

人工智能时代的算法厌恶： 研究框架与未来展望^{*}

罗映宇 朱国玮 钱无忌 吴月燕 黄 静 杨 智

摘要：理解人类对算法服务的负面反应，是有效应对算法管理挑战的关键前提，是推动新一代人工智能健康发展和推进数字社会建设的重要议题。学术界对算法厌恶问题展开研究并取得丰富成果，但多数仅针对单一领域的单一问题。本文运用系统文献综述法，采用“概念辨析—形成机制—理论边界—研究展望”的框架，对人工智能时代算法厌恶的研究进展进行梳理与解析。首先，对管理领域算法厌恶的核心文献进行系统性梳理，在辨析算法、人工智能、机器人三者关系的基础上，界定算法的内涵和外延，明确提出人工智能时代算法厌恶的定义。其次，构建理论研究框架，从算法视角和人类视角双维度提炼出算法厌恶的6种成因机制，并从个体、情境和算法3个维度阐述算法厌恶的12种理论边界。最后，提出具有科学价值和实践意义的六大未来研究议题，为深化算法厌恶理论和实践研究提供理论指南。本文拓展了算法厌恶理论体系研究的认知边界，对促进中国企业有效应对算法管理挑战、推动中国人工智能健康发展和数字社会建设具有重要启示。

关键词：人工智能 算法决策 算法厌恶 人机交互

DOI:10.19744/j.cnki.11-1235/f.2023.0117

一、引言

算法(Algorithms)是人工智能(Artificial Intelligence)发展的核心驱动因素，是企业实现数字化管理、国家实现智能社会建设目标的关键基础。随着机器学习、自然语言处理和计算机视觉等算法技术的迅速发展，人工智能由弱人工智能向强人工智能升级发展，人工智能新时代正在到来(郭凯明等,2023;阿明等,2021)^{①②③④⑤}。目前，算法逐渐渗透到社会运作的方方面面，成为推动企业数字化转型和国家治理现代化的重要力量。越来越多的企业采用算法来代理组织管理、优化产品服务设计和提升用户体验，以期从激烈的商业竞争中脱颖而出。不仅如此，社会组织和政府部门也将算法视作智能社会建设和国家远景目标实现的重要基础，《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》中多处提到要“迎接数字时代，激活数据要素潜能，推进网络强国建设，加快建设数字经济、数字社会、数字政府，以数字化转型整体驱动生产方式、生活方式和治理方式变革。”^⑥

算法的应用重塑了社会和经济系统的运作方式，推动了整个社会的数字化进程，人们和组织纷纷拥抱算法。例如，滴滴在发展中全面采用算法来优化乘客和司机的匹配过程，提升平台工作效率和用户体验(朱国玮等,2023)。爱彼迎通过“智能定价”算法为房东推荐定价方案，以应对市场需求变动(程、佛利,2019)。美团外卖的“超脑”系统，通过算法为骑手制定最优工作流程，极大提高了配送服务的效率和质量(陈龙,2020)。算法的应用促进了企业“降本增效”目标的实现。麦肯锡数据显示，仅使用算法就能为企业削减约10%~20%的成本、提升约10%的收益^⑦。

在数字化发展形势一片大好的背景下，企业在利用算法推进数字化转型的过程中仍面临着巨大挑战。数

^{*} 本研究得到国家自然科学基金“移动互联网时代的协同消费商业模式研究：理论模型与策略优化”(基金号：71871089)、国家社会科学基金重大招标项目“数据赋能激励制造企业创新驱动发展及其对策研究”(基金号：18ZDA063)、国家自然科学基金“消费者对医疗人机融合智能服务的接受机制与引导策略研究”(基金号：72302051)、湖南省自然科学基金“数字化技术赋能背景下低碳消费模式研究：理论机制与实践路径”(基金号：2023JJ30175)、湖南省哲学社会科学基金教育学专项重点课题“信息技术赋能视角下‘慕课’教师与学生协同发展研究”(基金号：18ZDJ04)的资助。吴月燕为本文通讯作者。

据显示,2022年数字化转型成效显著的企业所占比例仅为17%,中国企业数字化转型指数得分首次出现下降^⑧。对企业来说,完成数字化转型不仅需要解决创新投入的资本和技术壁垒,更重要的是如何应对算法投入使用后的“用户壁垒”,即组织、员工和消费者对算法的拒绝和回避。学者们发现,企业管理者总是更愿意依赖自身直觉而不采纳算法建议(艾伦、乔杜里,2022);零工工作者会抗拒平台算法提供的工作规划和绩效评估(默尔曼等,2021;陈龙,2020),甚至会从采用算法的平台离职(梁等,2022);消费者对基于算法提供医疗诊断和康复建议的虚拟机器人表现出不信任(隆戈尼等,2019)等。算法应用带来的“用户壁垒”不仅与企业满足消费者“期望升级”的目标相冲突,也给企业数字化转型发展设置了阻碍。

算法管理挑战主要源自算法风险引发的个体算法消极反应。人工智能时代的算法是基于大数据的、以机器学习为底层逻辑的算法,存在各种潜在风险,如算法结果存在不可预测性和不可解释性、算法服务过程中可能存在隐私泄漏、算法错误、数据偏差以及非道德决策等算法风险。鉴于算法风险会给社会稳定乃至国家安全带来严重后果,国家正在逐步完善算法治理相关政策、法规、制度。党的二十大报告提出要“推动战略性新兴产业融合集群发展,构建新一代信息技术、人工智能、生物技术、新能源、新材料、高端装备、绿色环保等一批新的增长引擎”,同时要“提高防范化解重大风险能力,严密防范系统性安全风险。”^⑨算法风险的防范与治理符合国家安全发展战略和需求。

然而,算法治理制度上的完善并不能完全消解个体对算法的消极反应。个体对算法的消极反应不仅与客观的算法风险有关,还受人们对算法的主观心理感知影响。因此,明晰用户对算法的消极反应及其背后原因和后果,能够帮助企业更好地破除算法“用户壁垒”,充分发挥算法的积极正向作用。例如,人们为何会拒绝和回避算法?这会导致什么后果?在哪些情况下这种消极反应会被增强或削弱?对这些问题的揭示与回答,不仅能够帮助企业清晰认识算法在推进组织发展过程中如何受到阻碍,为企业的数字化转型提供指导和启发,符合企业数字化转型的需要,还能为国家和政府提供促进全社会数字化的新指南,进一步推动国家开展算法风险防范和治理工作。基于此,本文拟以“人们对算法的消极反应”(即“算法厌恶”)为核心主题并对其进行研究,以期服务于企业、国家和社会的数字化发展工作。

在学术界,学者们将人们拒绝使用算法建议和服务的行为或行为倾向称之为算法厌恶(Algorithm Aversion)(狄弗斯特等,2015)。算法厌恶的现象引起了研究者的广泛关注,并成为近年来人机交互领域的热点研究话题。已有研究涉及了医疗服务、电子商务、组织决策、零工经济、自动驾驶、法律审判等领域;研究内容主要包括两类,一类是人们对不同类型算法服务的负面态度及行为反应,另一类是面对同一服务时,人们对算法和人类服务者的偏好反应差异等;研究也从不同的理论视角揭示了算法厌恶的发生机制及相应的边界条件等(李等,2019;威尔森等,2017;蓬托尼等,2021;隆戈尼等,2019;隆戈尼、齐安,2022;金姆、杜哈切克,2020;斯里尼瓦桑、沙利叶,2021;默尔曼等,2021;郁苏波夫等,2021)。

学者们对算法厌恶展开了多方面讨论,也取得了丰富的研究成果。然而,已有研究大多围绕单一领域的单一问题展开,存在研究相对零散、概念界定不清晰、研究结论矛盾等问题。首先,学术界对算法厌恶概念中“算法”的定义尚未形成公认且清晰的界定,出现不少研究将“算法”、“人工智能”、“机器人”等概念等同的现象。其次,随着研究的不断深入,出现了矛盾的研究结果和解释机制,亟待研究者进行系统梳理。例如,有研究指出人们认为算法是不完美的,而有的研究发现相反结论(隆戈尼等,2019;格兰努罗等,2021;狄弗斯特等,2015)。再次,与传统的研究背景相比,人工智能时代的算法基础更为复杂,算法表现形式更为多样,有必要展开新理论体系研究。例如,传统算法强调自动化,结果可预测且无法学习改进;而人工智能时代的算法则是以机器学习为基础,它可以学习和改进,结果也多不可预测。最后,国内对算法厌恶的研究有限,相关理论基础仍显薄弱。

作为一个新兴研究领域,算法厌恶仍存在广阔的探索空间,综述性研究有助于为特定领域提供系统且深刻的理解,助推该领域理论研究与管理实践的共同发展,尤其是对于新兴研究领域。本文通过系统总结人工智能时代算法厌恶的概念内涵和外延,正式提出人工智能时代算法厌恶的概念,并基于此构建算法厌恶研究

框架,揭示驱动机制,识别边界条件,同时提出具有科学价值的六大未来研究方向。本文对探索算法决策领域的理论创新、推进人机交互领域的理论研究,以及推动算法驱动的管理实践具有重要意义。

二、研究方法和研究进展

(一)研究主题界定

算法包括传统的一般算法和驱动人工智能的机器学习算法。很早就有学者关注到一般算法背景下的算法厌恶现象(米尔,1954),由于一般算法更多是通过数据和规则产生可预测的稳定结果,人们更多是由于不信任统计模型的结果而产生算法厌恶。随着技术的发展,机器学习算法展现出经验学习、思维模拟、类人智能反应等能力。在此背景下,人们强烈感受到自己身为人类的身份受到算法(或人工智能)的威胁,如工作被替代、隐私被窥探等(隆戈尼等,2019;格兰努罗等,2021),从而产生算法厌恶。为了呼应人工智能时代这一现实背景,本文将研究主题聚焦于机器学习算法驱动下,人与算法互动中产生的算法厌恶。人与算法互动时,往往是与算法赋能的实体机器人(如护理机器人)或虚拟机器人(如在线智能客服)进行互动,或者是与嵌入式算法(如搜索引擎算法)进行互动(格里克森、沃利,2020)。前者是与有形形式的互动,后者是与无形形式的互动,这两种互动可以统称为人与智能机器的互动,简称“人机交互”(详见第三部分“人工智能时代的算法厌恶定义”中关于算法的表现形式)。由此,本文进一步将研究主题确定为“人工智能时代下人机交互中的算法厌恶”(后文简称“人工智能时代的算法厌恶”)。

(二)研究现状与进展

人机交互兴起于数字平台的管理创新实践,并于近些年成为战略管理、市场营销、人力资源等领域热议的话题。随着机器学习算法的发展,学者们将注意力从算法技术发展逐渐转向人机交互中用户的反应和行为,即对于提供相同或类似的服务,包括决策建议、推荐或辅助服务,相比于由人类提供,由算法提供更可能受到用户的青睐还是厌恶。近年来,Science、Journal of Marketing、Information System Research、Management Science、Psychological Science、MISQ、《管理世界》等主流期刊围绕这一话题发表了一系列文章,揭示算法在不同情境中的应用。目前,全球学者围绕人机交互中的算法厌恶开展了卓有成效的研究,该主题是当前主流研究方向之一。

为了系统全面地展示人工智能时代算法厌恶相关主题的研究概况,本研究借鉴特兰菲尔德等(2003)、刘善仕等(2022)的系统文献综述法(Systematic Literature Review)以及刘洋等(2020)的文献研究方法对人工智能时代算法厌恶的研究成果进行总结和述评。

第一,文献检索。本文以“Algorithm(ic)Aversion”“Artificial Intelligence”“Human-Computer Interaction”“Algorithm(ic)Management”“Algorithm(ic)Decision”“Human-like Bot”“Anthropomorphism”等为关键词,在Web of Science、Emerald、Springer、EBSCO、CNKI等数据库中进行初步检索。鉴于狄弗斯特等(2015)发表的论文正式开启算法厌恶的研究,本文将检索时间确定为2015~2023年。在剔除社论、新闻、会议论文、工作论文、评论等类型文献后,初步获得1764篇文献。这些文献不仅包括算法厌恶,还包括算法决策领域的其他研究。为了初步了解学术界在算法决策领域投入的关注和取得的成果,本研究初步对1764篇文献所有作者单位的国别进行了分类统计,结果显示,算法决策领域研究中,美国、欧洲、中国学者发文量占据了全球发文量的90%以上。

第二,文献筛选。通过阅读初步检索获得的文献摘要和全文,发现这些文章的研究质量不一,且大量研究不属于本文聚焦的“算法厌恶”这一研究主题。为了严格控制文献质量且契合本文主题,进一步依据以下标准对文献进行筛选:(1)剔除与研究主题不符的文献,如剔除探讨算法技术、算法效能及其他方面的文献;(2)增加被遗漏的参考文献。最终,本文共获得符合条件的核心文献106篇,外文期刊来源包括Nature、Science、Management Science、Information System Research、Journal of Marketing等领域顶级期刊;中文期刊包括《管理世界》、《南开管理评论》等。统计结果显示,“人工智能时代的算法厌恶”主题下,美国、欧洲、中国学者的发文量仍占据全球发文量的90%以上,与算法决策领域文献的统计趋势一致。

第三,编码和分析。参考刘善仕等(2022)和刘洋等(2020)的文献计量分析方法,通过提取作者、年份、文

献来源、研究主题、研究方法、核心研究发现等关键内容,并对其进行初步梳理和编码(韦伯斯特、沃森,2002)。为确保编码的科学、严谨和准确,本文进一步利用 CiteSpace 软件进行知识图谱计量分析。共词结果(《管理世界》网络发行版附录附图1)显示,人工智能、算法、算法厌恶、决策、判断、拟人化、信任、用户接受度等是目前全球学者关于“人工智能时代的算法厌恶”这一领域关注的焦点。共被引(《管理世界》网络发行版附录附图2)分析结果显示,洛格等(2019)、狄弗斯特等(2015,2018)、卡斯特洛等(2019)、隆戈尼等(2019)、罗等(2019)、门德等(2019)、约曼斯等(2019)、黄和拉特斯(2018)等是“人工智能时代的算法厌恶”研究的核心共被引文献。结合文献计量分析结果,本文对文献编码的初步结果进行调整,之后由两位管理学领域专家共同讨论、斟酌后得出最终的文献编码结果(见《管理世界》网络发行版附录附表1)。

基于上述工作,本文发现:(1)人工智能算法厌恶研究领域,不仅美国、欧洲学者关注多,中国学者也投入大量关注,并且有丰富的研究成果发表;(2)发表该研究主题的期刊数量多,管理学、心理学等领域的顶级主流期刊均有发表,发表数量呈逐年上升趋势;(3)尽管中国学者发表了大量成果,但多数发表于国际期刊。总体而言,国内外学术界、实践界均关注到“人工智能时代的算法厌恶”这一研究主题的理论价值和实践意义,是人工智能领域的主流研究方向之一。

三、概念界定和研究框架

(一)算法厌恶

自米尔(1954)在临床预测中发现算法厌恶现象以来,研究者已在众多领域中观察到这一现象,并对算法厌恶提出了诸多定义。目前,狄弗斯特等(2015)提出的定义认可度最高,他们认为算法厌恶是在已知算法预测更为精确的前提下,人们依然倾向于依赖人类自身进行的决策的现象。但是,该定义存在一些局限。第一,该定义存在前提局限,狄弗斯特等(2015)是基于人们已知算法的表现优于人类的前提,一些研究表明人们在不清楚算法的优势时也会存在算法厌恶(隆戈尼等,2019)。第二,多数研究关注算法与人类的比较,并指出人们在对算法表现出厌恶后会转而接受人类建议或依赖自身作出决策。然而,人们在不同情境中对算法决策的态度也属于算法厌恶研究范畴(默尔曼等,2021;孙等,2022;隆戈尼等,2019;川口,2021;比格曼、格雷,2018;吉鲁克斯等,2022;基尔,2020;葛等,2021)。第三,狄弗斯特等(2015)提出的算法厌恶更多代指狭义上的算法(如嵌入式算法),而当前算法领域的研究有更为广义的内涵(如算法赋能的机器人或人工智能等),已有定义的内涵只能体现其中很小一部分研究。总体来说,已有关于算法厌恶的定义,无法囊括人工智能时代的算法厌恶研究。为解决上述问题,本研究将在讨论算法与人工智能、机器人之间的关系,梳理算法与人类交互逻辑的基础上,提出人工智能时代具有普适性的算法厌恶定义。

(二)算法、人工智能和机器人

为了理解算法、人工智能、机器人之间关系,本研究对人工智能时代算法的应用逻辑进行梳理。第一层:底层算法;第二层:应用方向;第三层:表现形式;第四层:应用场景(图1)。

根据算法的应用逻辑图,机器学习是人工智能时代算法的内核,嵌入式算法和机器人是人工智能中的表现形式,不同表现形式下的算法厌恶和机制不尽相同。因此,本研究将重点围绕算法表现形式与应用场景适配下的算法厌恶展开讨论。在论述核心内容之前,先对算法、人工智能、机器人的关系进行阐述。

1. 算法与人工智能

算法是驱动人工智能发展的关键因素,机器学习算法是现代人工智能的内核。依据“人工智能之父”约翰·麦卡锡的观点,人工智能是一种制造智能机器的科学和工程,它通过计算机来了解人们在同样任务中的生物和非生物反应,并利用机器模仿和学习人类思维来解决问题和制定决策(麦卡锡,2004)。传统的人工智能主要依赖符号知识表示和程序推理机制,而现代人工智能则建立在机器学习之上(罗素、诺维格,2010),并以此为基础发展了人工智能的内核。机器学习是区别现代人工智能与传统人工智能(如自动化)的重要概念。传统人工智能的运作结果是确定的,整个流程不涉及任何学习过程(瑞吉、希曼斯,2019)。但是现代人工智能

在自动化的前提下,还会根据经验和反馈进行学习和调整,决策结果多数不可预测,这一过程被称为机器学习(格里克森、沃利,2020)。

2. 人工智能与机器人

机器人是现代人工智能的表现形式之一。基于机器学习算法的现代人工智能主要有3种表现形式:嵌入式算法、算法赋能的虚拟机器人和算法赋能的实体机器人(格里克森、沃利,2020)。也就是说,机器人(实体和虚拟)与嵌入式算法相似,均是现代人工智能的表现形式,具体阐述如下。

