

# 人工智能发展对伦理的挑战及其治理对策

杨博文 伊彤 江光华

(北京科学学研究中心, 北京, 100089)

**摘要:** 人工智能、基因重组、干细胞移植等新兴技术逐渐被应用到经济社会的发展中。然而, 新技术如同“双刃剑”, 可能引发道德伦理等问题, 因此需要加强对新兴技术伦理治理的研究。文章聚焦人工智能, 分析新兴技术发展可能带来的伦理挑战及伦理治理存在的问题, 结合当前国内外人工智能伦理治理实践, 提出中国人工智能伦理治理的对策建议, 以期促进人工智能在发展中规范、在规范中发展。

**关键词:** 人工智能; 伦理挑战; 伦理治理

**中图分类号:** B82-057; TP18

**文献标识码:** A

**DOI:** 10.19881/j.cnki.1006-3676.2021.01.07

## The Ethical Governance and Challenges of Artificial Intelligence Development

Yang Bowen Yi Tong Jiang Guanghua

(Beijing Research Center for Science of Science, Beijing, 100089)

**Abstract:** Artificial intelligence, gene recombination, stem cell transplantation and other emerging technologies are gradually applied to the development of economy and society. However, new technologies are like “double-edged sword”, which may lead to moral and ethical problems. Therefore, it is necessary to strengthen the research on ethical governance of emerging technologies. Taking artificial intelligence as an example, this paper focuses on artificial intelligence, analyzing the ethical challenges and problems that may be brought by the development of emerging technologies, and puts forward suggestions for ethical governance of artificial intelligence in China based on the current practice of ethical governance of artificial intelligence at home and abroad, so as to promote the standardization and development of artificial intelligence in the process of

项目基金: 本文系科技部“面向2021—2035年国家中长期科技发展规划面向社会征集研究单位开展重大问题研究”项目“面向2035年的科技伦理问题与应对措施研究”研究成果。

作者简介: 杨博文, 女, 1983年生, 助理研究员, 研究方向: 科技政策与科技管理。伊彤, 女, 1967年生, 研究员, 研究方向: 科技创新战略、科技政策。江光华, 女, 1973年生, 副研究员, 研究方向: 科技文化、科技政策。

development.

**Key words:** Artificial intelligence; Ethical challenges; Ethical governance

现代科技发展具备的不确定性、高风险性等特点决定了科技伦理问题的产生具有必然性。如果把科学技术视为一辆正在行驶的列车，科技伦理就是指引导列车前行，用以保证科技研发与应用的方向正确的路标。对新兴技术伦理的研究，一方面要对新技术的发展应用予以规范，另一方面又要保护和促进“负责任的研究”。笔者以人工智能技术的发展为例，研究新兴技术伦理治理的对策。

## 一、人工智能发展带来的伦理挑战

人工智能是引领新一轮科技革命和产业变革的战略性技术，不仅能替代人类的体力劳动，更几乎可以完全替代人类的脑力劳动<sup>[1]</sup>。人工智能更像是智慧载体，它可以通过深度学习不断进化，它的颠覆性、复杂性以及社会关联性等特性所引发的伦理挑战和风险日趋凸显，如何应对人工智能等新技术引发的伦理挑战已经是相关研究人员所要面对的重要课题。

### （一）存在性风险对人类主体地位的挑战

随着信息技术、生物技术的发展，人机互动、人机协同甚至人机一体都成为人工智能发展的趋势。尽管现阶段对人工智能技术的开发、测试、应用等都还在一个基本可控的伦理框架内，但随着人工智能技术的能进一步发展，当具备无自主意识的人工智能做出超出科研人员预期的行为并产生相应的伦理影响时，现有的伦理框架及人类的思维本质将因此面临极大的挑战。如果技术的发展推动智能机器人越来越像人类，这将会对目前社会的人际关系、家庭结构产生重大影响，甚至颠覆人类在人类活动中的主体地位。机器人是否应享有人类的基本权利？是否应建立新的人机关系甚至人际关系的价值原则？人类的唯一主体地位或将因此受到挑战<sup>[2]</sup>。随着人工智能的强智能化发展，人类有可能逐渐难以参与智能生产的过程，甚至最终成为智能机器的零件，人与机器的关系可能发生扭转，

