

Doi: 10. 11835 / j. issn. 1008-5831. pj. 2023. 05. 002

欢迎按以下格式引用:杨红艳,卢思佳,徐拥军. 自然评价:人工智能驱动下的学术成果评价模式重构[J]. 重庆大学学报(社会科学版),2023(4):101-114. Doi: 10. 11835 / j. issn. 1008-5831. pj. 2023. 05. 002.



Citation Format: YANG Hongyan, LU Sijia. Natural evaluation: The reconstruction of the academic publication evaluation model driven by artificial intelligence[J]. Journal of Chongqing University (Social Science Edition), 2023 (4); 101 - 114. Doi: 10. 11835 / j. issn. 1008 - 5831. pj. 2023. 05. 002.

自然评价:人工智能驱动下的学术成果评价模式重构

杨红艳^{a,b}, 卢思佳^b, 徐拥军^{a,b}

(中国人民大学 a. 书报资料中心; b. 信息资源管理学院, 北京 100872)

摘要:学术成果评价作为科研项目评审、人才评价、期刊评价、机构评价等活动的基础,是学术评价体系的重中之重,也是科研管理的关键环节之一。当前,我国学术成果评价模式存在“同行评议黑箱”“引文动机模糊”“评价数据间接且片面”等弊端,阻碍了学术创新,也与当下建构自主知识体系的战略相悖。而近年来人工智能等技术在数据、算法、算力上的突破性发展,特别是判别式模型和生成式模型在机器翻译、文本分类、文本摘要、情感分析、问答系统等领域日益成熟的应用,为传统学术成果评价模式的革新带来契机,为实现更加科学、多元与智能的学术成果评价提供了充分可能。基于对现有学术成果评价模式不足与人工智能应用契合度的分析,文章提出学术成果评价新模式——自然评价。自然评价模式是人工智能与学术成果评价的深度融合,既是对以往同行评议、文献计量、网络计量、替代计量等传统评价模式的批判性继承与发展,也是一种突破与革新。具体而言,自然评价是基于学术共同体在各种学术活动中自然产生的全量化的学术痕迹大数据,将人工智能技术应用于学术成果知识内容和学术共同体学术痕迹数据的语义理解与自动分析中,从而动态形成评价判断,并服务于知识创新与学术进步的一种评价模式。因其评价过程以自然形成为主、人为干预很少,故而命名为“自然评价”。文章从学理层面探讨了人工智能驱动下自然评价的技术逻辑、价值遵循与未来展望。就技术逻辑而言,自然评价以各类学术活动中的自然产生的痕迹数据为基础,通过算法支撑智能抽取数据中的语义并生成评价判断,通过算力赋能提升评价的精准度和效率。就价值遵循而言,自然评价秉持质量为先、公正为基、全面为要的价值原则,力求突围“数字规训”陷阱,破除“人情主导”桎梏,克服“片面评价”束缚。就未来展望而言,自然评价展现出顺应开放科学时代趋势,优化学术创新生态环境,促进自主知识体系建构的图景。

基金项目:中国人民大学科学研究基金重大项目“面向自主知识体系建构的学术代表作评价研究”(23XNL018)

作者简介:杨红艳,中国人民大学书报资料中心评价研究部主任,编审;徐拥军,中国人民大学信息资源管理学院教授,博士研究生导师,书报资料中心主任。

通信作者:卢思佳,中国人民大学信息资源管理学院,Email:lusijia_lsj@163.com。

最后,尽管本文已从学理层面系统探讨了人工智能驱动下自然评价的技术逻辑、价值遵循和未来展望,论述了其在理论上可以呈现出更科学的评价结果,但其中更为具体的人工智能技术实现机制、学术共同体评价激励机制、人机关系协调机制、不同主体评价赋权机制等难点,还有待在未来进一步展开研究。

关键词:自然评价;学术成果评价;生成式人工智能;判别式人工智能;自主知识体系;ChatGPT

中图分类号:TP18;G311 **文献标志码:**A **文章编号:**1008-5831(2023)04-0101-14

随着 ChatGPT 等人工智能聊天机器人的出现,人类的知识生产方式和学术研究范式发生了较大变革,对现行科研成果及人才评价机制形成了一定的冲击和影响^[1]。而在学术发展中,学术评价具有诊断、指引、激励、导向等多方面功能。近年来,管理部门先后出台多项关于学术评价改革的重要政策^[2]。2018年,中共中央办公厅、国务院办公厅印发《关于深化项目评审、人才评价、机构评估改革的意见》,要求进一步优化科研项目评审管理机制、改进科技人才评价方式、完善科研机构评估制度、加强监督评估和科研诚信体系建设^[3];2020年,教育部、科技部印发《关于规范高等学校 SCI 论文相关指标使用 树立正确评价导向的若干意见》《关于破除高校哲学社会科学研究评价中“唯论文”不良导向的若干意见》^[4-5];2021年,国务院办公厅出台《关于完善科技成果评价机制的指导意见》^[6]。上述文件要求坚决克服学术评价中的“五唯”倾向以及“SCI 至上”的痼疾,扭转“以刊评文”“以刊评人”“过度量化”等不良评价导向。而在各类学术评价活动中,学术成果评价作为科研项目评审、人才评价、期刊评价、机构评价等活动的基础,是学术评价体系的重中之重,也是科研管理的关键环节之一。

当前我国学术成果评价模式存在“同行评议黑箱”“引文动机模糊”“评价数据间接且片面”等弊端,阻碍了学术创新,也与当下建构自主知识体系的战略相悖。与此同时,人工智能等新一代信息技术近年来在数据、算法、算力上的突破性发展,特别是判别式人工智能(Discriminative Artificial Intelligence)和生成式人工智能(Generative Artificial Intelligence)在机器翻译、文本分类、文本摘要、情感分析、问答系统等领域日益成熟的应用,为传统学术成果评价模式的革新带来契机。而关于人工智能在学术成果评价中的系统应用还鲜有探讨。为此,笔者基于对现有学术成果评价模式不足及人工智能应用的契合度分析,提出学术成果评价的全新模式——自然评价。自然评价模式(简称“自然评价”)是基于学术共同体在各种学术活动中自然产生的全量化的学术痕迹大数据,将人工智能技术应用于学术成果知识内容和学术共同体学术痕迹数据的语义理解与自动分析中,从而动态形成评价判断,并服务于知识创新与学术进步的一种评价模式。基于此,本研究从学理层面对人工智能驱动下自然评价的技术逻辑、价值遵循与未来展望进行深入探讨。