(1)嵌入式算法。嵌入式算法对用户来说是完全隐形的,它往往没有视觉呈现或身份,而是嵌入在不同类型的应用程序中(李等,2015),例如搜索引擎、决策系统等。嵌入式算法通过一系列运作对用户行为进行影响和控制,如推荐商品和服务、控制员工劳动过程等(伍德等,2019;甘迪尼,2019),但是用户很难感知到它的存在。

(2)算法赋能的虚拟机器人。算法赋能的虚拟机器人是一种没有实体的人工智能形式,一般拥有独特的身份或者化身,例如聊天机器人、语音助手等。它可能存在于任何电子设备上,并具有面部、身体、声音、语言能力等人类特征。虚拟机器人可以与人类进行交互,向人类提供咨询服务和建议(格里克森、沃利,2020)。

(3)算法赋能的实体机器人。算法赋能的实体机器人是一种以实体形式存在的机器人,具有多种功能以及不同的机械或类人表现形式,可以执行现场或远程任务,如办公楼和机场中的机器人(格里克森、沃利,2020)。

(三)人工智能时代的算法厌恶定义

算法厌恶源自人们与算法交互过程中产生的负面感知,发生在算法表现形式和应用场景的交互中。在狄弗斯特等(2015)的研究中,他们探究的是人们对统计模型预测结果的偏好,该研究中的算法是指嵌入式算法。嵌入式算法可以直接生成结果并对人类决策产生影响,如在投资决策、审计、组织管理等领域中,嵌入式算法直接给出决策建议(康默福特等,2022;默尔曼等,2021;刘等,2022)。

嵌入式算法具有无形性,人们往往无法主动与其产生较多互动,限制其向个体应用场景推广。企业开始尝试将算法嵌入到一些载体中,以算法赋能的虚拟或实体机器人形式呈现,促进人们与算法产生更多互动,学者们通过研究也发现了人们对算法赋能的机器人产生的厌恶现象。例如,人们对机器人(vs.人类)提供的产品推荐、医疗和客户服务等总是会更加抵触(张等,2021;隆戈尼、齐安,2022;拉普等,2021;陈等,2022;帕克等,2022;昆、权,2022)。这些研究指出,人们对机器人的厌恶是源自对机器人外表之下的嵌入式算法的负面感知,如无法学习、缺乏自动化意识等(金姆、杜哈切克,2020;海豪斯,2008)。因此,这类研究也属人工智能时代算法厌恶研究范畴。

与传统的算法厌恶相比,人工智能时代的算法具有以下两种新的特征:一是算法的基础。传统的算法强

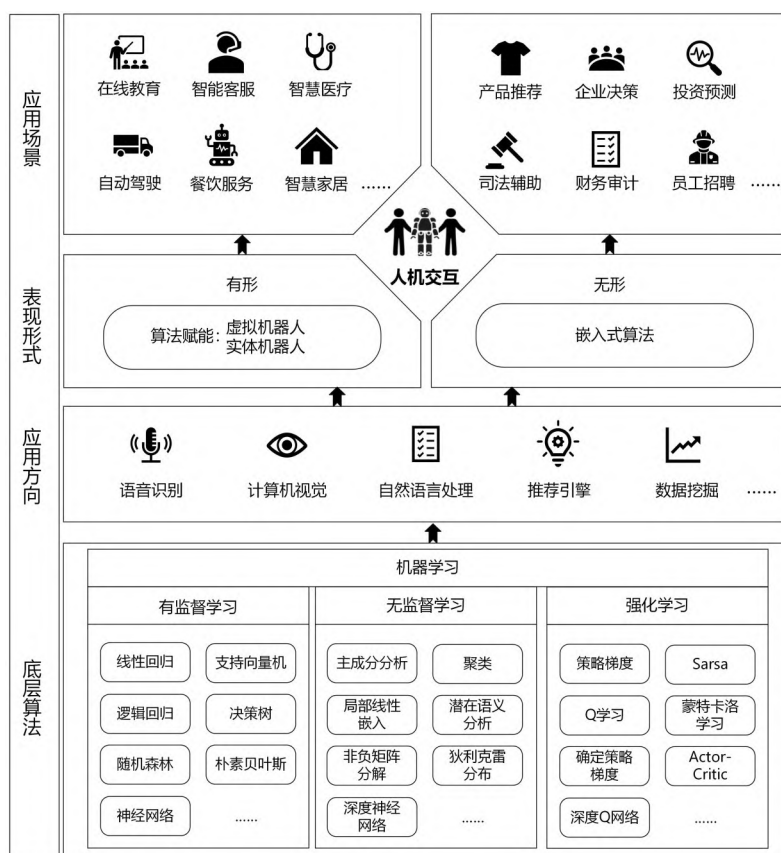


图1 人工智能时代算法的应用逻辑

注:作者根据资料进行归纳和整理。底层算法指机器学习算法,包括有监督学习、无监督学习和强化学习。

调自动化流程、结果可预测且无法学习改进;而人工智能时代的算法是以机器学习为基础,它可以学习和改进,结果也多是不可预测。二是算法的形式。传统的算法仅强调嵌入式算法,而人工智能时代的算法不仅包括嵌入式算法,还包括嵌入式算法赋能的有形体。据此,本研究正式提出算法厌恶的定义:人工智能时代的算法厌恶是指人类对嵌入式算法以及算法赋能的虚拟和实体机器人提供的建议和服务展示出的负面情绪、消极态度和回避倾向。

已有研究中,学者采用各种概念来指代算法,如人工智能、机器人、系统等。这主要与算法的表现形式有关,算法形式又与特定领域、行业和对对象产生适配。在使用嵌入式算法的领域,研究者会使用算法、算法平台、控制系统等概念;而在使用算法赋能的机器人的领域,研究者会使用人工智能、机器人、推荐代理等概念(具体见《管理世界》网络发行版附录附表2)。为了避免混乱,如无特殊说明,根据本文研究主题及人工智能时代算法厌恶的正式定义,本研究提到算法厌恶的对象时统一采用“算法”这一概念。

(四)算法厌恶的研究框架

基于对管理领域与研究主题相关的核心文献的梳理,研究发现人类与算法的交互过程中,驱动算法厌恶的机制主要有6种,包括朴素信念、算法错误、隐私担忧、过度自信、错误归因和自我效能感。这六大成因可归纳为两类视角。第一类视角是人类如何认识算法,简称“算法视角”。算法视角往往与人类所处的社会环境有关(莫林,2004)。例如,人们会根据媒体、电影提供的信息和生活中的经验,认为算法是刻板的、不能理解人类的情感(隆戈尼、齐安,2022)。基于此,本研究将人类视角定义为:在算法厌恶中,人们的注意力和认知聚焦于对算法所产生的感知,主要包括朴素信念、算法错误和隐私担忧。第二类视角是人们如何认识人类自身,简称为“人类视角”。人类视角往往与人类对自身的认知有关(莫林,2004),例如,人类往往(自信地)认为人比算法更有能力(班克、赫塔尼,2019;陈等,2022)。基于此,本研究将算法视角定义为:在算法厌恶中,人们的注意力和认知聚焦于人类对自身(或人类)所产生的感知,主要包括过度自信、错误归因和自我效能感。据此,本研究基于算法视角和人类视角的驱动机制,形成算法厌恶成因板块的主要内容。作为研究框架的关键要素,边界条件也是机制的重要组成部分。因此,本研究还从个体、情境和算法3个维度归纳了算法厌恶的边界因素。其中,个体因素(如性别、年龄)是不可操纵因素,情境因素(如道德、主客观)和算法因素(如算法拟人化、算法角色)是可操纵因素,具有较强的管理实践意义。

综上,研究从算法管理挑战出发,围绕“概念研究”“成因研究”“边界研究”3个核心内容展开。由于对算法厌恶类型的区分有助于深入了解算法厌恶的成因,因此,研究根据算法的表现形式,将其分为算法赋能的虚拟机器人、算法赋能的实体机器人和嵌入式算法,并形成“算法厌恶驱动机制与算法表现形式匹配”的纵横网络研究图(见本研究第五部分未来展望表1)。同时,本文还提出具有理论与实践意义的六大未来研究方向。基于此,形成本研究系统完整的算法厌恶研究框架(图2)。

四、算法厌恶的成因和边界

本研究从算法视角和人类视角对算法厌恶的成因进行归纳。算法厌恶的算法视角认为人们对算法的关注会产生算法厌恶,这一现象描述了人们如何评估、判断算法能力,以及对算法做出的行为进行推测的过程,可从朴素信念、算法错误和隐私担忧三方面展开论述。算法厌恶的人类视角认为人类对自我认知的关注会产生算法厌恶,这一现象描述了人们如何评估、判断、推测和归因自己能力的过程,可从过度自信、错误归因和自我效能感三方面展开论述。接下来,研究先从算法视角阐述算法厌恶的成因。

(一)算法厌恶的成因

1. 基于算法视角的算法厌恶成因

算法厌恶的算法视角机制主要包括三类:朴素信念、算法错误和隐私担忧。

(1)朴素信念。朴素信念(Lay Beliefs)是人们依据自身经历或环境对特定对象形成的常识性解释(莫尔登、德维克,2006;麦等,2019),如人们会认为“道德产品”等于“不那么结实”(麦等,2019)。过去的研究发现,

人们会对算法(包括嵌入式算法和算法赋能的机器人)形成两种特定的朴素信念。一种是人们认为算法是“机器人”,即算法具有机器人的特征:如刻板、无法理解人的情感等(道斯,1979;格鲁夫、米尔,1996);另一种是人们认为算法会对人类身份产生威胁,即算法会替代人的功能和自我价值,产生身份威胁感知(斯坦因、奥赫,2017;王欣等,2021)。产生算法厌恶的朴素信念的相关研究阐述如下。算法是“机器人”。与算法厌恶相关的第一个朴素信念机制是人类认为算法是“机器人”。与人类时常基于感性思维行动相反,算法的行为总是理性的、不会与外界产生情感交流,这与人们常说的“像机器人一样思考”相符(隆戈尼、齐安,2022)。在人们的观念中,算法经常与机器人的特征联系起来,并且外界的信息常常强化这种感知,例如在《星际迷航》、《终结者》等电影中,算法赋能的机器人被赋予了卓越的能力,但是他们无法体验人类的情感,这些客观经验塑造了算法是“机器人”的朴素信念。

受到算法是“机器人”观点的影响,人们常将机器人的特征代入到人机互动中并影响其行为。第一,人们会认为算法是无法学习和进步的(海豪斯,2008;赖希等,2023);当算法出现错误时,人们会认为算法无法从中吸取错误经验并重复犯错,导致人们对算法失去信心(狄弗斯特等,2015)。第二,人们会认为算法的运行逻辑是有序、刻板且稳定的,因而算法无法考虑不确定因素以及个体的唯一性(隆戈尼等,2019;格兰努罗等,2021)。第三,人们会认为算法缺乏同理心(格鲁夫、米尔,1996),因此对于需要情感体验的产品(如享乐品),人们更倾向于相信人类而非算法推荐(隆戈尼、齐安,2022);或在接受客户服务时,人们对机器人(vs. 人类)客服表现出粗鲁(vs. 温和)的态度(罗等,2019)。第四,人们会认为算法的行为是不受法律和社会规则约束的,并认为算法比人表现得更不道德,从而不愿意接受算法做出的与道德有关的决定(比格曼、格雷,2018;吉鲁克斯等,2022;博纳丰等,2016;基尔,2020;费格罗亚·阿米霍斯等,2022)。例如企业高管对算法建议表现出迟疑态度,这是因为企业需要确保其运营符合道德和伦理,而通常的观点则认为算法难以具备(或不具备)做出道德判断的能力(乔希、韦德,2020;泰尔坎普、安德森,2022)。第五,人们还会认为算法缺乏自动目标和自我意识(沃德等,2013;加维等,2023),因此算法在从事高解释水平的任务时,人们展示出更多的不信任(金姆、杜哈切克,2020)。

简而言之,在算法是“机器人”这一信念的影响下,人们会普遍认为算法具有机器人的特征,并将这些特征

代入到人机交互中,认为算法难以进行交互,从而拒绝采纳算法提供的建议。这种现象在需要专业性技能的领域,如医疗咨询和投资决策领域,显得尤为突出(张等,2021;乔德里等,2022;隆戈尼、齐安,2022;邹、施宾格,2018;金姆等,2022)。

算法会对人类身份产生威胁。算法厌恶相

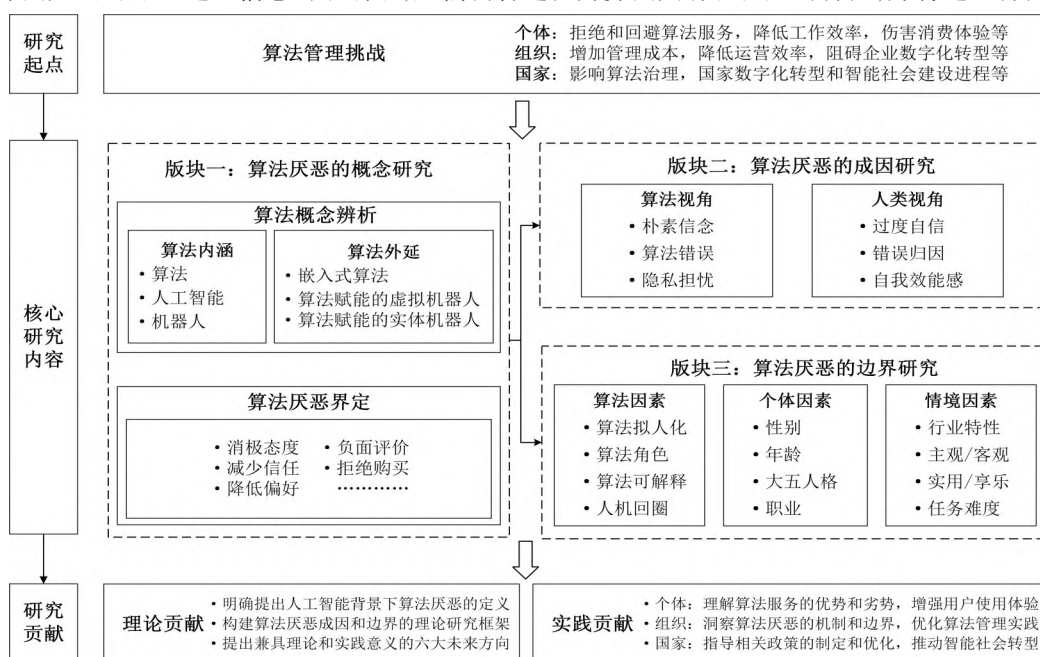


图2 人工智能时代算法厌恶的研究框架

注:(1)已有算法厌恶的研究中,不仅包括人类对算法和人类提供相同决策建议、服务时的不同反应(隆戈尼、齐安,2022),也有只关注算法,探究人类对不同类型算法提供决策建议、服务的反应(斯里尼瓦桑、沙利叶,2021)。(2)在已有关于算法厌恶的研究中,算法和人类在不同研究中被以不同的代称表示,其中,“算法”常见代称包括:算法、人工智能、控制系统、机器人等(狄弗斯特等,2015;隆戈尼、齐安,2022;梁等,2022),“人类”常见代称包括:自己、人类专家、服务者等(狄弗斯特等,2015;隆戈尼等,2019;隆戈尼、齐安,2022)。

关的第二个朴素信念机制是人类认为“算法会对人类身份产生威胁”,这一信念起源于恐怖谷理论。恐怖谷理论指出,当算法与人类的相似度达到一定程度时(大约70%相似),随着算法与人类的相似度继续提升,人类对算法的态度会急剧下降(侯、麦克多曼,2010)。

关于恐怖谷理论的形成原因有两种观点。第一种观点认为,人们的恐惧感源自身份的不确定性,当算法与人的外表以及行为的相似度越来越高时,人们会无法确定算法与人的差异,从而感到自身的存在受到威胁(侯、麦克多曼,2010)。例如,人们会认为高度拟人化的算法可作为生物实体替代人类,使人们感受到威胁(侯、麦克多曼,2010)。第二种观点则认为,算法拥有重塑行业的社会功能,对人们的社会功能价值产生替代威胁(李等,2019;威尔森等,2017;蓬托尼等,2021;亨特等,2022;黄、拉斯特,2018)。例如,王林辉等(2022)指出,算法技术会导致中国19.05%的劳动就业者面临高替代风险,并且年龄越大、教育年限越短和从事高程式化任务的劳动者越容易被替代。唐小飞等(2021)以孝悌先生的“差序格局”理论为基础,指出算法会产生社会功能替代和社会存在替代,值得一提的是,这项研究首次从中国文化视角探讨了算法厌恶的形成机理。

基于社会存在价值和社会功能价值的威胁替代机制,学者们开展了大量关于算法应用的研究。多数研究发现,使用算法会导致员工产生负面感知。例如,李等(2019)发现酒店使用智能机器人可能导致员工离职倾向的增加;唐等(2021)发现企业算法反馈系统可能导致员工认为自己随时可以被替代,从而对自身能力丧失信心,进而导致工作产出的下降;麦克里等(2021)在顾客服务的研究中发现,消费者担心机器人替代自己,因此对机器人客服的评价会更低等。一些学者在恐怖谷理论基础上探讨了改进方案。例如,通过塑造算法产品形象(协助者 vs. 替代者)来减轻消费者的心理抗拒,从而提升消费者对算法产品的使用意愿(王欣等,2021)。

综上,已有研究揭示了人们关于算法的朴素信念如何影响其对于算法的态度,这些研究多是考虑了外部环境中与算法相关的信息和经验如何影响人们的行为。接下来,研究将从算法行为的视角探讨算法厌恶的形成,这一部分聚焦探讨当算法发生错误时,人们的行为会发生何种改变,以及为什么会发生改变。

