

论“负责任的人工智能”理念下的全球企业治理

刘鑫怡¹, 司伟攀¹, 晏奇²

(1. 中国科学技术信息研究所, 北京 100038;

2. 北京旷视科技有限公司, 北京 100096)

摘要: 目前全球已发布多个人工智能伦理规范, 将“负责任的人工智能”设定为核心发展目标。企业在人工智能基础研究、技术研发和治理体系构建中发挥着重要作用, 是落实“负责任的人工智能”关键主体, 但这一理念表述抽象, 难以指导企业的日常实践工作。当前企业人工智能治理也面临成本与收益不匹配、组织架构不完备等问题, 阻碍治理目标的实现。“负责任的人工智能”是推进人工智能各环节活动的一种伦理治理机制, 与环境、社会及公司治理(ESG)的企业社会责任理念密切相关, 要求在人工智能全生命周期活动中形成“政府监管—行业服务—企业自律—公众参与”的协同治理网络, 以人工智能融合环境、社会及公司治理(AI+ESG)理念作为企业人工智能伦理治理核心, 将人工智能伦理原则转化为有效的实践规则, 并根据不同类型的特征分别采取渐进式或重构式的治理模式, 承担共同但有区别的人工智能治理责任。

关键词: 人工智能; 负责任; 伦理规范; 治理

中图分类号: F204 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2023.02.008

随着人工智能(AI)在生产生活中的广泛运用, 伦理治理问题日益复杂, 受到全球关注。全球范围内已有多项政策文件将“负责任的人工智能”作为人工智能伦理治理的目标或原则, 为人工智能健康发展指明了方向。例如, 经济合作与发展组织(OECD)于2019年5月发布《人工智能原则》, 通过负责任管理促进人工智能创新。2022年9月欧盟委员会发布《人工智能责任指令(草案)》, 特别关注消费者使用人工智能产品中的维权问题。中国于2019年6月发布《新一代人工智能治理原则——发展负责任的人工智能》, 强调和谐友好、公平公正、包容共享、尊重隐私等8项原则, 提出了人工智能治理的框架和行动指南。然而, 人工智能伦理原则抽象模糊, 承载着多种解释的可能性,

难以转化为日常实践。尤其是作为人工智能研发应用活动重要主体的企业, 由于伦理理念缺失、治理工具缺乏等原因, 在运营过程中产生了算法歧视、隐私泄露等问题。因此, 本文从“负责任的人工智能”内涵切入, 结合企业社会责任等理论, 研究企业“负责任的人工智能”治理内涵, 在分析国内外企业伦理治理现状及挑战的基础上, 构建企业“负责任的人工智能”治理框架和实施路径。

1 企业“负责任的人工智能”治理理念

1.1 “负责任的人工智能”研究现状及内涵

以“responsible artificial intelligence”为主题在Web of Science数据库检索, 相关文献共4383篇, 以此作为标题的文献共47篇, 发文时间集中于

第一作者简介: 刘鑫怡(1994—), 女, 硕士, 助理研究员, 主要研究方向为人工智能治理。

项目来源: 科技创新2030—“新一代人工智能”重大项目(2020AAA0105301); 中国科学技术信息研究所创新研究基金青年项目“人工智能企业落实伦理规范的机制研究”(QN2022-18)。

收稿日期: 2022-11-14

2019—2022年。最早提出这一概念的是瑞典于默奥大学的社会与伦理人工智能教授弗吉尼亚·迪纳姆（Virginia Dignum），她在2017年指出“负责任地开发人工智能是将价值观转化为技术要求，开发能处理伦理困境的方法”^[1]。还有学者探讨“负责任的人工智能”理论与实践产生差距的原因，建议从制度完善、软硬件研发等方面进行落实^[2]。以“负责任的人工智能”为关键词在中国知网数据库进行检索，文献数量共112篇，主要从人工智能风险治理、伦理框架等视角开展研究。例如，代金平等^[3]倡导构建“负责任的人工智能”社会治理体系；王韞等^[4]将责任方法嵌入人工智能应用设计创新；刘永安^[5]对责任缺口的成因和问题进行了分析等。

