

文章编号: 2095-1663(2024)02-0037-08 DOI: 10.19834/j.cnki.yjsjy2011.2024.02.06

大数据赋能研究生学术不端行为治理

——价值意蕴、现实梗阻与推进策略

邹太龙^{1,2}

(1. 广州大学 教育学院, 广州 510006; 2. 湖北民族大学 教师教育学院, 湖北 恩施 445000)

摘要:作为一种全新的思维方式和先进的技术手段,大数据赋能研究生学术不端行为治理的价值意蕴主要体现在学术成果不断数据化让抄袭造假行为难以藏身、大数据检测软件防控各类学术不端行为的发生、互联网黑名单系统完善学术不端行为惩戒机制、聚类分析有助于开展针对性强的学术道德教育。然而,大数据赋能研究生学术不端行为治理也面临着学术监督主体缺乏使用大数据的行为意向、各类科研数据库的开放共享程度亟待提高、科研成果价值开发和合理使用的矛盾加剧、大数据固有缺陷妨碍学术不端行为的治理等现实梗阻。为此,需要打通科研数据开放共享的融合渠道、研发基于大数据的学术不端智能监督系统、提高学术不端监督主体的数据素养;全方位撑起科研数据的安全保护伞。

关键词:大数据;研究生学术不端;ChatGPT;学术不端行为治理

中图分类号: G643

文献标识码: A

作为学术研究队伍的中坚力量,研究生是科学研究的直接参与者和重要贡献者,其培养质量和学术素养成为衡量一个国家创新能力的关键指标,直接决定着我国当前乃至未来的科技竞争力^[1]。新世纪以来,我国研究生教育质量不断提高,研究生在科技创新中的作用日益凸显,贡献度也持续攀升。据教育部统计数据显示,自然科学基金重点项目成员中超过50%为在读研究生,其中博士生比例超过30%,许多世界级的研究成果,主要研究人员都是我国自主培养的博士^[2]。但是近些年来,研究生学术不端行为也屡见不鲜,已成为发生学术不端行为的“重灾区”^[3],引发了社会大众的广泛热议、网络媒体的争相报道和国家部门的高度重视。研究生学术不端是学界之殇,它不仅妨碍研究生的学术成长和全面发展,也贬低学术尊严,败坏学术风气,玷污学术

生态,引发学术共同体危机,更会降低人才培养质量,严重制约国家的学术繁荣发展和民族的自主创新能力。

治理学术不端行为、加强学术规范教育、推进学术道德建设,高校依然任重道远。特别是迈入大数据时代,研究生学术不端行为更加隐匿复杂和花样百出,涌现出一些新变化、新特征和新问题,学术不端行为发生了变异和升级,增加了甄别、监测和惩戒的难度。例如,面对研究生使用ChatGPT代写的学术论文,传统的专家评鉴方式难以奏效,教育评价与学术诚信度认定变得愈加困难,由此催发的学术伦理问题的迫切性、系统性、使命感日益凸显^[4]。因此,为提高新时代研究生学术不端行为的治理效能,除了要沿用和完善行之有效的既有措施外,如加大学术诚信教育宣传力度、强化导师学术诚信正面引

收稿日期: 2023-09-11

作者简介: 邹太龙(1988—),男,土家族,湖北巴东人,广州大学教育学博士后流动站特聘副研究员,湖北民族大学教师教育学院副教授,北京师范大学未来教育研究中心特约研究员,教育学博士。

基金项目: 中国博士后科学基金第73批面上资助项目“大数据支持下德育评价变革的逻辑向度及其反拨机制研究”(2023M730762);国家社科基金项目“高校德育变革的大数据赋能与实现路径研究”(22FKSB038)

导、创新学术诚信实践活动平台、完善高校学术评价管理制度、加强学生学术诚信自我教育^[5],还应该利用大数据这个新思维、新技术和新工具对已有的或可能发生的学术不端问题采取应对措施,从而提高治理的针对性、实效性和时代感。

一、大数据赋能研究生学术不端行为治理的价值意蕴

作为一种全新的思维方式和先进的技术手段,大数据是研究生学术不端行为治理的一柄利器。二者的耦合既有必要性,也具可行性,说其必要,一是因为传统的学术监督方式在大数据时代已显得捉襟见肘,难以有效应对新问题和新的挑战;二是因为大数据需要进一步拓宽其应用范围、释放其巨大价值。说其可行,主要是因为针对抄袭剽窃、伪造篡改数据、署名不当、一稿多投多发、虚假同行评议等诸多学术不端行为,大数据确实可以大显身手。拥有大数据是时代特征,应用大数据是时代机遇,大数据赋能下的研究生学术不端行为治理将呈现出科学化、精准化、智能化和前瞻化的特征,是保障和提升研究生教育质量的重要手段。

