



# 人工智能伦理治理亟需迈向实践阶段

曹建峰

当今社会,数字产业化与产业数字化持续推动数字世界与物理世界深度融合,人类的生活和生产方式日益被重塑。人工智能算法与数据作为这场经济与社会领域变革的核心推动力,将是未来数年的话题制造者。毫无疑问,数据与人工智能算法的结合,嵌入恰当的应用场景,将带来显著的经济与社会效益。

但也必须重视人工智能算法在公平、安全、隐私、透明、责任等方面可能带来的问题。算法系统的输出结果可能对特定个体或群体造成歧视、不公平、排斥等,例如人脸识别算法被指在识别有色人种上准确性较差,广告算法、招聘算法被指排斥女性劳动者。算法分发可能助长虚假信息的传播、扩散,也能限制用户对信息的自由选择,进而造成“信息茧房”效应。同时,算法也可能被滥用或恶用,给个人权益、公共利益、国家安全等带来威胁,如过度采集与分析人脸信息、大数据杀熟、深度伪造等。

在这些背景下,人工智能的发展应用迫切需要伦理价值来提供引导与约束,这在国内外已经成为人工智能领域的一项基本共识。

## 三个阶段

2016年以来,伴随着人工智能技术的快速发展应用,人工智能伦理大致经历了三个发展阶段,当前迫切需要迈向“实践”阶段。

第一阶段从2016年开始,可以称之为原则大爆炸阶段。这个阶段的核心是各界纷纷提出或制定伦理原则。人工智能需要遵循一定的伦理原则为出发点,上到国家、政府机构与国际组织,下到科技公司、行业组织与学术团体,纷纷提出或制定AI伦理原则,国内外

的主流科技公司如微软、谷歌、IBM、腾讯等也都积极响应,陆续提出了各自的AI伦理原则。据不完全统计,相关的AI原则文件超过百份。总之,在这个阶段,各界都在积极倡导AI伦理原则,但缺乏必要的共识与具体的实践。

第二个阶段从2018年开始,可以称之为共识寻求阶段。这个阶段从2018年开始,可以称之为共识寻求阶段。这个阶段从2018年开始,可以称之为共识寻求阶段。这个阶段从2018年开始,可以称之为共识寻求阶段。这个阶段从2018年开始,可以称之为共识寻求阶段。

第三个阶段从2019年开始,可以称之为AI伦理治理实践阶段。在这个阶段,产业界开始思考如何实施、执行AI原则,探索可以把AI原则转化为实践的机制、做法、工具等。目前,谷歌、微软、IBM等科技公司已在积极推进AI伦理方面的落地工作,让AI原则操作化、落地化,真正融入、嵌入AI研发流程与业务应用。总之,AI伦理实践应成为当前以及未来AI伦理工作的核心方向,因为要真正把AI伦理落到实处,仅仅倡导AI伦理原则是不够的,下一阶段需要着力探索把抽象的AI原则“翻译”为或“转变”为具体的实践的治理路径。

## 五个路径

从原则到实践,是AI伦理领域的发展方向。目前,我国已提出了AI伦理相关的原则、框架等,如《新一代人工智能治理原则——发展负责任的人工智能》。一些科技公司也提出了类似的倡议,在此基础上,需要进一步探索AI伦理的落地实施方案,更多依靠伦理治理的相关实践来促进负责任、安全可信AI的发展应用。结合国内外的相关探索与研究,AI伦理治理主要有以下五个实践路径。

一是伦理委员会。如果说人工智能是未来智能社会的基石,那么可以说伦理就是使这一基石稳固的必要保障。所以,考虑到不同AI系统的不同影响,以及立法滞后、法律不健全等因素,科技公司需要在法律合规这一最低要求之外,积极履行伦理责任。伦理委员会是科技公司履行AI伦理责任的最基础机制。成立伦理委员会,对AI相关业务与应用进行必要的伦理审查,已经成为了科技行业的“必选项”,例如微软的AETHER委员会、谷歌的AI原则审查小组、IBM的AI伦理委员会等。伦理委员会的主要职责是制定AI伦理相关的内部标准与流程,并基于此对AI相关业务进行伦理审查,以识别、预防、消除AI相关应用的安全、公平、隐私等方面风险。在具体运作上,伦理委员会需要多元参与,即技术、法律、伦理等不同专业领域人士的协作配合;伦理委员会负责建立案例、标准、程序、工具、资源等,成为制度知识的资料库,发挥治理主体的作用。此外,伦理委员会比政府立法反应快,能够及时跟进技术创新与应用的快速发展。

