

责任式创新范式下的新兴技术创新治理解析 ——以人工智能为例

梅亮^{1,2}, 陈劲^{1,2}, 吴欣桐^{2,3}

(1. 清华大学 经济管理学院, 北京 100084; 2. 清华大学技术创新研究中心, 北京 100084;
3. 西南石油大学 经济管理学院, 成都 610500)

摘要: 从“责任式创新”这一新兴创新范式切入, 聚焦人工智能技术创新治理议题, 构建了责任式创新范式下人工智能技术创新治理分析框架(包括技术维、经济维、伦理维和社会维), 并系统分析了人工智能技术创新在理论和实践层面的双重性, 从而为人工智能技术创新的可持续发展提供借鉴。

关键词: 责任式创新; 新兴技术; 创新治理; 人工智能

中图分类号: F091.3540 文献标志码: A 文章编号: 1002-980X(2018)01-0001-07

1 研究背景与问题提出

以内燃机、原子能、信息科学、生物医药和互联网等为代表的科技创新推动了产业发展和社会变革^[1-2]。然而, 随着科技与社会的交互程度不断加深, 典型的科技创新在促进社会进步的同时产生了许多负外部性影响^[3]。以纳米科技、转基因、无人驾驶汽车、干细胞研究、生物医药技术、核能、机器人和军事安全技术等为代表的新兴科技被视为争议性技术创新, 引发了研究与政策层面对科技创新面向技术社会控制、伦理安全等方面的担忧^[4], 旧有的科技创新模式、政策法规、评估方法和科学社会契约关系等在面向新兴科技创新治理时存在制度缺失(institutional void)^[5], 科技创新的社会责任议题引发了学术研究者和政策制定者的关注^[6]。

“人工智能”作为绿色与智能时代的创新基石, 自 1956 年被 John McCarthy 提出后, 经历了技术层面、经济层面和社会层面的飞速发展, 逐步形成了弱人工智能时代、强人工智能时代和超人工智能时代的发展转型^[7]。智能语音技术、温度可识别的马克杯、国际象棋“深蓝”机器人、谷歌“可穿戴”设备、围

棋 AlphaGo 等的应用, 证实了人工智能的快速进化以及超越人类部分能力的科技创新和产业应用潜力。国际咨询与预测行业领导机构麦肯锡将人工智能与先进机器人定为 2025 年 12 大新兴科技创新发展领域。2016 年 3 月 21 日我国工业和信息化部、发展和改革委员会和财政部联合发布《机器人产业发展规划(2016—2020)》, 将弧焊机器人、真空(洁净)机器人和全自主编程智能工业机器人等十大标志性创新产品的应用与发展列入国家发展规划^①。

然而, 随着人工智能的高速发展, 科学家、产业巨头和政治家等开始纷纷担忧人工智能的潜在危机和发展隐患。2016 年 1 月 20 日, 英国物理学家 Stephen Hawking 发表评论, 认为人工智能的潜在危机正变得越来越大, 并指出: 人工智能的短期影响在于人类能否控制机器, 长期影响在于机器能否被控制。而后, 美国特斯拉公司创始人 Elon Musk 指出发展人工智能是在“召唤魔鬼”, 并表示将投资人工智能监控领域。微软总裁 Bill Gates 同样支持人工智能危机论的观点, 并认为人工智能已强大到需要警惕的阶段。2016 年 3 月, 以 Hawking、Musk 和 Gates 为首的全球数百名科技领域的专家签署公开联

收稿日期: 2017-11-09

基金项目: 国家自然科学基金青年科学基金项目“责任式创新的共性理论基础与作用机理研究”(71704090); 中国博士后科学基金第 61 批面上项目“责任式创新: 一个‘溯源—理论—实践’的整合框架研究”(2017M610097); 全国哲学社会科学基金项目重大项目“建设世界科技创新强国的战略比较与实现路径研究”(17ZDA082); 国家科技支撑计划项目“公共文化科技服务能力建设与绩效评估体系及共性技术研究”(2015BAK26B01)

作者简介: 梅亮(1987—), 男, 浙江舟山人, 清华大学经济管理学院在站博士后, 清华大学技术创新研究中心助理研究员, 研究方向: 开放式创新、责任式创新、创新政策; 陈劲(1968—), 男, 浙江余姚人, 清华大学经济管理学院教授、博士生导师, 研究方向: 创新管理、科技政策; 吴欣桐(1991—), 女, 四川南充人, 西南石油大学博士研究生, 清华大学技术创新研究中心访学博士生, 研究方向: 性别化创新、创新政策。

① <http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757018/c4746362/content.html>

名信,呼吁政府和社会对人工智能研究进行理性监管,以确保人工智能的科技创新不会失控,并警示人工智能的潜在危害^[8]。2017年9月4日,俄罗斯总统普京强调,领导人工智能的国家将会成为世界规则制定者,并担忧人工智能机会创造之外的社会危机^①。

