合研究所处的商务背景,加入“business”“marketing”“consumer”“e-commerce”“service”等关键词,以确保检索内容与特定应用场景紧密相关。

基于上述关键词,英文文献检索基于Web of Science核心合集,限定引文来源为SSCI,语种为英文,文献类型为论文或综述;中文文献检索则基于中国知网(CNKI)数据库,限定引文来源为CSSCI。通过对检索出的所有文献的标题、摘要和关键词进行人工筛查,剔除掉了与本文主题关联度较低的文献后,保留了161篇英文文献和156篇中文文献作为本文最终用于分析的样本。

1.2 研究样本分析

1.2.1 发文趋势

为了系统考察商务环境下的AI伦理问题的研究发展趋势,本文对现有相关文献的发文情况进行了统计与分析,并绘制了图1以呈现其变化过程(截至2025年4月22日)。图中黑色和灰色的折线和趋势线分别体现了源自Web of Science核心合集(SSCI引文索引)和CNKI数据库(CSSCI引文索引)的相关文献发文数量的变化趋势。整体来看,2018年之前,商务环境下的AI伦理问题相关研究仍然较少,但从2018年开始,国内外该领域研究开始显著增长,并从2023年开始呈现出迅速增长态势。对比来看,从2018年至2020年,虽然发表在国内外期刊上的相关文献数量暂时领先于国际期刊,且增长速度更快,但从2021年开始,国际期刊上的相关文献迅速增多,呈现出更快的发展趋势。由此可见,近年来国内外学者对商务环境下的AI伦理相关问题的关注度日益增加,该议题已逐步演变成为一个重要的研究方向。

1.2.2 研究演进

为分析商务环境下的AI伦理问题的研究演进过程,本文利用CiteSpace软件对英文与中文文献的关键词进行时间线分析,揭示研究焦点随时间的演进规律,分别见图2和图3。由图2可知,国外相关早期研究主要集中于AI技术的应用及用户行为相关问题,包括技术采纳、服务体验和行为机制。随着研究的深入,关注点逐渐扩展到AI的伦理、隐私、算法偏差及信任等社会影响议题。近年来,生成式AI、创意与创新以及技术对社会和组织的潜在影响逐渐成为研究热点。这一演进表明,国外研究主题随着时间呈现出阶段性拓展的趋势,从技术应用与用户行为到伦理与社会影响,再到创意与创新相关议题,研究视角涵盖心理学、计算机科学、管理学及社会学等多个领域。由图3可知,国内研究起步阶段以企业管理和商业实践中的AI应用为核心,主要关注零售、商业模式及企业管理等方面的议题。随着研究的发展,关注点逐渐延伸至企业创新、组织能力建设及产业结构调整

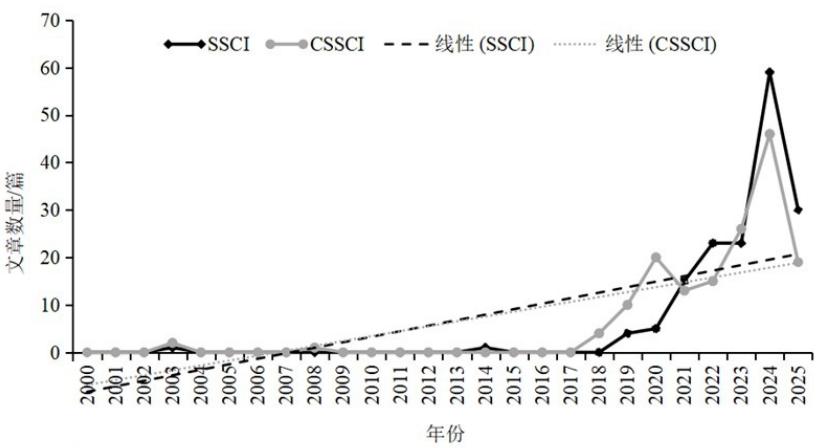


图1 商务环境下的AI伦理问题相关研究国内外发文数量变化趋势
Fig.1 Publication Trends on AI Ethics in Business Contexts: China and Abroad

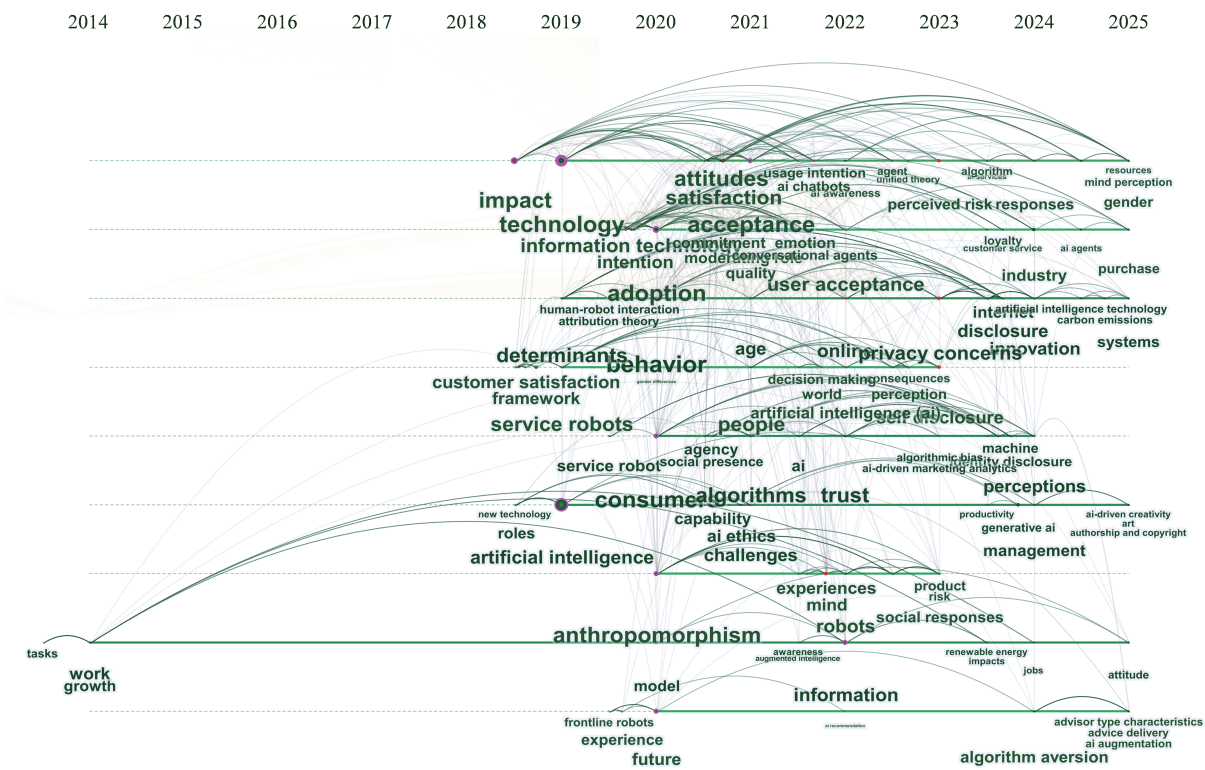


图2 国外关于商务环境下AI伦理问题的研究时间线
Fig.2 Timeline of Research on AI Ethics Issues in Business Contexts Abroad