“负责任”一词不仅是指承担责任，其更为深层次的意义在于通过人工智能伦理治理推动实现增进人类福祉的目标。“负责任的人工智能”是一种进行人工智能设计、开发、部署和使用的动态方法，能够提高人工智能的可信度^[6]。“负责任的人工智能”原则可以概括为公平、隐私、问责、透明和稳健，但学界对于其准确定义尚未达成共识^[7]。一些国家将“负责任的人工智能”纳入国家发展战略，例如，2021年英国出台的《国家人工智能战略》提出，与全球合作伙伴共同促进人工智能负责任发展。2022年日本发布的《人工智能战略2022》指出，日本应在“强大而负责任的人工智能”方面建立领导地位，具体措施包括努力实现可解释人工智能，促进网络安全和人工智能交叉领域的技术发展以提高可靠性等^[8]。

在实现“负责任的人工智能”理念的方法上，“负责任”是指从基础研究、应用研究到产品和服务的全方面负责任^[9]。这一观点也体现在中国的《新一代人工智能伦理规范》中，该文件以人工智能管理、研发、供应和使用这4类活动作为主线，强调将伦理道德融入人工智能全生命周期。有学者提出建立负责任的技术设计和伦理影响评估两个分析框架，将伦理方法落实到技术研发应用中^[10]。也有学者使用基于伦理和人权的协同方法，在分析人工智能技术潜在伦理风险和机遇的同时，确定相关主体的权利和责任^[11]。

总体来看，“负责任的人工智能”属于人工智能伦理治理的重要内容，其要求人工智能活动不仅要满足法律规范的底线，也要符合社会的伦理期待。它是

一种人工智能动态治理框架，包含可信、公平、非歧视和决策自主等多种伦理原则，在实践中应根据政府、企业等不同主体的目标定位而采用不同的实施机制。

1.2 企业“负责任的人工智能”研究现状及内涵

企业“负责任的人工智能”治理理念旨在运用伦理治理标准对企业商业活动进行道德评价，这一理念与企业社会责任理念的发展密切相关。根据卡罗尔的企业社会责任金字塔理论，企业社会责任融合了经济责任、法律责任、伦理责任以及慈善责任等多重义务，是企业自成立以来便需遵守的准则^[12]。之后，对于企业社会责任的要求发展为环境、社会及公司治理（ESG）理念，强调企业要关注非财务性企业评价，从追求自身利益到追求实现社会价值的最大化。这一治理理念与企业“负责任的人工智能”治理目标一致，一方面，要求企业以合法合规经营为基础，防范人工智能研发应用的风险，促进自身发展；另一方面，应推动环境、社会和经济的可持续发展，增进人类福祉。根据腾讯、商汤科技等人工智能企业发布的10余份年度ESG报告，“环境”标准主要包括节能减排、合理利用资源、可持续发展等要素；“社会”标准主要包括信息安全与隐私保护、员工多样性等；“公司治理”主要包括投资者关系管理、客户管理等内容。同时，上述企业在报告中详细阐述了人工智能伦理治理相关内容，包括建立人工智能风险防控机制、保障数据安全和提升算法可解释性等。由此可见，ESG报告不仅能展现企业履行社会责任的情况，也是衡量企业“负责任的人工智能”治理实践的重要标准。

在企业社会责任和伦理治理的语境下，企业“负责任的人工智能”治理的重点在于“责任”，需对其含义进行深入理解。根据《信息技术与道德哲学》的观点，出于某种原因有意采取某种行为并对此负责是“自然责任”；作为义务而具有某种制度性的角色使得个人需“为一个行为负责”，当个人命令代理人采取某些行为时，代理人应承担“道德责任”^[13]。然而，由于人工智能无法作为主体承担责任，所以当其作为道德代理人时，产生的影响都体现为对被代理人的道德奖惩和评价。同时，商业活动中存在设计者、销售者和使用者等众多被代理人，“负责任的人工智能”也体现为多利益相关方参与的集体责任。因此，企业在人工智能活动中要处理好3对关系：结果性责任与过程性责任、对