(一)学术成果不断数据化让抄袭造假行为难以藏身

不断数据化是大数据时代境遇中学术成果呈现、存储、传播和共享的大势所趋和必然走向,它克服了纸质化的诸多缺陷,具有全网互联、长久保留和易于提取的典型特征。正因为此,网络时代拙劣的复制粘贴已穷途末路,很容易被检测发现,过去许多难以发现的学术不端行为如今也可以轻而易举地被“网”打尽,其中十分普遍、危害最大的抄袭造假行为在数据化面前更是无所遁形,越来越容易被快速披露出来。舍恩伯格曾举例说到,很多欧洲政客(包括一名德国国防部部长)的抄袭行为因此被曝光,最终不得不引咎辞职^{[6][11]},这也可从近些年来国际上频频爆出的大量撤稿事件中窥见一斑。而更可怕的是,网络的全球性覆盖还让那些所谓的“高级”“隐蔽”的抄袭行为暴露无遗,比如:把译介的国外作品以自己名义发表、“巧妙”改写他人作品、对不同数据库的论文删减拼凑后再次发表、引用他人成果不注明出处、“注水”论文,等等。此前引起广泛关注和舆论热议的梁莹事件就凸显出了大数据在这方面的独特价值——学术污点没有“被遗忘权”,即便费尽九

牛二虎之力撤回论文,也并不意味着学术污点随之彻底擦除。恰恰相反,在教育数字化深入推进和学术成果日益数据化的背景下,这种类似“此地无银三百两”的做法更会欲盖弥彰,留下的数据足迹将长久存在下去。从这个角度看,大数据可以助力学术打假,有力整饬学术道德失范乱象,有效遏制学术不端行为的蔓延,从而营造一种恪守学术规范、弘扬学术道德、潜心学术研究、捍卫学术尊严的风清气正的学术生态。

(二)大数据检测软件防控各类学术不端行为的发生

目前,普遍公认的学术不端行为总体上可归为两类,一是包括伪造(Falsification)、篡改(Fabrication)、剽窃(Plagiarism)在内的科研造假行为(简称FFP行为),二是处于灰色领域的科研不当行为(Questionable Research Practices,简称QRP行为)^[7]。无论是哪一类学术不端行为的治理,大数据都可以大显身手,彰显其独特价值。首先,在治理FFP行为方面,利用大数据建立起来的反剽窃系统可以将学生的论文自动和已公开发表的文献数据库进行比对、查验和鉴定,根据抄袭行为的严重程度决定是否通过,从而遏制明目张胆的剽窃行为。另外,区块链的时间戳、分布式技术、不可篡改和匿名性等特征既可以保护研究者的知识产权,也让伪造和篡改数据变得异常困难^[8]。其次,在治理QRP行为方面,大数据也有广阔的应用前景。譬如,相关部门可以借助大数据强大的关联分析功能和自动挖掘技术研发学术作品风格鉴别系统,该系统可以对某位研究生的所有论文进行汇总整合,交叉检验,勾勒“画像”,一旦发现某篇文章的风格和之前截然不同,同时又是多人共同署名时,系统就会自动生成署名不当的报告单供监督主体参考,在此基础上判断是否存在“搭便车”的行为。再如,多源异构的大数据还可以快速准确地识别代人捉刀、虚假同行评审的行为,如稿件内容与作者研究方向之间的跨度过大、短时间内使用同一邮箱大量投稿、审稿人与投稿者隶属相同机构、审稿意见过于简单且评价过好等,这让传统的监督手段望尘莫及。总之,基于大数据建立起来的学术不端行为检测软件可以筑起学术道德的防火墙,形成一种“不能学术不端”的防范机制,有效降低各类学术不端行为的发生率。