等问题。近几年,研究进一步聚焦技术与产业融合、人机交互以及商业生态构建等宏观战略问题。整体来看,国内研究主题呈现出由企业实践到创新与组织能力,再到宏观战略与社会发展逐步演进的趋势,体现出研究主题的不断拓展和战略性提升。总体而言,中英文文献均显示AI伦理问题研究具有明显的时间动态性和主题演进规律。国外文献更侧重微观层面的用户行为和伦理风险,而国内文献则强调宏观层面的企业实践、产业升

级与政策导向。

2 AI伦理概念界定与研究进展分析

2.1 商务AI伦理的概念

在商务环境中,AI技术的广泛应用在提升运营效率与用户体验的同时,也带来了日益复杂的伦理挑战。本文在梳理已有文献的基础上,结合商务环境下AI技术的应用特征与场景,界定了商务AI伦理的概念,并总结出

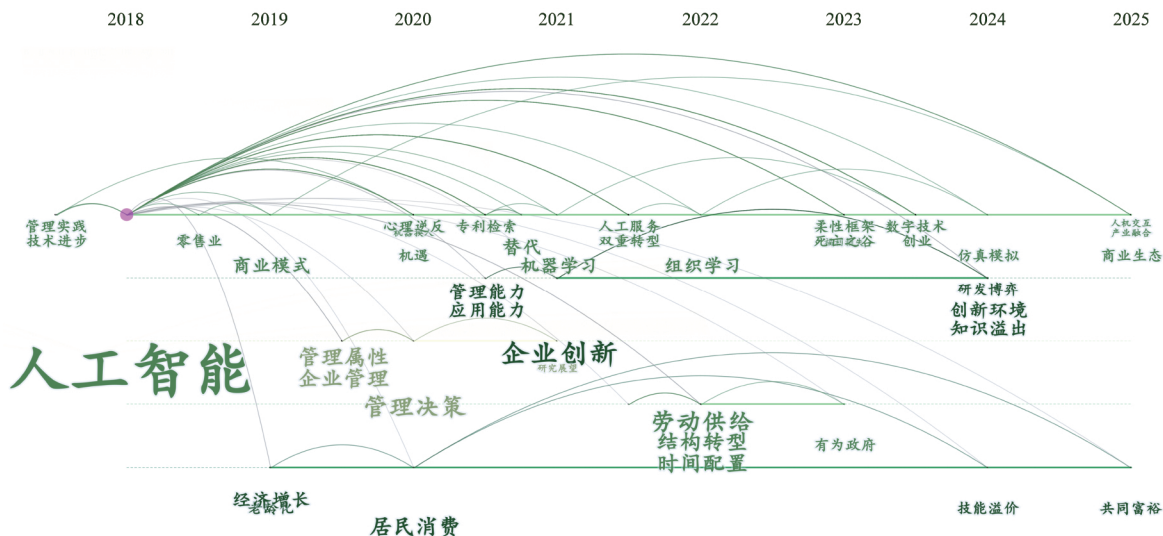


图3 国内关于商务环境下AI伦理问题的研究时间线

Fig.3 Timeline of Research on AI Ethics Issues in Business Contexts in China

其五大核心议题。具体而言,商务AI伦理指贯穿AI系统设计、开发、部署及应用全过程的道德规范体系,旨在规范AI在商业场景中的使用,保障个人、组织与社会的多方利益。其核心议题包括透明性、公平性、隐私性、可问责性与仁慈性,下面将系统阐述这些议题的内涵。

透明性要求企业和技术提供方向用户和监管机构清晰展示AI系统的运行机制、数据来源与决策流程,减少“黑箱”效应,使各方能够理解和监督AI在商业运营中的应用,从而建立信任和保障合规^[21,22]。为提升系统透明性,现有研究主要通过增强模型可解释性来实现。一方面,在模型构建阶段,利用注意力机制等技术手段^[23,24],使模型内部决策过程更加可理解;另一方面,通过事后可解释性方法,对已训练模型的输出进行分析与解释,如LIME^①与SHAP^②算法^[25,26]。研究表明,可解释性增强有助于提高用户对AI系统的信任与接受程度^[27]。在商业实践中,提升系统透明性与可解释性,有助于优化用户体验并促进AI技术的合理应用。

公平性强调AI系统在用户画像和个性化推荐等商业应用中,应有效规避因数据偏差、算法设计或部署策略引发的歧视与不公,确保不同客户群体在服务获取与资源分配中享有平等机会^[28]。它不仅涉及算法层面的去偏与歧视防控,也延伸至社会层面,既可能改善社会资源分配的公平性^[29],也可能带来就业结构和收入分配等方面的差异^[30,31]。近年来,相关研究主要集中于设计去偏算法和制定公平性指标^[28,32-34],在提升效率的

同时,关注不同群体的机会与权益,有助于推动公平原则在实际商业场景中的落实。

隐私性关注用户个人信息在商业智能处理、精准营销等环节中的安全保护,防止数据被未经授权收集、滥用或泄露,努力实现用户隐私权利与数据驱动业务发展的平衡^[21]。研究表明,AI技术的类型及其侧重点不同,会以不同方式影响用户的隐私感知与决策^[35,36]。具体而言,侧重数据收集和处理的AI主要影响用户对信息采集和使用的感知,而侧重输出结果的AI主要影响用户对系统决策或推荐结果的可见性及潜在后果的感知,从而作用于用户的隐私评估与行为选择。这一研究发现为企业在设计商业应用时平衡数据使用与用户隐私提供了有益参考。为应对这一挑战,学界积极探索包括差分隐私、联邦学习、数据沙箱和加密技术等在内的多样化的技术手段^[1,37]。

可问责性强调应明确各参与主体在算法设计、数据处理与决策执行等环节中的责任边界,建立可审计、可追踪的技术与制度机制,以便在出现决策失误或用户权益受损时,能够有效实现责任追究与及时纠正。在商业实践中,这不仅有助于保障用户权益、维护企业信誉^[38,39],也体现了组织在伦理和社会责任方面的承诺,使AI系统能够遵循社会规范与道德预期^[40,41]。相关文献进一步探讨了如何将问责机制融入企业流程中,通过技术和制度的结合提升责任落实的可操作性^[39,42]。

仁慈性体现了AI在商业应用中应秉持人本价值导

① LIME:局部可解释模型无关解释(Local Interpretable Model-Agnostic Explanations),通过在目标样本邻域内生成扰动样本,并基于这些样本按距离加权拟合一个可解释的代理模型,以近似处理复杂模型的局部决策边界,从而生成特征层面的预测解释。

② SHAP:沙普利加性解释(Shapley Additive Explanations),通过计算每个特征在所有可能特征组合中的边际贡献,为模型预测结果分配加性特征重要性值。SHAP 保证预测值与基准值之差可被所有特征的贡献值精确分解,并满足局部准确性、一致性与缺失性等公平性公理。

向,积极促进用户福祉与社会整体利益的提升。其核心在于保障技术进步服务于公共利益,确保商业创新在创造经济价值的同时,也承担相应的社会责任^[42,43]。有研究表明,AI系统在忽视社会福祉时可能对用户和社会造成损害,而将公共福祉纳入商业实践则有助于推动智能向善,促进各方整体利益^[40,42]。