一、研究回顾与框架构建

(一) 学术成果评价研究

同行评议和文献计量是当前广泛应用的两种学术成果评价模式。其中,同行评议主要指由科学系统内同行专家组成的团体,依据所设定的标准对学术成果进行考察^[7]。文献计量则是以学术成果数量、引用等数据为基础的计量评价^[8]。这两种方法在推动学术成果评价发展的同时,也面临诸多争议。其中,同行评议因评审专家范围小、主观性强^[9]、人情味浓^[10]、研究领域不完全对口^[11]、缺乏监督反馈机制^[12]等现实问题一直备受质疑,其评审结果的可靠性也因此被打上了问号。早在

1986年,泰伦斯·布鲁克斯(Terrence A. Brooks)就指出了引文分析法中对引文动机无区分的问题^[13]。文献计量也因引用行为的不充分性^[14]、随意性^[15]和“注水”问题^[16]等饱受诟病。

许多学者针对这些问题展开了深入探讨,提出了网络计量、替代计量等以社会影响和学术影响测度为主的新兴计量方法^[17],试图对现有评价模式进行优化,以使其能够更加科学和适用。其中,替代计量方法首先由普里姆-杰森(Priem Jason)在Twitter平台上提出^[18],并引发了研究热潮。与文献计量不同的是,替代计量采用下载量、收藏量、浏览量、转载量、点赞量、评论量等变量来衡量学术成果的质量^[19],提供了更多维的评价视角。伊曼纽拉-雷亚尔(Emanuela Reale)等还将学术成果影响力具体划分为科学影响力、政治影响力和社会影响力三个类别^[20]。不可否认的是,新兴评价方法在一定程度上确实弥补了传统同行评议和文献计量的缺陷。例如,替代计量不再局限于“小同行”的评价或是引文数据的分析,还将学术成果的社会影响力也纳入了评价范围^[21]。尽管学术成果评价模式已在不断修正,但仍多为细枝末节的修补,始终没有突破原有评价框架的限制。

结合近年兴起的新兴信息技术,部分学者对学术成果评价提出了新设想。例如,弗朗西斯科·隆扎诺(Ronzano Francesco)等提出利用自然语言处理和机器学习等人工智能技术来识别学术论文中的创新点、亮点等以评价学术成果价值^[22-23];由于引用内容分析能够客观揭示作者的引用意图,更加真实地考察被引文献的学术影响力^[24],万小军、李品等认为学术成果评价应转向多维度的引用内容与行为分析^[25-26],从语义层面评价被引文献的实际贡献;曾建勋提出应推动科研论文语义评价体系,开创基于语义内容的学术成果评价工具和模式^[27];杨红艳等剖析了大数据时代学术评价可能的变革,指出应开展基于大数据的学术成果评价^[28-29];此外,索传军、盖双双等提出可通过语料库、知识库和相关数据集的构建,实现基于认知计算的学术论文评价^[30]。但由于当时技术背景下尚存在数据难获取、质量低、成本高等问题^[31],其落地应用存在较大困难。

(二)人工智能发展现状

从19世纪50年代起发展至今,在科学家的不断推动下,人工智能历经逻辑推理、专家系统与深度学习三次发展浪潮^[32],已取得诸多突破性发展,呈现出跨界融合、人机协同、群智开放、自主操控等新特征。习近平总书记也多次作出重要指示,强调“要深入把握新一代人工智能发展的特点,加强人工智能和产业发展融合,为高质量发展提供新动能”^[33]。推动人工智能数次发展浪潮更迭的动力,在于数据、算法和算力三个要素的共同发展^[34]。目前对人工智能的界定存在多种解释,暂未达成统一定义。2021年,联合国教科文组织发布《人工智能伦理问题建议书》,提出可以将人工智能系统视为“能够以类似智能行为的方式来处理数据和信息的系统”,这种智能行为通常包括推理、学习、感知、预测、规划或控制等^[35]。综合学界主流观点与对人工智能三要素的考虑,本文所述“人工智能”可定义为以大数据和强计算能力为基础,在算法支撑之下而形成的能够按照一定目标模拟人类学习、推理、决策等思维活动的计算系统。

当下,人工智能正处于第三次浪潮的飞速发展阶段,广泛应用于医疗、教育、制造、自动驾驶等领域^[36]。数据要素方面,近数十年互联网的发展和普及为人工智能持续发展积聚了海量数据,提供了丰富的学习语料;算法要素方面,以深度学习为代表的算法模型的突破,机器学习、自然语言处理、人机交互技术等实践应用,使人工智能解决问题的能力不断提升。其中,卷积神经网络、循环神经网络、残差网络、迁移学习、集成学习等判别式人工智能模型的发展,则大大提高了自动分类、聚类、识别、预测的准确率和效率;循环神经网络、长短时记忆网络、变换器等生成式人工智能模型的

发展,使机器能够自主生成复杂的行为和策略。算力要素方面,GPU、NPU、FPGA 等各类人工智能专用芯片的诞生,云计算(并行计算)的出现,突破了运算能力限制,海量数据快速处理成为现实。再观照人工智能的现实应用情况:从1997年IBM公司“深蓝”击败了国际象棋世界冠军,到2016年谷歌旗下DeepMind公司的AlphaGo击败世界围棋冠军^[37],再到2022年底OpenAI公司发布ChatGPT,接入便捷、操作门槛低,而其所展现出的强大的对话理解、逻辑推理、多风格长文本生成以及程序代码自动生成等能力^[38],真正实现了人工智能的“去神秘化”“去壁垒化”^[39]。可以看到,在大量实践场景中,人工智能已从“可用”“能用”上升至“好用”甚至“惊艳”。

(三) 自然评价模式的框架构建

在人工智能驱动下开展学术成果评价模式创新,实现人工智能与学术评价的深度融合,是加强科研质量管理、营造公平学术创新环境的有力助推。因此,笔者基于现有学术成果评价模式的不足与学术评价“评判学术进展、鉴别学者贡献、规范学术行为、激发学者创造力”^[40]的目标旨归,综合考虑人工智能应用契合度,提出一种新的学术成果评价模式——自然评价。根据前文的定义,因评价过程以自然形成为主、人为干预很少,故而命名为“自然评价”。自然评价模式的实现须搭建学术共同体可以协同工作的数字化、网络化平台,并且面向广大学术共同体成员全样本、全流程、全量化地收集学术评价数据。

本文以人工智能“三要素”(数据、算法、算力)作为切入点,深入学术成果语义层面,梳理人工智能驱动下实现自然评价模式的技术逻辑,并围绕自然评价的价值遵循进行系统阐释,结合开放科学时代趋势、学术创新生态环境优化、自主知识体系建构展望自然评价的未来图景(图1)。