(三) 互联的黑名单系统完善学术不端行为惩戒机制

研究生学术不端行为治理是一项长期性、复杂性、艰巨性和系统性工程,按照责任主体和性质的不同,主要有环境润育、主体自律和制度他律三种治理路径,而环境润育和主体自律在本质上属于一种软性力量,其约束力和震慑力十分有限,效果也不甚理想。所以,有效的学术不端行为治理要更加注重制度他律的作用,大数据恰好在这方面有其用武之地。不同于生物记忆,数字化记忆可以实现永久存储,而且很容易被提取和查询,极大提高了学术不端行为被披露的可能性。一旦研究生的学术不端行为被科研诚信系统和学术不端黑名单系统所记录,就如同绑上了一颗无法拆除的定时炸弹。这些难以磨灭的学术污点会让学术不端者永远受困于过去,被学术共同体嗤之以鼻,在今后的学术发展甚至人生成长中处处碰壁。可见,“永久记忆”犹如一把时刻高悬头顶的“达摩克利斯之剑”,会对学术不端者产生巨大的心理震慑作用,一想到失范行为的沉重代价远远高于预期收益,学术不端行为的严重后果和高额成本倒逼他们抵制住各种诱惑而不再明知故犯、铤而走险,无形之中构筑起了一道抵制学术不端行为的坚实壁垒和心理屏障。而且,教育部门还可以依托大数据平台建立全国性的为学界高度认可的学术监督专家库,拓宽学术不端行为的投诉、举报和审查渠道。更重要的是,大数据强大的动态监测功能可以持续采集信息,实时更新状态,及时的反馈机制实现了研究生学术不端行为的常态化监测。总之,借助大数据建立学术诚信的长效监督机制和终身诚信档案管理系统,有利于完善“不敢学术不端”的惩戒机制,倒逼研究生时刻警惕,不敢越雷池半步,从而反推研究生确证学术道德规范权威,重塑学术共同体的价值信仰,唤回其对学术道德规范的敬重与遵从^[9]。

(四) 聚类分析有助于开展针对性强的学术道德教育

为了更好地贯彻和落实国家出台的有关学术道德建设的文件精神,大多数高校都制定了相应的学术道德规范条例。但是,由于没有充分把握研究生学术不端行为的实际情况、具体表现和时代特征,这些条例在一定程度上存在着界定不清晰、缺乏针对性、操作性不强等问题,未能在学术不端行为治理中发挥出最大功效。借助大数据强大的聚类分析功能

和可视化呈现技术,可以从多个维度去分析和总结研究生群体学术不端行为的突出特征和集中表现,根据性质和严重程度采取不同的应对措施和治理手段,从而提高学术道德教育的针对性和实效性。比如,抄袭比最高的部分常见于论文什么地方、参考文献是否存在列而未引和装点门面的现象、哪类学术不端行为发生最频繁,等等,从而全面深入地摸清当前学术规范和诚信教育存在的突出问题和亟待改进的“短板”,为高校开展针对性更强的学术规范和学术道德教育提供数据支撑和可靠建议。只有基于实情,高校才能有的放矢、对症下药、精准施策,强化学术规范和教育举措的有效供给,“结合学术规范核心问题、具体要求等相关内容以及实例进行全面而深入的阐述与解读,包括从正面明确、详细地阐明学术活动的基本要求及其原因,对禁止的行为结合举例进行详细的描述,对如何避免剽窃作出实际的指导等”^[10],有效防止“无知型”学术不端行为的发生,从而推动研究生学术道德教育朝着体系化、连续性和规范化方向迈进。

二、大数据赋能研究生学术不端行为治理的现实梗阻

大数据的广泛应用和有效嵌入推动研究生教育管理决策迈入了数据驱动的新范式,为研究生学术不端行为治理提供了有效的技术手段。在其支持下,研究生学术不端行为的监督主体更加多元、监督范围更加广泛、监督渠道更加畅通、监督效果更加显著、监督过程更具预测性。但是,由于大数据技术及其领域应用还不成熟,致使其在研究生学术不端行为治理中的巨大价值和独特功能目前还没有得到充分释放和最大实现,依然面临多重困境和挑战。

(一) 学术监督主体缺乏使用大数据的行为意向

美国教育部教育技术办公室发布的报告《通过教育数据挖掘和学习分析促进教与学》指出,大数据教育应用中的技术挑战可以通过基础设施建设和技术研发解决,但 80% 以上的限制和挑战却来源于“人”^[11]。由此表明,人是大数据在教育中落地生根的决定性因素,而重中之重在于人是否具备使用大数据的行为意向——相关主体倾向于使用某一技术的程度。尽管大数据的出现和应用为研究生学术不端行为的治理开启了一场寻宝游戏,而学术监督主体的大数据使用行为意向正是主宰这场游戏的关