内责任与对外责任、个体责任与集体责任。结果性责任, 即问责制, 是指企业具有一套识别、分析并给出归责结果的制度设计。过程性责任是指在人工智能研发应用的每个环节都能明确责任主体、负责对象、承担责任的方式和程序。根据世界企业永续发展委员会对企业利益相关方的分类, 对内责任是企业对员工、股东、用户等内部受益人履行职责; 对外责任是保障妇女、残障人士等特殊群体, 以及大众和社会整体的合法权益^[14]。个体责任是指企业内部不同人员在各自工作岗位上对人工智能的研发应用所承担的相应责任, 而集体责任是指人工智能的设计者、生产者、部署者及用户等在产业链中协同一致, 维护好人工智能生态圈。

当前, 人工智能已经成为全球新一轮科技革命和产业变革的核心驱动力, 对重构世界经济版图、重塑社会经济结构和变革人类生活方式等都产生了极为深远的影响。在这种时代背景下, 各国都十分重视通过人工智能带动国家相关行业的协同发展, 企业发挥着将国家人工智能发展政策、伦理治理原则等贯彻到人工智能研发、生产制造之中, 并进一步将相关产品市场化的作用。因此, 企业作为“政府—企业—市场”链条中的关键纽带, 其必然承担落实“负责任的人工智能”的重要责任。作为一种使人工智能技术的设计、实施和应用符合伦理原则和社会规范的综合治理框架, 企业“负责任的人工智能”治理理念覆盖人工智能全生命周期的活动, 为企业人工智能伦理治理指明方向。

2 企业“负责任的人工智能”治理实践

各国政府、国际组织、智库以及企业积极开展企业“负责任的人工智能”治理研究与实践, 积累了较为丰富的治理经验。一方面, 国际组织、各国政府及研究机构等发布了一系列政策文件和伦理指南, 构建了实施框架; 另一方面, 微软、谷歌等人工智能企业积极探索, 形成了一系列“负责任的人工智能”治理的经验做法。

2.1 各国政府、国际组织、智库等加强对企业“负责任的人工智能”治理的指导

多个国际组织、国家政府部门已发布相关文件或报告, 明确针对企业人工智能研发应用中的伦理治理活动提出要求。例如, 美国国防部发布的《负责任的人工智能指南》为国防领域人工智能企业提

供指导, 欧盟委员会发布的《人工智能法案》《人工智能责任指令(草案)》等要求企业强化人工智能风险治理。各国政府、国际组织、智库等对企业的指导主要包括以下3点:

第一, 重视伦理风险影响评估。美国国防部发布的《负责任的人工智能战略和实施途径》中提出, 使用综合的风险管理方法, 坚持安全、可靠、合乎伦理的标准^[6]。日本经济产业省《实施人工智能原则的治理指南》认为, 企业应明确人工智能治理目标与人工智能系统现状之间的差距, 评估此种差距造成负面影响的严重性、范围和发生频率^[15]。世界经济论坛发布的《增强人工智能领导能力: 人工智能高管工具包》指出, 企业部署人工智能系统时应考虑模型性能、安全、社会、经济等风险, 建立数据保护影响评估、风险评估、系统开发阶段伦理评估和系统部署前的偏见评估等4项评估机制^[16]。

第二, 涵盖人工智能全生命周期。世界经济论坛发布的《负责任地使用技术》指出, 颠覆性技术具有设计开发、部署销售以及使用3个主要阶段, 应在每一个阶段运用基于伦理和人权的协同方法, 明确识别不同利益相关方所承担的责任和影响力^[11]。美国国防部发布的《负责任的人工智能指南》指出, 人工智能治理涵盖规划、开发和部署3个阶段^[17]。经济学人智库发布的《保持领先地位: 负责任的人工智能商业案例》认为, 企业应在产品设计开发的最初阶段和产品发布售后阶段嵌入“负责任的人工智能”治理实践。如果仅通过回顾性评估来判断人工智能产品和服务是否负责, 可能会导致产品开发项目取消, 甚至召回产品等问题^[18]。

第三, 针对人工智能不同阶段采取不同治理方法。2022年世界经济论坛发布的《人工智能公平和包容的蓝图》白皮书分别阐述了人工智能研发应用的5个阶段的具体实施措施, 包括识别用例、模型设计、数据收集、模型开发测试以及监测部署^[19]。电气与电子工程师协会(IEEE)发布标准IEEE P7000《解决系统设计期间伦理问题的建模过程》, 针对人工智能的概念探索和研发两大环节, 指导工程师和技术人员确定自主性、公平性及隐私等多项人工智能相关伦理价值的优先次序, 继而将价值优先级作为系统设计的基础, 从源头降低伦理风险^[20]。