2.2 基于利益相关者视角的理论分析框架

本文借鉴利益相关者理论^[44,45],聚焦消费者、组织与社会三个核心利益相关者群体,分析商务环境中AI应用所面临的伦理挑战。由于各方在AI系统中的角色与利益诉求不同,所面临的伦理问题也存在显著差异^[21]。在个体层面,消费者作为AI产品与服务的直接使用者,主要关注信任机制、隐私保护与算法歧视等问题。在组织层面,企业与员工共同构成该维度的核心主体。企业作为AI系统的部署者和管理者,借助AI重构工作流程、提升效率的同时,也承担保护数据隐私、确保合规运营及维护员工权益等伦理责任;员工作为AI技术的使用者与协作者,则需适应人机协作,并面临技术替代带来的挑战与心理压力。在社会层面,AI的发展引发了社会公平挑战、就业结构重构等系统性问题。因此,本文将从消费者、组织与社会三个维度,系统探讨商务环境下AI所引发的伦理影响。

2.2.1 AI对消费者的伦理影响

通过对相关文献的系统梳理,本文构建了一个AI影响消费者的理论框架(图4)。该框架阐释了AI特征如何激发消费者的认知与情感反应,并进一步影响其经济与非经济层面的反馈。同时,引入了微观与宏观两个层面的边界条件,以刻画不同影响路径下的调节机制。

首先,AI特征可分为外在特征和内在特征两个维度。外在特征指消费者在与AI交互过程中可直接感知

的属性,如拟人化、情感表达与感官属性等。其中,拟人化表现为AI呈现出类人的社会线索的程度,如外观、语言和声音。研究表明,适度的拟人化可以增强消费者对AI的感知温暖,进而提升信任^[46]。然而,过高的拟人化程度可能引发“恐怖谷效应”,使消费者产生怪异感,反而降低接受度^[46,47]。情感表达则体现为AI通过文字、表情符号、图像等言语或非言语方式模拟并传达情绪。已有研究指出,AI表达情感关怀能够减轻顾客的期望违背,有助于提高满意度^[48]。感官属性则代表AI通过视觉、听觉、触觉等多感官渠道提供服务体验,有助于增强消费者的享乐价值^[49],从而增强使用意愿^[50]。

相较之下,内在特征则是反映AI的功能性表现,如专业性、个性化、隐私性、算法偏见与算法决策解释详细程度等。其中,专业性体现为AI在特定领域中展现出的知识水平与问题解决能力^[51],高专业性有助于增强消费者的感知质量,从而提升信任感^[52]。个性化反映AI根据用户需求提供定制服务的能力,高个性化能够提升消费者的感知有效性,进一步增强使用意愿^[53]。隐私性指AI对用户数据安全的保护能力,较高的隐私性可以降低消费者的感知风险,从而提升使用意愿^[54]。此外,算法偏见指的是AI算法在对不同群体(如种族、性别、国籍等)进行处理时,表现出系统性且具有歧视性的差异对待。相关研究表明,此类偏见会引发消费者对结果公平性的担忧,从而削弱其对AI系统的信任^[55]。算法决策解释详细程度指AI系统在向消费者说明其决策过程时所提供的信息深度和细节丰富程度。恰当的详细程度有助于提升消费者的感知理解,进而增强其信任和购买意愿^[56]。

其次,AI对消费者的影响机制主要通过认知反应和情感反应两种路径实现。认知反应是消费者对AI功能

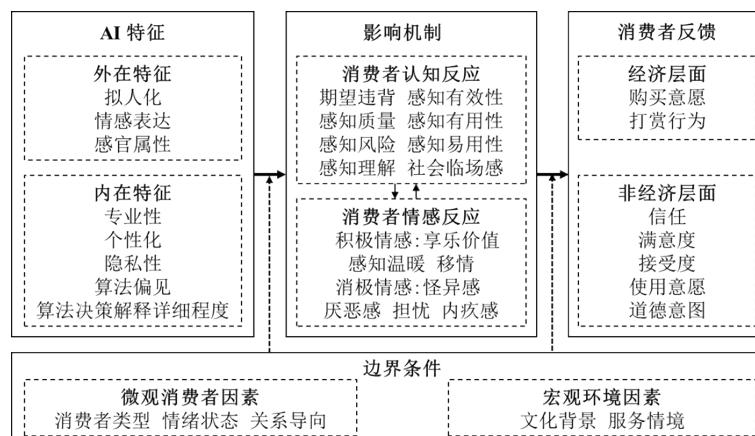


图4 AI对消费者的伦理影响理论框架

Fig.4 Theoretical Framework of the Ethical Impact of AI on Consumers

表现的理性评估过程,涵盖期望违背、感知质量、感知有效性、感知风险、感知理解、感知有用性、感知易用性以及社会临场感等方面。例如,拟人化通过增强熟悉感和社会连接性,有助于提升消费者的感知易用性与感知有用性,从而增强使用意愿^[47]。拟人化还有助于增强社会临场感,从而增强交互的真实性,进一步提升使用意愿^[47]。而情感反应则聚焦于消费者在与AI交互中的情感变化,包括积极情感如享乐价值、感知温暖、移情等,以及消极情感如怪异感、厌恶感、担忧、内疚感等。例如,适度的拟人化与感官交互有助于激发消费者的积极情感,进而提升信任^[46]。但若AI表现出过度拟人化或算法偏见问题,则可能触发消费者的消极情感,进而削弱使用意愿^[57,58]。

需要指出的是,尽管现有研究多将认知反应与情感反应作为独立路径加以考察,但心理学研究明确揭示二者之间存在密切的双向互动关系。一方面,认知评估过程决定个体的情绪反应与情感归因^[59];另一方面,情感状态也会通过改变注意力分配、记忆检索,影响后续的认知判断与行为选择^[60]。因此,本研究将AI特征引发的认知反应与情感反应构建为双向互动的内在机制,二者共同塑造消费者的行为反应。例如,AI的高专业性可能通过提升感知质量而产生信任,信任又会降低感知风险、增强消费者对AI的接受度。这种认知与情感之间的双向互动机制表明,将二者割裂考察的研究视角难以全面把握AI对消费者伦理影响的内在复杂性。

进一步,消费者产生的认知与情感反应最终将转化为一系列经济与非经济层面的反馈。在经济层面,包括购买意愿和打赏等。研究发现,AI对消费者的购买决策具有显著影响。例如,相较于真人主播,消费者较难对AI主播产生移情,因而其购买意愿通常更低^[61]。在非经济层面,包括信任、满意度、接受度、使用意愿、道德意图等。其中,道德意图指的是消费者有意识地倾向于做出符合道德规范的行为的意愿。有研究指出,消费者在与AI互动时,缺乏源自情感关怀的人际互动所引发的内疚感,因而可能降低其道德意图^[62]。

最后,AI对消费者影响的边界主要包括两类因素:消费者因素和环境因素。消费者因素包括消费者类型、情绪状态和关系导向等。在消费者类型方面,性别、年龄、意识形态等个体差异显著影响消费者对AI的态度。研究发现,女性消费者出于对隐私问题的担忧,使用AI的意愿较低^[63];年轻消费者则因对新技术的接受度较高,更愿意尝试使用AI^[54];技术乐观主义者倾向于信任