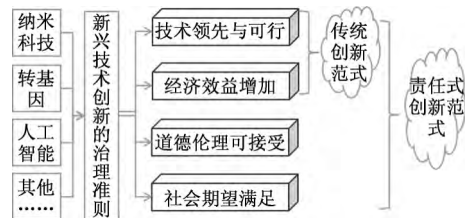
在理论层面,人工智能的快速演进同样引发了传统科学与系统观对科技创新的讨论。首先,从技术的社会控制视角,人工智能的技术兴起引发了对“科林格里奇”困境的讨论,即人类无法在人工智能技术兴起阶段有效预测其社会后果,当负外部性危机出现时,该技术因对经济与社会结构的深度嵌入而无法被有效管控^[9]。其次,从创新的双重性视角出发^[3],人工智能科技创新在催生新兴产业发展与社会关系重构的同时,往往引发道德伦理、环境保护和社会民生等方面的负向危机,如人工智能带来的工人失业危机、社会情感缺失,以及强人工智能对人类活动的替代风险等。再次,从道德伦理的视角,人工智能的社会影响分析聚焦于机器人伦理^[10]、信息与计算伦理^[11]等狭义范畴下的机器伦理研究,亟待系统化地开展人工智能伦理议题的讨论^[12]。最后,人工智能技术创新治理呼唤责任嵌入的制度设计,并通过更广泛的科学、技术与社会的交互,实现人工智能技术创新发展的社会满意^[13]。

由此,作为科学研究与产业创新的典型驱动型技术^[14],人工智能的兴起在推动产业创新、创造增长机遇、引导社会变革的同时,其技术创新涉及的治理议题及其分析亟待引入更为系统的创新范式框架^[13],从而将新兴技术发展的社会责任嵌入于创新治理的框架内,以有效实现欧盟“2020 智慧增长”战略面向新兴技术治理的两个目标:第一,人类有效界定技术创新的社会结果和社会影响;第二,支持特定技术创新并将之引导至社会满意的发展演进方向^[15]。

2 责任式创新范式下新兴技术创新治理分析框架的构建

基于新兴科学与技术的兴起以及科技创新与社会政策的交互性增强,科学与社会的关系从传统的“科学与社会”(science and society)、“社会中的科学”(science in society)开始向“科学伴随社会”和“科学为了社会”(science with and for society)的范式目标转移^[16]。新兴科学技术创新的治理也开始从科学共同体的内部自治转向更广泛社会层面的治理^[17]。2003年美国面向纳米科学领域的新兴

科技研究颁布法令,首次在国家创新政策层面提出“责任式发展”(responsible development),指出纳米科学与技术发展的核心是提高这一新兴科技的积极影响并管控其潜在危害,从而最大限度地实现社会发展的责任需要^[18]。与之相关,欧盟委员会针对新兴科技的可持续发展提出“责任式研究与创新”(responsible research and innovation),进一步在“地平线 2020 框架计划”中将该核心概念确定为“责任式创新”(responsible innovation),即面向未来目标而对当下科学与创新的集体管理^[19],并认为科学研究与技术创新必须有效反映社会需求和社会意愿、反射社会价值和社会责任,创新应表现为道德可接受、社会期望满足、安全和可持续等^[2,20]。作为研究与创新的新兴范式^[21-22],责任式创新以科技创新活动的双重性^②^[3]、创新结果面向未来的不确定性^[23]为基本假设,更为关注科学共同体之外更广泛情境下的技术创新治理议题^[23],强调在创新主体承认能力局限的前提下^[24],以更广泛的利益攸关主体参与^[19],展开对特定技术创新过程与结果的预测式治理^[25]和响应式治理^[18]。作为对创新活动的责任嵌入,责任式创新排除了对创新结果的简单预测,认为未来结果的不确定性会引发传统创新过程和治理模式的变革^[2,23],创新过程需要基于开放、透明的原则,有效解释创新的伦理规范、价值回报和发展意义,引导创新为未来挑战提供解决方案,实现社会期望满足和公共价值创造^[2,26]。在责任式创新范式下,新兴技术治理的评估准则,在传统创新范式聚焦技术先进性与可行性、经济效益增加的基础上^[27],进一步延伸至道德伦理可接受与社会期望满足两个标准^[2,27]。责任式创新范式下新兴技术创新治理分析框架如图 1 所示。



注:该图基于参考文献[2]进行修改完善。

图 1 责任式创新范式下新兴技术创新治理分析框架

3 责任式创新范式下人工智能技术创新治理框架与分析

人工智能概念最早在 1956 年美国达特茅斯会

① <https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world>

② 创新同时产生正向和负向的影响。

议上被提出,相关研究与实践之后的60余年时间里历经了孕育、初创、形成、发展和繁荣五个阶段的演进,如图2所示。作为知识与智能时代最重要的新兴技术和驱动型技术典型^[14],近年来人工智能受到学界和产业界的广泛关注,相关讨论主要涉及其技术、经济、社会等方面。然而,随着人工智能技术应用和产品研发的深入、人工智能新兴产业的兴起以及人工智能技术创新的迭代,人工智能在就业、个人隐私、机器人伦理、算法黑箱、国家和社会安全等方