二是伦理实践框架。除了对AI业务进行伦理审查,国外科技公司也在落实AI伦理相关的实践框架,以解决人工智能带来的歧视、不透明、不可解释、隐私等问题。例如,在算法透明性方面,谷歌推出了面向人工智能的“模型卡片”机制,IBM则推出了“AI事实清单”机制,这些机制类似于产品的说明书与食品的营养成分表,对人工智能模型相关的模型细节、用途、影响因素、指标、训练数据、评估数据、伦理考虑、警告与建议等方面进行解释说明,以便人们能够更好地理解、认知AI模型。再如,在隐私保护方面,联邦学习框架(federated learning)既能促进数据利用,又能很好地保护个人隐私。简言之,联邦学习是指在机器学习的过程中,各参与方可借助其他方数据进行联合建模,但却无需

共享其数据资源。借助联邦学习,可以解决数据不完整、数据不充分等问题,同时保护个人隐私及数据安全。联邦学习在车险定价、信贷风控、销售预测、视觉安防、辅助诊断、隐私保护广告、自动驾驶等方面具有很大应用前景。此外,AI伦理检查清单、AI公平性检查清单等机制也日益得到科技公司的重视,在确保AI符合伦理要求方面将发挥越来越重要的作用。

三是伦理工具。伦理工具侧重从技术上寻找针对透明性、可解释、公平性、安全性、隐私保护等问题的技术解决方案。此外,还包括针对人脸伪造在内的深度伪造的鉴别工具。从伦理工具的开源发展到商业化服务,大型科技公司和AI伦理创业公司正在弥补AI领域缺失的一环,为AI伦理落地提供了全新的思路。目前,谷歌、微软、IBM以及一些AI伦理创业公司等都在积极开发多元化的AI伦理工具并集成到云服务中,提供AI伦理服务(ethic as a service,简称EaaS),赋能客户与产业。EaaS将来有望成为云服务与云AI的标配。

四是标准认证。就像现在的隐私保护一样,人工智能伦理也可以采取标准认证的方式来推进,符合AI伦理标准的AI产品与服务可以申请相应的认证。目前,IEEE、ISO等国际标准化组织在积极组织制定相关的标准,并尝试推出认证项目。未来我国在这方面也需要积极投入,抢占标准高地,以标准认证的方式鼓励、促进可信的、负责任的AI的发展应用。

五是伦理培训。技术研发人员处在AI业务一线,是对技术负责的第一人,需要培养他们的伦理意识,帮助他们成为AI业务实际中积极践行伦理要求,把伦理要求嵌入产品开发设计与运作的流程。所以政府与企业也要对其技术人员加强伦理培训,高校则要加强对AI伦理相关教育培训体系的搭建。



## 三管齐下

对于AI治理,伦理治理、法律治理及技术治理都有各自的作用空间,不可偏废。但就当前而言,人工智能治理需要更多依靠伦理治理的方式。因为考虑到AI算法的复杂性与持续迭代性,在不宜草率推出强制性立法时,伦理治理对于应对AI算法应用带来的问题无疑是最合适也是最有效的方式。

AI伦理固然重要,是确保实现负责任、安全可信AI的重要路径,但也不能片面强调伦理及立法规制,以免影响AI技术的创新应用及其经济社会效益的最大化。

例如,过度强调个人信息与隐私保护,可能导致AI应用获取不到足够的可用数据,进而阻碍AI应用及其对个人与社会的价值;过度强调算法的透明性,可能降低算法的准确性或效率,进而阻碍规模化应用;片面强调在所有场景下都要求人类最终决策与控制,可能无法充分发挥AI解决既有决策环节中人类决策者相关的偏见、歧视等问题。而且很多高风险的AI应用,往往也是高价值的AI应用,所以要在权衡不同利益追求的基础上,做到伦理与发展的平衡,实现负责任的技术创新与应用。

(作者系华东政法大学数字法治研究院特聘研究员)