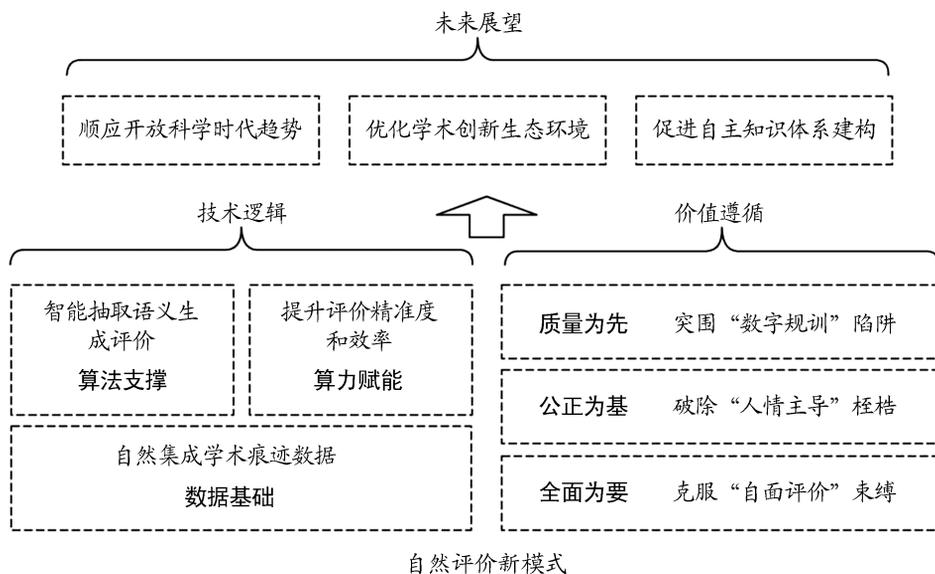


图1 自然评价新模式框架

二、人工智能驱动下自然评价的技术逻辑

当下人工智能的迅速发展,高度契合了学术成果评价对智能技术引入的迫切需求。就技术逻辑层面而言,人工智能驱动下的自然评价离不开数据、算法和算力的支持。自然评价以自然集成学术痕迹数据为基础,通过算法支撑智能抽取数据语义并生成评价判断,通过算力赋能提升评价的精准度和效率。

(一) 数据基础:自然集成学术痕迹数据

人工智能驱动自然评价的基础在于自然集成学术痕迹数据。即通过收集学者、期刊工作者等主体在学术活动中自然形成的海量学术痕迹数据,以备后续对其序化组织和语义分析,从而能够从学术共同体全同行、量化/语义双重维度推动对学术成果的全面评价。从形成主体视角看,可自然集成的学术痕迹数据主要包括四个方面。

1. 来自学者(作者或读者)的痕迹数据

以学术论文阅读、研究的场景为例,学者主要产生四类学术痕迹,分别为阅读痕迹、利用痕迹、评论痕迹与传播痕迹。其中,阅读痕迹包括阅读时间、部分内容停留时间、阅读笔记、收藏/下载、亮点/创新点标记、要点批注、内容质疑等;利用痕迹包括成果引用(需要区分引用动机)、内容摘取、关键词点击、相关内容扩展搜索等;评论痕迹包括点赞、评论、推荐、转发等;交流痕迹包括与作者的留言、私聊,与其他读者的评论互动等。

2. 来自期刊工作者的痕迹数据

以学术论文评审场景为例,期刊工作者主要产生三类学术痕迹,分别为评审意见痕迹、编辑反馈痕迹、交互信息痕迹。其中,评审意见痕迹包括评审结果、修改意见等;编辑反馈痕迹包括编辑修改意见等;交互信息痕迹包括邮件等沟通往来信息。

3. 来自实践工作者的痕迹数据

实践工作者是社会各领域从事具体实践工作并对学术场域有需求的用户或合作关系的群体,能够着重从学术成果的实践应用过程对其进行评价。来自实践工作者的学术痕迹主要包含成果转化痕迹、社会传播痕迹、成果反馈痕迹等。其中,成果转化痕迹包括学术成果被政策采纳的情况、转化为实际应用等;社会传播痕迹包括学术成果的社会传播广度、社会反响情况等;成果反馈痕迹包括实践工作者对于学术成果的应用效果的反馈,比如留言、点赞、投票、评论等。

4. 来自其他主体的痕迹数据

以评价活动、评价结果为主,包含学术成果质量评价痕迹、学术成果影响力痕迹、评价结果发布痕迹等相关数据,主要产生于学术成果第三方评价、媒体宣传报道等场景。其中,学术成果质量评价痕迹包括教学、科研、会议等多种场景下相关主体对于学术成果质量的评价判断;学术成果影响力痕迹包括引文、转载、媒体报道、政策采纳、企业应用等数据;评价结果发布痕迹包括专业评价机构、学会/协会、科研管理机构发布的多种评价成果。

(二) 算法支撑:智能抽取语义生成评价

算法依据人工智能所试图实现的目标而确立,同时也划定了目标实现的路径与方法^[41]。从算法要素来看,人工智能驱动自然评价的支撑在于依托判别式人工智能、生成式人工智能等系列算法的强大功能,智能抽取学术文献和上述自然集成的学术痕迹数据的语义数据,动态生成有评价功能的判断结果。在这个过程中,前者用于完成智能抽取学术亮点/创新点等判别式任务;后者用于承接智能生成评价结论等生成式任务。

1. 依托判别式人工智能从学术文献中智能抽取亮点、创新点等成果数据,作为生成评价结论的重要依据

以学术亮点/创新点为例,尽管人工标注准确度较高,但主观性强、成本高、效率低,无法解决海量学术文献亮点/创新点标注的需要^[42]。而判别式人工智能通过对上述集成的学者(作者、读者)

人工标注痕迹数据的学习,能够分析亮点的语言学特征、探寻其在论文中的分布规律,建立起不同类型数据输入到输出的映射关系,实现对论文亮点和创新点的自动识别和抽取。在此基础上,还可通过“从人类反馈中强化学习”(RLHF)机器学习方法,根据专家反馈信息不断调整参数和权重,进一步优化语义抽取模型性能。此外,智能抽取的亮点、创新点成果数据不仅可以作为评价依据,还可以作为学者的阅读辅助数据,帮助其快速准确把握文献要点。

2. 同时依托判别式人工智能与生成式人工智能,生成兼具量化指标与语义描述的综合性评价结果

依托前者,能够进一步对自然集成的学术痕迹数据和智能抽取的成果数据进行语义分析与综合。依托后者,则首先需要运用变分自编码器、生成对抗网络等算法构建适宜的生成式模型,并大量“投喂”人工评价数据进行模型训练。因此,在此过程中,可通过系统匹配,向关注和研究某类问题的学者广泛推送相关成果,鼓励其关注和评价留痕,以获取尽可能多人工评价数据用于模型训练。同时还可以再通过嵌入一定的赋权机制,分别赋予人工智能、“大同行”“小同行”、期刊工作者、实践工作者、其他评价机构等不同主体一定权重,在协调学术成果评价中人机关系的同时,尽可能地减少个别主体对评价结果的过度干预影响,从而最大限度保证评价结果的客观性、全面性。