键。也就是说,如果学术监督主体使用大数据的意识没有得到唤醒和提升,那大数据在研究生学术监督中的应用就无从谈起。

美国学者戴维斯(Davis F. D.)提出的技术接受模型(Technology Acceptance Model, TAM)理论认为,技术的系统使用直接受行为意向的控制,而行为意向则受有用性感知和易用性感知这两个中间变量的影响,前者是指使用某一技术对工作的改善程度,后者是指使用某一技术所需要的努力程度^[12]。很显然,只有当学术监督主体认为大数据具有较大价值(有用性感知),并且能得心应手地使用(易用性感知)时,他们才会付出切实行动。具体而言,有用性感知和易用性感知又分别受到绩效感知、社群因素和自我效能感、促成条件这四个次级因素的影响。遗憾的是,当前这四个次级因素并不能对学术监督主体的使用意向产生促进作用,具体表现在:其一,由于大数据尚处于起步探索阶段,其功能没有得到充分释放,学术监督主体还不完全清楚其在改进学术监督工作方面的独特价值,因而缺乏使用大数据的动力;其二,社会心理学揭示,当一个人处于群体环境时,行为通常会和群体保持一致^[13],但由于大多数高校没有基于大数据开展学术不端行为治理,学术监督主体也就没有足够的动力和行为参照,难以产生“我应该使用大数据”的行为意向;其三,作为一种新兴技术,大数据的使用门槛较高,这使得很多学术监督主体对其产生畏难甚至抵触情绪,自我效能感低下,以至于背负过多的技术压力和技术恐惧而成了“技术落伍者”;其四,大数据应用是一个极其复杂的过程,学术监督主体在使用中难免会遇到各种难题,但又缺乏可以让使用行为变得容易的促成条件,如大数据应用平台、硬件设备、培训服务和技术支持等,其易用性感知明显下降,就会出现“想用而难用”的困境。

(二) 各类科研数据库的开放共享程度亟待提高

大数据蕴含大价值,而大价值得以释放的前提和基础在于各类数据的融合共通,彼此关联的数据价值要远远高于孤立的数据,“大数据价值的实现,在于数据与数据的连接”^[14]。故此,只有不同层级、不同类型和不同部门之间的科研数据自由流通、互通有无,海量而分散的数据才能集腋成裘、交叉检验、相互佐证,从而提高研究生学术不端行为诊断和治理的准确性,以免出现漏洞甚至错误。但遗憾的是,尽管研究生的各种科研信息和资料在信息化建

设的推动下基本实现了数据化,但它们都在各自系统内独自存储,这些系统不仅横向上壁垒林立、彼此隔离,而且纵向上缺乏统筹、难以整合,俨然一个个分散而独立的岛屿,存在信息碎片化、静态化、简单化等不利因素,形成了一种信息封闭而无法释放其累积效应的“数据孤岛”现象。

多源异构的科研大数据是对研究生学术不端行为进行有效监督的基础条件,过去之所以监督困难,主要原因就在于用于参照比对的数据样本十分有限。所以从这个角度看,“数据孤岛”罪莫大焉,除了管理上存在高投入、高能耗、低效率等问题,更重要的是阻隔系统间的数据融合渠道,和整体大系统数据割断联系,缩小了诊断和监督的范围及说服力,严重制约科研数据的价值释放。概括而言,科研领域出现“数据孤岛”的原因主要有二:一是由于缺乏顶层设计、集中领导和统筹谋划,许多科研管理部门未能真正理解科研信息化建设的内涵和要求,往往基于部门内部的业务流程和工作需要建设相对分立、单向封闭的科研信息管理系统,其后果是科研数据的采集、存储、分析和应用都停留在部门层面而无法实现跨部门跨层级的广泛共享和综合应用。当前,为了垄断市场和追求利益最大化,一些网络数据经销商鼓动期刊独家授权,造成知网、万方、维普等数据库互不兼容,限制了基准数据库的检测范围^[15],各个期刊的投稿系统亦各自为政,离大数据时代对科研数据的集约化要求相差甚远。二是因为不少科研数据的拥有者或管理者过于担心,超出一定范围的数据共享可能引发数据安全、隐私泄露和知识产权等问题,为了避免数据开放共享后引发的难以预料的麻烦,他们干脆因噎废食拒绝共享科研数据。总之,各类科研数据库的封闭阻塞和流通不畅,严重影响了学术不端行为监督的全面性和有效性。

(三) 科研成果价值开发与合理使用的矛盾加剧

大数据的勃然兴起和广泛应用加剧了科研成果在价值开发与合理使用之间的矛盾,而且这一矛盾很难调和,给学术不端行为的治理带来了严峻挑战,让其陷入两难境地。一方面,开放共享由知识产权的相对权利属性决定,是科研成果发挥价值、释放功能的基本前提。从这一点看,科研成果开放共享程度越高,汇聚而成的数据就越全面,用于检测学术不端行为的数据基础就越坚实,监督效果自然就更好。但是,如果一味追求开放共享的最大化,势必又会威胁到知识产权的安全性,尤其是在数字化、全球互联