面的潜在风险引发了人们对人工智能技术创新治理的高度关注。如何规避人工智能技术创新治理陷入“科林格里奇困境”^[28],规范机器伦理道德与法律准则^[12,29],管控人工智能的社会风险和社会影响^[13],实现对人工智能发展争议性的有效治理等,成为研究与实践的重要议题。为此,责任式创新有助于系统评估人工智能技术创新的双重影响、有效匹配科技创新活动与社会发展目标,从而为科技创新的责任嵌入提供了治理框架选择。

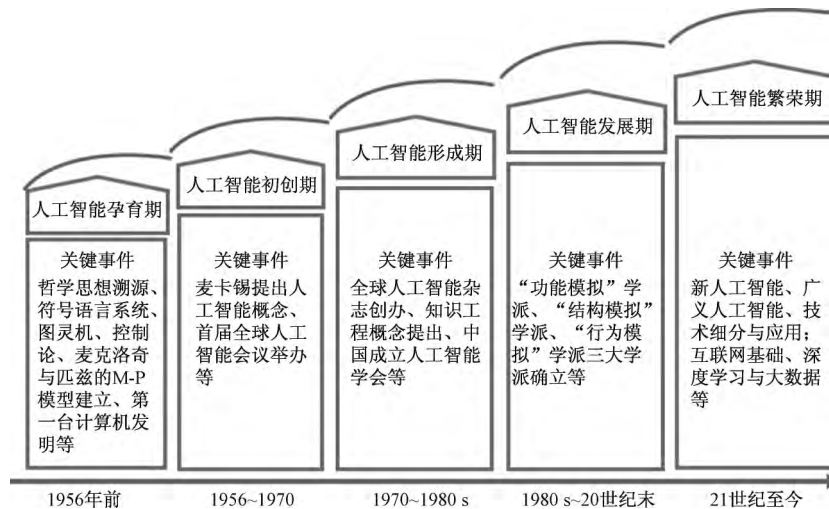


图2 人工智能发展简史

3.1 责任式创新范式下人工智能技术创新的双重性

1) 技术维。

“人工智能”是关于机器智能化的科学与工程^[30],涉及智能体的研究与设计,强调通过智能体对外部环境系统感知,最大限度地实现智能体行动结果的成功^[31]。自1956年人工智能概念被提出以来,人工智能研究发展演进为四个范畴:技术科学的仿人类思考(thinking humanly),即使得计算机和机器具备完全自主思考的能力^[32];理性思考(thinking rationally),即基于计算模型的心智研究^[33];仿人类行动(acting humanly),即对计算机开展人类活动的研究^[34];理性行动(acting rationally),即机器智能研究和智能体设计^[35]。随着技术研究的深入,人工智能作为典型的新兴技术,面向产品创新、服务创新和商业模式创新等领域涌现多样化的产业应用,如语音认知、手写识别、机器翻译、系统过滤、搜索引擎、邮箱应用、人脸识别、图像识别、棋牌游戏、无人驾驶汽车、智能机器人、语义分析、视频监控和情感计算等,从而为技术产业的兴起、技术价值的创造等发挥了积极作用。以弱人工智能领域的技术

创新为例,诸如智能搜索、智能推荐、智能排序和智能筛选等算法一级的技术创新通过嵌入互联网平台,促进了以互联网为基础的科技产业创新的迭代与演进。再如,以机器人为代表的创新应用,面向语音识别与认知的语音机器人,面向救灾、仿生和载人等领域的特种机器人等强人工智能领域的技术创新,在技术和产业之外逐步服务于人类生活的方方面面。科学家和技术创新者由此重审人与机器的关系,并试图从人机交互(human-computer interaction)^[36]、人工智能与人类共生^[37]等角度寻求未来人工智能技术创新的社会解答。然而,人工智能技术的兴起同时引发了技术负外部性、技术社会控制等方面的问题:人工智能研究项目中的自动化武器、机器人武器、大数据外泄等引发了国防安全威胁,机器人技术的快速演进引发了“技术奇点”临近危机下机器人对人类生存的全面威胁^[38]等。2017年8月21日,Elon Musk等116名人工智能与机器人专家联合发表公开信,呼吁联合国关注人工智能存在的潜在社会威胁,上书请求联合国禁止“杀人机器人”的研发^①。