# 多家互联网平台助力打击整治非法社会组织



本报记者 张维

铲除非法社会组织的行动,在互联网平台企业中已获得积极反响。

记者近日在采访中了解到,包括水滴、中航信、腾讯等互联网平台在内的企业充分发挥行业作用,结合自身优势、专业特长,积极配合做好搜集非法社会组织活动证据线索,遏制非法社会组织虚假信息网络散布传播,开展网上打击整治非法社会组织等工作。

3月20日,民政部、中央宣传部、中央政法委等18个部门深入贯彻党中央决策部署,联合开展为期3个半月的进一步打击整治非法社会组织专项行动正式启动。同时,民政部等22部门联合印发了《关于铲除非法社会组织滋生土壤 净化社会组织生态空间的通知》(以下简称《通知》)。这是多部门首次专门就打击整治非法社会组织工作联合出台的重磅文件,联合发文参与部门之多,涉及面之广,对策之精准,都是前所未有的。

“这是我国社会组织发展史上的一个重要里程碑。”中国人民大学法学院教授刘俊海评价道,“不仅标志着无缝对接的协同监管执法合作机制的建立健全,而且清晰地表明了国家和全社会对非法社会组织的零容忍态度。”

据了解,《通知》印发后,在广大社会组织中引起强烈反响。尤其是全国性社会组织积极行动,引领示范,组成了专项行动中社会组织力量的“第一方阵”,体现了应有的使命责任和担当。在本领域、本行业具有代表性、示范性、典型性的全国性行业协会商会,正在结合自身组织和所在行业情况,配合开展打击整治工作。

近年来,随着信息技术的进步,科技手段的发达,越来越多的非法社会组织为逃避监管和打击,不断变换活动方式,从传统线下为主,开始逐渐向线上转移。互联网平台是社会治理的重要主体。“互联网不是法外之地,不能为非法社会组织活动提供温床。”民政部有关负责人在3月29日召开的互联网平台协同打击整治非法社会组织工作座谈会上说。

值得注意的是,《通知》下发后,水滴、中航信、腾讯互联网企业积极表态并迅速行动起来。例如,水滴公司创始人兼CEO沈鹏认为:“社会组织在我国经济社会发展等方面发挥着越来越重要的作用,

已成为我国社会建设事业的重要力量。然而,一些非法社会组织披着社会组织的外衣,在贪婪敛财的同时,极大地破坏了社会组织的名誉和健康发展的环境,也给一些人民群众造成了不少经济损失。”

沈鹏表示,水滴筹、水滴保作为服务数亿用户的互联网科技公司,愿尽平台之力,不向非法社会组织提供广告、宣传、活动等服务,发挥信息搜集、数据检索等方面的技术优势,助力打击整治非法社会组织,防范化解社会风险隐患,捍卫人民群众正当权益。

“打击整治非法社会组织,需要共同治理,合力围歼。”北京市京师律师事务所主任、中国慈善联合会副监事长张凌霄建议,进一步加强法治宣传,依靠社会力量,营造共同防范非法社会组织存在的舆论氛围。同时,广泛发动群众投诉举报,实施精准打击,压缩非法社会组织的生存空间。进一步充实执法力量,加大打击力度,依法严惩,把集中整治常态化,加强网络巡检与排查,线上线下同步查处,多部门联合形成共同治理的围歼合力,不给非法社会组织留下活动空间。

清华大学公共管理学院副院长、教授邓国胜提出,下一步要完善相关法律法规,强化取缔非法社会组织的制度建设;加强民政部门执法队伍建设,提升执法能力;紧紧依靠公众和媒体,发挥社会监督作用。

## 2020年移动互联网报告发布

# 疫情带火“云面试”“云求职”

本报讯 记者张维

近日,极光大数据发布《用户行为变迁 行业垂直深耕——疫情下的2020年移动互联网报告》(以下简称《报告》)。《报告》显示,受疫情影响,2020年,求职招聘行为从线下转向线上,推动了线上招聘市场发展。疫情期间,招聘求职平台上线“云招聘”“云面试”等功能模块,为就业建立便捷通道,推动行业迅速回暖。

根据《报告》,2020年2月后,主流招聘App日活跃用户规模稳健回升。随着国内产业不断转型升级,新业态的发展催生了更多就业与用工需求。市场需求的变化也正倒逼招聘平台不断升级人工智能、大数据等技术能力,提升人才岗位匹配效率与服务品质。