(三)算力赋能:提升评价精准度和效率

人工智能驱动下自然评价精准度和效率的提升源于强大算力的赋能。自然评价的算力通过学术活动数字平台来实现。数字平台集成学术痕迹数据收集、智能抽取语义生成评价等多重功能,依靠对全样本、大数据的算力赋能以提升评价的精准度和效率。

1. 算力赋能以提升自然评价的精准度

一方面,算力通过确保各主体在学术活动中自然产生的学术痕迹数据均能够得以收集、保存、序化、分析,积累充分的评价依据,从数据获取的全面性层面提升自然评价的精准度。另一方面,算力通过确保对学术文献亮点/创新点等成果数据的语义抽取的准确性,促进评价的自动化和智能化,从数据理解的可靠性层面提升自然评价的精准度。

2. 算力赋能以提升自然评价的效率

一方面,通过增强芯片计算能力和优化算法模型能够提升平台算力,支持超大“码流”的内容传入评价平台云端进行存储和分发,支持大规模、宽广度、多粒度的原始学术痕迹数据的挖掘和分析,为学术评价数据收集、存储、分析全过程提供底层算力支持。另一方面,算力赋予自然评价以全时空生产力。在算力支持下,人工智能可以随时随地进行评价数据的收集和分析,为评价工作者在具体评价活动中的“离场”提供了空间,这在很大程度上减轻了人力评价负担。但这种“离场”并非代表评价责任完全转嫁于人工智能,而是允许学术共同体有更多精力从更高层面来思考学术成果评价导向和具体评价标准的优化——这实质上强化了学术共同体在学术成果评价中的主导者地位。

三、人工智能驱动下自然评价的价值遵循

人工智能驱动下自然评价的技术逻辑与价值原则交织互融——不仅需要数据、算法、算力作为驱动,还有利于秉持质量为先、公正为基、全面为要的价值原则,突围“数字规训”陷阱,破除“人情主导”桎梏,克服“片面评价”束缚,达成评价目标、评价方式与评价价值的有机统一。

(一)质量为先:突围“数字规训”陷阱

传统学术成果评价多建立于对发表期刊级别、引用量、发文期刊影响因子等指标进行统计的基础上,评价工作程序“简洁高效”,还能有效调动学者的科研生产积极性。但量化评价产生的各种数字在成为工具的同时,也逐渐形成了“规训”的力量。“规训是一种强制给定的技术结构,不仅具有干预、监视肉体的能力,而且还是知识生产的重要手段”^[43]。这集中体现于学术评价活动中各类“数字至上”现象。例如,某高校设置如下奖励政策:在SCI、SSCI等期刊上发文1篇,最高奖励10万元^[44]。而无论是对科研机构、高校、教师还是对学生的学术评价,均建立于其所发表学术成果的“期刊级别”“影响因子”“引用量”的基础上——评价变成计数比赛,质量则是其次。自然评价则将质量为先列为首要原则,突围评价的“数字规训”陷阱。

1. 坚持以评估学术成果的质量为核心

“数字规训”使得学术成果的价值评价受制于数字化编码,学者则深受数字化控制^[45]。这违背了学术生产的基本规律,导致学术成果数量繁荣却质量下降。尽管发文期刊级别、影响因子、引用量等量化数据的确能够在一定程度上反映出学术成果的质量水平,但如果“唯以数字论英雄”,则有失偏颇。相对于这些数字指标,学术痕迹数据中的语义数据和智能抽取的学术亮点/创新点等语义数据包含更多有助于科学评价学术成果质量的信息,因此,自然评价应在将量化数据纳入考量的基础上,赋予语义评价数据以相对更高的权重。

2. 采用多元机制评价学术成果质量

自然评价强调评价主体的多元性、评价方式的多元性与评价工具的多元性。例如,从个别同行评价转向基于学者、编者、管理者、实践者等多元主体的评价;从小规模样本的评价转向面向海量文献的大数据评价;使用前台评价工具与后台评价工具的多元组合,其中后台评价工具用于自然集成学术痕迹数据,汇聚海量学术成果评价数据,前台评价工具则用于对后台原始数据进一步智能抽取、语义挖掘和知识序化,最终面向用户呈现客观真实的评价结果。

3. 保护需要长期投入、坐“冷板凳”的研究

应用型研究更重视时效性,发文效率一般相对较高。而基础型研究则更需要长期积淀,常常“十年磨一剑”。尽管其在学术研究整体发展中起着基石性作用,但却不宜采用“计数式”评价,不可避免地现行评价体系中处于弱势地位。自然评价对这类研究成果,着重聚焦对其理论深度、创新价值的判断,以保护基础理论研究者免受巨大的量化评价压力,使其有更多时间和精力来长期深耕,让真正专注于学术研究本身的学者都能拥有成就感和获得感^[46]。

(二) 公正为基:破除“人情主导”桎梏

长期以来,学术成果评价的话语权“隐性”地掌握在少数权威机构和学者的手中——“以刊评文”现象盛行,但“好”期刊对知名机构、本单位或本系统相关机构学者多有额外偏爱,难以完全做到“以质论英雄”,从而损害正常的知识创新和传播。同时,不公正的学术评价也将损害学术共同体的权益。因此,自然评价最基础、最底层的价值原则便是公正,这要求破除“人情主导”桎梏。

1. 以人工智能平衡各主体的评价权力

依托于自然集成的海量数据、强大的算法和算力,人工智能在驱使传统学术评价模式走向自然评价模式的同时,也显现出“自主性”潜能:人工智能可以基于海量学术痕迹数据进行自主学习,研判和模拟学者的评价思维,甚至识别出其中的“违规”评价行为并过滤违规数据,呈现出既作为人类评价助手而又具有自主评价决策能力的算法评价形态。通过对海量评价数据的收集、抽取、分析、

挖掘和自主生成评价结果,人工智能在自然评价中也达成了技术权力的建构,从而成为了同各学术评价主体并行的评价者。这在一定程度上均衡了少数权威机构或学者的学术评价话语权,形成了对传统评价权威的补充,起到了权力的平衡作用。