和易于提取的大数据背景下,科研成果的传播、复制和获取变得更加容易、方便和高效,这也为科研数据的泄露和滥用提供了可乘之机。另一方面,科研成果具有知识产权的专有性和排他性特征,未经所有者许可不得擅自使用,但大数据日益强大的信息抓取能力和深度挖掘功能为侵权行为提供了技术手段,侵权变得愈发容易、频繁和隐蔽,知识产权保护难度加大^[16]。而且,在大数据和人工智能的背景下,侵权责任的界定变得模糊,追踪侵权者也比较困难。如此一来,研究生可能出于科研成果不被滥用的考虑,数据保护意识越来越强,过于强调科研成果及个人学术信息的保护,不愿意公开和共享,走向过分保护科研数据的极端,与知识产权的内在本质相抵牾,结果造成科研成果价值开发难以拓展和深化,削弱了治理学术不端行为的数据基础,可能导致“合理使用”的消失,这显然也背离了开放、自由和共享的网络精神。

(四)大数据固有缺陷妨碍学术不端行为的治理

大数据在防治研究生学术不端行为方面价值巨大、前景广阔,正因为此,我们可能过于相信它的威力而忽视其先天缺陷。对此,舍恩伯格曾发出过警告:“尽管大数据的力量是那么耀眼,但我们必须避免被它的光芒诱惑,并善于发现它固有的瑕疵”^{[6]247}。法国当代技术哲学家斯蒂格勒(Bernard Stiegler)提出的“技术药理学”效应也揭示了技术作为治疗和毒药所具有的“雅努斯”双面特征^[17]。大数据作为一种技术,具有代具、义肢的本质属性,是人的能力的一种延伸,可以用来防范和治理研究生学术不端行为,但利用不当反过来也会滋生研究生学术不端行为,对人类社会形成一种反噬。大数据是一把双刃剑,在研究生学术不端行为的治理上并非包治百病的万灵药,而是存在着不可避免的缺陷或不足,我们需要清醒认识到大数据的负面影响和潜在风险。

其实,不管应用于哪个领域,大数据自身都客观存在着不可调和的矛盾。体量庞大、类型丰富、传播迅速是大数据的基本特征,这些特征一方面让信息获取更为多元和便捷,这不仅空前提高了科研资料的搜集效率,也扩大了研究生学术不端行为的监督范围,甚至可以实现“样本=全体”的全域监测,以防出现监督死角和盲区,从而提高学术监督的全面性和准确性。但另一方面也为学术不端行为的滋生和繁衍打开了方便之门,极大地增加了学术不端行为

的监督难度。随着网络搜索引擎、文献数据库、翻译软件、反监测系统等信息技术的层出不穷和日益盛行,研究生在学术研究中进行复制、粘贴、剽窃、篡改、造假、伪造等学术不端行为变得更加便捷、容易、隐蔽和高级,涌出了一些过去难以发现或未曾出现的不端行为,呈现出日益智能化和隐形化的趋势^[18]。比如,大数据的不断投喂和深度学习让ChatGPT表现出了惊人的创造性,能够生成接近甚至超越人脑水平的学术论文,这不仅增加了学术造假的鉴别难度,还会导致学生批判性思维和创新能力的匮乏,毕竟“真正的创造行为是有意识地去创造规则,而不是来自偶然或随机的联想或组合”^[19]。此外,一些不法分子还专门利用大数据技术开发反监测软件和反抄袭系统,帮助研究生“巧妙”地规避,真可谓是“道高一尺,魔高一丈”。所以,从某种程度上说,大数据时代研究生学术不端行为频频出现且花样不断翻新,也有大数据及其特征的“功劳”。

三、大数据赋能研究生学术不端行为治理的推进策略

作为防范和治理研究生学术不端行为的一柄利器,大数据极大提高了学术监督的科学性、针对性、前瞻性和实效性,但现实中却面临科研数据共享困难、学术监督主体数据素养欠缺、学术不端检测系统滞后、科研信息容易泄露等诸多困境,严重掣肘大数据的作用发挥和价值释放。为此,需要从多个角度和不同层面探索大数据赋能研究生学术不端行为治理的进路或策略。

(一)打通科研数据开放共享的融合渠道

数据的全面、自然、动态、连续采集是大数据治理的基础性和先导性工作^[20],以多源异构的海量数据为基础的关联分析和预测是大数据的核心价值和独特优势,也是实现研究生学术不端行为治理科学化、智能化和前瞻化的关键所在。然而,大数据这一功能的充分释放却依赖于两个因素:一是全方位科研信息的数据化,二是各级各类科研数据库和信息采集平台之间的无障碍沟通和共享。也就是说,只有将零散孤立、残缺不全的科研数据汇聚而成一个庞大的数据池,大数据支持下的研究生学术不端行为治理才有坚实可靠的数据基础。否则,数据的局限和数据库的残缺会直接影响检测结果的全面性和准确性,无法满足大数据环境下对研究生学术监督