据人社部统计,2020年一季度的新增就业人数明显低于2019年,主要考虑为受疫情因素影响,多数企业均处于不稳定状态,员工更换工作的意愿也有所下降;同时,由于无法返校,高校毕业生求职也出现后延。2020年三四季度,随着高校毕业生就业情况逐步恢复,以及大量海外留学生回国就业,城镇新增就业人数逐步恢复。

《报告》指出,在此背景下,2020年许多线下招聘会取消,大量求职模式从线下转为线上,推动了线上招

聘市场发展。从2020年3月开始,各家在线招聘平台进入快速增长期。

《报告》显示,对于疫情下火起来的“云面试”“云求职”,用户整体的接受度比较高,有54.8%的用户认为线上面试更高效,23.7%的用户认为跟线下面试差不多,21.5%的用户认为线上面试效率低,不如线下直接面试。

此外,互联网求职在一线及新一线城市普及度更高。新一线城市的求职者占比最高,占比达40%,一线城市和二线城市的求职用户占比均为27.9%。一线城市和二线城市求职用户占比均为27.9%。从性别分布来看,互联网求职人群中男性占比更高,达58.9%;从年龄分布来看,互联网求职人群主要集中在25至34岁,其中25至29岁的用户占比34.1%,略高于30至34岁的用户。

报告指出,近年来国内经济强调增强的产业链供应链自主可控能力,就业平台连接人才与市场,需不断升级人工智能、大数据等技术能力,提升人才岗位匹配效率与服务品质,不断提升人工智能、大数据等技术能力,提升人才岗位匹配效率与服务品质,顺应市场变化。

法制网见习记者 邢国涵

近日,拼多多启动“多多读书月”活动,联合中信出版社、知识出版社、青岛出版社、磨铁图书、果麦文化等30余家出版社及图书出版公司等,针对社科、文艺、教辅等多个类目超过1000款图书进行百亿补贴大型直补。

据了解,“多多读书月”旨在推动全民阅读,消费者可在拼多多App搜索“多多读书月”直达,如《钢铁是怎样炼成的》《白

# 拼多多联合30余家出版社启动“多多读书月”

## 源头直补1000余款平价正版好书

《浮生六记》《巴黎圣母院》《瓦尔登湖》等均成为读书月重点补贴图书。“多多读书月”期间,拼多多还将推出“百万图书满减金”,源头直补1000余款平价好书。同时,著名历史学家、央视“百家讲坛”主讲人阎崇年等知名作家届时将做客百亿补贴直播间为读者荐书。

“平台更广大的用户有着旺盛的知识消费需求,尤其是三四线城市、农村地区图书购买增长迅猛,增长率已经远超一线城市。”“多多读书月”负责人介绍说,拼多多将持续加强优质图书的引进力度,截至目前已引进超过2700余家出版社、图书公司、渠道商入驻。

据了解,拼多多已与国内多家大型出版社、图书公司、渠道商等达成密切合作,双方通过定制、专供等多种形式,推出更多优质、平价的正版图书。如7.8元/本的正版《民法典》,1个月时间便拼出30万册;7.8元的正版彩图版《山海经》,上线20余天拼出近6万册。

同时,百亿补贴还推出名家荐书活动,邀约知名作家做客直播间为拼多多读者荐书,如郝景芳、沈石溪、杨红樱都通过直播形式为读者开书单。

“到目前为止,知识出版社拼多多旗舰店已经直播667场,累计观看人数超过1328万人。”知识出版社负责人介绍说,直播已成为知识出版社的增长引擎,推动店铺销售收入超过7000万元码洋,并迅速成长为启蒙认知类店铺榜单和少儿英语榜单的第一名。知

识出版社是一家有着四十多年历史的老牌出版社,隶属中国大百科全书出版社。

果麦文化拼多多电商负责人表示,果麦在拼多多平台保持了高爆发速度,今年第一季度已经完成去年全年三分之二的销售额,“拼多多的新闻读者群体,让我们更加了解图书市场的新需求,对于未来增长抱有极大的期待。”

推进全民阅读的同时,拼多多还不断加大盗版图书的打击力度。拼多多平台治理负责人表示,当下率先进入人工智能算法,基于海量数据进行挖掘和分析,从底层重构违规商家、商品识别、假货识别、劣品识别等模型库,对图书商品进行24小时线上巡检。