(2)算法错误。狄弗斯特等(2015)的研究发现,当被试者意识到算法预测出现错误后,即便足够的证据表明算法模型的表现优于人类预测,他们依然会展现出强烈的算法厌恶。然而,当人类的预测出现错误后,被试者对其依赖程度并未降低。因此,一些研究认为算法错误会导致算法厌恶。

算法错误导致的算法厌恶源自人们对算法能力的过高期待,这一观点与朴素信念所提的观点相悖。研究者发现,人们会对科技展现出高度的信任和期待,并相信算法的预测是完美的(狄弗斯特等,2015);因而,一旦算法预测发生了微小错误,他们就会因为期望受损认为算法预测并不可靠。与之相反,相比于算法,人们对于自身的期望是较低的,能够容忍自身出现错误,并认为自己能够改进和修正错误;故而当人类预测出现错误时,人们对人类决策表现出包容态度(狄弗斯特等,2015)。另一种观点则认为算法错误导致的算法厌恶源自人们对算法产生的预先期待。研究者发现,人们会对算法产生预先期待。这一期待可能产生于他们此前与算法交互的经验,或者是来自于身边人对算法的看法等(伯顿等,2020)。

(3)隐私担忧。由于算法会主动持续不断地收集和使用用户的隐私信息(如人口统计信息、Cookie、地理位置等),用户对算法的隐私担忧与日俱增,进而产生算法厌恶(梁等,2022)。但是在不同的情境中,隐私担忧的产生机制是不同的。一种观点认为算法会直接侵犯人们的隐私,甚至威胁人们对隐私的(心理)所有权,从而产生隐私担忧(齐希等,2021;默尔曼等,2021;梁等,2022)。例如,在梁等(2022)的研究中,他们发现零工工作者会担心平台利用算法系统侵犯自己的隐私,因此会更倾向于选择没有使用算法系统的零工平台。但是,当平台采取适当的隐私保护政策时,则能在一定程度上缓解零工工作者的隐私担忧,促进企业对零工工作者的雇佣。另一种观点认为算法对隐私信息的收集、存储和使用是不透明的,导致人们对算法产生隐私不确定性感知(俞、陈,2018;纳图尔等,2020)。例如在互联网保险行业,如何保障用户隐私是整个行业面临的关键问题之一,用户可能感知到平台未经授权使用用户信用、银行账户、身体健康状况等信息,用来关联合适的保险方案(阿拉德瓦尼,2001),用户在网络保险中(相比于传统保险)感知到更高的隐私威胁,这进一步降低了他们

对算法生成的保险方案的接受意愿(俞、陈,2018)。

上述研究主要从算法视角维度回顾了导致算法厌恶产生的三类成因。接下来,本研究将从人类对算法的关注转向人类对自我的关注,揭示人类视角下算法厌恶的成因。

2. 基于人类视角的算法厌恶成因

算法厌恶的人类视角机制主要包括三类:过度自信、错误归因和自我效能感。

(1)过度自信。已有研究发现,人们会自信地认为自己拥有比算法更高的决策能力,因此更依赖于自身的决策。例如,在一项借贷的研究中,刘等(2022)发现大部分决策者都倾向于低估算法提供的信息准确性,并认为自身的决策更为有效,从而产生算法厌恶。研究者认为,过度自信源自人类的直觉(戴恩、普拉特,2007)。人们通常将直觉性决策与创造力和洞见力相联系,而算法决策显得程序化。当面对一些需要创造性的活动时(如投资决策),人们经常会高估自己的能力并自信地认为他们不需要依赖算法(班克、赫塔尼,2019;陈等,2022)。另一种观点源自于人类最优理论,该理论指出人们会认为自身比其他事物(如算法)更加完美。因此,尽管个体明白算法预测准确性更高,但出于人类更优这一想法,他们依然更倚重人类预测。此外,也有研究发现人们对于他人(而非是自身)决策的过度自信会导致他人产生算法厌恶。例如,人们会自信地认为不使用算法辅助的专家专业性更强,这导致专家面对算法建议时变得更加谨慎(戴、塔尤尔,2022;穆罕默德等,2022)。

总体而言,设计者采用算法辅助决策的目的是帮助人们做出理性的行为;然而由于人类自身的“有限理性”,人们反而会做出非理性的行为。有趣的是,尽管先前的研究认为人们的能力普遍比算法更弱,但在一项对阿里巴巴仓库管理的研究中发现,仓库管理员比算法拥有更多的信息,其决策效果反而会更好(孙等,2022)。因此,研究者认为算法框架搭建者应该将人的特殊性纳入决策模型中,以进一步优化算法的代理结果。

(2)错误归因。归因理论认为,为了维持或提高自我价值,人们会将有利的结果归因于自己(内部归因),将不利结果归因于外部因素(外部归因)(考尔德、伯恩柯兰特,1977)。例如,人们将好的考试成绩归因为自己的天分和努力;而差的考试成绩归因为教师水平等外部因素。关于算法厌恶的研究也发现,错误归因也是人们产生算法厌恶的内在机制之一。例如,研究者发现人们更愿意接受人类提供的有利建议(叶尔钦等,2022),这是因为当人们收到一个好消息时,他们会做出内部归因,即相比于算法,人们会认为人类的建议更可能考虑他们的独特性;相反,当人们收到一个坏消息时,不论推荐者是算法还是人类,人们的接受度都不会有差异;这是因为人们会对不利的建议做出外部归因,从而削弱人类建议者的作用(叶尔钦等,2022)。

(3)自我效能感。自我效能感被定义为“个体对自己是否有能力完成某一行为所进行的推测和判断”(班杜拉,1982),它往往与个体的积极行为产生联系,例如自我效能感高的员工,其工作积极性和绩效也会越高(索萨等,2012)。

在人机交互中,自我效能感受损的个体可能产生算法厌恶,这一现象主要发生在零工经济领域。例如,一项对零工经济劳动过程的研究发现,零工工作者认为自己的行为时刻受到算法的监视,需要被迫接受系统安排的指令,这使得零工工作者的自我效能感降低(默尔曼等,2021),并产生算法厌恶。又如美团外卖研究中,骑手的选择受到算法的强制安排,其配送任务分配、路线时间规划都是由算法提供,骑手仅仅拥有形式上的拒绝权(李胜蓝、江立华,2020;谢小云等,2021)。算法平台的介入还导致了控制权的重新分配(刘善仕等,2022),例如将监督权力分配给消费者,从而对工作者实施更为严格的约束和评价,这使得零工工作者的自我效能感进一步受损(陈龙,2020),从而导致算法厌恶。

综上,本文讨论了过度自信、错误归因和自我效能感在算法厌恶研究中的作用;然而如前所述,算法厌恶是人类的一种非理性行为,对人类自身和企业都会产生不利影响。因而理解算法厌恶的边界,即哪些条件下算法厌恶会被强化或者弱化,对用户和企业具有重要意义。

(二)算法厌恶的边界

根据对文献的系统性梳理,算法厌恶的边界主要包括三大类因素:个体因素、情境因素、算法因素,其中个

理论述评

个体因素是难以操纵的因素,而情境因素和算法因素则是可操纵因素,具有较强的实践意义。接下来,本研究首先阐述个体因素在算法厌恶中的边界。

1. 个体因素

已有研究指出,个体因素会调节算法厌恶,这些个体因素包括性别、年龄、大五人格和职业。

(1)性别。性别对于算法厌恶的调节作用存在争议。以金融领域为例,研究表明相较于男性,女性在进行决策时对自身专业技能更缺乏自信,因此,女性更有可能考虑外部建议(洛伦索等,2020)。相反地,也有研究指出,女性认为算法决策是更低效的,因而更可能考虑自己的观点(阿劳霍等,2020)。

(2)年龄。年龄对于算法厌恶的调节作用也存争议。有研究发现年龄越大的人越不喜欢算法推荐(洛格等,2019);然而,也有研究发现人们的年龄与算法依赖呈正向关联,越高年龄段的人越可能采用算法的建议(阿劳霍等,2020)。有趣的是,当算法出现错误时,不同年龄段的人对算法展示出了相同的信任衰减和行为模式(洛格等,2019)。

(3)大五人格。大五人格认为人有5种核心特质:开放性、负责性、宜人性、外倾性以及神经质(鲍尔、普鲁斯,2015),这5种特质对算法厌恶的调节作用也不一致。开放性、责任心、宜人性程度越高,产生算法厌恶的可能性越低;外倾性、神经质程度越高,产生算法厌恶的可能性越高(范埃施等,2021;狄弗斯特、巴蒂,2020)。

(4)职业。劳动者从事的职业在任务属性、技能宽度等方面呈现异质性,从而影响其被算法替代的可能性。职业差异的存在导致人们承担的被替代风险烈度不一,因而产生了不同程度的风险感知。在上文朴素信念分析中提到,算法会对人类的社会功能价值产生替代威胁,进而引起人们的警惕乃至厌恶情绪。唐小飞等(2021)研究发现类人工智能机器人的社会价值替代正向影响价值感知,而社会存在价值替代比社会功能替代引发的风险感知程度更高。王林辉等(2022)对人工智能背景下职业可替代风险的研究指出,非程式化、注重思维、社交和综合能力的职业被算法替代的风险更小,因而从事职业具有这些特点的劳动者所产生的算法厌恶较小;相反,技能单一、算法可复现程度高的职业更容易被替代,因而这类劳动者会产生较强的算法厌恶。徐鹏和徐向艺(2020)对人工智能时代的企业管理变革进行探究,指出人工智能会产生越来越多的“人工智能员工”,导致管理和伦理等一系列挑战,这其中也包括了职业被替代引发的算法厌恶问题。

此外,关于个体特质,也有学者从职业专业性角度(算法熟悉度)对算法厌恶展开讨论,拥有丰富的专业知识的专家(对算法越熟悉),越可能对算法产生厌恶(伯顿等,2020)。也有研究指出,算法熟悉度与人们对算法的态度呈U型(艾伦、乔杜里,2022)。具体来说,当人们对算法的熟悉度为中等时,算法厌恶程度最低。这是因为算法熟悉度较高的人可能会经历更多的算法错误,导致算法厌恶(刘等,2019),而对于新手来说,算法是一个神秘、不透明的黑箱,他们会对算法保持怀疑(卡斯特维基,2016)。

2. 情境因素

在怎样的情景中算法厌恶会被激发或削弱是研究界关注的重要话题之一。已有研究发现的语境主要与朴素信念有关,强调人们对算法的能力评估与情境的适配性。基于对已有研究的梳理,本研究讨论4种具有代表性的情境因素。

(1)行业特性。阿西莫格鲁和雷斯特雷波(2020)基于美国数据得出,算法在高度机械化行业更多地产生消极作用,而轻型制造业却得益于算法协助。李磊等(2021)基于中国样本的实证分析显示,在一些资本密集型行业,算法对就业和绩效均形成积极影响,而在部分劳动密集型行业,算法会产生替代效应,引发就业问题。尽管研究结论有相异之处,但却共同反映了行业异质性带来的替代效应程度差异,进而可能导致不同程度的算法厌恶。何宇等(2021)也指出人工智能技术带来机遇的同时,也会引发社会生产力变革和全球价值链重构,对发展中国家的冲突尤其强烈,同时隐含着算法厌恶问题。

(2)主观/客观情境。当任务是难以预测和量化的主观任务时,人们会更多地排斥算法,因为人们会认为主观任务需要直觉,而算法不具备这一能力。例如,在涉及电影、书籍乃至约会对象的推荐等具有主观性的

任务时,人们会更信任人类建议而非算法建议(卡斯特洛等,2019)。相反,当任务是客观任务时,如数学计算,人们会减少对算法的抗拒,这是因为客观任务需要逻辑分析,而算法正好具备这一能力(卡斯特洛等,2019)。

(3)享乐/实用情境。当产品突出享乐性时,人们会产生算法厌恶,这是因为人们认为算法不具备理解消费者情感需求的能力;当产品突出实用性时,人们则会增加对算法的偏好,这是因为人们认为算法具备理解人们逻辑需求的能力(隆戈尼、齐安,2022)。

(4)任务难度。一般而言,任务难度越高,个体越倾向于使用算法来辅助或替代完成任务,也就是算法厌恶的程度越弱。依据认知心理学观点,任务难度一般取决于个体在进行任务中消耗的认知资源程度,消耗的认知资源越高,任务难度也越高(罗映宇等,2021)。例如,当任务涉及逻辑性问题(如对算力、逻辑推理能力需求高)时,个体处理这些任务需消耗大量认知资源,他们会更愿意接受算法的帮助,算法厌恶被削弱(隆戈尼等,2019;格兰努罗等,2021)。相比于一个菜鸟配送骑手,一个资历丰富的骑手对配送路线耳熟能详,他们更可能违背算法的指令,依据自己的经验进行配送(朱国玮等,2023)。此外,职业特性也会调节任务难度对算法厌恶的影响。例如,对普通人而言,开展创造性活动(如摄影、绘画)需要消耗大量认知资源,他们往往愿意接受或使用算法的辅助,如使用人工智能生成内容(Artificial Intelligence Generated Content,简称AIGC)来生成绘画;而对摄影师或画家而言,这一活动被视作是简单任务,他们往往不愿意接受算法的帮助,算法厌恶程度更高(张伯伦等,2018;卡斯特洛等,2019)。又如,相比于主观性强的职业(如画家、音乐家),一些客观性强的职业(如工程师、精算师)对算力需求更高,他们更乐于寻求算法帮助,算法厌恶程度削弱(恩卡尔等,2009;卡斯特洛等,2019;张等,2021)。

除上述研究,研究者还探讨了解释水平(高 vs. 低)、道德情境、推荐情境(替代 vs. 增强,确定需求 vs. 非确定需求)、决策结果(喜欢 vs. 不喜欢,随机结果 vs. 非随机结果,尴尬 vs. 不尴尬)等因素在算法厌恶中的边界(余伊琦等,2022;金姆、杜哈切克,2020;叶尔钦等,2022;吴继飞等,2022;霍特维尔、范多恩,2022;周等,2022;韩等,2022)。

3. 算法因素

算法因素主要探究算法拟人化、算法角色、算法可解释性以及人机回圈的调节效应。

(1)算法拟人化。算法拟人化分为外表拟人化和行为拟人化。外表拟人化指算法赋能的机器人在形式上与人的相似性(格里克森、沃利,2020;缪等,2022)。学者们指出外表拟人化会影响人们接受算法的建议(门德等,2019)。例如,当机器人的外表与人类相似时,人们会认为它们具有与人类相同的理解情感的能力,会更愿意接受它们的建议(隆戈尼、齐安,2022;亚姆等,2021;门德等,2019;普拉尔、范斯沃,2017);并且认为机器人与人类具有相同的思维模式,会提升对机器人服务的使用态度和意愿(普拉尔、范斯沃,2017)。然而,越来越多研究指出,算法拟人化会带来负面影响。例如,由于外表拟人化赋予了算法与人相同的“能力”,当愤怒的消费者遇到拟人化算法时,极可能加重消费者的负面反应。并且,消费者会认为拟人化算法具有与人一样的承担责任的能力,当拟人化算法发生错误时,人们更可能对算法以及使用算法的品牌产生负面评估(斯里尼瓦桑、沙利叶,2021);当服务失败时,拟人化算法也会加剧顾客对算法的厌恶感和负面态度(王海忠等,2021)。

行为拟人化指算法赋能的机器人在行为上与人的相似性(格里克森、沃利,2020),直接影响了人们对算法内在逻辑的感知。例如,一项研究将拟人化算法推荐与人的推荐效果进行比较,研究者向被试者呈现两种类型的推荐:低行为拟人化推荐(身体同步或嘴唇同步)、高行为拟人化推荐(身体和嘴唇同步),研究者发现相比于高行为拟人化,低行为拟人化的算法推荐更受欢迎(拉格尼等,2016)。综上,目前的研究认为完美、行为不犯错的算法更可能引起人们的高度不适。

(2)算法角色。用户对算法产生接受或拒绝行为,往往受算法在其行为中扮演的角色所影响。例如在企业管理中,员工会认为算法是控制者,因而会对算法指令做出反抗(默尔曼等,2021);在采用机器人的企业,员工

理论述评

会对自身的价值产生担忧,认为机器人会成为自己的替代者,这也会导致算法厌恶(李等,2019;威尔森等,2017;蓬托尼等,2021)。基于这一思考,一些研究者提出转变用户对算法角色的认知,从而影响用户反应。例如,将算法形象描述成“协助者”(vs.“替代者”)有助于提升消费者对产品的评价(王欣等,2021);或是将算法概念化为增强能力(vs.替代能力)的实体有助于提升消费者对推荐者的态度,并改变购买行为(隆戈尼、齐安,2022)。

探讨算法厌恶中的算法角色对企业管理具有重要意义,然而目前的研究对此关注甚少,导致这一现象的原因是当前研究对于算法厌恶中的算法角色划分及其理论基础仍不清楚。为扩展这一问题,本研究在研究展望中从算法的表现形式和行为感知两个方面尝试将算法角色划分为6个角色,并探讨概念形成和发展的方向。

(3)算法可解释性。算法决策是一个黑箱,人们并不清楚其内在逻辑和结果输出方式(卡斯特维基,2016)。心理学研究指出,虽然人与算法的决策都是不透明的,但是人可以通过自省的方式了解自身的决策过程,降低对自身决策不透明的感知;然而,对于算法决策,人们无法进行这一过程。对此,人们需要通过额外的方式增加对算法的了解,提升算法透明度。

增加算法的可解释性,建立透明的算法机制有助于削弱算法厌恶(巴比奇等,2021;罗谢洛、帕克尔,2022)。人们可以接受外部的训练来了解算法。例如,通过知识学习把握算法决策的重要性,指导如何正确与算法交互、理解数据以及结果输出的内涵(伯顿等,2020)。企业也可通过隐私政策(或其他方式)告知用户算法如何使用和输出结果。例如研究发现,零工工作者对于平台算法产生隐私担忧,往往会选择不使用算法平台的企业。然而,当平台向员工提供透明度的解释,告知他们平台如何收集和使用隐私信息时,可以缓解他们的隐私担忧和算法厌恶(梁等,2022)。但是,提升透明度也可能产生负面影响,如果平台使用的算法比较复杂时,增加透明度会增加用户的信任;但当平台使用的算法比较简单时,增加透明度则会降低用户的信任(雷曼等,2022)。