提出的协同性、关联性和快速性要求。

为夯实大数据驱动下研究生学术不端行为治理的数据基础,可以从三个方面推进科研数据的开放共享。其一,相关管理部门应加快科研信息的数据化建设。具体措施包括:制定科研信息数据化标准规范体系,对科研数据的格式、规格、口径或类型作统一规定,避免不同数据库因数据编码和格式不一致而出现“数据竖井”现象;全方位采集维度更多、频度更密、粒度更细的科研信息,尤其是要注重对研究过程的信息收集,全面覆盖研究生的教育背景、研究方向、科研经历、学术成果、学术道德教育情况和诚信记录等信息。其二,研究生管理部门要统一规划和部署,破除条块分割的管理体制,“解决研究生教育数据的采集和存储碎片化问题,打通研究生教育的‘数据孤岛’”^[21],建立一个互通共享、统一高效的以大数据技术为支撑的全国性科研信息管理平台,推动科研管理部门、科研机构、各大高校、学术组织、学术期刊、学术成果数据库的信息开放和共享,将不同层次、不同类型、不同专业的研究生科研数据有效融合,从而扩大用于检测的基准数据库的覆盖范围,便于学术监督主体对这些数据实现最大化利用。其三,推进高校、政府和大数据企业之间的数据开放和共享业务,一是在确保隐私安全的前提下,将研究生科研数据与政府部门公开的政务数据进行整合分析,为学术不端行为的识别提供参考;二是借助服务外包的形式,利用大数据企业的技术优势,研发使用性能好、安全系数高的统一科研数据管理服务平台。

(二)研发基于大数据的学术不端智能监督系统

在技术层面,目前我国的学术不端监督体制未能紧扣时代发展脉搏和科技革新潮流,特别是没有很好地将大数据、全媒体、区块链、ChatGPT生成式人工智能这些新兴技术有效利用起来,因而存在检索不全面、智能化不高的问题,以至于很多可以通过技术手段就能解决的学术不端行为依然大行其道。比如,国内常用的CNKI学术不端文献检测系统(AMLC)、万方论文相似度检测系统、维普通达论文检测系统、ROST反剽窃系统和CrossCheck等基本上都是对同一种语言的检测,而对跨语种和小语种的检测力有不逮。又如,由于大多数软件只能对结构化数据进行检测,而无法识别半结构化或非结构化数据,因而对PDF、图片、表格、公式、多样态数据集集成等形式的抄袭造假行为无能为力,这些功能缺陷在一定程度上为研究生学术不端行为提供了可乘

之机,也滋长了诸多隐性学术不端行为。

为了避免因技术原因而出现学术不端监督的漏洞,可以由教育部学位管理与研究生教育司牵头,加快研发和创建一个基于大数据、跨平台、多学科、综合性的全国学术不端智能监督系统。首先,与知名大数据公司合作,针对研究生学术不端行为的时代特征和演进趋势,对已有的学术不端检测系统进行更新换代或业务升级,填补系统漏洞,提高学术监督的有效性。其次,借鉴吸收国际Turnitin、GDcha、iThenticate和afeassign等数据库的算法、功能和特点,研发基于大数据的新型学术不端检测软件,进行全方位、跨语种、多层次的检测,为学术不端监督提供坚实可靠的技术支撑,实现“以其人之道还治其人之身”的目的。针对一稿多投,可以检测同一邮箱是否多次投寄相同稿件;针对代写代发的捉刀问题,既可以检测同一邮箱是否多次投寄署名作者不同的稿件,也可以查验作者电话号码与作者单位属地是否一致^[15];针对直接翻译外文据为己有的问题,可以通过跨语言机器翻译实现国内外学术资源数据库的交叉检验……再次,出台规定要求所有研究生的学术成果都要经过全国性学术不端检测系统的查重、比对和鉴定,威慑那些打算学术不端的研究生,从而降低学术不端行为的发生率^[22]。最后,利用大数据技术建立学术不端行为投诉举报平台,相关部门及时查阅、整理与核实举报信息,将存在学术不端行为的研究生纳入科研征信黑名单系统,并通知培养单位严肃处理。

(三)提高学术不端监督主体的数据素养

与传统的学术监督方式截然不同,大数据支持下的研究生学术不端行为治理更具循证性、技术性、专业性和挑战性,非常强调科研数据在学术监督中的价值和功用,这也就对学术监督主体的专业能力提出了新的更高要求,除了要拥有学术道德、业务能力、勤勉务实等传统素质外,还必须具备较高的数据素养,能够基于海量的科研数据做出正确合理的判断和决策(data-based decision)。

数据素养是人们有效且正当地发现、评估和使用信息和数据的一种意识和能力^[23],该定义指出了数据素养的四大基本构件:数据意识、数据技能、数据反思和数据伦理。其中,数据意识要求学术监督主体在思想深处高度认可科研数据和现代信息技术之于研究生学术不端行为治理的意义和价值,并能意识到科研数据缺乏对学术监督可能带来的负面影