2.2 领军企业探索“负责任的人工智能”治理机制

多项研究表明,企业已在实践中广泛采用“负责任的人工智能”进行伦理治理。北美地区和欧洲地区的部分企业开展了数据治理、伦理解决方案设计和风险控制等活动^[21]。西班牙电信公司采用设定价值边界、运用技术工具和分配职责等一系列措施,将人工智能原则转化为一套方法论体系^[22]。微软、谷歌、国际商业机器公司(IBM)等大型科技公司作为全球率先制定人工智能伦理原则或规范的公司,在“负责任的人工智能”方面同样具有领先的理念和实施机制。上述企业通过发布伦理指南及标准规范、成立伦理委员会等方式为其他人工智能企业提供了参考。具体措施主要包括以下4项:

第一,设立伦理委员会等专门机构。微软、谷歌、深度思维(DeepMind)等人工智能企业设立了高级别治理机构,在全面推行“负责任的人工智能”中发挥指导协调作用。伦理委员会通常由具有计算机科学、社会科学、伦理、政策等多学科背景的人员组成,其职能包括审查整个企业的项目提案、提供咨询建议和进行伦理影响评估等。伦理委员会的伦理审查贯穿项目运作全流程,例如,深度思维公司的人工智能系统AlphaFold团队在整个项目的运作中向机构审查委员会做了6次汇报,听取审查委员会的建议以防范伦理风险^[23]。

第二,设计伦理治理行动指南。为落实企业人工智能伦理原则,各企业进一步发布行动指南、实施标准等规范,细化责任要求,指导具体业务。国际商业机器公司发布《人工智能伦理在行动:推进可信人工智能的企业指南》,引导设计和开发人员在人工智能全生命周期中系统地考虑和重视人工智能伦理问题^[24]。商汤科技公司针对数据、算法及应用管理风险,建立了风险控制机制和可评估的人工智能治理落地方案。

第三,运用多种技术治理工具。企业注重采用技术手段加强人工智能伦理治理,提升人工智能可解释性、透明度、安全性和稳定性等。微软公司设计了技术工具、管理工具,以及按照应用场景需求的治理工具包。国际商业机器公司针对人工智能的可解释性、公平性、鲁棒性、透明性和隐私性分别提出技术解决方案,法国安盛集团探索确保公平性的技术工具,日本电气公司致力于保障人工智能系统的质量和稳定性。

第四,联合成立自治联盟及行业协会。2021年,元(Meta)公司与国际商业机器公司等10多家企业联合成立“数据信任联盟”,该组织设计了13类55个问题作为评估供应商的标准,涵盖人工智能训练资料、模型设计、偏差测试方法、偏差矫正和算法透明度提升与问责等,以避免招聘及员工发展中的算法偏见,负责任地部署智能技术。

2.3 企业在人工智能治理实践中面临的挑战

2021年波士顿咨询公司的调查显示,在1000多个大型组织中,只有不到一半的组织具备成熟的“负责任的人工智能”计划^[25]。目前,企业人工智能治理仍面临着运营成本、组织结构和政策法规等诸多挑战。

2.3.1 内部因素:运营成本和组织架构问题阻碍治理进程

一方面,企业践行“负责任的人工智能”成本与收益较难论证,导致企业对其的重视程度不足。虽然这一理念的实施可以从增加产品差异化、提升品牌声誉和满足客户期待等方面提升企业价值,但是目前尚未带来明确的经济效益。相反,企业需在原来的组织架构、工作流程以及业务往来上做出较大调整,甚至舍弃某些项目,例如,谷歌公司出于人工智能技术可能用于战争的担忧,放弃竞标美国国防部数十亿美元的云计算合同^[26]。同时,数据和算法是人工智能系统运行的基础,解决数据源缺乏、数据质量不高、算法偏见、可解释性和透明度不佳等问题也需要花费一定的成本,企业对人工智能治理投资能否得到预期回报存在顾虑。