当前,拼多多已与京版十五社反盗版联盟、少儿出版反盗版联盟、中国出版协会、新华文轩等机构签署知识产权保护等协议,联合推进打击侵权盗版图书。2020年10月,拼多多还与英中贸易协会(CBBC)签署合作备忘录,加强双方知识产权合作,该协会有超过400家会员,其中包括牛津大学出版社、大英图书馆、多家英国顶尖大学等。

“我们希望通过持续一整月的‘多多读书月’活动,进一步推动与出版社、知名作者的深度合作,将更多优质、平价的正版好书带给读者。”据“多多读书月”负责人透露,下一步,还将推动国内知名作者进驻拼多多,在全民阅读月为读者推荐更多的好书。

# 漯河中院二审判决:职业打假人非普通消费者

## 以牟利为目的属变相经营不支持10倍索赔

法制网见习记者 邢国涵

一买家花3560元网购200盒创伤药,然后以产品质量问题诉至法院,向商家索赔35600元。

近日,河南省漯河市中级人民法院针对这起“职业索赔”案作出判决:以索赔为目的进行的购买商品活动,购买商品是索赔中的一个环节,其行为整体具有营利性,属于变相的经营行为,不支持10倍索赔。

2019年10月22日,原告张某某花费3560元网购200盒创伤药,随后以该药生产厂家已注销,产品系假冒伪劣商品为由,起诉网店经营者漯河某百货店,要求该百货店退还货款3560元,支付10倍赔偿金35600元。

张某某主张其是日常生活需要购买,而漯河市源汇区人民法院一审查明,张某某在多次诉讼中以商品质量问题为由主张10倍索赔,并非普通消费者。

源汇区法院认为,根据消费者权益保护法第二条的规定:消费者是为生活消费需要购买、使用商品或者接受服务,与消费者为了生活需要而使用商品的目的不同,以索赔为目的进行的购买商品活动,购买商品是索赔中的一个环节,其行为整体具有营利性,属于变相的经营行为。

结合张某某另有多起购买商品后索赔的情形,源汇区法院不采信张某某普通消费者以生活消费为目的购买涉案商品的主张,仅判决百货店退还张某某货款3560元,并对其要求支付10倍赔偿金的诉讼请求不予支持。

张某某不服,上诉至漯河中院。漯河中院支持一审法院观点,

亦认为以索赔为目的进行的购买商品活动属于变相的经营行为,判决“驳回上诉,维持原判”。

近年来,全国以“打假”“维权”为名发起的“职业索赔”恶意投诉举报每年超100万件,“职业索赔”逐渐呈现出团伙化、专业化、规模化、程式化的特征和趋势,不仅严重困扰企业正常的生产经营秩序,影响营商环境,且“职业索赔人”滥用投诉举报、信息公开、复议诉讼、监察投诉等权利,挤占了有限的行政资源和司法资源。

在今年全国两会上,全国人大代表、山西省市场监督管理局副局长李志强建议尽快修订消费者权益保护法,明确将以营利为目的而购买、使用商品或者接受服务的公民、法人或其他组织排除在消费者之外,将“知假买假”排除在正常消费行为之外。在李志强看来,消费者权益保护法中对于消费者的概念过于简单,针对职业打假人的“知假买假”行为,可明确其作为普通民事合同行为,受民法调整,不适用消费者权益保护法的惩罚性赔偿。

李志强在实践中接触到了大量的职业打假案件,他发现知假买假的行为有形成商业化的趋势,越来越多的职业打假人、打假团队,其动机并非为了净化市场,而是利用惩罚性赔偿为自身牟利或借机对商家进行敲诈勒索,背离了社会监督的初心。

李志强建议,针对职业打假案件建立司法机关、公安机关与行政机关的协调机制,对职业打假人“知假买假”、恶意申请、敲诈勒索、缠讼滥诉行为,建立较为统一的联动整治机制,建立共享的负面行为“黑名单”,有效遏制职业打假人的牟利性打假蔓延行为,适时借助行政法和司法解释,公布指导性案例等形式,逐步遏制职业打假人的牟利性打假行为。

以“打假”“维权”名义举报、起诉商家意图牟利的“职业索赔”现象,近年来被严格规制。去年1月1日实施的《市场监督管理投诉举报处理暂行办法》明确规定,对以营利为目的的“职业举报”不再受理。