AI,而卢德派^①则可能表现出较强的抵触态度^[64]。在情绪状态方面,例如愤怒状态下,消费者对AI拟人化的满意度会显著降低^[15]。在关系导向方面,研究表明,对公共导向型消费者,AI的积极情感表达有助于提升满意度,而对交换导向型消费者而言,这种影响不存在甚至相反^[16]。环境因素方面,消费者所处的文化背景、服务情境等外部环境因素,均会显著影响其对AI应用的反馈结果。其中,在文化背景层面,集体主义文化背景(如中国)的消费者更容易信任AI机器人,而个体主义文化背景(如美国)的消费者则相对保守^[46]。在服务情境方面,消费者对功能型服务中的AI的使用意愿更强。例如,机场自助值机系统通过智能排队优化和个性化座位推荐等功能,提高了办理效率,从而促进采纳;而在享乐型服务中,如酒店语音点餐系统、房间智能助手或迎宾机器人,AI的作用主要在于提升客户体验,消费者使用意愿相对有限^[65]。此外,在医疗等高风险情境中,消费者对AI的信任程度也显著低于购物等低风险情境^[66]。

2.2.2 AI对组织的伦理影响

基于AI组织应用的相关文献,本文从员工特征、AI特征和企业特征三个核心维度出发,构建理论框架(图5),讨论AI如何影响员工认知、情感与行为,并作用于企业绩效。同时,从微观、中观与宏观层面探讨了AI应用效果的关键边界条件。

AI企业应用中,员工行为的变化主要由员工特征与AI特征共同作用,通过影响认知反应与情感反应的双向互动机制,最终塑造其行为表现。其中,员工特征中的“AI意识”尤为关键,它指员工对AI、智能机器人等新兴技术对其未来职业发展影响的看法^[67]。根据压力交易理论,员工会对这种感知做出认知评估^[68-71]。若将其视为挑战性压力源,可能激发内在动机,表现出如服务创新等积极行为^[68]。反之,若将其视为阻碍性压力源,则更可能引发厌恶感、恐惧感和工作不安全感等消极情感,导致关键资源耗竭。根据资源保护理论,自我耗竭的员工将采取资源节约行为如抑制服务创新行为、建言减少^[68,69],甚至采取如违反公司规定与离职等消极行为^[70,71]。

AI特征方面,可供性与其承担的功能角色也显著影响员工生产行为。可供性指技术功能属性为员工提供的完成某种任务或目标的可能性^[72]。有学者将AI聊天机器人的可供性划分为个性化、自动化与可用性三个维度。在强监管环境中,这些功能属性受严格的数据存储

①“卢德派”一词源自19世纪初英国的工人运动,代表反对机械化生产对传统手工业冲击的群体。在当代语境中,常用以指代对新兴技术持有强烈不信任或抵触态度的个体或群体。

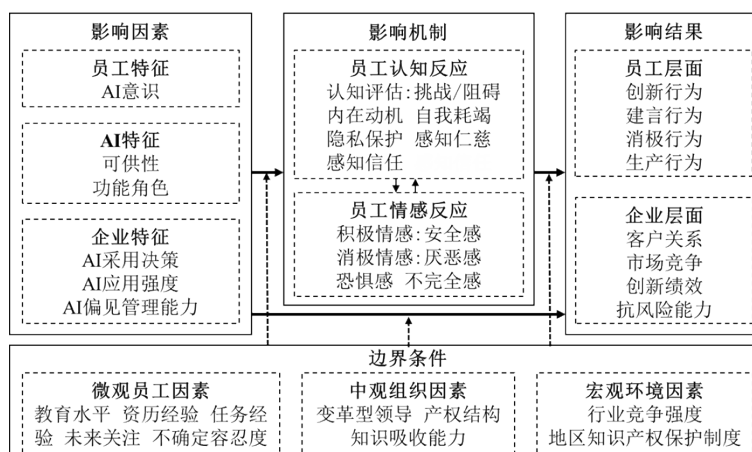


图5 AI对组织的伦理影响理论框架

Fig.5 Theoretical Framework of the Ethical Impact of AI on Organizations

与分析原则限制,会增强员工对隐私保护和数据安全的感知,进而提升其对AI的仁慈与信任感知,最终降低人机协作中的道德顾虑^[72]。此外,AI承担不同功能角色时会对员工绩效产生不同效果。当AI作为辅助者,协助员工处理重复性结构化任务时,员工得以专注于更高层次非结构化工作,提升创造力与工作绩效^[73]。而当AI作为管理者,承担绩效评估与工作反馈功能时,员工可能因高质量的反馈提升绩效,也可能因低AI信任损害绩效^[74]。

企业特征方面,AI采用决策、AI应用强度和AI偏见管理能力均显著影响企业绩效。具体而言,AI采用决策是提升企业绩效的重要前提。在客户关系维度,AI采用通过实现营销个性化显著增强了客户满意度及客户获取能力^[75]。在市场竞争维度,AI能够实现精准客户细分及定向沟通优化,扩大企业市场份额^[76]。在企业创新维度,AI采用通过丰富企业知识库、优化产品开发业务流程以及提升资源配置效率等关键机制,有效促进了企业创新^[77]。AI落地企业后,AI应用强度作为衡量企业AI技术应用深度的特征,是提升企业抗风险能力的关键因素。研究表明,AI应用强度高的企业展现出更强韧性,能在较短的事件窗口内抵消自然灾害导致的大部分估值损失^[78]。然而,随着AI深度应用,其中潜在的数据偏见、模型偏见与部署偏见等算法偏见,可能对客户关系造成损害,并对公平性这一核心伦理议题构成直接挑战。因此,提高企业的AI偏见管理能力,成为将公平性伦理理念转化为具体管理实践的关键路径。通过确保不同客户群体在服务获取与资源分配中享有平等机会,偏见管理能力不仅有效保障了消费者权益,也切实回应了技术伦理对公平价值的关切,进而为企业赢得持续的市场信任、构建长期竞争优势奠定坚实基础^[28]。

尽管AI在多个维度提升企业绩效,但其实际成效依赖于特定的员工特征、组织条件与外部环境因素。员工特征中,较高教育水平或拥有丰富组织层面知识(高资历经验)的员工,因其知识深度和批判性思维,更易识别AI缺陷,可能导致更强的AI厌恶和更低的信任度,从而于AI中获益更少^[79,80]。相比之下,在特定任务上熟练度高、具备丰富实践知识的员工(高任务经验),其掌握的实践常识与AI形成互补,进而更高效地利用AI提升生产力^[80]。员工的心理特质也发挥调节作用:高不确定容忍度和高未来关注的员工更可能视AI为挑战性因素,从而展现出更积极的行为倾向^[68,71]。组织条件中,变革型领导通过强化员工信任感与工作安全保障,有效缓解AI厌恶及其引发的消极行为^[72]。企业的知识吸收能力和产权结构也会影响AI应用效果。高知识吸收能力促进外部技术知识的系统性内化,增强AI技术在企业创新中的效能^[81]。国有产权结构依托其特有的制度性资源调配优势与抗风险能力,支撑AI驱动的企业创新^[81]。外部环境中,行业竞争强度和制度环境会影响AI应用效果。高竞争强度行业的内部知识溢出和竞争对手快速模仿较强,会削弱AI带来的独特竞争优势,降低AI对企业绩效的积极影响^[75]。反之,健全的知识产权保护制度通过提高技术模仿难度和增强研发激励,显著增强了AI对企业创新的促进作用^[81]。