2. 提前规避人工智能作为评价主体可能带来的新的“不公正”问题

人工智能在为学术成果评价带来超高效率的同时,也因其强大的评价功能而聚合形成新的权力。这是因为,尽管人工智能自身具有一定自主学习和决策能力,但究其根本,最底层的学术痕迹数据集成机制、语义抽取机制、评价生成机制等均是人为设计,其中所运用的算法也是由人所设计——评价机制和评价算法的设计者能够在很大程度上左右评价结果。因此,人工智能不仅应视作自然评价的驱动者,也应视作关键的评价主体对其予以监督,提前规避算法带来的不公正问题,确保平等对待所有学者。由此,人机协同评价成为可能。

3. 采取多种措施约束潜在的不当评价行为

在规制人工智能评价机制和评价算法之外,还需设计面向学者的评价行为规范体系。自然评价模式下一篇学术成果所接受的评价来自多元主体,特别是众多大小同行学者。在此情况下,可能会出现类似当下娱乐产业中“打榜”式的流量化运作而产生大量虚假评价数据,这将会冲击评价结果的公信力,不利于营造求真求实的学术环境。因此,应通过管理或技术的途径采取有力措施约束不当评价行为,包括实名注册、技术识别与过滤、提高评价行为失范成本,并引导学者树立正确评价价值观,以避免类似马太效应的现象在人工智能支持下愈演愈烈。

(三) 全面为要:克服“片面评价”束缚

学术成果评价的积极作用在于促进学术的发展进步,但如若导向有偏,也将滋生唯论文、唯帽子、唯职称、唯学历、唯奖项等问题,助长“学术泡沫”“学术腐败”^[47],出现诸如购买论文版面、找同行评审专家“打招呼”、组团“刷引用数据”等乱象。而产生这些问题的其中一个关键因素便是学术成果评价的片面性。因此,自然评价的重要原则之一是克服“片面评价”束缚,拥抱多元立体的评价理念,从评价对象、评价主体、评价流程、评价结果表征等多方面对学术成果进行全面评价。

1. 全主体、全成果评价

首先是全主体评价。对学术成果的评价,既应包含编辑对稿件价值的判断,也应包含其他学者对论文的综合评价数据,还应集成智能抽取的语义数据,方能形成对学术成果的全方面评价。由有限数量的精英专家评价向全体学术同行评价转变,由学术共同体主导学术成果评价,实现评价话语权的重组和优化,规避评价中的偏颇,并用学术成果评价引导学术生态良好发展。这对于读者、作者、审稿者、期刊出版商等均有重要意义。其次是全成果评价。挪威高等教育机构协会曾提出基于学术成果表现的“挪威模型”(Norwegian Model),对评价对象的全面性提出了要求——应涵盖全国各领域结构化、可验证、有效的全部学术成果记录^[48]。自然评价在此基础上,进一步倡导应覆盖预印本、论文、著作、研究报告等全种类评价对象,且不拘泥于文本,还包括数据、图片、视频、音频等知识。依托学术共同体的“云评价”模式,基于其专业属性与能力开展全样本评价和数据分析,形成针对所有学术成果的评价大数据,层层精选,逐步定位优秀文献。

2. 全程嵌入学术创新系统

受时间成本与人力成本限制,当前学术成果评价通常是在特定时间节点的评价,且通常为“一次性”评价。人工智能驱动下的自然评价模式允许学术共同体在任何接触评价对象的节点产生评

价数据,促进学术成果评价动态性发展,完成评价从时点评价向动态评价、学术生产全流程评价的转变。学术共同体可以在学术生产流程的各个环节中参与评价,例如审稿评价、二次转载评价、读者评价、学术轨迹评价、突破性理论评价等。从研究开始那一刻,成果全生命周期都要接受学术共同体的监督和遴选。在自然评价模式下,学术评价与学术创新将形成良好互动,同时,科学的评价结论也能为学者开展研究提供有效参考,例如帮助识别研究热点、难点、前沿等,使学者能够更好地集中精力于“创新”环节。

3. 评价结果表征量化和语义并行

强调共性与普适性的量化评价指标有利于提高学术成果评价的效率,但却会抹去其中复杂而内隐的诸多影响因素,如学科差异、领域差异,以及是否有“名校”“名学者”等光环等。在全量化评价下,被评价成果的这些差异通通被隐去,其学术价值简化为一个个简单的指标、数据^[49],在高效率的同时也导致了学术成果评价的“失真”。因此,应坚持内部学术共同体评价的根本性,并参照外部数理统计的普适量化评价,借助外部数理统计量化评价中效度数的科学解释来佐证或补充内部主体性评价的结论。换言之,就是内部评价赋值于外部评价的“系数”,实现多维互动的立体式评价,回归学术成果评价注重创新和价值的初衷。

四、人工智能驱动下自然评价的未来展望

(一) 顺应开放科学时代趋势

自然评价需要建立于资源开放流动的基础之上,开放科学的发展改变了传统学术交流体系,促使学术成果评价朝着开放化方向重塑,而自然评价将能够很好地承担这一转向职责。未来,人工智能驱动下自然评价必然顺应着开放科学的时代趋势,有利于净化学术空气、清除学术不端;保障学术信息资源自由流动、加速科学研究进展是开放科学的发展初衷。两者构成了相互促进的关系,共同推动学术生态的健康发展。

一方面,开放获取平台从评价端改善学术生产、学术传播。通过开放数据、开放期刊、开放各类设施资源,更多学者能够以低成本或无成本接触和利用学术文献,产生更多元的学术痕迹数据,进而提高自然评价的覆盖面和准确度。此外,在预印本等开放获取平台透明评审和开放评议渠道的支持下,学者能够及时掌握最新研究动态,避免开展重复研究。

另一方面,开放评价平台接收学术共同体的广泛监督。封闭性评价通常是聘请特定数量的专家、基于特定目的或规则、在特定范围内进行学术价值判断。而基于开放平台,学术成果评价可以覆盖更多的学术共同体成员,收集更加广泛的评价数据与观点集合,接受学术共同体的广泛监督,改变传统评价模式中“暗箱操作”“人情关系”“左右为难”等状况。开放的评价平台对创新研究更加友好和包容^[50],并从评价主体多元性和评价结果的综合性,保证了评价结果的科学、专业、公正、客观。

最后,开放式同行评议为学术再生产提供契机。在尊重作者和评审者意愿基础上,可公开学术成果的评审者和作者身份(可选择性进行出版前开放同行评议或出版后开放同行评议),最终将审稿人的建议、作者的修改及回复、公众评议的结果等信息,在相关主体均知情的前提下最大化向公众开放。开放同行评议打破了作者与审稿人间交流的壁垒,作者和审稿人可以针对论文存在的不足展开有效沟通,实名的审稿人会更加认真对待评议工作,给出更多具有创新性的建议,同时大众