(4)人机回圈。决策者在进行人机交互时,需要对算法抱有信心才会采纳其建议(伯顿等,2020),而信心的缺失一部分来源于人类对算法错误的认知。若能够参与修正算法错误或最小化算法错误发生的可能性,人们将更倾向于依赖并不完美的算法建议(狄弗斯特等,2015),这是因为参与修正或降低算法错误可能性可以提高人们对算法模型的信心。这类提升信心的模式可加强使用者体验感与参与感,增进理解,构成一种被称为“人机回圈”(Human in the Loop)的人机交互方式。人机回圈被定义为:将个人的判断融入算法系统流程中去,构成流程运行闭环的人机交互方式(富格纳等,2021)。人机回圈可作为削弱算法厌恶的一项重要策略。

人机回圈具有两个目标:一是让个人参与算法决策系统流程,使算法决策更为有效。例如一项研究表明,允许对算法进行修正即可大幅降低算法厌恶感,即使修正权是受到限制的(狄弗斯特等,2018;川口,2021)。那么,个体拥有的权限越高,他们对算法的接受度是否也会越高?答案是否定的。研究指出,人们在意的不是能够修正多少,而是能否修正(狄弗斯特等,2018)。因此,只要给予人类一定的算法修正权限,就能降低人类的算法厌恶。

二是使算法更多地考虑人类偏差,提升算法决策效果,从而削弱算法厌恶。例如,孙等(2022)在阿里巴巴的物流包装效率的研究中发现,算法的决策效率总是与预期的结果存在偏差,其原因可能是员工比算法掌握了更多物流包装相关的知识。据此,该研究提出将个体的偏差纳入到算法的优化之中,可以较好地提升算法表现,使得员工更可能采纳算法(孙等,2022)。此类对算法的完善,促进了算法修正错误并减少算法错误的可能性,提高了算法预测准确性,在降低算法厌恶方面有重要作用。

综上所述,本研究从两个方面归纳了人机回圈对削弱算法厌恶的贡献。一是允许个人参与算法决策程序,提供输入,但由人类负责最终结果(伯顿等,2020),即人类辅助算法学习。二是机器学习算法考虑人类偏差并提升决策效果,从而利用算法决策改进人工任务(孙等,2022),即算法学习辅助人类。两者的有机结合对

于克服算法厌恶有重要的指导意义。

五、人工智能时代算法厌恶的未来研究议题

尽管已有的研究对算法厌恶现象及其内在机理展开了丰富的讨论,并从多个维度探讨了算法厌恶效应边界,但现有研究仍存在较丰富的拓展空间。基于对已有文献的梳理、述评与讨论,结合理论不足与现实困境,本研究提出算法厌恶领域具有科学价值及实践意义的六大未来研究议题。

(一)立足新视角、采取新方法,进一步深化算法厌恶的成因研究

本文基于“算法厌恶类型与注意视角的匹配”对算法厌恶的内在机制进行了梳理。研究有以下发现:第一,现有研究主要从机器人特征和身份威胁两种朴素信念考虑算法厌恶的成因,如传统的研究认为机器人是刻板、无法学习的(道斯,1979;格鲁夫、米尔,1996)。然而朴素信念并不稳定,易受自身经历和社会环境的影响(莫尔登、德维克,2006),如近期火爆的人工智能生成内容正在逐渐改变人们对AI的刻板认知。因而,未来研究需要考虑朴素信念的不稳定性,适时调整已有观点。

第二,现有算法厌恶研究主要从管理学、心理学视角讨论成因,缺乏经济学视角的研究。根据经济学观点,对于厌恶风险的经济主体(如劳动者、企业家、消费者),算法风险必然会产生算法厌恶。因此,未来研究可将隐私担忧、算法错误等纳入算法风险概念,全面构建不确定决策视角下的算法风险决策框架,并从经济学视角对算法厌恶成因提供更普适的解释。例如,将算法风险分为内部风险(如算法模型风险、数据风险等)和外部风险(如金融风险、社会风险等),并以此构建算法风险决策框架,评估和测量算法厌恶,实施算法风险管理。

第三,现有研究对算法厌恶的解释机制存在矛盾性问题。例如,关于算法是否客观的问题,有研究认为算法是客观的,因为它是机器人,采用理性思维(隆戈尼等,2019;格兰努罗等,2021);也有研究认为算法是片面、不客观的,因为算法是以人们提供的数据为基础,而人们的行为带有主观色彩(比格曼、格雷,2018;吉鲁克斯等,2022;博纳丰等,2016;基尔,2020)。针对矛盾问题,已有研究考虑了算法熟悉度的边界作用(罗等,2021)。未来研究可以进一步识别更多边界条件,并引入神经科学设备开展交叉研究,如通过脑电设备记录个体复杂的认知活动并输出为脑电信号(施卓敏、张珊,2022),为解决算法厌恶中的解释机制悖论提供一种新颖且科学的视角。

第四,学者们围绕算法与应用场景的匹配开展了相关研究并获得一些成果,但仍有一些模块值得未来探索。例如,已有文献对嵌入式算法和算法赋能机器人的厌恶机制展开了相对丰富的研究(富格纳等,2022),但还有一些板块值得进一步探索,如算法赋能的虚拟和实体机器人在隐私担忧、过度自信、错误归因等机制下的探索。表1总结了算法厌恶驱动机制与算法表现形式的匹配研究及未来可能可以进一步探索的方向。

(二)进一步关注算法决策黑箱,探索削弱算法厌恶的理论方案

在人工智能时代,基于算法的数字化管理、平台治理已经成为未来发展的重要方向。然而,人们在多数情况下并不清楚算法的运作逻辑和结果输出依据(卡斯特维基,2016),算法歧视、算法偏见等负面现象频频发生。因此,针对算法决策“黑箱”进行有针对性的阐述或解释,提出相应的解决方案,对企业管理和社会治理显得尤为重要(肖红军,2022;雷,2020;裴嘉良等,2022)。

第一,研究者应明确算法黑箱的表现形态,以提出相应的应对方案。例如,作为符号代码的算法“黑箱”,可以通过展开代码培训、提高普通公众的代码素养来解决。第二,研究者应关注算法可解释性的双刃剑效应。例如,有研究发现当平台使用的算法比较简单时,增加透明度会降低用户信任,而算法较为复杂时,增加透明度会提升用户信任(雷曼等,2022)。第三,研究者可对解释方式的效果予以进一步讨论。算法可解释性包括解释算法能力和解释算法决策,研究者需要明确不同解释方式的适用边界。例如,当用户是以结果为导向

表1 算法厌恶驱动机制与算法表现形式的匹配研究及未来探索方向

	算法视角			人类视角			其他机制
	朴素信念	算法错误	隐私担忧	过度自信	错误归因	自我效能感	
嵌入式算法	√	√	√	√	√	√	...
虚拟机器人	√	√	○	○	○	√	...
实体机器人	√	○	○	○	○	○	...

注:该项匹配结果仅考虑目前研究的观点。√表示与已有研究相匹配的机制;○表示未来可能可以进一步探索的方向。

理论述评

向时,企业向该类用户解释算法决策还是解释算法能力的效果更好,而当用户是以体验为导向时,是否又会有所不同?未来研究(特别是人力资源领域)需要在各种情境中考虑解释算法能力还是算法决策的问题。第四,研究者可将决策黑箱纳入算法风险决策框架中,从经济学角度对算法风险以及算法厌恶进行测量,进而采用行为干预(如隐私政策、启动效应等)来管理算法风险和算法厌恶。例如,研究者发现零工工作者总是在隐私收益和损失之间做出权衡,他们一方面按照平台设定的规则进行工作,另一方面担心平台侵犯隐私;而当平台提供隐私政策来解释算法对隐私的使用模式时,他们对平台表现出更高的信任(梁等,2022)。

(三)关注算法厌恶量化研究,推动算法厌恶程度指数构建,促进算法厌恶后果研究

目前算法厌恶的研究主要采用二分法(算法 vs. 人类)的定性方案展开,如在面对同样的服务或建议时,用户更倾向于接受算法还是人类的服务或建议(默尔曼等,2021;孙等,2022;隆戈尼等,2019;川口,2021;比格曼、格雷,2018;吉鲁克斯等,2022;基尔,2020;葛等,2021),鲜有学者对算法厌恶进行量化研究。关注算法厌恶量化研究、推动算法厌恶程度指数构建,并促进算法厌恶后果的研究,有利于更深入理解算法厌恶的成因和边界,还能够为企业评估算法厌恶策略有效性提供定量测度工具,为降低算法厌恶程度提供更有效的理论指导。

算法厌恶量化研究可从底层理论探索与实证研究两方面同时展开,以相互印证和改进。算法厌恶量化研究的底层理论构建是指通过给定一系列假设,从数学上证明算法厌恶函数的存在,并对其性质加以研究,为算法厌恶提供数学角度的扎实证据。然而在算法厌恶函数存在的前提下,依然可能存在两方面问题。第一,数学上只能证明算法厌恶函数的存在性,但无法具体写出其解析式;第二,算法厌恶函数存在解析式,但根据解析式得到的具体结果与现实情况大相径庭。若存在第一个问题,则说明底层理论仅能为算法厌恶的存在性提供依据,但管理实践意义较弱。若存在第二个问题,则说明由于假设过强导致理论结果与现实情况不符。为了印证或者避免理论构建的无效性或者难以推进,算法厌恶量化实证研究应该与底层理论构建同步进行。

算法厌恶的量化实证研究可以考虑通过构建算法厌恶量化模型的方式来计算用户的算法厌恶程度指数,即当用户接收到算法提供的服务或建议时,基于用户的一系列信息,构建量化模型计算用户的算法厌恶程度。具体而言,通过纳入用户信息(如消费者/员工),包括人口统计特征(如性别、年龄、教育背景、家庭特征、人格特质等)和历史行为信息(如历史购物、浏览行为、社会活动轨迹、在线虚拟活动信息等)及其动态变化等,以及外部环境特征(如政治、经济、社会、技术等)及其动态变化等,并对这些指标和信息赋予一定权重,计算得出用户的算法厌恶程度指数。基于用户的算法厌恶程度指数,结合具体研究或者管理问题,探究有效降低(或逆转)用户算法厌恶程度的手段和策略。此外,研究者也可开展算法厌恶量表开发研究,设计适用的算法厌恶测量工具,以进一步推进算法厌恶的量化实证研究。

未来研究还需进一步关注算法厌恶所产生的负面后果。已有研究主要从算法厌恶现象出发,重点揭示在不同情境中产生的算法厌恶的机制以及削弱算法厌恶的边界条件(李等,2019;威尔森等,2017;蓬托尼等,2021;隆戈尼等,2019;隆戈尼、齐安,2022;金姆、杜哈切克,2020;斯里尼瓦桑、沙利叶,2021;默尔曼等,2021;郁苏波夫等,2021)。但是,鲜有研究关注到算法厌恶对个体、组织和社会产生的后果。例如,外卖骑手对平台算法管理产生的算法厌恶是否会对影响企业的员工招募、算法使用、企业绩效甚至战略规划产生影响?未来研究可以结合算法厌恶的量化模型对算法厌恶后果进行实证研究,揭示不同算法厌恶程度与相应负面后果的关系;也可聚焦利用算法开展管理实践的企业或组织部门,以案例研究方法对算法厌恶后果展开探究。

(四)发展算法角色理论框架,多方面、多领域拓展算法角色的应用研究

人工智能算法角色在算法厌恶领域有重要影响,但已有研究存在较大缺口,亟需发展算法角色理论框架,并进行多方面、多领域拓展。本文从以下两部分提出算法角色理论议题:首先,依据解释水平理论和行为动机理论,本文初步提出一个算法厌恶背景下的算法角色理论框架。其次,依据算法角色理论框架,本文进一

步提供两个应用研究方向,以对该领域的研究提供引导。

1. 算法角色理论构建

算法表现形式可分为有形(实体机器人和虚拟机器人)和无形(嵌入式算法),无论是其表象形式还是可理解性均从具体向抽象发展,而用户在与算法交互的过程中,对算法动机的感知极为重要。因此,本文在解释水平理论和行为动机理论基础上,提出一个基于用户视角的算法角色理论初探方案(图3)。

在水平视角,本研究通过解释水平理论讨论算法表现形式与算法厌恶的关系。解释水平理论将客体的表征形式分为高解释水平和低解释水平。在高解释水平下,人们会采用本质的、抽象的特征描述事物,抽象程度高,如算法。在低解释水平下,人们会采用表面的、具体的特征描述事物,抽象程度低(特罗普、里伯曼,2010),如机器人。因而,算法表现形式的划分正好与解释水平的概念相契合,即有形表现形式(机器人)对应低解释水平,无形表现形式(嵌入式算法)对应高解释水平。本研究认为算法的无形形式比有形形式更容易产生算法厌恶。这是因为:首先,道德判断是算法厌恶产生的重要因素,已有研究指出,解释水平会影响人们的道德判断,解释水平越高,人们越可能认为其存在道德问题(阿吉拉尔等,2013);其次,解释水平与个体的风险感知判断有关,人们认为高解释水平的表征具有更高的风险(特罗普等,2007);最后,解释水平还会影响人们对能力的估计,对于高解释水平的表征,人们会产生认知偏差,并高估自己的能力(索德堡等,2015),这也可能增强人们的算法厌恶。

在垂直视角,本研究探讨人们对算法的行为动机感知来影响其态度。行为动机理论将动机分为利己和利他,表达了用户对主体行为目的感知(马纳、盖略特,2007)。当用户认为算法的行为动机是有利于自身时,例如,当用户认为算法是来指导、辅助他时,会对算法产生积极的态度(王欣等,2021)。然而,当用户认为算法的行为动机是有利于企业时,例如取代动机,或是剥削动机,会对算法产生消极的态度(默尔曼等,2021)。这一概念从垂直维度对算法厌恶进行了区分。

基于上述讨论,根据算法表现形式和用户对算法行为动机的感知(利己 vs. 利他),本研究将算法角色分为以下6种:控制者、推荐者、协助者、指导者、功能替代者和功能互补者。基于理论初步分析,算法角色的算法厌恶程度由低到高分4个等级。功能互补者的算法厌恶程度最低,协助者和指导者其次,功能替代者和推荐者再次,控制者的算法厌恶程度最高,下面结合算法角色理论初探方案对这6种角色进一步论述。

嵌入式算法一般嵌套在系统中,直接提供决策结果,虚拟机器人和实体机器人都是算法嵌入到有形体的表现形式。根据行为动机可分为指导者(利己)和控制者(利他)。指导者主要是指教练系统,它通过算法帮助人们实现特定的行为目的。例如,企业管理者利用算法评估员工的工作过程,并向他们提供指导建议(罗等,2021;马利克等,2022)。控制者一般指企业的算法平台,它通过算法对个体的行为进行控制,并代理其绩效考核等评价功能。例如,零工经济平台使用算法控制众包骑手的配送路线,并根据所收集的工作信息决定骑手的绩效等级(默尔曼等,2021)。协

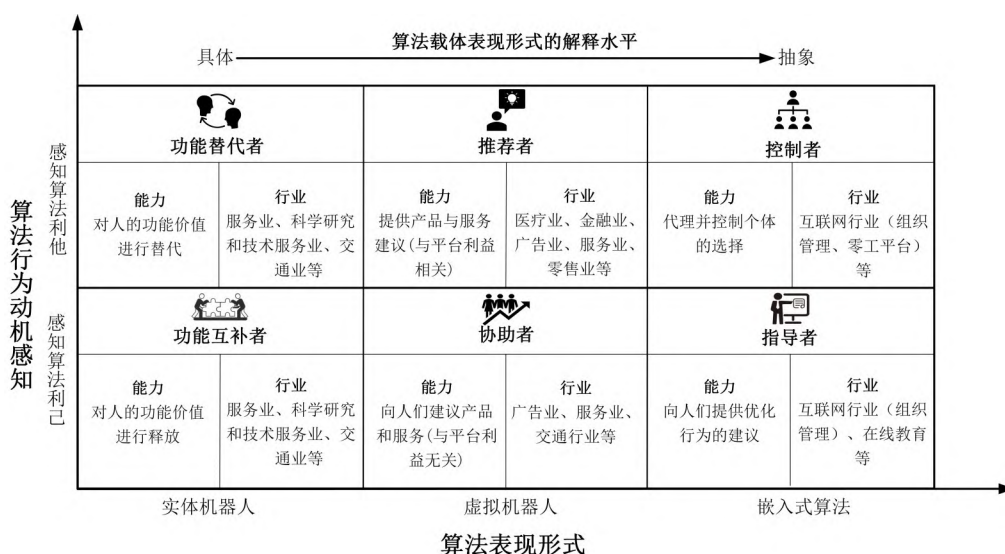


图3 用户视角下算法角色类型初探

理论述评

助者一般是指与用户进行交互的服务机器人,它通过算法程序帮助用户理解技能和服务内容并提供建议。例如,Siri语音助手根据用户偏好为用户制定路线规划。推荐者也是指服务机器人,它与协助者的不同点在于:推荐者提供的建议不仅涉及顾客利益,也涉及平台利益。例如保险机器人客服会根据用户信息提供保险服务方案(皮克特等,2022),保险服务方案同时涉及顾客和平台利益。功能替代者是智能机器人实现与人相似的功能价值,并取代部分人工作的角色。例如无人驾驶车辆使相关从业者的功能价值被替代(李等,2019;威尔森等,2017;蓬托尼等,2021)。功能互补者是智能机器人实现与人相似的功能价值,将人们从冗余、繁琐的任务中释放出来的角色。例如,酒店管理者采用智能机器人实现服务智能化,缓解员工工作压力(李等,2019)。两种角色的差异在于:智能机器人从事的工作对人类工作的重要性。如果智能机器人替代了人类认为重要的工作,就可能产生威胁;如果智能机器人替代了人类认为不重要的工作,人类反而会认为工作流程得到了优化。