人工智能由此引发对人类传统伦理道德判断的颠覆，人类将第一次真正面对存在性风险。

### （二）数据泄露和信息泛滥带来的挑战

人工智能技术的基础是基于海量数据的算法设计和深度学习，而数据的采集和使用若管理不到位会导致数据安全问题。当前互联网应用软件采集的个人信息维度多样，社交通信、移动支付、网络购物、卫星导航、智能手环等智能工具采集了用户多方面的数据，将一个人的喜好、行踪、受教育程度、职业、财富、健康等多种信息整合归纳，多维度的数据轨迹足以以为个人画像。这些信息如果被非法利用就会构成对他人隐私权的侵犯。2019年8月，一款名为“ZA0”的AI换脸社交软件在年轻人中流行开来，用户只需提供一张正面照就能将其他视频中人物的脸换为自己的脸。但该软件的用户隐私协议没有对用户阐明数据泄露的风险，甚至认为该软件对使用者的肖像权具有“全球范围内免费、不可撤、永久可转授权”，而如果侵权了明星肖像，最后责任都在用户。此外，当前的监控设备也因其便捷性被广泛使用，商场、医院、学校甚至快递收取点都有监控系统。这些设施在便利工作生活、辅助安全保障的同时，收集并存储的信息并没有被合理监管，导致公民的个人隐私也存在隐患。

### （三）权利与责任认定问题带来的挑战

人类对人工智能的担忧之一，就是它的高度智能化系统如果被不正当利用或者失控可能会对人类社会产生威胁，随之而来的必定是技术研究与应用引发的责任归属问题。目前已经有一些医院采用算法读取患者影像，凡是机器判断为阳性的片子，医生再通过人工诊断进行复核，诊断出来是阴性的，医生不再检查。一旦因算法出错导致误诊，特别是出现假阴性误诊，谁应该签字承担责任？再如，随着谷歌、特斯拉和Uber等智能无人驾驶致人死亡事件的相继发生，关于机器人谁应当对此负责的伦理争论愈演愈烈。麻省理工学院进行的“道

德机器”测试曾提出无人驾驶的伦理困境：通过模拟行驶中不常见的一些情景，大量采集公众选择的数据样本，“训练”一辆无人驾驶汽车的人工智能系统对这些情景作出判断。这些判断是基于公众的意见，简单来看是由公众对这些情景进行了判断，这些判断没有正确答案，但形成了无人驾驶汽车行动的依据，而由此产生的责任划分则难以明确。因此，如果事先没有适当的权责划分，容易导致人工智能产品脱离社会伦理、法律规范，其可能扰乱经济社会秩序，造成更大危害。

#### （四）人工智能的岗位替代带来的挑战

人工智能以其智能化、高效化，成为提高工作效率、提升经济效益的重要方式，实现对产业结构和社会分工模式的重构，有可能造成大量人员的结构性失业，从而影响社会公平。当前，许多领域的工作岗位已经逐渐尝试应用人工智能技术。例如，制造大厂富士康在多个工厂和多条生产线已经实现了自动化，实现了24小时的无人工厂，仅在昆山工厂就减员6万人。麻省理工学院发布研究报告称，过去20年，美国有36万—67万工作岗位被机器人夺走，未来10年，还将有350万个岗位被人工智能替代<sup>[3]</sup>。虽然这并非人类开发者刻意为之，但是人工智能技术的应用对人类就业和财富分配观念带来的影响，对社会公平可能带来的损害，都引发了潜在的伦理风险。

## 二、中国人工智能伦理治理的实践探索

为促进人工智能的健康发展，中国一直以来都很重视对人工智能的伦理、法律、社会影响等问题的研究，积极研究制定相关政策法规，以全球化的视角推动人工智能治理。但随着人工智能技术的快速应用与发展，与之匹配的伦理治理体系还未建立，相关政策还需不断完善，科研人员的伦理自觉还有待提升，社会影响的应对方式还需探索。