读者也可以给出一些修改意见,这就使得学术成果评议在很大程度上得到质的改善。与此同时,学者还可以通过开放的评价数据了解到同行最新的研究关注点,并有可能在这种开放对话式的同行评议中找到合作伙伴,促进学术再生产。

(二) 优化学术创新生态环境

党的二十大报告提出,要“加快实施创新驱动发展战略”,“坚持创新在我国现代化建设全局中的核心地位”^[51]。自然评价通过有机嵌入学术创新系统,能够促进学术资源合理分配,营造求真求实的研究氛围,保护和促进学术再生产,进而优化学术创新生态环境。

一方面,以包容的自然评价营造求真求实的研究氛围。学术的本质在于求真。以数字符号、量化指标为取向的学术成果评价实施起来更加简便高效,满足了学术行政部门绩效问责的初衷,但却有失学术求真求实的初心和价值,甚至导致趋同、内卷、异化的学术氛围^[45],学术成果评价作为指挥棒之同时也成为了一道学术“枷锁”。自然评价模式在数据、算法和算力的支持下嵌入学术生产全程,集成了源于广大学术共同体价值判断和来自人工智能自动抽取语义而形成的多元评价数据,其产生的评价结果,能够更加真实、客观、全面地呈现出学术成果的质量,进而推进知识创新进程。

另一方面,以公正的自然评价促进学术资源的合理分配。广义层面的学术资源是指支持学术研究的各类有形或无形条件的总和,包括期刊版面、课题资助、社会声誉等^[52]。关于学术评价如何作用于学术资源分配,有学者通过分析指出,学术资源的传输构成一条利益链“行政权力部门(掌握资源)—学术评价者(分配、管理资源)—科研机构及工作者(开展学术研究)”,学术评价则构成了其中的中介节点^[53]。学术成果评价是其他所有类型学术评价(对学者、机构、期刊等的评价)的最底层基础,学术成果评价的不公正将会导致学术资源分配失衡,违背以评价引领学术研究健康发展、鼓励学者知识创新的初心。自然评价模式的运用,使得学术评价主体更多元、标准更科学,分散了原本集中的学术评价话语权,将促使学术成果评价回归于以评价诊断学术问题、监督学术成效、促进学术创新^[54]的初衷,进而促进学术资源的合理分配,以维护学术创新生态的良好发展。

(三) 促进自主知识体系建构

百年未有之大变局对建构中国自主知识体系而言是巨大的机遇。国家的发展与世界的变化为学术研究提供了不竭的实践和问题源泉,也为学者做出具有中国特色、国际影响的“真研究”“好成果”提供了强大动力和广阔空间^[55]。与此同时,学术成果评价作为研究的“指挥棒”,对于自主知识体系建构的深度、广度起着决定性影响。自然评价模式的推广应用,将有助于形成自主学术评价体系,为学者开展自主研究创造更好的评价环境,进而促进我国自主知识体系建构。

一方面,以质量为评价标准,扭转学术“伪国际化”怪圈。近年来,我国的学术评价标准一直在“向外看齐”。无论是对科研院所、高校等科研机构学术水平的评价体系,还是学者职称、“帽子”等的评定标准,都体现出这样一种理念:在国际上权威期刊、权威平台发表学术成果,才算获得“学术国际化”入场券,才称得上被国际学术同行所认可^[56]。例如,部分高校在人才引进政策中明确规定,某类人才需至少发表3篇被SCI收录的前5%期刊的论文^[45]。所谓“学术国际化”,实质上乃是对西方学术评价体系的臣服。但越来越多学者注意到,如此盲目崇拜国际权威期刊、围着西方学术评价体系转圈的“学术国际化”,实质上是一种评价标准的错位和学术标准的功利化偏移^[46]。同时,这种评价标准上的“向外看齐”,也直接导致了学术成果“向外流失”的严重问题。人工智能驱动下的自然评价以质量为先,不拘泥于是否发表于所谓的国际权威期刊,而是从评价端鼓励学术“质”的

创新、社会价值的彰显,从而使学术研究脱下“伪国际化”的外衣,而是以“质”服人,赢得真正的国际学术话语权。

另一方面,推进学术评价自主,为学者开展自主研究创造更好的评价环境。在自然评价模式下,将不再把国际期刊发文视作“学术国际化”的硬指标,而是探索中国学术成果评价的自主道路。首先是研究内容自主。学者不必为追捧“国际化”而刻意“国际化”,以致在研究的侧重、内容、风格等都“思他人之所思、想他人之所想”^[57],而是可以在放眼看世界的同时生长于本土,“思人所思”的同时“想我所想”,为自主知识体系建构积累丰厚成果。其次是研究语言自主。不同的语言代表着不同的思维方式、不同的文化背景。在国际学术界最通行的语言为英文,SCI、SSCI、AHCI等国际权威期刊很少接纳非英语的研究成果和学术期刊^[55]。以英文发表成果能获得最广泛传播和认可,而这种传播和认可的优势使得许多非英语母语国家的学者只能使用英文来发表成果,又进一步加强了英文学术语言优势^[58]。这背后所反映的不仅仅是语言霸权,还有科研霸权、文化霸权。当前,我国已具备提升中文国际影响力的宏观基础,在自然评价模式的推动下,中国学者更加不必再“削足适履”,为发表国际期刊而使用英文等外语来思考、写作,而是可以使用自己最熟悉的母语中文来开展中国的研究。最后是研究发表自主。由于不必再刻意将成果发表于国外权威期刊,学者可以将之发表于中国的高质量学术期刊。而随着高质量成果的积累,中国的高质量学术期刊也将能够在国际学术界占据更核心的地位,进而吸引更多优质自主研究成果稿源,反哺自主知识体系建构。

五、结语

本研究提出的自然评价模式是人工智能与学术成果评价的深度融合,既是对以往同行评议、文献计量、网络计量、替代计量等传统评价模式的批判性继承与发展,也是一种突破与革新,为实现更加科学、多元与智能的学术成果评价提供充分可能。在质量为先、公正为基、全面为要的价值遵循下,自然评价以其强大的数据基础、算法支撑、算力赋能“另辟蹊径”,在理论上可以呈现出更科学的评价结果,让真正有创新、有贡献、自主性强的学术成果能够在评价中脱颖而出。需要说明的是,尽管本文已从学理层面系统探讨了人工智能驱动下自然评价的技术逻辑、价值遵循和未来展望,但其中更为具体的人工智能技术实现机制、学术共同体评价激励机制、人机关系协调机制、不同主体评价赋权机制等难点,还有待在未来进一步展开研究。

参考文献:

- [1] 蒋华林. 人工智能聊天机器人对科研成果与人才评价的影响研究——基于 ChatGPT、Microsoft Bing 视角分析[J]. 重庆大学学报(社会科学版), 2023(2): 97-110.
- [2] 中共中央办公厅 国务院办公厅印发《关于深化职称制度改革的意见》[EB/OL]. (2017-01-08) [2023-03-09]. http://www.gov.cn/zhengce/2017-01/08/content_5157911.htm#1.
- [3] 中共中央办公厅 国务院办公厅印发《关于深化项目评审、人才评价、机构评估改革的意见》[EB/OL]. (2018-07-03) [2023-03-09]. http://www.gov.cn/zhengce/2018-07/03/content_5303251.htm.
- [4] 教育部 科技部印发《关于规范高等学校 SCI 论文相关指标使用 树立正确评价导向的若干意见》的通知[EB/OL]. (2020-02-18) [2023-03-09]. http://www.gov.cn/zhengce/zhengceku/2020-03/03/content_5486229.htm.
- [5] 教育部印发《关于破除高校哲学社会科学研究评价中“唯论文”不良导向的若干意见》的通知[EB/OL]. (2020-12-10) [2023-03-09]. http://www.moe.gov.cn/srcsite/A13/moe_2557/s3103/202012/t20201215_505588.html.
- [6] 国务院办公厅关于完善科技成果评价机制的指导意见[EB/OL]. (2021-07-16) [2023-03-09]. <http://www.gov.cn/>

zhengce/content/2021-08/02/content_5628987.htm.

- [7] 郭碧坚, 韩宇. 同行评议制——方法、理论、功能、指标[J]. 科学学研究, 1994(3): 63-73.
- [8] 刘作义, 陈晓田. 科学研究评价的性质、作用、方法及程序[J]. 科研管理, 2002(2): 33-40.
- [9] 万昊, 谭宗颖, 朱相丽. 同行评议与文献计量在科研评价中的作用分析比较[J]. 图书情报工作, 2017(1): 134-152.
- [10] 王平, 宋子良. 同行评议制的固有缺点与局限性[J]. 科技管理研究, 1994(4): 22-26, 13.
- [11] 刘爱玲, 王平, 宋子良. 科技奖励评审过程的研究[J]. 科学学研究, 1997(1): 49-55.
- [12] 杨瑞仙, 李贤, 李志. 学术评价方法研究进展[J]. 情报杂志, 2017(8): 106-112.
- [13] BROOKS T A. Evidence of complex citer motivations[J]. Journal of the American Society for Information Science, 1986, 37(1): 34-36
- [14] MACROBERTS M H, MACROBERTS B R. Problems of citation analysis: a critical review[J]. Journal of the American Society for Information Science, 1989, 40(5): 342-349.
- [15] 马凤, 武夷山. 关于论文引用动机的问卷调查研究——以中国期刊研究界和情报学界为例[J]. 情报杂志, 2009, 28(6): 9-14, 8.
- [16] 周春雷. h指数合作式注水缺陷与对策[J]. 图书情报知识, 2009(3): 109-112.
- [17] 刘廷元, 刘纾曼. 网络计量学和替代计量学的挑战及其社会影响的稳健和非稳健计量评价[J]. 情报学报, 2022(10): 1044-1058.
- [18] PRIEM J. I like the term #articlelevelmetrics, but it fails to imply * diversity * of measures. Lately, I'm liking #Altmetrics [EB/OL]. (2017-02-27) [2022-12-10]. <https://twitter.com/jasonpriem/status/25844968813>.
- [19] 陈敏, 王轶. 破“五唯”政策视角下的学术成果评价研究[J]. 重庆大学学报(社会科学版), 2021(4): 60-70.
- [20] REALE E, AVRAMOV D, CANHIAL K, et al. A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research[J]. Research Evaluation, 2018, 27(4): 298-308.
- [21] 余厚强, 邱均平. 替代计量指标分层与聚合的理论研究[J]. 图书馆杂志, 2014(10): 13-19.
- [22] RONZANO F, SAGGION H. Knowledge Extraction and Modeling from Scientific Publications[C]//International Workshop on Semantic, Analytics, Visualization. Cham: Springer, 2016: 11-25.
- [23] YANG W. Evaluative language and interactive discourse in journal article highlights[J]. English for Specific Purposes, 2016, 42: 89-103.
- [24] SMALL H, KLAVANS R. Identifying Scientific Breakthroughs by Combining Co-citation Analysis and Citation Context [C]//Proceedings of the 13th international conference of the International Society for Scientometrics & AMP; Informetrics. Leuven: International Society for Scientometrics & AMP; Informetrics, 2011: 783-793.
- [25] WAN X J, LIU F. Are all literature citations equally important? Automatic citation strength estimation and its applications [J]. Journal of the Association for Information Science and Technology, 2014, 65(9): 1929-1938.
- [26] 李品, 杨建林. 大数据时代哲学社会科学学术成果评价: 问题、策略及指标体系[J]. 图书情报工作, 2018(16): 5-14.
- [27] 曾建勋. 推动科研论文语义评价体系建设[J]. 数字图书馆论坛, 2021(11): 1.
- [28] 杨红艳. 顺势而为: 基于大数据的学术评价新模式探析[J]. 甘肃社会科学, 2018(3): 76-81.
- [29] 杨英伦, 杨红艳. 学术评价大数据之路的推进策略研究[J]. 情报理论与实践, 2019(5): 62-66, 152.
- [30] 索传军, 盖双双, 周志超. 认知计算——单篇学术论文评价的新视角[J]. 中国图书馆学报, 2018(1): 50-61.
- [31] 杨红艳. 大数据时代学术评价的数据化难点及其应对[J]. 现代情报, 2020(11): 136-143.
- [32] 曾海军, 张钰, 苗苗. 确保人工智能服务共同利益, 促进教育系统变革——《人工智能与教育: 政策制定者指南》解读[J]. 中国电化教育, 2022(8): 1-8.
- [33] 新华社. 习近平: 推动我国新一代人工智能健康发展[EB/OL]. (2018-10-31) [2023-03-25]. <http://www.scio.gov.cn/31773/31774/31783/Document/1640446/1640446.htm>.
- [34] 刘珊, 黄升民. 人工智能: 营销传播“数算力”时代的到来[J]. 现代传播(中国传媒大学学报), 2019(1): 7-15.
- [35] UNESCO's Recommendation on the Ethics of AI [EB/OL]. [2023-03-26]. <https://montreal.ethics.ai/unescos-recommendation-on-the-ethics-of-ai/>.