另一方面,企业内部组织架构不健全,实施人员和职能分工不明确。推动人工智能伦理治理需要由企业高管部署、人工智能伦理办公室等机构进行协调。然而,受到人员选择、员工接受程度和运营成本等因素影响,部分企业未设立人工智能治理部门,或新设机构在实践中没有有效发挥作用。例如,2019年谷歌伦理委员会的一些成员因不当言论遭到数千名员工集体抵制,或因运营军事人工智能项目引发了谷歌公司的内部分歧,导致伦理委员会在成立短短9天之后解散^[27]。这一事件使得外界对伦理委员会的实效产生质疑。在岗位职能方面,目前主要由法务总监、隐私官等非技术高管推动“负责任的人工智能”实践,

而首席技术官和研发团队在其中的作用和职责并不明确, 不利于在设计研发环节落实人工智能伦理的相关原则。

2.3.2 外部因素: 治理理念的多样化和政策法规的缺失存在矛盾

从治理理念来看, 由于人工智能影响的复杂性、跨学科交流鸿沟、专业知识缺乏、职能分工复杂以及技术工具激增等原因^[28], 全球已发布的诸多人工智能伦理规范存在重叠、差异或矛盾的问题, 导致“负责任的人工智能”在实践中具有差异性。虽然许多组织认同公平性、可解释性和透明度等人工智能伦理原则, 但他们对共同原则具有不同的解释, 使得在执行过程中产生了困难。例如, 一些组织认为正义是减少不必要的偏见, 而另一种观点认为正义是尊重多样性、包容性和平等性^[29]。

从政策法规的发布情况来看, 各国仍以倡导性的人工智能伦理原则或伦理规范为主, 缺乏明确的支持性政策法规和问责监督机制, 使得企业目前以自发开展“负责任的人工智能”实践为主, 缺少向相关政府部门提出建议的渠道, 不利于形成良好的人工智能治理生态。

2.3.3 内外因素交织: 企业对人工智能全生命周期治理的不足

人工智能管理、研发、供应和使用的每个环节均面临不同的负责任需求, 但目前人工智能治理以单一环节治理为主, 各相关主体未形成联动协作。从企业内部管理来看, 原有的治理方法、人员职责等可能无法契合人工智能研发与应用的发展需求, 进而导致企业缺乏对人工智能全生命周期的治理。从人工智能产业生态来看, 产业链上下游相关主体均应在自己职责范围内负责事务工作, 但处于运营环节和使用环节的采购方可能缺乏相关“负责任的人工智能”的理论知识与实践经验, 导致不当使用或滥用人工智能产品与服务的情况。例如, 房地产商使用人脸识别系统随意收集个人信息, 侵害消费者权益。

3 企业“负责任的人工智能”治理框架与路径的构建

“负责任的人工智能”作为一种使人工智能技术的设计、实施和应用符合伦理和社会规范的综合治理框架, 契合时代发展需要, 得到各国国家政策

的大力支持, 但也面临着企业成本与收益不匹配、组织架构不健全等问题, 需要构建全生命周期的多层治理框架, 在此基础上, 不同类型的企业基于自身需求探索适合自身特色的治理路径。

3.1 企业“负责任的人工智能”治理框架

以中国相关政策文件、国际组织发布的共识性规范文件为依托, 以实现企业自身发展与增进人类福祉为目标, 企业“负责任的人工智能”治理框架可从以下3个层次进行设计:

第一, 构建人工智能融合环境、社会及公司治理(AI+ESG)的企业治理原则及标准。在人工智能治理中融合环境保护、社会责任及公司治理理念, 将人工智能系统的透明度、可解释性和稳定性等指标纳入环境、社会及公司治理评价体系中, 建立“负责任的人工智能”客观衡量标准, 借助环境、社会及公司治理对企业绩效和投资的正面影响, 更好地建立用户信任, 增强对“负责任的人工智能”的认同感, 推动企业可持续发展。

第二, 覆盖人工智能全生命周期的内部治理。由企业活动产生的责任原则、责任主体的分工协作、责任环节、预防机制及救济措施等负责任要素共同构成治理的共识性框架。根据管理、研发、供应和使用4类主要活动, 表现为绿色化的基础设施、负责任的数据处理、负责任的模型设计、负责任的开发部署、负责任的审查评估以及问责制等。

第三, 形成四方协同治理网络。企业人工智能治理涉及开发商、采购方、用户、研究机构、政策制定者、国际组织和投资方等多利益相关方, 推动形成“政府监管—行业服务—企业自律—公众参与”的协同治理网络, 以完善相关政策法规、搭建利益相关方协作平台等方式创建人工智能发展的良好生态。