2.2.3 AI对社会层面的伦理影响

AI的社会影响主要从社会治理、收入分配、社会权力三个核心维度重塑社会公平,理论框架见图6。在社会治理层面,政府机构AI应用的影响具有双重性。一方面,AI可以通过改善公共服务供给(如提高服务供给精准性)、加强政府内部管理(如加速审批流程)和辅助政府决策(如基于数据的政策模拟与评估),潜在地改善社

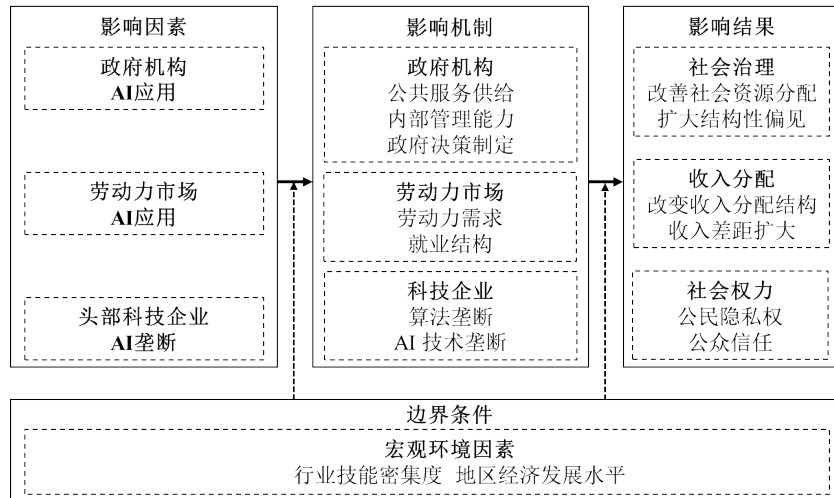


图6 AI对社会的伦理影响理论框架

Fig.6 Theoretical Framework of the Ethical Impact of AI on Society

会资源分配的公平性,惠及更广泛群体^[29]。另一方面,若算法设计、训练数据或实施场景存在偏差,AI可能复制并放大历史性不平等或结构性偏见,导致公共服务对弱势群体的可及性下降,从而削弱AI化公共服务的普惠效果,对平等权利构成新的威胁^[82]。

在收入分配层面,劳动力市场AI应用影响了收入差距^[83]。在就业结构上,创造效应与替代效应并存。一是替代效应减少了劳动力需求,引发了技术性失业风险,如工业机器人替代低技能工人完成重复性工作、生成式AI替代知识工作者进行决策与协作^[84,85]。二是创造效应增加劳动力需求,既通过促进企业规模扩张或新业务开发间接创造就业^[86],也催生如提示工程师等新职业形态^[87]。然而,由于AI对不同技能劳动力需求的替代性存在差异,整体上呈现“就业极化”:高技能与低技能岗位增长,中等技能岗位萎缩^[88,89]。就业结构的变化进一步推动了收入分配结构的改变。随着对与AI技能互补的劳动力需求的增长,收入分配愈发向高技能、高学历群体倾斜,进一步加剧收入不平等^[30,31]。这一趋势在经济发达地区、技能密集型行业中尤为明显^[31,90],进一步增加了收入差距对社会公平治理提出的挑战。

在社会权力层面,头部科技企业对AI技术与数据资源的集中垄断,可能影响公民权利与社会公平^[91]。数据规模、算法能力、专利与资本壁垒使得大型科技企业在市场竞争与技术领先上占据优势,进而强化其经济权力与信息控制力。此外,头部科技企业对平台算法(如信息推荐机制)的设计能够塑造信息流与议程设置,影响公众认知与舆论生态。由于商业目标与公共利益之间存在固有张力,若监管不足或缺乏透明机制,可能损害公民隐私权、削弱问责性并侵蚀公众信任,从而对社

会公平的制度基础和公民权利构成长期风险。

3 未来研究方向

本文基于对现有文献的系统梳理与分析,界定了商务AI伦理的基本概念,总结了其核心议题,并构建了商务环境下AI伦理问题的理论研究框架,进而从AI技术的包容性、可控性和可持续发展方面,探讨了商务环境下AI伦理问题未来的研究方向。

3.1 提高AI技术的包容性

在AI技术快速发展的背景下,坚持“以人为本”的技术开发与设计理念已成为一项关键原则。未来的研究应注重增强对AI的包容性设计和应用,确保技术成果能够惠及不同群体,回应多元社会的需求。首先,未来可以探讨如何在AI系统的设计、开发和部署过程中,确保来自不同背景、能力、性别和文化的人群都能平等地参与其中,并公平地从中受益。研究可以进一步探讨AI系统的包容性指标设计,例如如何衡量AI系统是否公平地为所有群体提供服务。其次,在追求包容性的过程中,可能会以牺牲效益为代价。因此,未来的研究可以聚焦于开发平衡成本与效益的算法,通过定量评估资源投入与社会回报之间的关系,优化决策路径,实现最佳平衡。这个过程中,融合社会学和经济学的理论框架,可为AI的公平性和效益平衡提供更丰富的视角。最后,未来的研究也可以关注AI技术如何赋能弱势群体,推动社会服务的创新。探究如何通过数据的驱动以及智能工具的使用,让AI为边缘人群提供更加精准、高效和可持续的社会服务解决方案。

3.2 增强AI技术的可控性

随着AI技术在生活中的快速部署,如何防范其潜在

风险并引导其向善发展,已成为亟待解决的关键议题。这就需要进一步增强对AI技术的理解,并提高对其全生命周期的可控性。首先,未来可以继续探索AI的可解释性问题,尤其是在AI技术开发与应用中提升系统透明度,使用户与监管者能够更好地理解AI的决策逻辑、处理流程和输出结果。未来研究可聚焦于构建直观的解释模型或设计高效的反馈机制,以增强消费者信任、优化使用体验,进而扭转其对AI的消极态度。其次,在制定决策时不应过度或完全依赖AI,丧失人类的自主判断。因此,未来可以探究如何在AI技术的使用过程中识别出关键的决策环节。例如,研究可以聚焦于开发决策支持系统,帮助辨识哪些环节必须由人类做出最终判断,而哪些环节可以由AI辅助。通过引入“人机协同”的方法,可以设计机制,以确保在关键决策中,AI技术仅作为辅助工具,而不会完全替代人类判断。此外,未来的研究应着力解决AI技术的责任归属问题,结合法律和伦理学的框架,建立有效的责任追踪机制,确保在出现问题时可溯源、可问责。最后,未来还可以探讨如何在制度层面制定灵活的AI监管与伦理规范框架,特别是在技术快速发展的背景下实现动态调整和适应性监管。研究可以关注如何在AI技术生命周期的不同阶段进行差异化监管,以确保符合公共利益和伦理要求。同时,可以探讨如何利用数据分析实时监测AI的社会影响,为政策调整提供依据。

3.3 推动AI技术的可持续发展

在推动AI技术进步的同时,还应关注其在促进人类社会可持续发展中的作用,并在这一过程中有效应对伦理挑战,从而更好地增强人类福祉,促进技术的长久、健康发展。首先,未来研究可以探讨AI技术如何赋能商业模式创新,如AI在循环经济、共享经济等领域中的应用,以推动可持续的经济转型,促进绿色AI的应用和发展。其次,未来可以研究如何借助AI和智能治理系统推动低碳转型,并探索AI在环境保护和社会可持续发展方面的潜力。例如,AI在城市交通调度、能源管理、建筑节能、废物处理等领域应用,可提高城市运行效率,助力绿色智慧城市建设。最后,还可结合中国实践背景,探讨如何运用AI技术解决地区发展不平衡、区域经济发展不平衡等问题。借助AI技术可以精准识别区域发展瓶颈,提供科学的决策依据,从而助力制定更加公平有效的发展政策,推动实现区域协调发展和社会整体进步。

参 考 文 献

[1] Stanford University Human-Centered Artificial Intelligence. The 2025 AI index report. (2025-04-07)/[2025-05-26]. <https://hai.stanford.edu/>

ai-index/2025-ai-index-report.