- [36] 中国信息通信研究院. 人工智能白皮书(2022年)[R/OL]. (2022-04-12)[2023-03-26]. <http://www.caict.ac.cn/kxyj/qwfb/bps/202204/P020220412613255124271.pdf>.
- [37] 李帅. ChatGPT: 多维思考与审慎应用[N]. 中国社会科学报, 2023-03-21(06).
- [38] 任文岱. ChatGPT 热度不减——专家建议人工智能领域制定行业安全标准[N]. 民主与法制时报, 2023-03-22(03).
- [39] 蒲清平, 向往. 生成式人工智能——ChatGPT 的变革影响、风险挑战及应对策略[J]. 重庆大学学报(社会科学版), 2023(3): 102-114.
- [40] 李剑鸣. 自律的学术共同体与合理的学术评价[J]. 清华大学学报(哲学社会科学版), 2014(4): 73-78.
- [41] 贾开, 蒋余浩. 人工智能治理的三个基本问题: 技术逻辑、风险挑战与公共政策选择[J]. 中国行政管理, 2017(10): 40-45.
- [42] 索传军, 于果鑫. 学术论文研究亮点的语言学特征与分布规律研究[J]. 图书情报工作, 2020(9): 104-113.
- [43] GAITHER G H. Measuring up: the promises and pitfalls of performance indicators in higher education. ASHE-ERIC higher education report, 1994(66): 2.
- [44] 沈文钦, 毛丹, 蔺亚琼. 科研量化评估的历史建构及其对大学教师学术工作的影响[J]. 南京师大学报(社会科学版), 2018(5): 33-42.
- [45] 张卓, 刘冬冬. 高校教师学术评价的数字规训及其突围[J]. 大学教育科学, 2023(1): 74-82.
- [46] 胡钦太. 中国学术国际话语权的立体化建构[J]. 学术月刊, 2013(3): 5-13.
- [47] 张保生. 学术评价的性质和作用[J]. 学术研究, 2006(2): 10-15.
- [48] 刘强, 陈云伟, 张志强. 用于科技评价的挪威模型方法与应用综述[J]. 数据分析与知识发现, 2019(5): 41-50.
- [49] 余利川, 李佳源, 段鑫星. 大学学术评价的技术治理“失灵”与理性复归[J]. 西南大学学报(社会科学版), 2022(3): 171-183.
- [50] 刘益东. 从同行承认到规范推荐——开放评价引发的开放科学革命与人才制度革命[J]. 北京师范大学学报(社会科学版), 2020(3): 29-41.
- [51] 习近平. 高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——习近平同志代表第十九届中央委员会向大会作的报告摘登[N]. 人民日报, 2022-10-17(02).
- [52] 徐拥军, 陈晓婷, 杨红艳, 等. 学术资源反垄断视角下知识创新促进策略研究[J]. 情报资料工作, 2023(2): 53-63.
- [53] 朱剑. 科研体制与学术评价之关系——从“学术乱象”根源问题说起[J]. 清华大学学报(哲学社会科学版), 2015(1): 5-15.
- [54] 宋丽萍, 王建芳. 开放科学环境下负责任评价实现路径研究[J/OL]. 中国图书馆学报: 1-12. <http://kns.cnki.net/kcms/detail/11.2746.G2.20230217.0935.002.html>.
- [55] 胡海波. 努力建构中国自主的知识体系[N]. 中国社会科学报, 2022-06-28(08).
- [56] 朱剑. 学术评价、学术期刊与学术国际化——对人文社会科学国际化热潮的冷思考[J]. 清华大学学报(哲学社会科学版), 2009(5): 126-137.
- [57] 罗志田. 史学前沿臆说[J]. 四川大学学报(哲学社会科学版), 2008(4): 32.
- [58] 文秋芳. 学术国际话语权中的语言权问题[J]. 语言战略研究, 2021(3): 76-85.

Natural evaluation: The reconstruction of the academic Publication evaluation model driven by artificial intelligence

YANG Hongyan^{a,b}, LU Sijia^b, XU Yongjun^{a,b}

(*a. Information Center for Social Sciences; b. School of Information Resources
Management, Renmin University of China, Beijing 100872, P. R. China*)

Abstract: As the basis for the evaluation of scientific research projects, talents, journals, and institutions, academic publication evaluation is an important part of the academic evaluation system and a key

step in scientific research management. The current academic publication evaluation model has some disadvantages such as “black box of peer review”, “ambiguous motivation of citation” and “indirect and one-sided evaluation data”, which hinders academic innovation and also runs counter to the current strategy of building an independent knowledge system. The breakthrough development of AI in data, algorithms, and arithmetic power in recent years, especially the increasingly mature application of discriminant and generative models in machine translation, text classification, text summary, dialogue system, etc. has brought a great opportunity for innovation of the traditional academic publication evaluation model and provided a possibility to realize a more scientific, pluralistic, and intelligent academic publication evaluation. From the analysis of the inadequacy of the existing evaluation models and the fit of AI applied to it, the paper proposes a new model of academic publication evaluation, namely Natural Evaluation. Natural Evaluation is an integration of AI and academic publication evaluation, which is both a critical inheritance and a breakthrough of the traditional evaluation models such as peer review, bibliometrics, webometrics, and alternative metrics. Specifically, Natural Evaluation is an evaluation model based on the fully quantified academic trace data naturally generated by the academic community in various scholarly activities, and AI technology is applied to the semantic understanding and automatic analysis of the knowledge content of academic results and academic trace data of the community, to dynamically form evaluation judgments and serve knowledge innovation. The paper discusses the technical logic, value compliance, and prospect of Natural Evaluation driven by AI from the doctrinal level. In terms of technical logic, Natural Evaluation is based on the naturally generated trace data from various scholarly activities, and the algorithm supports the intelligent extraction of semantics in the data and generates evaluation judgments, and enhances the accuracy and efficiency of evaluation through arithmetic power empowerment. In terms of value compliance, Natural Evaluation upholds the value principles of quality first, impartiality, and comprehensiveness, and seeks to break out of the trap of “digital discipline”, and break the shackles of “human favor dominance” and “one-sided evaluation”. In terms of the future outlook, Nature Evaluation presents a picture of responding to the trend of the open science era, optimizing the ecological environment of academic innovation, and promoting the construction of independent knowledge systems. Finally, although the technical logic, value compliance, and prospect of Natural Evaluation have been systematically discussed from the theoretical level, and it can be theoretically proved to present more scientific evaluation results, the specific difficulties of AI technology implementation, academic community evaluation incentive, human-machine relationship coordination, and evaluation empowerment of different subjects are yet to be further researched in the future.

Key words: Natural Evaluation; academic publication evaluation; generative artificial intelligence; discriminative artificial intelligence; independent knowledge system; ChatGPT

(责任编辑 彭建国)