3.2 不同企业的人工智能治理路径

在中国《新一代人工智能治理原则发展负责任的人工智能》的指导下, 企业践行“负责任的人工智能”既存在普适性、共识性, 也可能受业务类型、规模大小、资源能力、技术成熟度和风险敏感度等因素影响, 应采取不同的治理路径, 承担共同但有区别的责任。例如, 根据所处人工智能产业链环节的不同, 可将企业分为基础层企业、研发层企业 and 应用层企业。基础层企业注重数据收集处理、计算基础设施应用等方面的伦理治理。研发层企业重点

关注算法模型设计、基础数据集构建和研发模型部署等环节；应用层企业的业务同时聚焦产品研发和市场化两个阶段，除关注研发端治理外，还注重向用户提供必要的使用指导，包括可能的个人信息收集、数据存储安全和风险防控措施等方面的内容。因此，在治理模式、治理形式、治理工具和归责程序上，不同能力的企业应基于自身需求，探索适合自身特色的实践路径。

3.2.1 渐进式与重构式的治理模式

渐进式治理模式是指将人工智能开发和使用的伦理框架嵌入企业现有的制度和程序中，积极将伦理原则纳入商业伦理制度和日常运营中。可参考现有良好经验做法，明确内部人员的相应职责，由政府、行业协会等协助进行伦理培训、评估项目伦理风险等。

重构式治理模式强调，企业应构建一套完整的负责任治理战略和组织架构，形成跨部门的协作和跨职能团队的广泛沟通，对人工智能项目做出负责任的评估。进一步将供应商、客户和其他合作伙伴纳入负责任的生态系统中，提升公众信任度。

3.2.2 集中指导与分散执行的治理形式

企业高层自上而下指导人工智能伦理原则和伦理治理指南的实施，保证指导方法的一致性，分享最佳实践案例并明确各部门协作机制。各部门在分析现有工作结构和流程的基础上提出对人工智能治理的个性化需求和建设，以完善治理体系，提高治理效率。

3.2.3 恰当的治理技术工具和问责程序

依据自身资源和能力，企业可选择购买适合企业的技术工具，或根据企业需求定制技术工具，还可以由企业自行研发可识别、监测、诊断系统稳健性、公平性、隐私保护等相关工具。在问责程序上，企业可设定最终的负责人，例如，开发人员或批准发布应用程序或模型的人员；或是采取正式的审查流程，设定所涉及的利益相关方的责任范围，使问责流程透明化、公开化。

4 结语

由于人工智能技术具有技术属性和社会属性高度融合的特征，企业在设计、研发、部署人工智能等环节开展业务时易产生新的伦理问题，需要创新治理

理念、细化治理路径和践行企业社会责任。企业“负责任的人工智能”治理理念融合环境、社会及公司治理标准，作为一种人工智能治理的系统性方法和实践，以可持续发展、包容共享和敏捷治理为核心，从而实现人工智能发展中环境利益、社会不同群体利益以及企业自身发展利益的平衡。

鉴于当前企业人工智能治理存在的一系列挑战，“负责任的人工智能”治理理念要求政府、行业、企业和公众形成四方协同治理网络，与企业内部治理形成联动效应，塑造人工智能发展良好生态。多方主体将结合自身发展，运用政策法规、伦理规范和技术标准手段等多种方式，对人工智能进行积极引导、风险控制和源头治理。随着人工智能技术的不断发展，企业对“负责任的人工智能”治理理念认识的不断深化，领军企业的良好实践经验也将带动其他缺乏治理体系和治理能力的企业探索形成“负责任的人工智能”治理体系，推动人工智能伦理治理理念在社会经济活动中落实落地。■