2. 算法角色的应用研究方向

基于上述理论框架,本文进一步提供两个发展算法角色的应用研究方向。应用方向一:转变算法角色,削弱算法厌恶。对已有研究的总结发现,算法角色与应用领域具有一定的适配性。例如,在人力资源领域,研究者主要从控制者角色展开算法厌恶的研究;而在营销领域,研究者主要从推荐者角色展开研究。未来研究可以在多种领域中尝试转变用户对算法角色的感知,从而克服算法厌恶:(1)人力资源领域,在员工心中塑造算法是有益于工作的指导者而非是冷血、剥削的控制者的角色;或是在员工心中塑造算法是功能互补者而非是功能替代者的角色;(2)市场营销领域,在消费者心中塑造算法是帮助他们获得积极消费体验的指导者而非是机械的推荐者角色。

应用方向二:多视角探讨算法角色的边界。第一,从垂直维度,本研究认为用户感知算法是有益于自我还是他人主要归为以下3个因素:(1)取代能力的重要性;(2)决策是否涉及利益;(3)是否影响了决策自由。未来的研究可考虑从个体异质性和情境特殊性角度探讨研究角色的边界。例如,对于控制感需求低的人,他们是否会将控制者和指导者视作相同的角色?第二,从水平维度,相比于机器人,嵌入式算法更为抽象,这可能增加用户对嵌入式算法的厌恶。因而,增加透明度是削弱人们对嵌入式算法负面态度的理论方案之一。未来的研究可讨论如下议题:(1)增加透明度是否会削弱人们对嵌入式算法的算法厌恶?(2)增加透明度对算法厌恶的削弱效应在机器人中是否依然有效?(3)未来研究可基于本文提出的理论框架,从不同视角拓展和完善算法角色理论框架。

(五)探索和丰富人工智能时代下算法青睐的概念内涵及作用机制

在最近的研究中,越来越多的研究者关注到一种与算法厌恶相反的现象,即算法青睐(Algorithm Appreciation)。洛格等(2019)是较早发现这一现象的学者,他们将算法青睐定义为人们在做决策时倾向于接受算法建议和服务的现象(洛格等,2019;尤等,2022)。例如,对于一些具有具体的外部准确性标准的领域,如投资决策、体育预测等,人们会产生更高的算法适用性感知,并引发算法青睐(洛格等,2019)。当人们理解算法运作的原理时,人们会更青睐于算法提供的决策建议(约曼斯等,2019)。通过对已有文献的梳理,发现算法青睐的研究从数量上远低于算法厌恶,学者对于算法青睐成因的理解仍处于初步阶段。未来研究还需提炼和界定人工智能时代背景下的算法青睐的概念内涵,并拓展该领域研究,以作为算法厌恶的姊妹篇形成人工智能时代算法决策研究全貌。

研究者可以借鉴本研究框架探索算法青睐的成因。例如,研究者可以考虑在人们的认知中,算法何时比人类表现得更好,也许在涉及主观性和情感的领域,人们更可能表现出算法青睐。又如,算法错误的研究指出人们会对发生错误的算法展示厌恶;然而也有研究指出,人们对完全精准、不会犯错的算法也会产生抗拒,这是因为用户会产生被“揣测”的感觉(金姆等,2019)。例如,美国曾报道一项著名的隐私侵犯事件,美国第二大零售商塔吉特使用算法向一位父亲的女儿邮寄婴儿服装和孕妇优惠券,即使塔吉特的算法预测是准确的,但是仍然受到社会的强烈抵制。因而,未来的研究可以在涉及隐私的领域考虑算法出现一些错误反而会导致用户青睐的可能性。值得注意的是,现有研究对于算法厌恶和算法青睐的边界仍未清晰。一些研究认为它们是

同一行为的两个维度,例如青睐算法 vs. 厌恶算法,而非是厌恶算法 vs. 厌恶人类。本研究认为这一问题需要展开严格讨论。依据定义,算法厌恶代表人们对算法决策结果的厌恶倾向;而算法青睐代表人们对算法决策结果的积极偏好;不厌恶算法并非表示青睐算法。并且,算法厌恶研究中人们对代理者(算法或人)的接受程度普遍要低于算法青睐。未来的研究需要进一步讨论这两者的差异,并将算法厌恶和算法青睐的相关研究进行区分。

(六)发展中国背景下的算法厌恶理论体系和数字中国建设的相关研究

党的二十大报告明确指出要“加快发展数字经济,促进数字经济和实体经济深度融合,打造具有国际竞争力的数字产业集群”^⑩。算法作为数字经济的关键要素,需充分考虑中国背景的特殊性,以发展中国特色算法风险治理和算法厌恶理论。

第一,考虑中国数字经济应用场景的特殊性。例如,在零工经济领域构建算法厌恶理论研究体系。2022年,中国有2亿人口参与灵活就业,这一数据在2036年可能达到4亿^⑪,这使得“零工经济”成为中国组织管理的重点话题(裴嘉良等,2021)。通过算法管理,企业实现了对大量零工工作者劳动过程的匹配、控制和优化;但这也激发了零工工作者对算法代理的抗拒。与西方不同,中国的零工经济有其自身的独特性,零工工作者产生算法厌恶的机理亟需进一步探索。一方面,与西方参与零工经济的人群相比,中国的灵活就业群体走向年轻化,同时,灵活就业的边界逐渐扩大,和传统职业有一定的交叠。另一方面,随着社会形态的变化,中国的零工经济可能会延伸到一些新兴职业中,使得研究情境变得多元且更加复杂。未来研究需要考虑中国人口结构、职业类型以及文化背景的特殊性,在零工经济领域发展中国本土化的算法厌恶理论研究。另外,中国正在推进智慧城市和数字乡村融合发展,如何在城乡协同的数字经济应用中,探索不同经济主体对算法管理的反应,促进城乡融合发展中的数字融合,也是值得探讨的话题。

第二,考虑中国社会历史文化的特殊性。算法的训练数据主要源自西方社会数据或混合群体数据,其行为模式更符合西方逻辑。与西方社会相比,中国的文化演进更为连续和深厚,中国社会具有深厚的文化传统和风俗习惯,这可能导致算法的决策产生偏见。例如,儒家思想对中国社会产生深远影响,强调家庭、孝道、和谐等价值观,而西方社会价值观较为多元和个人主义。这种价值观的差异可能导致西方算法决策逻辑难以适配中国情境,从而产生算法厌恶。未来研究可以考虑中国社会历史文化特征,揭示产生算法厌恶的历史文化因素,并借鉴中国传统文化理论(如费孝通先生提出的“差序格局”理论),提出克服算法厌恶的理论方案,推动算法技术的应用和发展。

第三,考虑中国经济、人口的特殊性。目前,人口老龄化问题正日益凸显,破除老年人与算法的数字鸿沟,促进两者的数字融入,是推动高质量发展的重要话题。一方面,未来研究可以在算法与老年人群交叉的应用领域,如医疗保健、智能辅助技术、社会福利、养老服务等领域,探讨老年人群对算法服务的积极和负面反应。例如,在医疗服务中,老龄人群是更容易接受嵌入式算法还是机器人提供的诊疗建议?另一方面,研究者也可探讨老年人群特征(如学历、城乡地域、家庭等)对算法技术的反应,并探索合适的方案促进老年人对算法技术的接受。例如,相比于低学历老年人,高学历老年人是否更愿意接受来自算法的服务?这种行为偏见是否会受到外部因素(如政策、代际支持等)的调节?总之,中国数字化实践场景为未来算法厌恶的研究提供了广阔的讨论空间,研究者需扎根于中国数字经济发展的土壤,构建中国特色的算法风险治理和算法厌恶理论体系,推进中国数字经济的高质量发展。

六、总结

近年来,算法厌恶是被管理学、社会学、心理学等多个学科热议的话题之一。随着机器学习算法愈发成熟,越来越多的研究者将目光转向人机交互中的算法决策,并聚焦于算法厌恶这一应对算法管理挑战的重要议题。自2015年起,学术界涌现了算法厌恶研究的浪潮,大量研究出现在国内外顶级期刊上,推动其成为当前研究的主流方向之一。基于此,本研究以算法管理挑战为出发点,以“人工智能时代的算法厌恶”为研究主

题,提出算法厌恶研究框架。研究首先采用系统文献综述法对2015~2023年间发表的与主题相关的百余篇核心文献进行系统梳理和总结,以此为基础提出人工智能时代算法厌恶的定义,并对算法厌恶的成因和边界展开探究,研究构建了算法厌恶的研究框架。在该框架中,算法厌恶的成因从算法视角(朴素信念、算法错误和隐私担忧)和人类视角(过度自信、错误归因和自我效能感)两个方面展开论述,算法厌恶的理论边界则从个体、情境和算法3个维度展开论述。研究还基于理论和现实需要,提出具有科学价值和现实意义的6项未来研究议题,以期深化未来对算法厌恶的研究。

本文对算法厌恶理论研究以及管理实践做出如下贡献。首先,本研究对算法的内涵和外延进行了清晰界定和阐述,明确提出人工智能时代算法厌恶的定义。其次,本文是较早提出并构建算法厌恶研究框架的文献。研究从算法视角和人类视角双维度论述算法厌恶机制,并从不可操纵因素(个体因素)和可操纵因素(情境因素和算法因素)两部分论述算法厌恶的理论边界。本文从算法厌恶的成因深化、削弱算法黑箱引发的算法厌恶的理论方案、算法厌恶量化研究下的算法厌恶指数构建及后果研究、算法角色理论框架的发展与应用研究、算法青睐的概念内涵及作用机制研究,以及中国背景下的算法厌恶理论体系构建等6个议题,提出具有科学价值和实践意义的六大未来研究方向。特别地,本研究基于解释水平理论和行为动机理论,创造性地构建了算法厌恶中的算法角色框架,该框架从科学性和实践性角度描述了算法角色在算法厌恶中的作用,有助于学者厘清算法厌恶的内涵,并推动算法角色相关的研究。最后,本研究向管理者提供了一个理解算法厌恶成因和应对策略的系统框架,帮助管理者明确算法厌恶的发生情境,以及采取怎样的措施能够削弱甚至逆转算法厌恶。研究对促进中国企业有效应对算法管理挑战、推动中国人工智能健康发展、推进中国数字社会建设,均具有重要意义^⑩。

(作者单位:罗映宇、朱国玮、钱无忌、黄静、杨智,湖南大学工商管理学院;吴月燕,福州大学经济与管理学院)

注释

①观点来源:中华人民共和国中央人民政府,《人工智能时代已经来临》, https://www.gov.cn/xinwen/2018-03/25/content_5277158.htm, 2018年。

②观点来源:中华人民共和国国家发展和改革委员会,《人工智能时代:蓝图这样变为现实》, https://www.ndrc.gov.cn/xwdt/ztlz/szhzhbxd/yjjy/202006/t20200623_1231816.html, 2020年。

③观点来源:新华网,《未来已来——多国专家谈人工智能与人类劳动》, http://www.news.cn/tech/2023-05/01/c_1129583707.htm, 2023年。

④观点来源:Bill Gates,《The Age of AI has begun》, <https://www.gatesnotes.com/The-Age-of-AI-Has-Begun>, 2023年。

⑤观点来源:新华网,《AI时代,我们准备好了吗》, http://www.news.cn/fortune/2023-04/12/c_1129514012.htm, 2023年。

⑥⑨观点来源:中华人民共和国中央人民政府,《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》, https://www.gov.cn/xinwen/2021-03/13/content_5592681.htm, 2021年。

⑦麦肯锡研究报告:《Global survey: The state of AI in 2021》, 2021 <https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021>, 2021年。

⑧观点来源:埃森哲,《2022中国企业数字化转型指数》, 2022年。

⑩中国信通院,《中国数字经济发展研究报告(2023年)》: http://www.caict.ac.cn/kxyj/qwfb/bps/202304/t20230427_419051.htm, 2023年。

⑪观点来源:腾讯, <https://new.qq.com/rain/a/20220723A01F5F00>, 2022年。

⑫中外文人名(机构名)对照:阿明(Ameen);爱彼迎(Airbnb);程(Cheng);佛利(Foley);艾伦(Allen);乔杜里(Choudhury);默尔曼(Mohlmann);梁(Liang);隆戈尼(Longoni);狄弗斯特(Dietvorst);李(Li);威尔森(Wilson);蓬托尼(Puntoni);齐安(Cian);金姆(Kim);杜哈切克(Duhachek);斯里尼瓦桑(Srinivasan);沙利叶(Sarial-Abi);郁苏波夫(Jussupow);格兰努罗(Granulo);米尔(Meehl);格里克森(Glikson);沃利(Woolley);特兰菲尔德(Tranfield);韦伯斯特(Webster);沃森(Watson);洛格(Logg);卡斯特洛(Castelo);罗(Luo);门德(Mende);约曼斯(Yeomans);黄(Huang);拉斯特(Rust);孙(Sun);川口(Kawaguchi);比格曼(Bigman);格雷(Gray);吉鲁克斯(Giroux);基尔(Gill);葛(Ge);麦卡锡(McCarthy);罗素(Russell);诺维格(Norvig);瑞吉(Raj);希曼斯(Seamans);李(Lee);伍德(Wood);甘迪尼(Gandini);康默福特(Commerford);刘(Liu);张(Zhang);拉普(Rapp);陈(Chen);帕克(Park);昆(Kyung);权(Kwon);海豪斯(Highhouse);莫林(Morin);班克(Banker);赫塔尼(Khetani);莫尔登(Molden);德维克(Dweck);麦(Mai);道斯(Dawes);格鲁夫(Grove);斯坦因(Stein);奥赫(Ohler);赖希(Reich);博纳丰(Bonneton);费格罗亚·阿米霍斯(Figueroa-Armijos);乔希(Joshi);韦德(Wade);泰尔坎普(Telkamp);安德森(Anderson);沃德(Ward);加维(Garvey);乔德里(Chowdhury);邹(Zou);施宾格(Schiebinger);侯(Ho);麦克多曼(MacDorman);亨特(Hunt);唐(Tong);麦克里(McLeay);伯顿(Burton);齐希(Cichy);俞(Yu);纳图尔(Al-Natour);阿拉德瓦尼(Aladwani);戴恩(Dane);普拉特(Pratt);戴(Dai);塔尤尔(Tayur);穆罕默德(Mahmud);考尔德(Calder);伯恩柯兰特

(Burnkrant); 叶尔钦 (Yalcin); 班杜拉 (Bandura); 索萨 (Sousa); 洛伦索 (Lourenço); 阿劳霍 (Araujo); 鲍尔 (Power); 普鲁斯 (Pluess); 范埃施 (Van Esch); 巴蒂 (Bharti); 卡斯特维基 (Castelvecchi); 阿西莫格鲁 (Acemoglu); 雷斯特雷波 (Restrepo); 张伯伦 (Chamberlain); 恩卡尔 (Önköl); 霍特维尔 (Holthöwer); 范多恩 (Van Doorn); 周 (Zhou); 韩 (Han); 缪 (Miao); 亚姆 (Yam); 普拉尔 (Prah); 范斯沃 (Van Swol); 拉格尼 (Ragni); 巴巴奇 (Babic); 罗谢洛 (Parent-Rocheleau); 帕克尔 (Parker); 雷曼 (Lehmann); 雷格纳 (Fuegener); 雷 (Rai); 特罗普 (Trope); 里伯曼 (Lieberman); 阿吉拉尔 (Aguilar); 索德堡 (Soderberg); 马纳 (Maner); 盖略特 (Gailliot); 马利克 (Malik); 皮克特 (Peukert); 尤 (You)。