#### （一）主要进展

##### 1. 建立人工智能治理委员会

为推进人工智能发展，保持中国在人工智能等新技术革命中的后发优势，从2015年开始，国家通过顶层设计，有力地保障了人工智能的快速发展。2019年，中国成立了由科学技术部、国家发展和改革委员会、工业和信息化部、中国科学院、中国工程院、中国共

产党中央军事委员会科学技术委员会、中国科学技术协会等15个部门构成的新一代人工智能发展规划推进办公室，随后成立了国家新一代人工智能治理专业委员会，加强了对人工智能领域科技伦理的监督。

##### 2. 发布相关治理原则和规范

《国务院关于印发新一代人工智能发展规划的通知》中提出，要形成我国人工智能法律法规、伦理规范和政策体系。国家标准化管理委员会发布《人工智能标准化白皮书（2018版）》，从国家层面明确了中国人工智能为人类服务原则和责任原则<sup>[4]</sup>。2019年6月，新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》。由于人工智能技术正处于快速发展期，过早地从法制角度进行强制性约束可能会影响技术的发展和运用<sup>[5]</sup>，因此这些政策仅是原则性意见或通知，尚无具体管理制度。中国人工智能学会伦理道德专业委员会正在计划针对不同行业的人工智能设立和制定种种伦理规范，如智能驾驶规范、数据伦理规范、智慧医疗伦理规范、智能制造规范、助老机器人规范等<sup>[6]</sup>。

针对一些具体技术应用领域，国家也出台了相关管理规范。2019年11月底，国家互联网信息办公室、文化和旅游部、国家广播电视总局联合印发了《网络音视频信息服务管理规定》，明确提出“利用基于深度学习、虚拟现实等的虚假图像、音视频生成技术制作、发布、传播谣言的，应当及时采取相应的辟谣措施，并将相关信息报网信、文化和旅游、广播电视等部门备案”等规定，加强对数据隐私的管理。2018年4月，工业和信息化部、公安部和交通运输部等三部门发布了《智能网联汽车道路测试管理规范（试行）》，对测试主体、测试驾驶人、交通违法和事故处理等内容进行了明确规定，各省、市级政府相关主管部门可以组织开展智能网联汽车道路测试工作。

##### 3. 推动人工智能科研界建立共识

2019年5月25日，北京智源人工智能研究院正式成立了人工智能伦理与安全研究中心。北京智源人工智能研究院与北京大学、清华大学、中国科学院自动化研究所、中国科学院计算技术研究所等单位共同发布了《人工智能北京共识》。该文件提出，在人工智能的研发、使用、治理等方面应遵循有益于人类命运

共同体构建和社会发展的15条原则。2020年9月14日，北京智源人工智能研究院发布了中国首个针对儿童的人工智能发展原则《面向儿童的人工智能北京共识》，作为《人工智能北京共识》针对儿童群体的实施细则，提出包括“以儿童为中心”“保护儿童权利”“承担责任”“多方治理”4大主题的19条细化原则。

## （二）存在的问题

### 1. 人工智能伦理治理体系亟待建立

在人工智能领域，技术飞速发展，但与其相匹配的管理、运行、保障、监督机制仍显滞后和发展缓慢，政府对人工智能伦理监管的顶层设计尚处于起步阶段。与人工智能类似的大数据领域已基本完成制度建设，在全国形成了协同共治的大数据管理机制，各省、自治区、直辖市都设置了由中央统一领导的大数据政务服务管理局。而在人工智能领域，目前国家新一代人工智能治理专业委员会正逐步探索构建治理体系，政府对于技术伦理的治理仍处于探索阶段，治理组织体系、监管机制、治理方案等都还不够成熟。