参考文献：

- [1] DIGNUM V. Responsible artificial intelligence: designing AI for human values[J]. *ICT discoveries*, 2017(1): 1-8.
- [2] GOLBIN I, RAO A S, HADJARIAN A, et al. Responsible AI: a primer for the legal community[C]//2020 IEEE International Conference on Big Data(Big Data), December 10-13, 2020. Atlanta: IEEE, 2020: 2121-2126.
- [3] 代金平, 李杨. 构建负责任的人工智能社会治理体系[J]. *山东社会科学*, 2022(8): 147-153.
- [4] 王韞, 徐迎庆. 负责任的人工智能与设计创新[J]. *包装工程*, 2021, 42(6): 1-6, 36.
- [5] 刘永安. 在有意义的人类控制下发展负责任的人工智能[J]. *自然辩证法通讯*, 2022, 44(9): 95-101.
- [6] U.S. Department of Defense. Responsible artificial intelligence strategy and implementation pathway[EB/OL]. (2022-06-22)[2022-10-03]. https://insidecybersecurity.com/sites/insidecybersecurity.com/files/documents/2022/jun/cs2022_0138.pdf.
- [7] ARRIETA A B, DÍAZ-RODRÍGUEZ N, DEL SER J, et al. Explainable artificial intelligence(XAI): concepts taxonomies opportunities and challenges toward responsible AI[J]. *Information fusio*, 2020, 58(1): 82-115.
- [8] 統合イノベーション戦略推進会議. AI戦略2022[EB/

- OL]. (2022-04-22)[2022-10-03]. https://www8.cao.go.jp/cstp/ai/aistrategy2022_honbun.pdf.
- [9] 王秋蓉, 于志宏. 发展负责任的人工智能: 访清华大学文科资深教授、清华大学苏世民书院院长薛澜 [J]. 可持续发展经济导刊, 2019(7): 13-18.
- [10] PETERS D, VOLD K, ROBINSON D, et al. Responsible AI: two frameworks for ethical design practice[J]. IEEE transactions on technology and society, 2020, 1(1): 34-47.
- [11] World Economic Forum. Responsible use of technology[EB/OL]. (2019-09-19)[2022-10-03]. <https://www.weforum.org/whitepapers/responsible-use-of-technology>.
- [12] CARROLL AB. The pyramid of corporate social responsibility: toward the moral management of organizational stakeholders[J]. Business horizons, 1991, 34(4): 39-48.
- [13] 尤瑞恩·范登·霍文, 约翰·维克特. 信息技术与道德哲学 [M]. 赵迎欢, 宋吉鑫, 张勤译. 北京: 科学出版社, 2013: 199.
- [14] 殷慧芬, 侯怀霞. 企业社会责任法律问题的新发展 [M]. 北京: 法律出版社, 2021: 28.
- [15] Expert Group on How AI Principles Should be Implemented AI Governance Guidelines WG. Governance guidelines for implementation of AI principles Ver.1.1[EB/OL]. (2022-01-28)[2022-10-14]. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf.
- [16] World Economic Forum. Empowering AI leadership: AI C-Suite toolkit[EB/OL]. (2022-01-12)[2022-10-14]. <https://www.weforum.org/reports/empowering-ai-leadership-ai-c-suite-toolkit>.
- [17] U.S. Department of Defense. Defense Innovation Unit Publishes' Responsible AI guidelines'[EB/OL].(2021-11-15)[2022-10-14]. <https://www.defense.gov/News/News-Stories/Article/Article/2847598/defense-innovation-unit-publishes-responsible-ai-guidelines/>.
- [18] Economist Intelligence Unit. Staying ahead of the curve: the business case for responsible AI[EB/OL]. (2020-10-08)[2022-10-14]. <https://pages.eiu.com/rs/753-RIQ-438/images/EIUStayingAheadOfTheCurve.pdf>.
- [19] World Economic Forum. A blueprint for equity and inclusion in artificial intelligence[EB/OL]. (2022-06-29)[2022-10-14]. <https://www.weforum.org/whitepapers/a-blueprint-for-equity-and-inclusion-in-artificial-intelligence>.
- [20] IEEE Std 7000™-2021. IEEE standard model process for addressing ethical concerns during system design[S]. U.S.:IEEE, 2021:31, 39.
- [21] WANG Y C, XIONG M R, OLYA H. Toward an understanding of responsible artificial intelligence practices[C]//January 7-10, 2020. Hawaii: HICSS, Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020: 4962-4971.
- [22] BENJAMINS R, BARBADO A, SIERRA D. Responsible AI by design in practice[EB/OL]. (2019-12-20)[2022-10-03]. <https://arxiv.org/abs/1909.12838>.
- [23] Fortune. How to use A.I. responsibly: lessons from DeepMind and Intel[EB/OL]. (2022-09-20)[2022-10-03]. <https://www.jollypostie.com/store/fortune/responsible-ai-lessons-from-deepmind-and-intel-632a0b6e05f6c>.
- [24] IBM. AI ethics in action: An enterprise guide to progressing trustworthy AI[EB/OL]. (2022-03-31)[2022-10-14]. <https://www.ibm.com/downloads/cas/4DPJK92W>.
- [25] Boston Consulting Group. Are you overestimating your responsible AI maturity?[EB/OL]. (2021-03-30)[2022-10-03]. <https://www.bcg.com/publications/2021/the-four-stages-of-responsible-ai-maturity>.
- [26] FOX Business. Google pulls out of race for billion dollar Pentagon defense contract[EB/OL]. (2018-10-08)[2022-10-14]. <https://www.foxbusiness.com/technology/google-pulls-out-of-race-for-billion-dollar-pentagon-defense-contract>.
- [27] Financial Times. Google scraps ethics council for artificial intelligence[EB/OL]. (2019-04-05)[2022-10-14]. <https://www.ft.com/content/6e2912f8-573e-11e9-91f9-b6515a54c5b1>.
- [28] SCHIFF D, RAKOVA B, AYESH A, et al. Principles to practices for responsible AI: closing the gap[EB/OL].(2020-06-08)[2022-10-03]. <https://arxiv.org/abs/2006.04707>.
- [29] JOBIN A, IENCA M, VAYENA E. Artificial intelligence: the global landscape of ethics guidelines[J]. Nature machine intelligence, 2019, 1(9): 389-399.