参考文献

- (1) 陈龙:《“数字控制”下的劳动秩序——外卖骑手的劳动控制研究》,《社会学研究》,2020年第6期。
- (2) 何宇、陈珍珍、张建华:《人工智能技术应用与全球价值链竞争》,《中国工业经济》,2021年第10期。
- (3) 罗映宇、孙锐、王艺翔、韦志颖:《认知资源:解释社会助长效应的新视角》,《心理科学》,2021年第5期。
- (4) 李磊、王小霞、包群:《机器人的就业效应:机制与中国经验》,《管理世界》,2021年第9期。
- (5) 李胜蓝、江立华:《新型劳动时间控制与虚假自由——外卖骑手的劳动过程研究》,《社会学研究》,2020年第6期。
- (6) 刘善仕、裴嘉良、葛淳棉、刘小浪、湛一潘:《在线劳动平台算法管理:理论探索与研究展望》,《管理世界》,2022年第2期。
- (7) 刘洋、董久钰、魏江:《数字创新管理:理论框架与未来研究》,《管理世界》,2020年第7期。
- (8) 裴嘉良、刘善仕、崔勋、瞿皎皎:《零工工作者感知算法控制:概念化、测量与服务绩效影响验证》,《南开管理评论》,2021年第6期。
- (9) 裴嘉良、刘善仕、崔勋、张志朋、葛淳棉:《算法控制能激发零工工作者提供主动服务吗?——基于工作动机视角》,《南开管理评论》,2022年,网络首发。
- (10) 施卓敏、张珊:《神经营销ERP研究综述与展望》,《管理世界》,2022年第4期。
- (11) 唐小飞、孙炳、张恩忠、梅发贵:《类人智能机器人社会价值替代与风险态度研究》,《南开管理评论》,2021年第6期。
- (12) 王海忠、谢涛、詹纯玉:《服务失败情境下智能客服化身拟人化的负面影响:厌恶感的中介机制》,《南开管理评论》,2021年第4期。
- (13) 王林辉、胡晨明、董直庆:《人工智能技术、任务属性与职业可替代风险:来自微观层面的经验证据》,《管理世界》,2022年第7期。
- (14) 王欣、朱虹、姜帝、夏少昂、肖春曲:《人工智能产品“协助者”与“替代者”形象对消费者评价的影响》,《南开管理评论》,2021年第6期。
- (15) 吴继飞、朱翊敏、刘颖悦、梁嘉明:《智能客服厌恶效应的诱因、心理机制与边界研究》,《南开管理评论》,2022年第3期。
- (16) 肖红军:《算法责任:理论证成、全景画像与治理范式》,《管理世界》,2022年第4期。
- (17) 谢小云、左玉涵、胡琼晶:《数字化时代的人力资源管理:基于人与技术交互的视角》,《管理世界》,2021年第1期。
- (18) 徐鹏、徐向艺:《人工智能时代企业管理变革的逻辑与分析框架》,《管理世界》,2020年第1期。
- (19) 余伊琦、冯羽、盖嘉:《不确定情境下消费者对人工智能和人工服务的运气感知与相对偏好》,《南开管理评论》,2022年第3期。
- (20) 朱国玮、黄静、罗映宇:《算法管理下的共享经济阴暗面:概念框架与展望》,《南开管理评论》,2023年,网络首发。
- (21) Acemoglu, D. and Restrepo, P., 2020, “Robots and Jobs: Evidence from US Labor Markets”, *Journal of Political Economy*, Vol.128, No.6, pp.2188~2244.
- (22) Ameen, N., Tarhini, A., Reppel, A. and Anand, A., 2021, “Customer Experiences in the Age of Artificial Intelligence”, *Computers in Human Behavior*, Vol.114, No.1, 106548.
- (23) Aguilar, P., Brissino, S., Fernández-Dols and José-Miguel, 2013, “Psychological Distance Increases Uncompromising Consequentialism”, *Journal of Experimental Social Psychology*, Vol.49, No.3, pp.449~452.
- (24) Al-Natour, S., Cavusoglu, H., Benbasat, I. and Aleem, U., 2020, “An Empirical Investigation of the Antecedents and Consequences of Privacy Uncertainty in the Context of Mobile Apps”, *Information Systems Research*, Vol.31, No.4, pp.1037~1063.
- (25) Aladwani, A. M., 2001, “Online Banking: A Field Study of Drivers, Development Challenges, and Expectations”, *International Journal of Information Management*, Vol.21, No.3, pp.213~225.
- (26) Allen, R. and Choudhury, P., 2022, “Algorithm-augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion”, *Organization Science*, Vol.33, No.1, pp.149~169.
- (27) Araujo, T., Helberger, N., Kruijemeier, S. and de Vreese, C. H., 2020, “In AI We Trust? Perceptions about Automated Decision-making by Artificial Intelligence”, *AI & Society*, Vol.35, No.3, pp.611~623.
- (28) Babic, B., Gerke, S., Evgeniou, T. and Cohen, I. G., 2021, “Beware Explanations from AI in Health Care”, *Science*, Vol.373, No.3, pp.284~286.
- (29) Bandura, A., 1982, “Self-efficacy Mechanism in Human Agency”, *American Psychologist*, Vol.37, No.2, pp.122~147.
- (30) Banker, S. and Khetani, S., 2019, “Algorithm Overdependence: How the Use of Algorithmic Recommendation Systems Can Increase Risks to Consumer Well-being”, *Journal of Public Policy & Marketing*, Vol.38, No.4, pp.500~515.
- (31) Bigman, Y. E. and Gray, K., 2018, “People are Averse to Machines Making Moral Decisions”, *Cognition*, Vol.181, No.12, pp.21~34.
- (32) Bonnefon, J.-F., Shariff, A. and Rahwan, I., 2016, “The Social Dilemma of Autonomous Vehicles”, *Science*, Vol.352, No.13, pp.1573~1576.
- (33) Burton, J. W., Stein, M. K. and Jensen, T. B., 2020, “A Systematic Review of Algorithm Aversion in Augmented Decision Making”, *Journal of Behavioral Decision Making*, Vol.33, No.2, pp.220~239.
- (34) Calder, B. J. and Burnkrant, R. E., 1977, “Interpersonal Influence on Consumer Behavior: An Attribution Theory Approach”, *Journal of Consumer Research*, Vol.4, No.3, pp.289~300.

nal of Consumer Research, Vol.4, No.1, pp.29~38.

(35) Castelo, N., Bos, M. W. and Lehmann, D. R., 2019, "Task-dependent Algorithm Aversion", *Journal of Marketing Research*, Vol.56, No.5, pp.809~825.

(36) Castelvocchi, D., 2016, "Can We Open the Black Box of AI?", *Nature*, Vol.538, No.1, pp.20~23.

(37) Chamberlain, R., Mullin, C., Scheerlinck, B. and Wagemans, J., 2018, "Putting the Art in Artificial: Aesthetic Responses to Computer-generated Art", *Psychology of Aesthetics Creativity and the Arts*, Vol.12, No.2, pp.177~192.

(38) Chen, X., Cho, Y. H., Dou, Y. and Lev, B., 2022, "Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data", *Journal of Accounting Research*, Vol.60, No.2, pp.467~515.

(39) Cheng, M. and Foley, C., 2019, "Algorithmic Management: The Case of Airbnb", *International Journal of Hospitality Management*, Vol.83, No.10, pp.33~36.

(40) Chowdhury, S., Budhwar, P., Dey, P. K., Joel-Edgar, S. and Abadie, A., 2022, "AI-employee Collaboration and Business Performance: Integrating Knowledge-based View, Socio-technical Systems and Organisational Socialisation Framework", *Journal of Business Research*, Vol.144, No.6, pp.31~49.

(41) Cichy, P., Salge, T. O. and Kohli, R., 2021, "Privacy Concerns and Data Sharing in the Internet of Things: Mixed Methods Evidence from Connected Cars", *MIS Quarterly*, Vol.45, No.4, pp.1863~1892.

(42) Commerford, B. P., Dennis, S. A., Joe, J. R. and Ulla, J. W., 2022, "Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence", *Journal of Accounting Research*, Vol.60, No.1, pp.171~201.

(43) Dai, T. and Tayur, S., 2022, "Designing AI-augmented Healthcare Delivery Systems for Physician Buy-in and Patient Acceptance", *Production and Operations Management*, Vol.31, No.12, pp.4443~4451.

(44) Dane, E. and Pratt, M. G., 2007, "Exploring Intuition and Its Role in Managerial Decision Making", *Academy of Management Review*, Vol.32, No.1, pp.33~54.

(45) Dawes, R. M., 1979, "The Robust Beauty of Improper Linear Models in Decision Making", *American Psychologist*, Vol.34, No.7, pp.571~582.

(46) Dietvorst, B. J., Simmons, J. P. and Massey, C., 2015, "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err", *Journal of Experimental Psychology: General*, Vol.144, No.1, pp.114~126.

(47) Dietvorst, B. J. and Bharti, S., 2020, "People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error", *Psychological Science*, Vol.31, No.10, pp.1302~1314.

(48) Dietvorst, B. J., Simmons, J. P. and Massey, C., 2018, "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them", *Management Science*, Vol.64, No.3, pp.1155~1170.

(49) Figueroa-Armijos, M., Clark, B. B. and Veiga, S. P. d. M., 2022, "Ethical Perceptions of AI in Hiring and Organizational Trust: The Role of Performance Expectancy and Social Influence", *Journal of Business Ethics*, Vol.186, No.1, pp.179~197.

(50) Fuegener, A., Grahl, J., Gupta, A. and Ketter, W., 2022, "Cognitive Challenges in Human-artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation", *Information Systems Research*, Vol.33, No.2, pp.678~696.

(51) Fuegener, A., Grahl, J., Gupta, A. and Ketter, W., 2021, "Will Humans-in-the-loop Become Borgs? Merits and Pitfalls of Working with AI", *Management Information Systems Quarterly*, Vol.45, No.3, pp.1527~1556.

(52) Gandini, A., 2019, "Labour Process Theory and The Gig Economy", *Human Relations*, Vol.72, No.6, pp.1039~1056.

(53) Garvey, A. M., Kim, T. and Duhachek, A., 2023, "Bad News? Send an AI. Good news? Send a Human", *Journal of Marketing*, Vol.87, No.1, pp.10~25.

(54) Ge, R., Zheng, Z., Tian, X. and Liao, L., 2021, "Human-robot Interaction: When Investors Adjust the Usage of Robo-advisors in Peer-to-peer Lending", *Information Systems Research*, Vol.32, No.3, pp.774~785.

(55) Gill, T., 2020, "Blame It on The Self-driving Car: How Autonomous Vehicles Can Alter Consumer Morality", *Journal of Consumer Research*, Vol.47, No.2, pp.272~291.

(56) Giroux, M., Kim, J., Lee, J. C. and Park, J., 2022, "Artificial Intelligence and Declined Guilt: Retailing Morality Comparison between Human and AI", *Journal of Business Ethics*, Vol.178, No.4, pp.1027~1041.

(57) Glikson, E., and Woolley, A. W., 2020, "Human Trust in Artificial Intelligence: Review of Empirical Research", *Academy of Management Annals*, Vol.14, No.2, pp.627~660.

(58) Granulo, A., Fuchs, C. and Puntoni, S., 2021, "Preference for Human (vs. Robotic) Labor is Stronger in Symbolic Consumption Contexts", *Journal of Consumer Psychology*, Vol.31, No.1, pp.72~80.

(59) Grove, W. M. and Meehl, P. E., 1996, "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-statistical Controversy", *Psychology, Public Policy, and Law*, Vol.2, No.2, pp.293~323.

(60) Han, E., Yin, D. and Zhang, H., 2022, "Bots with Feelings: Should AI Agents Express Positive Emotion in Customer Service?", *Information Systems Research*, Forthcoming.

(61) Highhouse, S., 2008, "Stubborn Reliance on Intuition and Subjectivity in Employee Selection", *Industrial and Organizational Psychology*, Vol.1, No.3, pp.333~342.

(62) Ho, C. C. and MacDorman, K. F., 2010, "Revisiting the Uncanny Valley Theory: Developing and Validating an Alternative to The Godspeed Indices", *Computers in Human Behavior*, Vol.26, No.6, pp.1508~1518.

(63) Holthöwer, J. and van Doorn, J., 2023, "Robots Do Not Judge: Service Robots Can Alleviate Embarrassment in Service Encounters", *Journal of the Academy of Marketing Science*, Vol.51, No.4, pp.767~784.

(64) Huang, M. H. and Rust, R. T., 2018, "Artificial Intelligence in Service", *Journal of Service Research*, Vol.21, No.2, pp.155~172.

- (65) Hunt, W., Sarkar, S. and Warhurst, C., 2022, "Measuring the Impact of AI on Jobs at the Organization Level: Lessons from a Survey of UK Business Leaders", *Research Policy*, Vol.51, No.2, 104425.
- (66) Joshi, A. and Wade, M., 2020, "The Building Blocks of an AI Strategy", *MIT Sloan Management Review*, Vol.10, No.1, pp.08~20.
- (67) Jussupow, E., Spohrer, K., Heinzl, A. and Gawlitza, J., 2021, "Augmenting Medical Diagnosis Decisions? An Investigation Into Physicians' Decision-making Process with Artificial Intelligence", *Information Systems Research*, Vol.32, No.3, pp.713~735.
- (68) Kawaguchi, K., 2021, "When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business", *Management Science*, Vol.67, No.3, pp.1670~1695.
- (69) Kim, J. H., Kim, M., Kwak, D. W. and Lee, S., 2022, "Home-tutoring Services Assisted with Technology: Investigating the Role of Artificial Intelligence Using a Randomized Field Experiment", *Journal of Marketing Research*, Vol.59, No.1, pp.79~96.
- (70) Kim, T., Barasz, K. and John, L. K., 2019, "Why Am I Seeing This Ad? The Effect of Ad Transparency on Ad Effectiveness", *Journal of Consumer Research*, Vol.45, No.5, pp.906~932.
- (71) Kim, T. W. and Duhachek, A., 2020, "Artificial Intelligence and Persuasion: A Construal-level Account", *Psychological Science*, Vol.31, No.4, pp.363~380.
- (72) Kyung, N. and Kwon, H. E., 2022, "Rationally Trust, But Emotionally? The Roles of Cognitive and Affective Trust in Laypeople's Acceptance of AI for Preventive Care Operations", *Production and Operations Management*, Forthcoming.
- (73) Lee, M. K., Kusbit, D., Metsky, E. and Dabbish, L., 2015, "Working with Machines", Paper presented at the Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.
- (74) Lehmann, C. A., Haubitz, C. B., Fügner, A. and Thonemann, U. W., 2022, "The Risk of Algorithm Transparency: How Algorithm Complexity Drives the Effects on Use of Advice", *Production and Operations Management*, Vol.31, No.9, pp.3419~3434.
- (75) Li, J. J., Bonn, M. A. and Ye, B. H., 2019, "Hotel Employee's Artificial Intelligence and Robotics Awareness and Its Impact on Turnover Intention: The Moderating Roles of Perceived Organizational Support and Competitive Psychological Climate", *Tourism Management*, Vol.73, No.1, pp.172~181.
- (76) Liang, C., Peng, J., Hong, Y. and Gu, B., 2022, "The Hidden Costs and Benefits of Monitoring in the Gig Economy", *Information Systems Research*, Vol.34, No.1, pp.297~318.
- (77) Liu, P., Du, Y. and Xu, Z., 2019, "Machines Versus Humans: People's Biased Responses to Traffic Accidents Involving Self-driving Vehicles", *Accident Analysis & Prevention*, Vol.125, No.8, pp.232~240.
- (78) Liu, Y. J., Chen, Z., Meng, J. and Wang, Z., 2022, "What's in a Face? An Experiment on Facial Information and Loan Approval Decision", *Management Science*, Vol.69, No.4, pp.2263~2283.
- (79) Logg, J. M., Minson, J. A. and Moore, D. A., 2019, "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment", *Organizational Behavior and Human Decision Processes*, Vol.151, No.1, pp.90~103.
- (80) Longoni, C., Bonezzi, A. and Morewedge, C. K., 2019, "Resistance to Medical Artificial Intelligence", *Journal of Consumer Research*, Vol.46, No.4, pp.629~650.
- (81) Longoni, C. and Cian, L., 2022, "Artificial Intelligence in Utilitarian vs. Hedonic Contexts: The 'Word-of-machine' Effect", *Journal of Marketing*, Vol.86, No.1, pp.91~108.
- (82) Lourenço, C. J., Dellaert, B. G. and Donkers, B., 2020, "Whose Algorithm Says So: The Relationships Between Type of Firm, Perceptions of Trust and Expertise, and the Acceptance of Financial Robo-advice", *Journal of Interactive Marketing*, Vol.49, No.1, pp.107~124.
- (83) Luo, X., Tong, S., Fang, Z. and Qu, Z., 2019, "Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases", *Marketing Science*, Vol.38, No.6, pp.937~947.
- (84) Luo, X., Qin, M. S., Fang, Z. and Qu, Z., 2021, "Artificial Intelligence Coaches for Sales Agents: Caveats And Solutions", *Journal of Marketing*, Vol.85, No.2, pp.14~32.
- (85) Mahmud, H., Islam, A. N., Ahmed, S. I. and Smolander, K., 2022, "What Influences Algorithmic Decision-making? A Systematic Literature Review on Algorithm Aversion", *Technological Forecasting and Social Change*, Vol.175, No.1, 121390.
- (86) Mai, R., Hoffmann, S., Lasarov, W. and Buhs, A., 2019, "Ethical Products=Less Strong: How Explicit and Implicit Reliance on the Lay Theory Affects Consumption Behaviors", *Journal of Business Ethics*, Vol.158, No.3, pp.659~677.
- (87) Malik, A., Budhwar, P., Mohan, H., NR. and Srikanth., 2022, "Employee Experience—the Missing Link for Engaging Employees: Insights from an MNE's AI-based HR Ecosystem", *Human Resource Management*, Forthcoming.
- (88) Maner, J. K. and Gailliot, M. T., 2007, "Altruism and Egoism: Prosocial Motivations for Helping Depend on Relationship Context", *European Journal of Social Psychology*, Vol.37, No.2, pp.347~358.
- (89) McCarthy, J., 2004, "What is Artificial Intelligence", URL: <http://www-formal.stanford.edu/jmc/whatisai.html>.
- (90) McLeay, F., Osburg, V. S., Yoganathan, V. and Patterson, A., 2021, "Replaced by a Robot: Service Implications in the Age of The Machine", *Journal of Service Research*, Vol.24, No.1, pp.104~121.
- (91) Meehl, P. E., 1954, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, American Psychological Association.
- (92) Mende, M., Scott, M. L., van Doorn, J., Grewal, D. and Shanks, I., 2019, "Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses", *Journal of Marketing Research*, Vol.56, No.4, pp.535~556.
- (93) Miao, F., Kozlenkova, I. V., Wang, H., Xie, T. and Palmatier, R. W., 2022, "An Emerging Theory of Avatar Marketing", *Journal of Marketing*, Vol.86, No.1, pp.67~90.
- (94) Mohlmann, M., Zalmanson, L., Henfridsson, O. and Gregory, R. W., 2021, "Algorithmic Management of Work on Online Labor Platforms: When Matching Meets Control", *MIS Quarterly*, Vol.45, No.4, pp.1999~2022.