### 2. 相关法律法规不健全

人工智能相关技术发展及迭代迅速，政府难以明确监管对象和监管范围。且由于技术应用分布广泛，技术发展、产业发展、社会发展等各类政策都可能涉及伦理问题。目前，相应的法律、道德规范尚未出台，这导致伦理道德风险防范在人工智能一环的体制缺失<sup>[7]</sup>。现有涉及人工智能的政策法规主要以原则性共识为主，随着新技术的应用，中国政府逐步在无人驾驶、数据隐私等方面制定了相应规定以指导相关技术发展，但关于人工智能领域伦理治理的系统性政策法规体系尚未建立。

### 3. 科研人员的科技伦理自觉有待提升

近年曝出的科技伦理热点事件反映出中国存在科研人员在科研工作中不太重视科技伦理，对伦理知识掌握不足，缺少对新技术开发应用可能产生伦理问题的实质性探讨，不能自觉用伦理道德规范科研工作等问题。据中国科学技术协会2018年发布的一项有关科技工作者科技伦理的调查报告显示，自认为对科研诚信之外的科技伦理规范了解较多的科技工作者仅占总体的5%。以人工智能技术的发展为例，如果缺乏对技术伦理的重视，缺少有效措施的推进，不能让伦理成为保障人工智能有

序发展的规则，则有可能引发人工智能广泛应用带来的负面效应，发生严重的社会伦理事件。

### 4. 缺乏有效的社会公众沟通机制

随着新技术的研发开始触及人类自身能力的改变，并且影响人类的社会生活，公众的社会伦理意识正在觉醒。公众出于朴素的社会伦理道德观念会对新兴技术产生怀疑和恐惧，甚至反对新技术的应用。2020年9月，在加拿大安大略省北部的人行横道上，波士顿动力公司生产的机器狗在半夜肆无忌惮地乱跑，路人纷纷拍下视频上传到Twitter之后，众多网友表示很担心这种机器人会被用作不良用途。任何一项科学技术研究都不能独立于社会监督。2020年6月，IBM出于对种族歧视、隐私安全等敏感社会伦理问题的考虑，宣布放弃人脸识别相关技术与产品，同时IBM也提出应开展全国性对话，讨论美国执法部门是否应该和如何部署人脸识别技术。

## 三、人工智能伦理治理的国际经验

人工智能正在深刻改变人类社会生活、改变世界，世界各国都非常重视对相关技术的伦理研究，探索科技伦理治理方式。欧洲各国重视科技伦理对技术开发应用的规范，积极发布相关伦理道德准则；美国主要以企业自律自治为主，政府为相关研究机构提供了较为宽松的环境。

### （一）推动形成全球化的人工智能伦理准则

人工智能可能带来的伦理挑战是全球性的，是各国共同面临的问题，世界各国都需要负起责任。2019年5月，OECD成员国批准了人工智能原则即《负责任地管理可信的AI的原则》，提出了包容性增长、可持续发展和福祉、以人为本的价值和公平等伦理原则<sup>[8]</sup>。在此基础上，2019年6月，G20通过了联合声明以及《G20人工智能原则》，这是第一个由各国政府签署的人工智能原则，各国政府由此形成了基本共识，即研发或使用人工智能“应该尊重法律、人权和民主价值观”。

### （二）制定法律法规引导和规范技术发展

各国通过研究制定与人工智能治理相关的法律法规，逐步对技术伦理进行规范引导。例如，2016年美国在《为人工智能的未来做好准备》中提出对人工智能技术的监管原则和监管重点等内容，强调推动技术



创新、保持国家核心竞争力是发展重点。英国一直重视科技伦理治理,在《人工智能:未来决策制定的机遇与影响》中,强调人工智能发展的法律和伦理风险,政府强调公民权利的保护,坚持在技术无害的前提下开发和应用,并制定了企业问责机制<sup>[9]</sup>;2018年,发布《数字宪章》,提出要尊重并妥善使用个人数据,特别要保护儿童的线上安全、线下权利。欧盟2019年发布的《人工智能道德准则》指出应发展“可信AI”,提出了人类监督、隐私和数据管理等具体政策,可操作性较强。日本AI学会的伦理委员会在2016年发布人工智能研究人员应该遵守的伦理指标草案,引导和规范科研人员正确处理人工智能进一步发展导致的伦理道德、安全问题<sup>[10]</sup>。