① http://news.xinhuanet.com/world/2017-08/21/c_1121518687.htm

2)经济维。

人工智能科技创新驱动新兴产业的崛起,涌现了细分行业与利基市场的产品创新、服务创新、商业模式创新^[14],也引发产业创新周期进程中传统企业商业模式的转型,以及以人工智能为新兴焦点产业与传统产业的资源融合与价值重构。例如:AI 面向医疗产业形成的医疗影像诊断、医学病例分析以及人工智能专家诊断系统等;AI 面向通信领域形成的海量数据规划与 5G 研发;AI 智能化芯片的技术创新与产业应用;AI 引导无人驾驶汽车的开发;AI 面向非技术行业的商业模式创新——包括金融领域的财务、法律、股票机器人应用,安防领域基于人脸识别技术的门禁、打卡和基于感知智能技术的身份认证等^[39]。数据显示:截至 2017 年 6 月 31 日,全球

人工智能企业总计 2542 家,企业数量呈指数级增长趋势。以机器人市场为例:2017 年全球机器人市场规模预期为 232 亿美元,2012—2017 年的平均增长率达到 17%,其中中国的市场规模预期为 63 亿美元,2012—2017 年的平均增长率达到 28%^①。然而,人工智能在驱动新兴产业涌现、重构传统产业价值分配的同时产生了经济负外部性效应,突出表现为人工智能和机器人革命带来巨大的失业风险,民众个体和企业经营活动的价值回报受到挑战。基于对全球 352 名人工智能专家的调研资料,牛津大学未来人文研究所认为,语言翻译、论文写作、驾驶等人类工作岗位将在可预期的时间内完全被人工智能取代(如表 1 所示),这会冲击和破坏潜在的传统经济价值与分配机制。

表 1 人工智能对人类工作的替代时间

工作类型	人工智能替代的“预估时间点”(区间)
劳动力完全自动化(全部人类工作)(Full Automation of Labor) (All Human Jobs)	约 2137—2139 年
人工智能研究者(AI Researcher)	约 2104 年
高阶机器智能(所有人类工作任务)High Level Machine Intelligence(All Human Tasks)	约 2060—2062 年
数学研究(Math Research)	约 2058—2060 年
外科医生(Surgeon)	约 2053 年
普特南数学竞赛(Putnam Math Competition)	约 2049—2052 年
《纽约时报》畅销书写作(Write <i>New York Times</i> Best-Seller)	约 2048—2051 年
零售业职员(Retail Salesperson)	约 2029—2030 年
训练师 Go (Same Training as Human)	约 2028—2029 年
5 公里城市赛跑(5km Race in City)(Bipedal Robot vs. Human)	约 2028 年
卡车司机 Truck Driver	约 2027 年
流行音乐制作 Generate Top 40 Pop Song	约 2027 年
游戏行动解释 Explain Own Actions in Games	约 2026 年
中学论文写作 Write High School Essay	约 2026 年
文字阅读与语音转换 Read Text Aloud(Text-to-Speech)	约 2025 年
雅达利游戏 All Atari Games	约 2025 年
所有乐高组装 Assemble Any LEGO	约 2024 年
电话银行运营商 Telephone Banking Operator	约 2024 年
翻译者 Translate(vs. Amateur Human)	约 2024 年
演讲速记 Transcribe Speech	约 2024 年
星际争霸 Starcraft	约 2022 年
折洗衣服 Fold Laundry	约 2022 年
世界扑克系列赛 World Series of Poker	约 2020 年
愤怒的小鸟 Angry Birds	约 2019 年

注:根据“世界经济论坛”中牛津大学未来人文研究所发布的材料整理,其中替代预估时间是基于所有专家预测时间计算的均值。

3)伦理维。

人工智能与机器人研究与创新的兴起引发了学界和实践界对人工智能道德伦理可接受性的广泛讨论,人工智能与机器人的伦理责任议题被关注^[40,41],具体包括人权伦理、责任伦理、道德地位伦理、代际伦理、环境伦理以及大数据伦理与隐私等^[42]。以联合国人权宣言为基础,研究与实践呼唤

人类责任与机器人责任伦理的标准同等性,主要包括正义与团结、相互尊重与合作、信任与容忍、基本人性准则、公平的国际秩序、公共事务的有意义参与、意见与言论自由、身心健康、孩童与老年者权利、教育艺术与文化、反暴力与尊重生命等^[43]。学界将人工智能与机器人等的广泛应用描述为人类文明史上具有重要影响的社会伦理试验,强调人工智能的

① 数据来源:《中国机器人产业发展报告》(2017)。

道德代码与伦理嵌入的重要性^[44]，并通过预测式治理、科学家之外的公众参与治理、响应式治理、高阶自省^[19-20,45]、科学与社会整合框架^[46]等将人工智能技术创新治理的伦理议题嵌入于责任式创新范畴内。在实践领域，全球各国也通过构建相关制度机制——如制定并颁布人工智能政策与法规、人工智能创新发展准则条例等——实现人工智能伦理端的责任嵌入。典型如联合国发布的《关于机器人伦理的研究报告》，国际标准制定组织 IEEE 发布的《合伦理设计：利用人工智能和自主系统最大化人类福祉的愿景》，美国政府在“国家人工智能研究与发展战略计划”中提出的公平性、透明性和追责性标准^[47]，以及 Hawking、Elon Musk 等推动全球近 1000 名研究员签署并发布的“人工智能 23 条准则”（或称阿西洛玛人工智能原则）等^[48]。