- [2] Tata Consultancy Services. From potential to performance by design: TCS AI for business study—key findings. (2025-05-26)/[2025-05-26]. <https://www.tcs.com/insights/global-studies/ai-business-study-key-findings-report>.
- [3] Hunkenschroer AL, Luetge C. Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics*, 2022, 178(4): 977—1007.
- [4] Weidener L, Fischer M. Teaching AI ethics in medical education: A scoping review of current literature and practices. *Perspectives on Medical Education*, 2023, 12(1): 399—410.
- [5] Kumar P, Chauhan S, Awasthi LK. Artificial intelligence in healthcare: Review, ethics, trust challenges & future research directions. *Engineering Applications of Artificial Intelligence*, 2023, 120: 105894.
- [6] 黄崑, 徐晓婷, 黎安润泽, 等. 近5年图情SSCI期刊人工智能伦理研究文献分析与启示. *现代情报*, 2021, 41(6): 161—171.
- Huang K, Xu XT, Li ARZ, et al. Study of the literature of artificial intelligence ethics in library and information from 2015—2019. *Journal of Modern Information*, 2021, 41(6): 161—171. (in Chinese)
- [7] 刘鸿宇, 苗芳艳, 彭拾, 等. 人工智能伦理研究的知识图谱分析. *情报杂志*, 2019, 38(7): 85—93.
- Liu HY, Miao FY, Peng S, et al. Artificial intelligence ethics research: A knowledge map analysis. *Journal of Intelligence*, 2019, 38(7): 85—93. (in Chinese)
- [8] Huang CW, Zhang ZQ, Mao BF, et al. An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 2023, 4(4): 799—819.
- [9] Fügner A, Grahl J, Gupta A, et al. Cognitive challenges in human-artificial intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research*, 2021, 33(2): 678—696.
- [10] Adamopoulos P. The spillover effect of fraudulent reviews on product recommendations. *Management Science*, 2024, 70(12): 8818—8832.
- [11] Huang M-H, Rust RT. The caring machine: Feeling AI for customer care. *Journal of Marketing*, 2024, 88(5): 1—23.
- [12] Bauer K, von Zahn M, Hinz O. Expl(AI)ned: The impact of explainable artificial intelligence on users' information processing. *Information Systems Research*, 2023, 34(4): 1582—1602.
- [13] Chen ZN, Chan J. Large language model in creative work: The role of collaboration modality and user expertise. *Management Science*, 2024, 70(12): 9101—9117.
- [14] Mende M, Scott ML, van Doorn J, et al. Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, 2019, 56(4): 535—556.
- [15] Crollic C, Thomaz F, Hadi R, et al. Blame the bot: Anthropomorphism and anger in customer-chatbot interactions. *Journal of Marketing*, 2022, 86(1): 132—148.
- [16] Han E, Yin D, Zhang H. Bots with feelings: Should AI agents express positive emotion in customer service? *Information Systems Research*, 2023, 34(3): 1296—1311.
- [17] Beck BB, Wuyts S, Jap S. Guardians of trust: How review platforms can fight fakery and build consumer trust. *Journal of Marketing Research*, 2024, 61(4): 682—699.

- [18] Bauer K, Gill A. Mirror, mirror on the wall: Algorithmic assessments, transparency, and self-fulfilling prophecies. *Information Systems Research*, 2023, 35(1): 226—248.
- [19] Lee S, Shin D, Kwon KH, et al. Disinformation spillover: Uncovering the ripple effect of bot-assisted fake social engagement on public attention. *MIS Quarterly*, 2024, 48(3): 847—872.
- [20] Lambrecht A, Tucker C. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science*, 2019, 65(7): 2966—2981.
- [21] Telkamp JB, Anderson MH. The implications of diverse human moral foundations for assessing the ethicality of artificial intelligence. *Journal of Business Ethics*, 2022, 178(4): 961—976.
- [22] 李浩, 郭春红. “开卷有益”还是“闭卷藏珍”? 算法不透明性对员工创造力的双刃剑效应. *南开管理评论*, 2025; <https://link.cnki.net/urlid/12.1288.F.20250207.1049.002>.
- Li H, Guo CH. “Open the book for benefits” or “Close the book to hide treasures”? The double-edged sword effect of algorithm opacity on employee creativity. *Nankai Business Review*, 2025; <https://link.cnki.net/urlid/12.1288.F.20250207.1049.002>. (in Chinese)
- [23] Niu ZY, Zhong GQ, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing*, 2021, 452: 48—62.
- [24] 倪渊, 许愿清, 张健, 等. 基于多模态可解释模型的文化艺术品价值评估研究. *系统工程理论与实践*, 2025; <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=XTLL2025091200D&dbname=CJFD&dbcode=CJFQ>.
- Ni Y, Xu YQ, Zhang J, et al. The research on the value evaluation of cultural and artistic works based on multimodal interpretable models. *Systems Engineering-Theory & Practice*, 2025; <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=XTLL2025091200D&dbname=CJFD&dbcode=CJFQ>. (in Chinese)
- [25] Ribeiro M, Singh S, Guestin C. “Why should I trust you?”: Explaining the predictions of any classifier. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. San Diego, California. Stroudsburg, PA, USA: ACL, 2016: 97—101.
- [26] Lundberg SM, Lee SI. A unified approach to interpreting model predictions// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, New York: Curran Associates Inc., 2017: 4768—4777.
- [27] 胡保亮, 王嘉雯, 闫帅. 可解释性对用户接受AI用于知识创造的影响研究. *科研管理*, 2024; <https://link.cnki.net/urlid/11.1567.G3.20241227.1103.003>.
- Hu BL, Wang JW, Yan S. The research on the impact of interpretability on users’ acceptance of AI for knowledge creation. *Science Research Management*, 2024; <https://link.cnki.net/urlid/11.1567.G3.20241227.1103.003>. (in Chinese)
- [28] Akter S, Sultana S, Mariani M, et al. Advancing algorithmic bias management capabilities in AI-driven marketing analytics research. *Industrial Marketing Management*, 2023, 114: 243—261.
- [29] van Noordt C, Misuraca G. Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European union. *Government Information Quarterly*, 2022, 39(3): 101714.
- [30] Lankisch C, Prettnier K, Prskawetz A. How can robots affect wage inequality? *Economic Modelling*, 2019, 81: 161—169.
- [31] 潘馨悦, 王乐, 张紫琼, 等. 数字经济、产业智能化与不同技能劳动力间的收入差距. *系统管理学报*, 2024, 33(2): 472—487.
- Pan XY, Wang L, Zhang ZQ, et al. Digital economy, industrial intelligence, and income gap between different skilled laborers. *Journal of Systems & Management*, 2024, 33(2): 472—487. (in Chinese)
- [32] John-Mathews JM, Cardon D, Balagué C. From reality to world. A critical perspective on AI fairness. *Journal of Business Ethics*, 2022, 178(4): 945—959.
- [33] Mehrabi N, Morstatter F, Saxena N, et al. A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 2022, 54(6): 1—35.
- [34] Pessach D, Shmueli E. A review on fairness in machine learning. *ACM Computing Surveys*, 2023, 55(3): 1—44.
- [35] Martin KD, Zimmermann J. Artificial intelligence and its implications for data privacy. *Current Opinion in Psychology*, 2024, 58: 101829.
- [36] 郝乐. AI人机交互用户个性化推荐中隐私信息披露影响因素研究. *情报理论与实践*, 2024, 47(7): 69—80.
- Hao L. Research on the influencing factors of privacy information disclosure in personalized recommendations of AI human-computer interaction users. *Information Studies: Theory & Application*, 2024, 47(7): 69—80. (in Chinese)
- [37] Meurisch C, Mühlhäuser M. Data protection in AI services. *ACM Computing Surveys*, 2022, 54(2): 1—38.
- [38] Busuioc M. Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 2021, 81(5): 825—836.
- [39] Tóth Z, Caruana R, Gruber T, et al. The dawn of the AI robots: Towards a new framework of AI robot accountability. *Journal of Business Ethics*, 2022, 178(4): 895—916.
- [40] Mikalef P, Conboy K, Lundström JE, et al. Thinking responsibly about responsible AI and ‘the dark side’ of AI. *European Journal of Information Systems*, 2022, 31(3): 257—268.
- [41] Stahl BC. Embedding responsibility in intelligent systems: From AI ethics to responsible AI ecosystems. *Scientific Reports*, 2023, 13: 7586.
- [42] Hermann E. Leveraging artificial intelligence in marketing for social good—An ethical perspective. *Journal of Business Ethics*, 2022, 179(1): 43—61.
- [43] Moon MJ. Searching for inclusive artificial intelligence for social good: Participatory governance and policy recommendations for making AI more inclusive and benign for society. *Public Administration Review*, 2023, 83(6): 1496—1505.
- [44] Freeman RE. *Strategic Management: A Stakeholder Approach*. Boston: Pitman, 1984.
- [45] Parmar BL, Freeman RE, Harrison JS, et al. Stakeholder theory: The state of the art. *The Academy of Management Annals*, 2010, 3(1): 403—445.
- [46] Li Y, Zhou XY, Jiang X, et al. How service robots’ human-like appearance impacts consumer trust: A study across diverse cultures and service settings. *International Journal of Contemporary Hospitality Management*, 2024, 36(9): 3151—3167.
- [47] Blut M, Wang C, Wunderlich NV, et al. Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, 2021, 49(4): 632—658.
- [48] Zhang JB, Chen Q, Lu JD, et al. Emotional expression by artificial intelligence chatbots to improve customer satisfaction: Underlying