响;数据技能是指学术监督主体具有较强的科研数据采集、存储、共享、分析和决策等能力,能够有效利用科研数据对学术不端行为进行识别和鉴定,借助大数据学术搜索引擎开展覆盖全网的深度挖掘和智能检索,最大程度地揭露学术不端;数据反思要求学术监督主体对技术拜物教保持必要的谨慎,站在科技哲学的高度审视大数据的使用限度,不能过于依赖大数据或智能程序对人类智慧的替代^[24],须知机器不具备人脑对文本背后的深邃思想和文化内涵的欣赏力和理解力,避免滑入“技治主义”和“数据独裁”的陷阱;数据伦理是指学术监督主体具备对科研数据的整个应用过程主动进行伦理审视、防止科研数据泄露或滥用的能力。上述素养的提升需要组织和个人的携手共进,共同努力,学术监督机构和研究生管理部门可以通过跨学科人才引进、委托高校培养、岗前岗后一体化培训、与大数据企业合作等方式,组建一支既熟悉学术监督工作特点、需求和规律,同时又能熟练掌握大数据技术的监督队伍。个人层面的学术监督主体则可以针对数据素养结构的实际情况查漏补缺,同时基于实践中的具体业务需求寻找同事间的协作互助。简言之,只有学术不端监督主体尽快成为“数据脱盲者”(data literate),才能营造一种“数据驱动治理”的新型文化氛围,大数据支持下的研究生学术不端治理也才能由理念变为现实,进而增强学术监督的实效性。

(四)全方位撑起科研数据的安全保护伞

确保科研数据的安全是科研管理部门和学术监督主体义不容辞的责任,是保护知识产权推动高质量发展的内在要求,这不仅影响大数据在研究生学术不端行为治理中的持续应用和长远发展,也直接关系到广大科研人员的切身利益。为了防止科研数据泄露造成不可挽回的损失,相关主体应建立健全科研数据安全保障体系,全方位撑起科研数据的安全保护伞。

首先,在认识上树立一种“数据应用、安全第一”的意识,不同于其他领域的数据,科研数据的敏感性更强、滥用风险更大、危害后果更严重,相关的学术监督主体要充分认可和深刻理解科研数据安全之于科研工作者的的重要性——只有消除科研数据泄露的后顾之忧,他们才愿意共享和开放数据,也才能实现数据价值开发与隐私保护之间的良性循环和合理张力。其次,在管理上建立和完善科研数据的监管机制,具体措施包括:根据科研数据的性质、类型和实

际应用需求合理规定其存储及使用周期;按照重要性和敏感度对科研数据进行分级分类管理,明确不同用户的访问权限、操作权限与管理权限;对机密数据库实行多部门或多人共同管理;成立专门的数据审查委员会对科研数据的采集、共享、挖掘和分析等各个环节进行全程监督,力求做到数据应用的合理合法、公开透明、安全高效。再次,从传统技术和新兴技术两个维度双管齐下,为科研数据保驾护航。一是进一步升级信息加密、入侵检测防火墙、病毒和漏洞扫描、身份认证、访问门槛、匿名化和模糊化处理等传统的数据保护技术,二是与时俱进,积极引进和研发一些新兴的数据保护技术,比如区块链技术、自动销毁技术、数据脱敏技术和分布式访问等。最后,虽然大多数国家及其高校几乎都出台了保护知识产权的政策或法规,但它们尚不足以应对新出现的因数据使用而导致的伦理问题^[25],为此,需要尽快将《科研数据保护法》的制定提上立法日程,对科研数据的产权归属、共享范围、使用方式和安保举措等核心议题做出明确规定。与此同时,对倒卖、泄露、盗用科研数据来牟取不当利益的个人或组织予以严厉惩处,情节严重者,移交司法机关处理。需要注意的是,在新型技术层出不穷、网络环境日益复杂和数据泄露风险倍增的大数据时代,传统的、孤立的保护措施已顾此失彼,只有采取组合式的联防策略才能收到事半功倍的效果。

参考文献:

- [1] 初景利,解贺嘉,张冬荣,等. 研究生对学术不端相关问题认知的调查与分析[J]. 研究生教育研究, 2022(4): 60-65.
- [2] 洪大用. 加快推进研究生教育高质量发展 着力造就拔尖创新人才[N]. 中国教育报, 2022-11-25(001).
- [3] 郭跃, 濮燕屏. 论研究生学术不端行为治理规范体系的构建[J]. 学位与研究生教育, 2017(1): 23-28.
- [4] 令小雄, 王鼎民, 袁健. ChatGPT 爆火后关于科技伦理及学术伦理的冷思考[J]. 新疆师范大学学报(哲学社会科学版), 2023(4): 123-136.
- [5] 柳礼泉, 陈方芳. 构建研究生学术诚信教育长效机制的“五化”[J]. 研究生教育研究, 2016(2): 24-28.
- [6] 维克托·迈尔-舍恩伯格, 肯尼斯·库克耶. 大数据时代[M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013: 111, 247.
- [7] Nicholas H S. Fostering Integrity in Research: Definitions, Current Knowledge, and Future Directions[J]. Science and Engineering Ethics, 2006(1): 54.