4) 社会维。

责任式创新范式下人工智能技术创新治理框架最后延伸至人工智能的社会维度，并在更宽泛的科学与社会交互关系演进以及人工智能科技创新的社会情境中反映创新的责任意义^[13]。在理论层面，Brundage 从责任式创新的角度探讨了人工智能技术创新，强调人工智能与人机交互创造社会价值，并认为探索人工智能技术创新治理应在狭义的技术范畴与机器伦理研究之外，关注人工智能技术创新更广泛的社会影响，强调人工智能科技创新的本质目标应是社会满意，并深入讨论了人工智能技术创新治理的三个基本方面：第一，在人工智能的设计阶段，考虑并预测技术创新的社会影响；第二，在人工

智能的应用阶段，系统反思人工智能技术创新在特定应用领域的理论与实践意义；第三，引入公众参与，明确公众对人工智能技术创新的期望以及公众在人工智能发展中所需了解的信息和知识^[13]。实践界同样关注人工智能在促进生产率和社会运转效率提升的同时，在社会公平、资源分配、国家安全和文明福祉等方面具有的相关社会属性。2016 年 12 月联合国召开常规武器审查会议，针对“人工智能自动化武器的社会安全专项议题”展开讨论。在 123 个成员国中，19 个成员国明确呼吁禁止人工智能自动化武器的研发，并强调了其投入使用会给国际安全带来的负向影响。《未来简史》一书的作者 Yuval Harari 则担忧人工智能发展导致的失业问题等会引导社会基本结构趋向少数社会精英与多数无价值民众极端不平等的二元属性^[48]。2017 年 4 月 27 日霍金在全球移动互联网大会上提出“人工智能崛起可能是人类文明终结”的论断，指出了人工智能在自动驾驶、自主武器和隐私安全等方面的短期社会危害，以及长期视角下人工智能系统失控带来的重大社会危害^①。此外，人工智能产品责任^[47]、国防安全威胁^②、机器情感学^[49]、聊天机器人和性爱机器人等^[50]引发的社会心理危机和健康隐患等，需要研究进一步关注人工智能技术创新的社会责任议题，从而实现人工智能这一争议性技术创新对社会发展期望的满足。

基于此，本文基于责任式创新范式下新兴技术创新治理分析框架，对人工智能创新的双重性进行了分析总结，见表 2。

表 2 责任式创新范式下人工智能技术创新治理框架的分析总结

核心维度	正向效应	负向效应
技术维	弱人工智能技术创新，如智能搜索、智能推荐、智能排序等； 强人工智能技术创新，如语音机器人、特殊用途机器人（救灾、仿生载人等）	自动化武器、机器人武器；技术奇点；杀人机器人
市场维	机器人市场增长空间； 新兴市场创业、投资、人才储备； AI+医疗、AI+通信、AI+芯片、AI+金融、AI+安防等	失业危机；旧有产业价值链破坏与价值流失；经济负外部性
伦理维	机器人伦理报告；人工智能伦理设计；阿西洛玛人工智能 23 条原则	人权伦理、责任伦理、道德地位伦理、代际伦理、环境伦理、大数据伦理与隐私等
社会维	人工智能的社会价值创造； 人机共生与协同的价值涌现	国防与社会安全；社会不平等；社会心理危机；文明终结

4 结论与政策启示

人工智能经过近年来的快速发展已成为新兴技术驱动社会转型变革的重要实例^[13]。然而，人工智能面向机器信任与伦理、武器应用、算法黑箱、产品

问责、大数据隐私风险、就业替代等负外部性议题引发研究与实践对这一争议性技术创新及其双重性社会控制的广泛关注^[4]，也引导旧有科技创新治理范式向科技与社会深度嵌入下创新责任的制度化过程转移^[2,18-19,51]。本文聚焦于人工智能技术创新治理

① <http://tech.qq.com/a/20170427/018205.htm>

② 哈佛大学《人工智能与国家安全》报告

议题,从责任式创新的角度,构建了包括技术、经济、伦理和社会四维度的人工智能技术创新治理分析框架,并系统分析了人工智能技术创新的双重性。本研究的政策启示如下:

第一,从责任式创新范式视角,针对以人工智能技术创新为代表的争议性技术创新及其治理问题,提供了评估准则和分析框架。传统创新范式思维下人工智能技术创新本身潜在的负外部性未引起重视,其制度规范和治理模式也存在严重的制度缺失与制度空白(institutional void)^[5]。尽管近年来以人工智能技术创新为代表的新兴技术创新的争议性引发了一定讨论^[4],但是已有研究多聚焦于狭义的技术危机层面^[38]或机器伦理层面^[40-41],对人工智能技术创新及其治理问题缺乏系统而全面的分析^[13],亟待从技术、经济、伦理和社会多个方面深度讨论诸如人工智能等争议性技术的治理与管控方法^[52]。作为创新的新兴范式^[19,21-22],责任式创新在传统创新范式所关注的技术先进性与可行性、经济效益增加的基础上,进一步强调科技创新的道德伦理可接受以及社会期望满足等准则^[2,18,27],为人工智能技术创新提供了一种系统评估与治理的范式框架^[13],引导人工智能技术创新向社会满意的方向演进^[15,25]。

第二,从责任式创新范式视角,提供新兴技术创新治理的政策启示。责任式创新概念源自欧盟发布的“地平线 2020”计划,意指通过对现有科学与创新进行集体管理以探索未来^[19],并由此成为美国和欧盟科技政策中的“流行语”(buzzword)^[53]。Sutcliffe 从政策层面对责任式创新进行了系统总结,认为其核心目的在于实现创新的社会价值,强调研究与创新全过程中社会因素的融合以及异质性成员主体的包容,通过开放研究与创新过程,展开短期与长期时域内对特定创新活动在技术、经济、伦理和社会等方面的系统评估^[25]。作为一种创新面向政策层面的顶层设计战略^[15],责任式创新范式下新兴技术创新治理的根本在于实现创新的社会满意^[15]。人工智能作为典型的新兴技术,其技术创新治理已成为国际性议题,研究与实践呼唤责任嵌入的治理框架,实现人工智能面向技术、经济、道德伦理以及社会等全方位的治理。责任式创新范式下新兴技术创新治理分析框架为人工智能技术创新的双重性^[3]提供了一种思考启示,即在全球治理层面政策需要引入更广泛的利益攸关者,并通过科学与社会的交互以及政策制定者、科研创新者、民众对创新过程的协同参与,明确人工智能面向技术、经济、伦理和社会的发展机遇与潜在风险,进一步通过前瞻性的研究评估与响应性的过程监管^[20],引导人工智能向责任式创

新的过程演进,最终实现该新兴技术创新面向人类的社会满意,创造公共价值溢出。当前以美国和中国为代表的全球人工智能发展大国越来越意识到人工智能创新双重性的重要性,人工智能技术创新治理的国际间合作、监督、共享机制,人工智能创新治理的跨制度、跨文化协调机制等,也为人工智能技术创新演进的责任嵌入提供了未来的政策参考。

参考文献

- [1] OWEN R, BAXTER D, MAYNARD T, et al. Beyond regulation: risk pricing and responsible innovation[J]. *Environmental Science & Technology*, 2009, 43 (18): 6902-6906.
- [2] 梅亮, 陈劲. 责任式创新: 源起, 归因解析与理论框架[J]. *管理世界*, 2015(8): 39-57.
- [3] JONAS H. *The Imperative of Responsibility: In Search of An Ethics for The Technological Age*[M]. University of Chicago Press, 1984.
- [4] EUROBAROMETER S. Social values, science and technology[R]. Eurobarometer Special Report, 2005.
- [5] HAJER M. Policy without polity? Policy analysis and the institutional void [J]. *Policy Sciences*, 2003, 36 (2): 175-195.
- [6] 梅亮, 陈劲, 李福嘉. 负责任创新: 内涵辨析与启示[J]. *自然辩证法研究*, 2017(2): 49-53.
- [7] 王天一. *人工智能革命: 历史、当下与未来*[M]. 北京时代华文书局, 2017.
- [8] 刘园园. 霍金等千名专家敦促禁止“杀人机器人”[N]. *科技日报*, 2015-08-03.
- [9] COLLINGRIDGE D. *The Social Control of Technology* [M]. London: Pinter, 1980.
- [10] LI N P, ABNEY K, BEKEY G A. *Robot Ethics: The Ethical and Social Implications of Robotics* [M]. MIT Press, 2011.
- [11] FLORIDI L. *The Cambridge Handbook of Information and Computer Ethics* [M]. Cambridge University Press, 2010.
- [12] COELHO H, COSTA A R, LOPES L S, et al. On the intelligence of moral agency[Z]. *Encontro Português de Inteligência Artificial*, 2009.
- [13] BRUNDAGE M. *Artificial Intelligence and Responsible Innovation*[M]//MLLER V C. *Fundamental issues of artificial intelligence*. Springer International Publishing, 2016: 541-552.
- [14] TEECE D. Profiting from innovation in the digital economy: standards, complementary assets, and business models in the wireless world. 2017, Forthcoming.
- [15] VON SCHOMBERG R. A vision of responsible research and innovation[M]//OWEN R, BESSANT J, HEINTZ M. *Responsible innovation: managing the responsible emergence of science and innovation in society*. Wiley, 2013: 51-74.
- [16] MEJLGAARD N, BLOCH C. Science in society in Eu-