- mechanism and boundary conditions. *Tourism Management*, 2024, 100:104835.
- [49] Li ML, Yin DX, Qiu HL, et al. Examining the effects of AI contactless services on customer psychological safety, perceived value, and hospitality service quality during the COVID - 19 pandemic. *Journal of Hospitality Marketing & Management*, 2022, 31(1):24—48.
- [50] Chi OH, Gursoy D, Chi CG. Tourists' attitudes toward the use of artificially intelligent (AI) devices in tourism service delivery: Moderating role of service value seeking. *Journal of Travel Research*, 2022, 61(1):170—185.
- [51] 张紫琼, 吴玉婵, 叶强, 等. 数字人营销: 理论框架、研究进展与未来方向. *中国管理科学*, 2025, 33(1):259—272.
Zhang ZQ, Wu YC, Ye Q, et al. Digital human marketing: Theoretical framework, research progress, and future directions. *Chinese Journal of Management Science*, 2025, 33(1):259—272. (in Chinese)
- [52] Alboqami H. Trust me, I'm an influencer! - Causal recipes for customer trust in artificial intelligence influencers in the retail industry. *Journal of Retailing and Consumer Services*, 2023, 72:103242.
- [53] Lin XL, Shao B, Wang XQ. Employees' perceptions of chatbots in B2B marketing: Affordances vs. disaffordances. *Industrial Marketing Management*, 2022, 101:45—56.
- [54] Koh LY, Xia ZY, Yuen KF. Consumer acceptance of the autonomous robot in last-mile delivery: A combined perspective of resource-matching, perceived risk and value theories. *Transportation Research Part A: Policy and Practice*, 2024, 182:104008.
- [55] Omrani N, Riviello G, Fiore U, et al. To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. *Technological Forecasting and Social Change*, 2022, 181:121763.
- [56] Nizette F, Hammedi W, van Riel ACR, et al. Why should I trust you? Influence of explanation design on consumer behavior in AI-based services. *Journal of Service Management*, 2025, 36(1):50—74.
- [57] Gutuleac R, Baima G, Rizzo C, et al. Will virtual influencers overcome the uncanny valley? The moderating role of social cues. *Psychology & Marketing*, 2024, 41(7):1419—1431.
- [58] 杜维, 陈鑫. 运气还是能力——基于自我认知视角的智能服务算法偏差对消费者使用意愿的影响研究. *南开管理评论*, 2025, 28(3):136—147.
Du W, Chen X. Luck or ability - The research on the impact of intelligent service algorithm bias on consumers' willingness to use from the perspective of self-perception. *Nankai Business Review*, 2025, 28(3):136—147. (in Chinese)
- [59] Lazarus RS. *Emotion and Adaptation*. Oxford: Oxford University Press, 1991.
- [60] Forgas JP. Affect and cognition. *Perspectives on Psychological Science*, 2008, 3(2):94—101.
- [61] 王翠翠, 徐静, 秦海生, 等. 电商主播类型对消费者购买行为的影响: AI主播和真人主播. *管理科学*, 2023, 36(4):30—43.
Wang CC, Xu J, Qin HS, et al. Effect of E-commerce anchor types on consumers' purchase behavior: AI anchors and human anchors. *Journal of Management Science*, 2023, 36(4):30—43. (in Chinese)
- [62] Giroux M, Kim J, Lee JC, et al. Artificial intelligence and declined guilt: Retailing morality comparison between human and AI. *Journal of Business Ethics*, 2022, 178(4):1027—1041.
- [63] Hu YO, Min H. The dark side of artificial intelligence in service: The “watching-eye” effect and privacy concerns. *International Journal of Hospitality Management*, 2023, 110:103437.
- [64] Youn S, Jin SV. “In A.I. we trust?” The effects of parasocial interaction and technopian versus luddite ideological views on chatbot-based customer relationship management in the emerging “feeling economy”. *Computers in Human Behavior*, 2021, 119:106721.
- [65] Della Corte V, Sepe F, Gursoy D, et al. Role of trust in customer attitude and behaviour formation towards social service robots. *International Journal of Hospitality Management*, 2023, 114:103587.
- [66] Frank DA, Chrysochou P, Mitkidis P, et al. Navigating uncertainty: Exploring consumer acceptance of artificial intelligence under self-threats and high-stakes decisions. *Technology in Society*, 2024, 79:102732.
- [67] Brougham D, Haar J. Smart technology, artificial intelligence, robotics, and algorithms (STARA): Employees' perceptions of our future workplace. *Journal of Management & Organization*, 2018, 24(2):239—257.
- [68] Liang XD, Guo GX, Shu LL, et al. Investigating the double-edged sword effect of AI awareness on employee's service innovative behavior. *Tourism Management*, 2022, 92:104564.
- [69] He CQ, Xiong HY, Cai WJ, et al. How does AI awareness affect employees' voice behavior in the service industry? A transactional theory of stress perspective. *International Journal of Contemporary Hospitality Management*, 2025, 37(5):1662—1680.
- [70] Zhang YJ, Wang JJ, Zhang JW, et al. To be right on the button: How and when hotel frontline service employees' AI awareness influences deviant behavior. *International Journal of Hospitality Management*, 2025, 126:104090.
- [71] Yu HY, Shum C, Alcorn M, et al. Robots can't take my job: Antecedents and outcomes of gen Z employees' service robot risk awareness. *International Journal of Contemporary Hospitality Management*, 2022, 34(8):2971—2988.
- [72] Wang XQ, Lin XL, Shao B. Security and privacy protection in developing ethical AI: A mixed-methods study from a marketing employee perspective. *Journal of Business Ethics*, 2025, 200(2):373—392.
- [73] Jia N, Luo XM, Fang Z, et al. When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, 2024, 67(1):5—32.
- [74] Tong SL, Jia N, Luo XM, et al. The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 2021, 42(9):1600—1631.
- [75] Mu JF, Zhang JZ. Artificial intelligence marketing usage and firm performance. *Journal of the Academy of Marketing Science*, 2025, 53(4):1081—1134.
- [76] Mikalef P, Gupta M. Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 2021, 58(3):103434.
- [77] 李玉花, 林雨昕, 李丹丹. 人工智能技术应用如何影响企业创新. *中国工业经济*, 2024(10):155—173.
Liu YH, Lin YX, Li DD. How does the application of AI technologies affect firm innovation. *China Industrial Economics*, 2024(10):155—173. (in Chinese)