- (95) Molden, D. and Dweck, C., 2006, "Finding 'Meaning' in Psychology: A Lay Theories Approach to Self-regulation, Social Perception, and Social Development", *American Psychologist*, Vol.61, No.3, pp.192~203.
- (96) Morin, A., 2004, "A neurocognitive and Socioecological Model of Self-awareness", *Genetic, Social, and General Psychology Monographs*, Vol.130, No.3, pp.197~224.
- (97) Önköl, D., Goodwin, P., Thomson, M., Gönül, S. and Pollock, A., 2009, "The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments", *Journal of Behavioral Decision Making*, Vol.22, No.4, pp.390~409.
- (98) Parent-Rochelleau, X. and Parker, S. K., 2022, "Algorithms as Work Designers: How Algorithmic Management Influences The Design of Jobs", *Human Resource Management Review*, Vol.32, No.3, 100838.
- (99) Park, E. H., Werder, K., Cao, L. and Ramesh, B., 2022, "Why Do Family Members Reject AI in Health Care? Competing Effects of Emotions", *Journal of Management Information Systems*, Vol.39, No.3, pp.765~792.
- (100) Peukert, C., Bechtold, S., Batikas, M. and Kretschmer, T., 2022, "Regulatory Spillovers and Data Governance: Evidence from the GDPR", *Marketing Science*, Vol.41, No.4, pp.746~768.
- (101) Power, R. A. and Pluess, M., 2015, "Heritability Estimates of the Big Five Personality Traits Based on Common Genetic Variants", *Translational Psychiatry*, Vol.5, No.7, pp.e604.
- (102) Prah, A. and Van Swol, L., 2017, "Understanding Algorithm Aversion: When Is Advice from Automation Discounted?", *Journal of Forecasting*, Vol.36, No.6, pp.691~702.
- (103) Puntoni, S., Reczek, R. W., Giesler, M. and Botti, S., 2021, "Consumers and Artificial Intelligence: An Experiential Perspective", *Journal of Marketing*, Vol.85, No.1, pp.131~151.
- (104) Ragni, M., Rudenko, A., Kuhnert, B. and Arras, K. O., 2016, *Errare Humanum Est: Erroneous Robots in Human-robot Interaction*, Paper presented at the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).
- (105) Rai, A., 2020, "Explainable AI: From Black Box to Glass Box", *Journal of the Academy of Marketing Science*, Vol.48, No.1, pp.137~141.
- (106) Raj, M. and Seamans, R., 2019, "Primer on Artificial Intelligence and Robotics", *Journal of Organization Design*, Vol.8, No.1, pp.1~14.
- (107) Rapp, A., Curti, L. and Boldi, A., 2021, "The Human Side of Human-chatbot Interaction: A Systematic Literature Review of Ten Years of Research on Text-based Chatbots", *International Journal of Human-Computer Studies*, Vol.151, No.102630.
- (108) Reich, T., Kaju, A. and Maglio, S. J., 2023, "How to Overcome Algorithm Aversion: Learning from Mistakes", *Journal of Consumer Psychology*, Vol.33, No.2, pp.285~302.
- (109) Russell, S. J. and Norvig, P., 2010, *Artificial Intelligence: A Modern Approach*, Pearson Education, Inc.
- (110) Soderberg, C. K., Callahan, S. P., Kochersberger, A. O., Amit, E. and Ledgerwood, A., 2015, "The Effects of Psychological Distance on Abstraction: Two Meta-analyses", *Psychological Bulletin*, Vol.141, No.3, pp.525~548.
- (111) Sousa, C. M., Coelho, F. and Guillaumon-Saorin, E., 2012, "Personal Values, Autonomy, and Self-efficacy: Evidence from Front-line Service Employees", *International Journal of Selection and Assessment*, Vol.20, No.2, pp.159~170.
- (112) Srinivasan, R. and Sarial-Abi, G., 2021, "When Algorithms Fail: Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors", *Journal of Marketing*, Vol.85, No.5, pp.74~91.
- (113) Stein, J. P. and Ohler, P., 2017, "Venturing into the Uncanny Valley of Mind—the Influence of Mind Attribution on the Acceptance of Human-like Characters in a Virtual Reality Setting", *Cognition*, Vol.160, No.2, pp.43~50.
- (114) Sun, J., Zhang, D. J., Hu, H. and Van Mieghem, J. A., 2022, "Predicting Human Discretion to Adjust Algorithmic Prescription: A Large-scale Field Experiment in Warehouse Operations", *Management Science*, Vol.68, No.2, pp.846~865.
- (115) Telkamp, J. B. and Anderson, M. H., 2022, "The Implications of Diverse Human Moral Foundations for Assessing the Ethicality of Artificial Intelligence", *Journal of Business Ethics*, Vol.178, No.4, pp.961~976.
- (116) Tong, S., Jia, N., Luo, X. and Fang, Z., 2021, "The Janus Face of Artificial Intelligence Feedback: Deployment Versus Disclosure Effects on Employee Performance", *Strategic Management Journal*, Vol.42, No.9, pp.1600~1631.
- (117) Tranfield, D., Denyer, D. and Smart, P., 2003, "Towards a Methodology for Developing Evidence-informed Management Knowledge by Means of Systematic Review", *British Journal of Management*, Vol.14, No.3, pp.207~222.
- (118) Trope, Y. and Liberman, N., 2010, "Construal-level Theory of Psychological Distance", *Psychological Review*, Vol.117, No.2, pp.440~463.
- (119) Trope, Y., Liberman, N. and Wakslak, C., 2007, "Construal Levels and Psychological Distance: Effects on Representation, Prediction, Evaluation, and Behavior", *Journal of Consumer Psychology*, Vol.17, No.2, pp.83~95.
- (120) Van Esch, P., Black, J. S. and Arli, D., 2021, "Job Candidates' Reactions to Ai-enabled Job Application Processes", *AI and Ethics*, Vol.1, No.2, pp.119~130.
- (121) Ward, A. F., Olsen, A. S. and Wegner, D. M., 2013, "The Harm-made Mind: Observing Victimization Augments Attribution of Minds to Vegetative Patients, Robots, and the Dead", *Psychological Science*, Vol.24, No.8, pp.1437~1445.
- (122) Webster, J. and Watson, R. T., 2002, "Analyzing the Past to Prepare for The Future: Writing a Literature Review", *MIS Quarterly*, Vol.26, No.2, pp.7~23.
- (123) Wilson, H. J., Daugherty, P. and Bianzino, N., 2017, "The Jobs That Artificial Intelligence Will Create", *MIT Sloan Management Review*, Vol.58, No.4, pp.14~16.
- (124) Wood, A. J., Graham, M., Lehdonvirta, V. and Hjorth, I., 2019, "Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy", *Work Employment Society*, Vol.33, No.1, pp.56~75.

- (125) Yalcin, G., Lim, S., MJ van Osselaer, S. and Puntoni, S., 2022, "Thumbs Up or Down: Consumer Reactions to Decisions by Algorithms Versus Humans", *Journal of Marketing*, Vol.59, No.4, pp.696~717.
- (126) Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H. and Gray, K., 2021, "Robots at Work: People Prefer and Forgive Service Robots with Perceived Feelings", *Journal of Applied Psychology*, Vol.106, No.10, pp.1557~1572.
- (127) Yeomans, M., Shah, A., Mullainathan, S. and Kleinberg, J., 2019, "Making Sense of Recommendations", *Journal of Behavioral Decision Making*, Vol.32, No.4, pp.403~414.
- (128) You, S., Yang, C. L. and Li, X., 2022, "Algorithmic Versus Human Advice: Does Presenting Prediction Performance Matter for Algorithm Appreciation?", *Journal of Management Information Systems*, Vol.39, No.2, pp.336~365.
- (129) Yu, T. W. and Chen, T. J., 2018, "Online Travel Insurance Purchase Intention: A Transaction Cost Perspective", *Journal of Travel & Tourism Marketing*, Vol.35, No.9, pp.1175~1186.
- (130) Zhang, L., Pentina, I. and Fan, Y., 2021, "Who Do You Choose? Comparing Perceptions of Human vs Robo-advisor in the Context of Financial Services", *Journal of Services Marketing*, Vol.35, No.5, pp.634~646.
- (131) Zhou, Y., Fei, Z., He, Y. and Yang, Z., 2022, "How Human-chatbot Interaction Impairs Charitable Giving: The Role of Moral Judgment", *Journal of Business Ethics*, Vol.178, No.3, pp.849~865.
- (132) Zou, J. and Schiebinger, L., 2018, "AI Can Be Sexist and Racist—it's Time to Make It Fair", *Nature*, Vol.559, No.3, pp.324~327.

Algorithm Aversion in the Era of Artificial Intelligence: Research Framework and Future Agenda

Luo Yingyu^a, Zhu Guowei^a, Qian Wujia^a, Wu Yueyan^b, Huang Jing^a and Yang Zhi^a

(a. Business School, Hunan University; b. School of Economics and Management, Fuzhou University)

Abstract: Understanding the negative reactions of individuals towards algorithmic services is a pivotal prerequisite for effectively addressing the challenges of algorithm management. It is also a crucial topic for promoting the healthy development of the new generation of artificial intelligence and advancing the construction of the digital society. Although the academic community has made significant progress in researching algorithm aversion, most studies have been limited to a singular domain with a singular issue. Our paper employs a systematic literature review methodology and utilizes a framework comprising concept clarification, formation mechanisms, theoretical boundaries, and research prospects to comprehensively analyze and synthesize the research progress of algorithm aversion in the era of artificial intelligence. Firstly, we systematically review the core literature on algorithm aversion in the management domain. Based on the differentiation and analysis of the relationship among algorithms, artificial intelligence, and robots, we establish the connotation and extension of algorithms and provide a clear definition for algorithm aversion in the era of artificial intelligence. Secondly, we construct a theoretical research framework, extracting six causative mechanisms of algorithm aversion from both algorithmic and human perspectives, and elucidating twelve theoretical boundaries of algorithm aversion from the dimensions of individuals, contexts, and algorithms. Finally, we propose six future research topics that hold scientific value and practical significance, offering a theoretical guide for further advancing the research on algorithm aversion theory and practice. Our study expands the cognitive boundaries of researches on the theoretical system of algorithm aversion and provides important insights for facilitating addressing algorithm management challenges for Chinese enterprises, promoting the healthy development of artificial intelligence, and advancing the construction of a digital society in China.

Keywords: artificial intelligence (AI); algorithm decisions; algorithm aversion; human-computer interaction

=====

(上接第 168 页)

The Power of Standards: Empirical Evidence from China's Standard Essential Patents

Long Xiaoning^{a,b} and Zhang Meiyang^c

(a. The Intellectual Property Research Institute, Xiamen University; b. The Belt and Road Research Institute, Xiamen University;

c. School of Law and Economics, Zhongnan University of Economics and Law)

Abstract: Strengthening the internationalization of standard-essential patents (SEPs) and promoting the formation of international intellectual property rules are crucial for building a powerful country with intellectual property rights and a unified national market, thus realizing high-quality economic development. Despite extensive debates on the significance of technical standards, the study of their effects on innovation remains incomplete. Using datasets of Chinese patents, listed companies, and industrial enterprises, this paper empirically analyzes how technical standards impact patent value and innovation from the perspectives of market power and knowledge spillovers by employing the difference-in-differences approach. The findings are as follows: (1) After being declared as standard-essential, patents experience a significant increase in value, as evidenced by increased citations, licenses, and market value. (2) The patent value enhancement stems from the expanded influence following standard adoption, indicated by increased citations by patents from different technical fields, patents in complex subjects, and by non-SEPs; (3) The returns to the technology contributor are at least partly due to the market power of the standard, manifested in more benefits for small-scale innovators, increased market concentration and fewer new enterprises; (4) Standards generate knowledge spillover effects among industry followers, resulting in improved innovation quantity, quality, and geographic distribution. Thus, this study not only expands our understanding of the effects of standards on innovation and industrial upgrading, but also provides empirical support for the government to simultaneously promote international standardization and strengthen rule-based processes to prevent potential monopolistic consequences.

Keywords: standard-essential patent; technical standards; innovation; market power; knowledge spillover

Algorithm Aversion in the Era of Artificial Intelligence: Research Framework and Future Agenda

Luo Yingyu^a, Zhu Guowei^a, Qian Wujia^a, Wu Yueyan^b, Huang Jing^a and Yang Zhi^a

(a. Business School, Hunan University; b. School of Economics and Management, Fuzhou University)

Summary: In academia, scholars refer to the behavior or inclination of individuals rejecting algorithmic recommendations and services as algorithm aversion. In recent years, algorithm aversion has emerged as a hot topic in the field of human-computer interaction, and scholars have extensively explored when and why algorithm aversion occurs. Nonetheless, most existing studies have focused on a singular issue within a specific domain, leading to scattered research, unclear conceptual boundaries, and contradictory findings.

This paper employs a systematic literature review method and adopts a framework that encompasses concept clarification, formation mechanisms, theoretical boundaries, and research prospects to systematically review and summarize over a hundred core publications related to algorithm aversion from 2015 to 2023. Firstly, the paper systematically reviews the core literature on algorithm aversion in the management domain, delineates the relationships among algorithms, artificial intelligence, and robots, defines the connotation and extension of algorithms, and provides a clear definition of algorithm aversion in the era of artificial intelligence. Secondly, it constructs a theoretical research framework that identifies six causative mechanisms of algorithm aversion from both algorithmic (lay beliefs, algorithmic errors, and privacy concerns) and human perspectives (overconfidence, attribution errors, and self-efficacy beliefs). Furthermore, the paper elaborates on twelve theoretical boundaries of algorithm aversion from three dimensions: individuals (gender, age, Big Five personality traits, occupation), contexts (industry characteristics, subjective and objective contexts, hedonic and utilitarian contexts, task difficulty), and algorithms (anthropomorphism, algorithmic roles, algorithm interpretability, human in the loop). Lastly, the authors propose six research agendas, providing theoretical guidance for advancing research in algorithm aversion theory and practice.

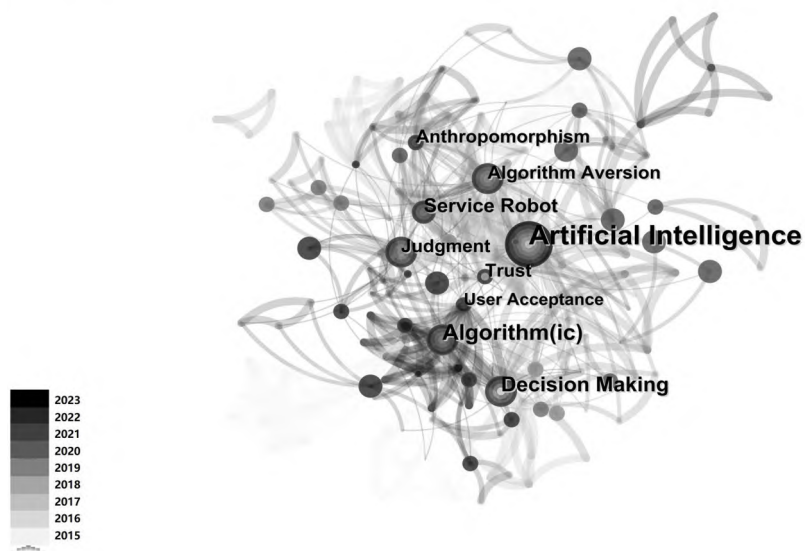
This paper contributes to the theoretical research on algorithm aversion and management practice in the following aspects. Firstly, it provides a clear definition and clarification of the connotation and extension of algorithms, establishing a clear understanding of algorithm aversion in the era of artificial intelligence. Secondly, this paper is one of the early works to propose and construct a research framework for algorithm aversion. It examines the mechanisms of algorithm aversion from both algorithmic and human perspectives and discusses the theoretical boundaries of algorithm aversion, encompassing uncontrollable factors (individual factors) and controllable factors (contextual and algorithmic factors). Moreover, it addresses six research topics related to the causes of algorithm aversion, theoretical boundaries, quantitative and consequential investigations, theoretical construction of algorithmic roles, sibling studies on algorithm appreciation, and theoretical system construction for local research. Especially, based on the theory of construal level and motivation, this research innovatively develops the framework of algorithmic roles in algorithm aversion, which describes the role of algorithmic roles in algorithm aversion, contributing to scholars' clarification of the connotation of algorithm aversion and promoting research related to algorithmic roles. Lastly, this study offers a systematic framework for managers to understand the causes of algorithm aversion and strategies for response, assisting managers in identifying the contexts in which algorithm aversion occurs and implementing measures to weaken or even reverse algorithm aversion. This research is of great significance in aiding Chinese enterprises in effectively coping with algorithm management challenges, promoting the healthy development of Chinese artificial intelligence, and advancing the construction of the Chinese digital society.

Keywords: artificial intelligence (AI); algorithm decisions; algorithm aversion; human-computer interaction

JEL Classification: M12, M21, M31, M50

附录:人工智能时代的算法厌恶:研究框架与未来展望

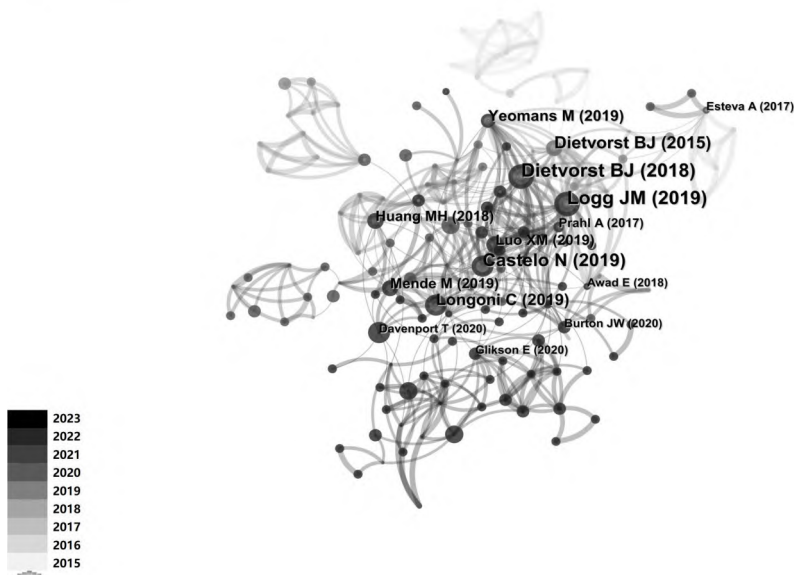
Copyright © 2023, China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