### (三) 通过建立共识机制,加强科研界自律

美国在研究制定人工智能相关法案的同时,主要通过科研界自身的共识机制来加强对人工智能科技伦理的监管。美国智库未来生命研究院在2017年召开的人工智能专题会议上提出了“阿西洛马人工智能原则”,呼吁全世界发展人工智能时要确保AI有益于人类社会的安全发展。电气电子工程师协会(IEEE)2017年发布的《人工智能设计的伦理准则(第二版)》,对价值嵌入、研究和设计等伦理事项提出了具体的建议,如提出当前技术发展还不足以赋予人工智能法律地位。这些原则的提出是人工智能领域科研人员达成的共识,通过加强技术伦理的规范为后续标准制定等工作提供参考。

### (四) 实行企业自我约束的治理方式

美国人工智能技术全球领先,但是对人工智能相关的伦理治理谨慎立法和监管,主要通过企业自我约束,探索对人工智能伦理挑战的监管。IBM、微软、谷歌、亚马逊等高科技公司发出倡议,提出确保“负责任地使用AI”。这些具有代表性的大公司还发起成立各类组织,研究人工智能发展应用的治理手段。例如,由亚马逊、微软、脸书、谷歌DeepMind以及IBM联合成立非营利组织人工智能合作组织(The Partnership on AI,简称PAI),推进公众对人工智能技术的理解;领英联合创始人Reid Hoffman与Omidyar Network联合成立人工智能伦理与治理基金,研究人工智能治理方法<sup>[10]</sup>;谷歌公司成立“全球技术顾问委员会”,请外部专家监督公司在应用人工智能等新技术时遵循伦

理准则的依从性。

## 四、对策建议

中国在人工智能领域已进入领跑者行列,要发展负责任的人工智能,中国政府应发挥更多的引导作用,在多个方面形成人工智能伦理治理机制,必须系统构建人工智能伦理治理体系。加强对人工智能社会影响、伦理问题等的研究,促进多主体协同互动,参与全球合作的人工智能伦理治理。

### (一) 构建人工智能伦理治理体系

构建以国家人工智能伦理治理委员会为核心的层次分明、规范有序、协调一致、公开透明的人工智能伦理治理体系。组建国家人工智能伦理治理委员会,委员会由人工智能技术专家、社会伦理学专家、法律专家、产业界人士、政府管理人员等共同组成,研究形成人工智能伦理挑战的治理方案,明确各级伦理治理委员会的职责。在国家新一代人工智能创新发展试验区建立区域性人工智能伦理治理委员会,研究人工智能伦理风险,探索伦理治理的新对策。国家及试验区所在地方政府设立人工智能伦理治理监督机构,通过公开透明的规章、政策、标准和法律制定、审理与批准、监测及信息报告等程序,依法依规对伦理治理委员会的伦理风险评估、伦理审查等工作进行检查管理。

### (二) 健全人工智能伦理治理政策法规

建立健全科技伦理治理政策法规,引导和规范人工智能技术的研发与应用。一是加强伦理规范制定。建立人工智能技术应用安全标准,建立数据收集、利用与保护的基本规则秩序,防范并打击数据滥用行为,打击人工智能相关的新型犯罪,研究探索推动人工智能健康应用的规则制度。二是将科技伦理原则融入科技政策。将以人为本、公平正义、和谐协调、前瞻性责任等科技伦理原则反映到科技政策中,正确处理好具体政策之间的短期利益和长远利益、局部利益和整体利益的价值冲突等伦理问题。三是逐步推进人工智能立法,在智能机器人、智能设备、智能运载产品等应用广泛的领域,率先启动立法工作。开展与人工智能应用相关的法律问题研究,探索追溯机制和问责制度的建立,特别是研究人工智能应用涉及的法律主体及相关权利、义务和责任等的界定。对自动驾驶、服