- rope[J]. Science and Public Policy, 2012, 39(6): 695-700.
- [17] CSONKA A I L. Responsible science in societies[J]. Responsible Innovation, 2014(57): 12-35.
- [18] OWEN R, MACNAGHTEN P, STILGOE J. Responsible research and innovation: from science in society to science for society, with society[J]. Science and Public Policy, 2012, 39(6): 751-760.
- [19] STILGOE J, OWEN R, MACNAGHTEN P. Developing a framework for responsible innovation[J]. Research Policy, 2013, 42(9): 1568-1580.
- [20] OWEN R, BESSANT J, HEINTZ M. Responsible Innovation: Managing The Responsible Emergence of Science and Innovation in Society [M]. John Wiley & Sons, 2013.
- [21] 梅亮, 陈劲. 创新范式转移——责任式创新的研究兴起[J]. 科学与管理, 2014, 34(3): 3-11.
- [22] 梅亮, 陈劲, 盛伟忠. 责任式创新——研究与创新的新兴范式[J]. 自然辩证法研究, 2014, 30(10): 83-89.
- [23] STAHL B C. Responsible research and innovation; the role of privacy in an emerging framework[J]. Science and Public Policy, 2013, 40(6): 708-716.
- [24] WYNNE B. Public uptake of science; a case for institutional reflexivity[J]. Public Understanding of Science, 1993, 2(4): 321-337.
- [25] SUTCLIFFE H. A report on responsible research and innovation [R]. MATTER and the European Commission, 2011.
- [26] VAN DEN HOVEN J. Value sensitive design and responsible innovation [M]//OWEN R, BESSANT J, HEINTZ M. Responsible innovation: managing the responsible emergence of science and innovation in society. Wiley, 2013: 75-83.
- [27] VAN DEN HOVEN J. Options for Strengthening Responsible Research and Innovation; Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation[M]. Publications Office of the European Union, 2013.
- [28] WEBER J. Making worlds: epistemological, ontological and political dimensions of technoscience[J]. Poiesis & Praxis, 2010, 7(1/2): 17-36.
- [29] ZEIDE J S. A critical review of legal issues in artificial intelligence[M]//DE SALVP D A, LIEBOWITZ J. Managing artificial intelligence and expert system. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1989: 122-137.
- [30] MCCARTHY J. Generality in artificial intelligence[J]. Communications of the ACM, 1987, 30(12): 1030-1035.
- [31] RUSSEL S, NORVIG P. Artificial Intelligence: A Modern Approach[M]. Prentice Hall, 2003.
- [32] HAUGELAND J. Artificial Intelligence The Very Idea [M]. A Bradford Book, 1985.
- [33] MCDERMOTT D, CHARNIAK E. Introduction to Artificial Intelligence[M]. Reading: Addison-Wesley, 1985.
- [34] RICH E, KNIGHT K. Artificial Intelligence [M]. McGraw-Hill, 1991.
- [35] POOLE D, MACKWORTH A, GOEBEL R. Computational Intelligence; A Logical Approach[M]. Oxford University Press, 1998.
- [36] DIX A. Human-computer interaction [M]//LIU Ling, ÖZSU M T. Encyclopedia of database systems. Springer, 2009: 1327-1331.
- [37] MUSK E. A. I. -Human Symbiote" Will Prevent A. I. Domination[EB/OL]. [2016-09-19]. <https://www.inverse.com/article/21157-elon-musk-ai-human-symbiote-neural-lace>.
- [38] VINGE V. The coming technological singularity; how to survive in the post-human era[C]//Vision 21: interdisciplinary science and engineering in the era of cyberspace. NASA Conference Publication, 1993: 11-22.
- [39] 腾讯研究院. 2017 中美人工智能创投现状与趋势研究报告[EB/OL]. [2017-08-30]. http://www.sohu.com/a/168236578_635113.
- [40] MURPHY R, WOODS D D. Beyond Asimov: the three laws of responsible robotics[J]. IEEE Intelligent Systems, 2009, 24(4): 14-20.
- [41] BODEN M, BRYSON J, CALDWELL D, et al. Principles of robotics: regulating robots in the real world[J]. Connection Science, 2017, 29(2): 124-129.
- [42] 涂序彦, 韩力群, 王洪泊. 广义人工生命[M]. 北京: 科学出版社, 2011.
- [43] ASHRAFIAN H. Artificial intelligence and robot responsibilities; innovating beyond rights[J]. Science and Engineering Ethics, 2015, 21(2): 317-326.
- [44] 段伟文. 人工智能的道德代码与伦理嵌入[N]. 光明日报, 2017-09-04.
- [45] 梅亮, 陈劲, 李福嘉. 责任式创新: 一个“内涵-理论-方法”的整合框架研究[J]. 科学学研究, 2018, forthcoming.
- [46] RODRIGUEZ H, FISHER E, SCHUURBIERS D. Integrating science and society in European framework programmes; trends in project-level solicitations [J]. Research Policy, 2013, 42(5): 1126-1137.
- [47] 杨婕. 全球人工智能发展的趋势及挑战[J]. 世界电信, 2017(2): 15-19.
- [48] 贾开. 人工智能时代: 发展逻辑与治理挑战[N]. 学习时报, 2017-08-23.
- [49] 涂序彦. 智能机器人, 情感机器人, 拟人机器人[J]. 华中科技大学学报: 自然科学版, 2004(S1): 1-4.
- [50] LEVY D. Love and Sex With Robots: The Evolution of Human-Robot Relationships[M]. Harper Collins, 2009.
- [51] 梅亮, 陈劲. 负责任创新: 时域视角的概念、框架与政策启示[J]. 科学学与科学技术管理, 2016, 37(5): 17-23.
- [52] GEORGE G, HOWARD-GRENVILLE J, JOSHI A, et al. Understanding and tackling societal grand challenges through management research[J]. Academy of Management Journal, 2016, 59(6): 1880-1895.
- [53] BENSUADE-VINCENT B. The politics of buzzwords at the interface of technoscience, market and society: the case of ‘public engagement in science’[J]. Public Understanding of Science, 2014, 23(3): 238-253.