- [78] Han MZ, Shen HC, Wu J, et al. Artificial intelligence and firm resilience: Empirical evidence from natural disaster shocks. *Information Systems Research*, 2025.
- [79] Kim JH, Kim M, Kwak DW, et al. Home-tutoring services assisted with technology: Investigating the role of artificial intelligence using a randomized field experiment. *Journal of Marketing Research*, 2022, 59(1): 79—96.
- [80] Wang WG, Gao GD, Agarwal R. Friend or foe? Teaming between artificial intelligence and workers with variation in experience. *Management Science*, 2023, 70(9): 5753—5775.
- [81] 孙慧, 罗添, 夏学超. 人工智能如何影响企业创新质量. *产业经济评论*, 2025(2): 5—26.
Sun H, Luo T, Xia XC. How does artificial intelligence influence the quality of corporate innovation? *Review of Industrial Economics*, 2025(2): 5—26. (in Chinese)
- [82] 岳平, 苗越. 社会治理: 人工智能时代算法偏见的问题与规制. *上海大学学报(社会科学版)*, 2021, 38(6): 1—11.
Yue P, Miao Y. Social governance: Problems and regulation of algorithm bias in the age of AI. *Journal of Shanghai University(Social Sciences Edition)*, 2021, 38(6): 1—11. (in Chinese)
- [83] Acemoglu D, Restrepo P. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 2019, 33(2): 3—30.
- [84] Zhang XC, Sun MR, Liu JX, et al. The nexus between industrial robot and employment in China: The effects of technology substitution and technology creation. *Technological Forecasting and Social Change*, 2024, 202: 123341.
- [85] Hui X, Reshef O, Zhou LF. The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. *Organization Science*, 2024, 35(6): 1977—1989.
- [86] Babina T, Fedyk A, He A, et al. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 2024, 151: 103745.
- [87] Giordano V, Spada I, Chiarello F, et al. The impact of ChatGPT on human skills: A quantitative study on twitter data. *Technological Forecasting and Social Change*, 2024, 203: 123389.
- [88] Goos M, Manning AL, Salomons A. Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, 2014, 104(8): 2509—2526.
- [89] 王永钦, 董雯. 中国劳动力市场结构变迁——基于任务偏向型技术进步视角. *中国社会科学*, 2023(11): 45—64+205.
Wang YQ, Dong W. The structural transformation of China's labor markets: A task-biased technological change perspective. *Social Sciences in China*, 2023(11): 45—64+205. (in Chinese)
- [90] Wang S, Wang YZ, Li CY. AI-driven capital-skill complementarity: implications for skill premiums and labor mobility. *Finance Research Letters*, 2024, 68: 106044.
- [91] 丁玮, 於兴中. 美国大科技公司反垄断及其权力规制. *中国政法大学学报*, 2022(1): 149—160.
Ding W, Yu XZ. Antitrust and power regulation of major technology companies in the United States. *Journal of CUPL*, 2022(1): 149—160. (in Chinese)

Ethical Challenges of Business Artificial Intelligence: A Theoretical Framework Based on Stakeholder Perspective

Ziqiong Zhang¹ Xin Su¹ Xueyan Wang¹ Zhaohua Deng^{2*}

1. School of Management, Harbin Institute of Technology, Harbin 150001, China

2. School of Management, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract Artificial intelligence (AI), while driving business model innovation and enhancing consumer experiences, has also raised widespread ethical concerns. Its complexity and systemic nature demand theoretical breakthroughs. Existing research is largely confined to single scenarios, lacking a systematic theoretical framework for addressing ethical issues in business AI. This study conducts a comprehensive analysis of 161 English-language articles from the Web of Science Core Collection and 156 Chinese-language articles from the CNKI database. First, it defines the fundamental concepts of business AI ethics and identifies five core issues. Second, drawing on a stakeholder perspective across the levels of consumers, organizations and society, it examines the influencing factors, underlying mechanisms, outcomes, and boundary conditions of AI ethical issues. Based on these analyses, this study constructs theoretical frameworks for business AI ethics. Finally, potential future research directions in this field are proposed. This paper establishes a framework that provides systematic theoretical support for a deeper understanding of business AI ethics. It also offers practical insights for enterprises to deploy AI in compliance with regulations and for policymakers to refine ethical governance systems.

Keywords business AI; AI ethic; stakeholders perspective; theoretical framework; AI technological inclusiveness and controllability

张紫琼 哈尔滨工业大学经济与管理学院教授。围绕“平台经济、消费者行为、在线生成内容”开展研究,在 *Management Information Systems Quarterly*、*Tourism Management*、*Journal of Travel Research* 等期刊发表录用论文70余篇。主持国家自然科学基金青年科学基金项目(B类)、重点项目及教育部新世纪优秀人才支持计划等项目。获第九届高等学校科学研究优秀成果奖(人文社会科学)三等奖1项、黑龙江省自然科学奖二等奖1项、黑龙江省科技进步奖二等奖1项等。

邓朝华 华中科技大学管理学院教授。围绕“医疗健康管理、AI风险管理、平台管理”开展研究,在 *Information Systems Journal*、*Journal of the Association for Information Science and Technology*、*Decision Support Systems*、《管理世界》等期刊发表论文60余篇,主持国家自然科学基金项目4项,获湖北省社科优秀成果奖三等奖3项、湖北省自然科学奖三等奖1项、陕西高等学校科学技术奖一等奖1项、陕西省科技进步奖二等奖1项。

(责任编辑 贾祖冰 张强)

* Corresponding Author, Email: zh-deng@hust.edu.cn