务类机器人、工业机器人等已经逐渐试验应用的人工智能技术，要加快研究相关安全法规，为新技术的发展应用提供法律依据和保障。

### （三）完善人工智能伦理自律机制

政府主管部门与企业、学术团体等共同研究制定人工智能的伦理准则，鼓励和支持通过行业自律，推动伦理审查、自律组织建立、行业标准制定等的发展。支持高校、科研院所、大型科技企业等建立起内部伦理审查机制，强化科技伦理管理主体责任。提高各学术团体的监督意识，推动自律规范深入科研人员内心、内化到科研活动中。在高等学校开设科技伦理课程，培养大学生遵循科技伦理、严守学术规范的良好习惯。针对企业关于人工智能、大数据等技术潜在风险认识不足的问题，发挥政府的引导、监管、评价等作用，对企业开发应用人工智能产品与服务的安全进行规范。鼓励企业设立伦理委员会，加强对人工智能科研人员的伦理责任意识培养。在项目选题时要对技术伦理风险进行评估，淘汰或暂停可能损害大多数公众利益、危及社会基本伦理规范的科研项目；在研发过程中要关注该项目的社会影响，如发现可能挑战伦理秩序的要及时中断；在项目完成交付使用前，应当对该研究成果开展广泛论证和充分评估，并将新技术应用可能带来的负面效应向社会公众进行充分说明，为政府制定相关决策提供建议。

### （四）建立人工智能伦理多方协商机制

由政府主管部门牵头，组织跨学科领域的专家、企业代表、用户和社会公众等利益相关者，建立多主体共同参与的协商机制，对科技伦理进行研讨并达成共识。一是参与伦理研究和顶层设计，明确行业伦理原则，兼顾行业创新发展。二是建立健全国家科技计划伦理管理制度，对于即将立项的科技计划项目，围绕其“该不该干”的实质性伦理问题进行深入探讨，各利益相关者通过公共协商来塑造科技创新的路径和方向。三是对于前沿、复杂技术可能引发的伦理问题，围绕判定科研项目能否开展的标准问题（如人脸识别的伦理标准），推动社会多方进行协商、交流、合作，把对不同利益相关者的考量纳入创新过程，使科技创新承担起应尽的社会责任。

### （五）参与全球人工智能伦理治理

主动参与国际人工智能伦理议题的讨论，在战略层面加以布局，树立中国人工智能技术伦理规范健康发展的国际形象。中国的人工智能技术全球领先，在科技安全风险管控、治理等重大国际共性问题上要积极发声，提出中国方案，加强国际交流，凝聚全球共识。同时，要积极参与全球人工智能伦理标准和规则的制定，推动建立全球共同的人工智能开发与应用的伦理框架，确保人工智能相关技术能更好地促进全人类和全世界的发展。

## 参考文献：

- [1] 何哲. 人工智能技术的社会风险与治理 [J]. 电子政务, 2020 (09) : 2-14.
- [2] 李伦. 人工智能与大数据伦理 [M]. 北京: 科学出版社, 2019.
- [3] 李增军, 李梦阳. 人工智能的若干伦理问题 [J]. 中国发展观察, 2020 (01) : 71-73.
- [4] 陶海燕, 周吉银. 浅议中国人工智能伦理治理的对策 [J]. 中国医学伦理学, 2020, 33 (06) : 703-707.
- [5] 杜严勇. 加强人工智能技术的伦理治理研究 [N]. 社会科学报, 2020-01-09 (05) .
- [6] 倪伟波. 多维共建“负责任”的科技 [J]. 科学新闻, 2019 (04) : 15-19.
- [7] 王晓丽, 徐鑫钰. 人工智能的道德风险及其治理 [J]. 华南理工大学学报 (社会科学版), 2020, 22 (02) : 10-17.
- [8] 曹建峰, 方龄曼. 欧盟人工智能伦理与治理的路径及启示 [J]. 人工智能, 2019 (04) : 39-47.
- [9] 王勉, 黄颖. 人工智能技术伦理治理现状及趋势 [J]. 电子商务, 2019 (07) : 6-9.
- [10] 李修全, 蒋鸿玲. 美日欧政府发展人工智能的新举措及对我国的启示 [J]. 全球科技经济瞭望, 2017 (10) : 34-36.