附图1 人工智能时代算法厌恶关键词共现图

注:为了保证图像清晰,提高文献计量的可视化程度,本图只保留共线较多的关键词。

Copyright © 2023, China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



附图2 人工智能时代算法厌恶文献共被引图

注:为了保证图像清晰,提高文献计量的可视化程度,本图只保留共被引较多的文献。

附表1 算法厌恶的研究主题、代表文献、作者/年份、文献来源、类型、研究方法和研究观点^①

算法厌恶研究主题	核心文献	作者/年份	文献来源	类型	研究方法	研究观点
定义	Algorithm aversion: people erroneously avoid algorithms after seeing them error	狄弗斯特等(2015)	Journal of Experimental Psychology: General	嵌入式算法	实验室实验	在已知算法预测更为精确的前提下,人们依然倾向于依赖人类自身进行的决策的现象。
	Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them	狄弗斯特等(2018)	Management Science	嵌入式算法	实验室实验	人们在认识到算法预测存在一定程度的不完美后有意识地排斥算法决策的现象。
朴素信念	Artificial intelligence in utilitarian vs. hedonic contexts: The “word-of machine” effect	隆戈尼、齐安(2022)	Journal of Marketing	算法赋能的虚拟机器人	实验法	人们会认为算法是难以理解用户情感的,因此更愿意接受算法推荐的实用品,而非是享乐品。
	Artificial intelligence and persuasion: a construal-level account	金姆、杜哈切克(2020)	Psychological Science	嵌入式算法	实验室实验 在线调查	人们会认为算法缺乏自主目标和意图,因此当算法阐述高解释水平的说服语句时,人们表现出更高的不信任。
	The building blocks of an AI strategy	乔希、韦德(2020)	MIT Slogan	-	理论述评	人们会认为算法决策是更不道德的,因此企业管理者不愿意采用算法的辅助决策。
	Man versus machine: Resisting automation in identity-based consumer behavior	梁格等(2018)	Journal of Marketing Research	嵌入式算法 算法赋能的机器人	实验法	当身份动机成为消费的驱动因素时,人们会产生算法厌恶。
	Bad news? Send an AI. Good news? Send a human	加维等(2022)	Journal of Marketing	嵌入式算法 算法赋能的虚拟机器人	实验法	人们还会推断算法代理在提供有利于自己的条件时缺乏自私意图,在提供有利于消费者的条件时缺乏善意意图。
	Consumers and artificial intelligence: An experiential perspective	蓬托尼等(2021)	Journal of Marketing	-	理论述评	作者从代理行为视角阐述了算法对消费者身份威胁的后果和形成机理。
	The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance	唐等(2021)	Strategy Management Journal	嵌入式算法	田野实验	人工智能的反馈可能让员工认为自己的工作有可被替代的风险,使他们不愿意接受算法的反馈,进而降低了工作产出。
	人工智能技术、任务属性与职业可替代风险:来自微观层面的经验证据	王林辉等(2022)	管理世界	嵌入式算法	二手数据	人工智能技术会引发职业替代风险,非程式化认知交互型社群可替代风险最低,年龄越大和受教育年限最短的劳动者可替代风险越高。
	人工智能产品“协助者”与“替代者”形象对消费者评价的影响	王欣等(2021)	南开管理评论	算法赋能的虚拟机器人	在线调查	人们会认为算法(人工智能)产生身份认同威胁,将人工智能产品的形象从替代者转变为协助者,能够减轻消费者对人工智能产品的抗拒。
	类人智能机器人社会价值替代与风险态度研究	唐小飞等(2021)	南开管理评论	算法赋能的机器人	在线调查	基于“差序格局”理论,作者将社会功能价值替代分为智力替代和体力替代;将社会存在价值替代分为自我替代、亲情替代和友情替代。
算法错误	Algorithm aversion: people erroneously avoid algorithms after seeing them err	狄弗斯特等(2015)	Journal of Experimental Psychology: General	嵌入式算法	实验室实验	人们意识到算法预测出现错误后,即便足够的证据表明算法模型的表现优于人类预测,他们仍会展现出强烈的算法厌恶倾向。
	Trust in automation: Integrating empirical evidence on factors that influence trust	霍夫、巴希尔(2015)	Human Factors	-	文献述评	在使用算法的过程中,越早被发现的算法错误会对算法信任感产生越消极的影响。
隐私担忧	The Hidden Costs and Benefits of Monitoring in the Gig Economy	梁等(2022)	Information System Research	嵌入式算法	田野实验	零工工作者会对算法平台产生隐私担忧,因此更愿意选择不使用算法平台的企业。
	Privacy Concerns and Data Sharing in the Internet of Things: Mixed Methods Evidence from Connected Cars	齐希等(2021)	MISQ	算法赋能的机器人	田野实验	无人驾驶会让驾驶员(心理上)担心隐私的所有权失去控制,也就是削弱了隐私的心理所有权感知
过度自信	Algorithm overdependence: How the use of algorithmic recommendation systems can increase risks to consumer well-being	班克、赫塔尼(2019)	Journal of Public Policy & Marketing	嵌入式算法	在线调查	人们经常会高估自己的能力,因此人们自信他们不需要依赖诸如自动化或算法等决策工具。
	What's in a Face? An Experiment on Facial Information and Loan Approval Decision	刘等(2022)	Management Science	嵌入式算法	实验室实验	人们会因为过度自信而产生算法厌恶。

		Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data	陈等 (2022)	Journal of Accounting Research	嵌入式算法	二手数据	由于认知限制和行为偏差,人们更倾向于接受人类对盈余管理的预测。
	错误归因	Thumbs up or down: Consumer reactions to decisions by algorithms versus humans	叶尔钦等 (2022)	Journal of Marketing Research	嵌入式算法 算法赋能的虚拟机器人	田野实验 在线调查 半结构化访谈	与算法相比,消费者更容易内化由人提供的有利的决策结果,但也容易外化一个不利的决策结果,无论决策者的类型如何。
	自我效能感	Algorithmic management of work on online labor platforms: When matching meets control	默尔曼等 (2021)	MISQ	嵌入式算法	案例研究	算法平台的介入让零工工作者认为自己时刻受到监视,需要被迫接受系统的安排,这使得员工的自我效能感降低。
		数字化时代的人力资源管理: 基于人与技术交互的视角	谢小云等 (2022)	管理世界	-	理论述评	算法平台让骑手无从选择,其配送任务分配、路线规划和时间规划都是由算法提供,骑手仅仅拥有形式上的拒绝权。
	个体因素	Whose algorithm says so: The relationships between type of firm, perceptions of trust and expertise, and the acceptance of financial robo-advice	洛伦索等 (2020)	Journal of Interactive Marketing	算法赋能的虚拟机器人	在线调查 实验室实验	女性相较于男性而言,在进行决策时对其专业技能更缺乏自信,因此女性比男性更倾向于接受算法的建议。
		Algorithm aversion: people erroneously avoid algorithms after seeing them err	狄弗斯特等 (2015)	Journal of Experimental Psychology: General	嵌入式算法	实验室实验	人们先与算法交互经验会影响他们使用算法决策建议。
		Machines versus humans: the impact of AI chatbot disclosure on customer purchases	罗等 (2019)	Management Science	算法赋能的虚拟机器人	实验法	对于有丰富知识的个体(如经济学家),他们对自身预测更有信心,并且不会选择依赖算法帮助。
	情境因素	Human - Robot Interaction: When Investors Adjust the Usage of Robo-Advisors in Peer-to-Peer Lending	葛等 (2022)	Information System Research	嵌入式算法	二手数据	在 P2P 借贷中,那些在线下受到违约行为的用户会更不愿意采取机器助手的建议。
		People are averse to machines making moral decisions	比格曼、格雷 (2018)	Cognition	嵌入式算法 算法赋能的虚拟机器人	实验室实验 在线调查	当面对道德任务时,人们对人类展示出更积极的态度(与算法相比)。
		Preference for human (vs. robotic) labor is stronger in symbolic consumption contexts	格兰努罗等 (2021)	Journal of Consumer Psychology	嵌入式算法 算法赋能的虚拟机器人	实验法	相比于高象征价值产品,人工智能更多地与低象征价值的产品产生联系。
	边界条件	Overcoming Algorithm Aversion: People will use Imperfect Algorithms if They can (Even Slightly) Modify Them	狄弗斯特等 (2018)	Management Science	嵌入式算法	实验室实验 在线调查	向用户提供修改算法的权力能够使他们更愿意接受算法的推荐。
		The Hidden Costs and Benefits of Monitoring in the Gig Economy	梁等 (2022)	Information System Research	嵌入式算法	田野实验	提升算法透明度和向零工工作者提供信息控制权有助于他们克服算法厌恶。
	算法因素	Blame the Bot: Anthropomorphism and Anger in Customer - Chatbot Interactions	卡洛里克等 (2022)	Journal of Marketing	算法赋能的虚拟机器人	实验法	当顾客以愤怒的情绪状态进入聊天机器人主导的服务交互时,聊天机器人拟人化对顾客满意度、整体企业评价以及随后的购买意愿都有负面影响。然而,对于处于非愤怒情绪状态的顾客来说,情况并非如此。
		When Algorithms Fail: Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors	斯里尼瓦桑、沙利叶 (2021)	Journal of Marketing	算法赋能的虚拟机器人	实验室实验 在线调查	智能客服的拟人化特征会提升人们对算法的伤害责任感知。
		服务失败情境下智能客服化身拟人化的负面影响: 厌恶感的中介机制	王海忠等 (2021)	南开管理评论	算法赋能的虚拟机器人	在线调查	在服务失败情境下,高形式拟人化的化身会加剧顾客对智能客服的厌恶感和负面态度。
		Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations	孙等 (2022)	Management Science	嵌入式算法	二手数据	将人类行为的偏差考虑到算法决策中,以进一步提升算法的效果(被动的过程)。
		Overcoming Algorithm Aversion: People will use Imperfect Algorithms if They can (Even Slightly) Modify Them	狄弗斯特等 (2018)	Management Science	嵌入式算法	实验室实验 在线调查	人们主动参与算法的改进过程,能够提升人们对算法模型的使用(主动的过程)。

注: (1) 由于篇幅有限, 本表仅考虑代表性研究。(2) 实验法指研究中同时使用田野实验、实验室实验和在线调查的研究。

附表2 人工智能时代的算法厌恶:算法的表现形式、判定标准、代称、研究领域、应用行业和交互对象

表现形式	判定标准		代称	研究领域	应用行业	交互对象
嵌入式算法	√		算法、算法平台、控制系统、人工智能系统、教练、数据模型等。	应用心理学、工商管理(人力资源管理、市场营销)、金融学等。	服务业、餐饮业、住宿业、互联网行业等。	消费者、管理者、普通职员、零工工作者、服务者、投资者等。
虚拟机器人	自动化	×	人工智能、聊天机器人、机器人客服、机器人、人工智能代理、推荐代理、推荐者等。	工商管理(市场营销)、电子商务、社会心理学等。	医疗行业、金融行业、餐饮页、广告业、零售业等。	消费者、医生、病人等。
	算法赋能	√				
实体机器人	自动化	×	机器人、智能机器人、人工智能等。	计算机应用技术、工商管理(人力资源管理)等。	服务业、交通、科学研究和技术服务业等。	服务人员、零工工作者等。
	算法赋能	√				

注:(1)此处嵌入式算法指机器学习算法。(2)判定标准:指本研究关注的算法形式,只有符合标准的才纳入本文研究。(3)研究领域:根据中国大学学科分类下二级学科分类。(4)应用行业:根据国民经济行业分类—中华人民共和国国家标准。5.交互对象:算法厌恶中的角色众多,本表中暂且考虑算法厌恶中具有代表性的六种应用场景中的八个角色,分别是:市场营销领域(消费者)、组织管理领域(管理者和普通职员)、零工经济领域(零工工作者)、服务领域(服务者)、医疗咨询领域(医生和病人)、投资领域(投资者)。

资料来源:根据文献和资料整理归纳。

注释

①中外文人名(机构名)对照:狄弗斯特(Dietvorst);隆戈尼(Longoni);齐安(Cian);金姆(Kim);杜切哈克(Duhachek);乔希(Joshi);韦德(Wade);梁格(Leung);加维(Garvey);蓬托尼(Puntoni);唐(Tong);霍夫(Hoff);巴希尔(Bashir);梁(Liang);齐希(Cichy);班克(Banker);赫塔尼(Khetani);刘(Liu);陈(Chen);叶尔钦(Yalcin);默尔曼(Mohlmann);洛伦索(Lourenço);罗(Luo);葛(Ge);比格曼(Bigman);格雷(Gray);格兰努罗(Granulo);卡洛里克(Crolic);斯里尼瓦桑(Srinivasan);沙利叶(Sarial-Abi);孙(Sun)。

参考文献

- (1)唐小飞、孙炳、张恩忠、梅发贵:《类人智能机器人社会价值替代与风险态度研究》,《南开管理评论》,2021年第6期。
- (2)王海忠、谢涛、詹纯玉:《服务失败情境下智能客服化身拟人化的负面影响:厌恶感的中介机制》,《南开管理评论》,2021年第4期。
- (3)王林辉、胡晟明、董直庆:《人工智能技术、任务属性与职业可替代风险:来自微观层面的经验证据》,《管理世界》,2022年第7期。
- (4)王欣、朱虹、姜帝、夏少昂、肖春曲:《人工智能产品“协助者”与“替代者”形象对消费者评价的影响》,《南开管理评论》,2021年第6期。
- (5)谢小云、左玉涵、胡琼晶:《数字化时代的人力资源管理:基于人与技术交互的视角》,《管理世界》,2021年第1期。
- (6)Banker, S. and Khetani, S., 2019, “Algorithm Overdependence: How the Use of Algorithmic Recommendation Systems Can Increase Risks to Consumer Well-being”, *Journal of Public Policy & Marketing*, Vol.38, No.4, pp.500~515.
- (7)Bigman, Y. E. and Gray, K., 2018, “People are Averse to Machines Making Moral Decisions”, *Cognition*, Vol.181, No., pp.21~34.
- (8)Chen, X., Cho, Y. H., Dou, Y. and Lev, B., 2022, “Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data”, *Journal of Accounting Research*, Vol.60, No.2, pp.467~515.
- (9)Cichy, P., Salge, T. O. and Kohli, R., 2021, “Privacy Concerns and Data Sharing in the Internet of Things: Mixed Methods Evidence from Connected Cars”, *MIS Quarterly*, Vol.45, No.4, pp.1863~1892.
- (10)Crolic, C., Thomaz, F., Hadi, R. and Stephen, A. T., 2022, “Blame the Bot: Anthropomorphism and Anger in Customer-chatbot Interactions”, *Journal of Marketing*, Vol.86, No.1, pp.132~148.
- (11)Dietvorst, B. J., Simmons, J. P. and Massey, C., 2015, “Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err”, *Journal of Experimental Psychology: General*, Vol.144, No.1, pp.114~126.
- (12)Dietvorst, B. J., Simmons, J. P. and Massey, C., 2018, “Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They can (even slightly) Modify Them”, *Management Science*, Vol.64, No.3, pp.1155~1170.
- (13)Garvey, A. M., Kim, T. and Duhachek, A., 2023, “Bad News? Send an AI. Good News? Send a Human”, *Journal of Marketing*, Vol.87, No.1, pp.10~25.
- (14)Ge, R., Zheng, Z., Tian, X. and Liao, L., 2021, “Human-robot Interaction: When Investors Adjust the Usage of Robo-advisors in Peer-to-peer Lending”, *Information Systems Research*, Vol.32, No.3, pp.774~785.
- (15)Granulo, A., Fuchs, C. and Puntoni, S., 2021, “Preference for Human (Vs. Robotic) Labor is Stronger in Symbolic Consumption Contexts”, *Journal of Consumer Psychology*, Vol.31, No.1, pp.72~80.
- (16)Hoff, K. A. and Bashir, M., 2015, “Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust”, *Human factors*, Vol.57, No.3, pp.407~434.
- (17)Joshi, A. and Wade, M., 2020, “The Building Blocks of an AI Strategy”, *MIT Sloan Management Review*, Vol.10, No.1, pp.08~20.
- (18)Kim, T. W. and Duhachek, A., 2020, “Artificial Intelligence and Persuasion: A Construal-level Account”, *Psychological Science*, Vol.31, No.4, pp.363~380.
- (19)Leung, E., Paolacci, G. and Puntoni, S., 2018, “Man Versus Machine: Resisting Automation in Identity-based Consumer Behavior”, *Journal of Marketing Research*, Vol.55, No.6, pp.818~831.
- (20)Liang, C., Peng, J., Hong, Y. and Gu, B., 2022, “The Hidden Costs and Benefits of Monitoring in The Gig Economy”, *Information Systems Research*, Vol.34, No.1, pp.297~318.

- (21) Liu, Y. J., Chen, Z., Meng, J. and Wang, Z., 2022, "What's in a Face? An Experiment on Facial Information and Loan Approval Decision", *Management Science*, Vol.69, No.4, pp.2263~2283.
- (22) Longoni, C. and Cian, L., 2022, "Artificial Intelligence in Utilitarian Vs. Hedonic Contexts: The 'Word-of-machine' Effect", *Journal of Marketing*, Vol.86, No.1, pp.91~108.
- (23) Lourenço, C. J., Dellaert, B. G. and Donkers, B., 2020, "Whose Algorithm Says So: The Relationships Between Type of Firm, Perceptions of Trust and Expertise, and the Acceptance of Financial Robo-advice", *Journal of Interactive Marketing*, Vol.49, No.1, pp.107~124.
- (24) Luo, X., Tong, S., Fang, Z. and Qu, Z., 2019, "Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases", *Marketing Science*, Vol.38, No.6, pp.937~947.
- (25) Mohlmann, M., Zalmanson, L., Henfridsson, O. and Gregory, R. W., 2021, "Algorithmic Management of Work on Online Labor Platforms: When Matching Meets Control", *MIS Quarterly*, Vol.45, No.4, pp.1999~2022.
- (26) Puntoni, S., Reczek, R. W., Giesler, M. and Botti, S., 2021, "Consumers and Artificial Intelligence: An Experiential Perspective", *Journal of Marketing*, Vol.85, No.1, pp.131~151.
- (27) Srinivasan, R. and Sarial-Abi, G., 2021, "When Algorithms Fail: Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors", *Journal of Marketing*, Vol.85, No.5, pp.74~91.
- (28) Sun, J., Zhang, D. J., Hu, H. and Van Mieghem, J. A., 2022, "Predicting Human Discretion to Adjust Algorithmic Prescription: A Large-scale Field Experiment in Warehouse Operations", *Management Science*, Vol.68, No.2, pp.846~865.
- (29) Tong, S., Jia, N., Luo, X. and Fang, Z., 2021, "The Janus Face of Artificial Intelligence Feedback: Deployment Versus Disclosure Effects on Employee Performance", *Strategic Management Journal*, Vol.42, No.9, pp.1600~1631.
- (30) Yalcin, G., Lim, S., MJ van Osselaer, S. and Puntoni, S., 2022, "Thumbs Up or Down: Consumer Reactions to Decisions by Algorithms Versus Humans", *Journal of Marketing*, Vol.59, No.4, pp.696~717.