(下转第76页)

[34] 中国信息通信研究院. 开源生态白皮书 [EB/OL]. [2022-11-01]. http://www.caict.ac.cn/kxyj/qwfb/bps/202109/t20210918_390059.htm.

Policy Analysis and Practical Cases for Open Sharing of Research Software in France

ZHAI Jun, FAN Wei-hua, ZUO Yun-hao, KOU Hong-hong

(School of Maritime Economics and Management of Dalian Maritime University, Dalian, Liaoning 116026)

Abstract: Since 2021, France has introduced new policies on open government, open science and data, all of which take open software and source code as key fields. The systematic analysis of them can provide references for China to optimize policies for open science. Using the methods of web survey and literature analysis, taking the opening and sharing of research software in France as the main line, this paper introduces the main objectives and key measures of these policies, shows the progress of practice, and then extracts useful enlightenment for China. During the period of the global open science movement of “equal emphasis on software and data”, as a powerful country in science and a leading country in open science, France has taken the lead in the world in the open sharing of research software. It is worth learning from and using for reference in promoting the unified opening of data and software, building an open science infrastructure for software and promoting international cooperation.

Keywords: open science; research software; open sharing policy; open source software; source code

(上接第66页)

Global Companies Governance Mechanism on Responsible Artificial Intelligence

LIU Xin-yi¹, SI Wei-pan¹, YAN Qi²

(1. Institute of Scientific and Technical Information of China, Beijing 100038;

2. Megvii Technology Limited, Beijing 100096)

Abstract: At present, a number of ethical norms of artificial intelligence (AI) have been issued around the world, and “responsible AI” has been set as the core development goal. Companies play an important role in AI basic theory research, technology research and development and governance system construction, becoming the key subject to implement “responsible AI”. However, the term of “responsible AI” is abstract, and it is difficult to guide the daily practice of companies. AI governance of companies also faces some problems such as the mismatch between cost and benefit, incomplete organizational structure, etc., which hinder the realization of governance. “Responsible AI” is an ethical governance mechanism to promote all aspects of AI activities, which is closely related to the corporate social responsibility concept of “environment, society and corporate governance” (ESG). It requires a collaborative governance network of “government supervision–industry service–enterprise self-discipline–public participation” to be formed in the whole life cycle activities of AI. It takes “AI+ESG” as the core of AI ethical governance of companies, and transforms AI ethical principles into effective and responsible practice rules. According to different types and characteristics, companies adopt progressive or reconfigurable governance models, and undertake common but differentiated AI governance responsibilities.

Keywords: artificial intelligence; responsible; ethical norms; governance