- [8] 曾玲,张辉洁,冉明会,等.人工智能时代科技期刊应对学术不端问题的研究进展[J].中国科技期刊研究,2020(3):270-275.
- [9] 肖健,黄继滔.研究生学术不端惩罚的价值定位与实践向度[J].学位与研究生教育,2021(3):27-31.
- [10] 朱纓.学术规范教育给力吗?——关于研究生学术规范教育现状的调查分析[J].学位与研究生教育,2014(1):23-27.
- [11] Bienkowski M, Feng M, Means B. Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief[J]. Office of Educational Technology, US Department of Education, 2012.
- [12] Davis F D. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results[D]. Massachusetts Institute of Technology, 1985.
- [13] 宋超英.组织行为学[M].兰州:甘肃人民出版社,2002:152-153.
- [14] 车品觉.决战大数据:驾驭未来商业的利器[M].杭州:浙江人民出版社,2014:175.
- [15] 王文福.网络时代期刊论文隐形学术不端挖掘策略[J].中国科技期刊研究,2016(7):677-682.
- [16] 李伦,孙保学,李波.大数据信息价值开发的伦理约束:机制框架与中国聚焦[J].湖南师范大学社会科学学报,2018(1):1-8.
- [17] 喻岚.数字时代的“技术药理学”和“一般器官学”——斯蒂格勒解读技术的雅努斯双面[J].安徽师范大学学报(人文社会科学版),2020(3):68-73.
- [18] 张重毅,方梅.科技论文隐性学术不端行为判别特征分析[J].中国科技期刊研究,2019(1):24-28.
- [19] 赵汀阳.人工智能的神话或悲歌[M].北京:商务印书馆,2022:16.
- [20] 吴刚,陈桂香.高校大数据治理运行机制:功能、问题及完善对策[J].大学教育科学,2018(6):34-38,66.
- [21] 王战军,蔺跟荣.动态监测:大数据驱动的研究生教育管理新范式[J].研究生教育研究,2022(2):1-8.
- [22] 龙献忠,陈方芳,刘绍云.论构建研究生学术道德教育的“三不”机制——基于“三不”反腐机制思想的启示[J].研究生教育研究,2018(1):21-25.
- [23] Mandinach E B, Gummer E S. A Systemic View of Implementing Data Literacy in Educator Preparation[J]. Educational Researcher, 2013(1):30-37.
- [24] 刘平,杨志辉.人工智能构建科技期刊智慧出版模式[J].中国科技期刊研究,2019(5):462-468.
- [25] Slade S, Prinsloo P. Learning Analytics[J]. American Behavioral Scientist, 2013(57):1510-1529.

Postgraduate Academic Misconduct Governance Empowered by Big Data ——Value Implication, Realistic Obstruction and Promotion Strategies

ZOU Tailong^{1,2}

(1. School of Education, Guangzhou University, Guangzhou 510006, China;

2. School of Teacher Education, Hubei Minzu University, Enshi 445000, Hubei, China)

Abstract: As a new way of thinking and an advanced technical means, the value implication of postgraduate's academic misconduct governance empowered by big data is mainly reflected in: the constant digitization of academic achievements makes it difficult for plagiarism and forgery to cheat; the big-data detection software prevents the occurrence of various academic misconducts; the interconnected blacklist system improves the punishment mechanism for academic dishonest behavior; and the cluster analysis benefits targeted academic moral education. However, the governance of postgraduate's academic misconduct empowered by big data is still confronted with the following realistic obstacles: the academic supervising entity lacks behavioral intention of using big data, the degree of openness and sharing of various research databases needs to be improved urgently, the contradiction between the value excavation and the rational use of scientific achievements is intensified, and The inherent defects in big data hinder the academic misconduct governance. Therefore, it is necessary for us to open up an integrated scientific research data channel that is accessible and to be shared, develop a data-based intellectual academic misconduct supervision system, improve the data literacy of academic supervising entities, and build up across the board a research data security protection system.

Keywords: big data; postgraduate's academic misconduct; ChatGPT; governance of academic misconduct