(下转第 43 页)

- namics[J]. *The Review of Economic Studies*, 1982, 49 (3):315-333.
- [77] CASAMATTA C, GUEMBEL A. Managerial legacies, entrenchment, and strategic inertia[J]. *Journal of Finance*, 2010, 65(6):2403-2436.
- [78] CHEMMANUR T J, CHENG Y, ZHANG T. Human capital, capital structure, and employee pay: an empirical analysis[J]. *Journal of Financial Economics*, 2013, 110 (2):478-502.
- [79] CHEN G. Initial compensation of new CEOs hired in turnaround situations[J]. *Strategic Management Journal*, 2015, 36(12):1895-1917.
- [80] GERAKOS J J, PIOTROSKI J D, SRINIVASAN S. Which U. S. Market interactions affect CEO pay? Evidence from UK companies[J]. *Management Science*, 2013, 59(11):2413-2434.
- [81] KRAUSE R, WHITLER K A, SEMADENI M. Power to the principals! An experimental look at shareholder say-on-pay voting[J]. *Academy of Management Journal*, 2014, 57(1):94-115.
- [82] WERNER S, TOSI H L, GOMEZ-MEJIA L. Organizational governance and employee pay: How ownership structure affects the firm's compensation strategy[J]. *Strategic Management Journal*, 2005, 26(4):377-384.
- [83] WOO-JIN C, HAYES R M, HILLEGEIST S A. Financial distress risk and new CEO compensation[J]. *Management Science*, 2016, 62(2):479-501.
- [84] ALMAZAN A, SUAREZ J. Entrenchment and severance pay in optimal governance structures[J]. *Journal of Finance*, 2003, 58(2):519-548.

Research Progress on Determinants of CEO Compensation

Huang Jie¹, Chen Guoquan²

(1. Department of Economics and Management, European Business School, Oestrich-Winkel 65375, Germany;

2. School of Economics and Management, Tsinghua University, Beijing 100084, China)

Abstract: This paper reviews 65 articles published in the top international journals from 2002 to 2017, and comprehensively categorizes these articles by the unit of analysis, the elements of analysis and the method of analysis respectively. The results show as follows: the studies in the field of finance and accounting seem to be saturated and approaching a steady status, while the studies in the field of management and sociology is booming; the most considered elements of CEO compensation are from the constructs of corporate governance and managerial labor market; qualitative research method and simulation method may be applied more in the future research in this field.

Keywords: CEO compensation; corporate governance; agency theory; managerial power theory

(上接第 7 页)

Innovation Governance of Emerging Technology from Responsible Innovation Perspective: Lessons from Artificial Intelligence

Mei Liang^{1,2}, Chen Jin^{1,2}, Wu Xintong^{2,3}

(1. School of Economics and Management, Tsinghua University, Beijing 100084, China; 2. Research Centre for Technological Innovation, Tsinghua University, Beijing 100084, China; 3. Southwest Petroleum University, Xichuan 610500, China)

Abstract: This paper focuses on the governance of controversial technological innovation "artificial intelligence" from the perspective of responsible innovation, and constructs the emerging technology's responsible innovation governance framework composed of technological, economic, ethical and social dimensions. Based on this framework, it systematically reviews the innovation duality of artificial intelligence in both theories and practices, and provides the implications for the development sustainability of artificial intelligence innovation.

Keywords: responsible innovation; emerging technology; innovation governance; artificial intelligence