Advances in Philosophy 哲学进展, 2023, 12(5), 890-897
Published Online May 2023 in Hans. https://www.hanspub.org/journal/acpp
https://doi.org/10.12677/acpp.2023.125153

Hans汉斯

强人工智能的伦理困境探究与伦理治理途径

张舒婷

河北师范大学马克思主义学院，河北 石家庄

收稿日期：2023年4月25日；录用日期：2023年5月15日；发布日期：2023年5月29日

摘　要

随着第四次科技革命的到来，机器人向智能化发展。在未来，强人工智能在发展的同时还面临着诸多伦理困境，诸如，强人工智能的发展伦理困境，包括崇拜人工智能导致人的存在异化、人类自身的片面性发展、发展人工智能技术的不确定性等；强人工智能的责任伦理困境与道德主体地位问题；强人工智能与人类产生的情感伦理困境等。面对强人工智能时代的伦理困境，其解决路径有：一、开展中国传统“以道驭术”的伦理思想教育；二、确立强人工智能“以人为本”的伦理原则；三、强人工智能时代的多元协同伦理治理。通过这些伦理治理进路，将发展出一个更加完善的强人工智能时代，为人机共存的美好前景做好铺垫。

关键词

强人工智能，伦理困境，伦理治理

Research on Ethical Dilemmas and Ethical Governance Approaches of Strong Artificial Intelligence

Shuting Zhang

School of Marxism, Hebei Normal University, Shijiazhuang Hebei

Received: Apr. 25th, 2023; accepted: May 15th, 2023; published: May 29th, 2023

Abstract

With the arrival of the fourth revolution of science and technology, robots are moving towards the intelligence. At the same time, strong artificial intelligence faces large numbers of ethical dilemmas. Such as strong artificial intelligence in the development of ethical dilemmas, including wor-

文章引用: 张舒婷. 强人工智能的伦理困境探究与伦理治理途径[J]. 哲学进展, 2023, 12(5): 890-897.
DOI: 10.12677/acpp.2023.125153

ship of artificial intelligence leads to human existence alienation, human beings' one-sidedness, developing artificial intelligence technology is uncertain, etc.; The ethical dilemma of strong artificial intelligence's responsibility and the status of moral subject; Strong artificial intelligence and the ethical dilemma of human emotion. In the face of ethical dilemmas in the era of strong artificial intelligence, the solutions are as follows: first, to carry out the traditional Chinese ethical thought education of "Daoism"; Second, establish the ethical principle of "people-oriented" for strong artificial intelligence; Third, multiple collaborative ethical governances in the era of strong artificial intelligence. Through these ethical governance approaches, a more perfect era of strong artificial intelligence will be developed, paving the way for a better prospect of man-machine coexistence.

Keywords

Strong Artificial Intelligence, Ethical Dilemma, Ethical Governance

1. 人工智能的历史发展趋势

1956 年夏天，麦卡锡、明斯基等科学家在美国达特茅斯学院开会研讨“如何用机器模拟人的智能”，首次提出“人工智能”(Artificial Intelligence，简称 AI)这一概念，标志着人工智能学科的诞生。

迄今为止，人类对于人工智能的研究还处于“弱人工智能时代”。“所谓弱人工智能，是指那些无法用人的思想推理、处理问题的智能机器，它们不具有像人一样的思考方式，只是在机械地、重复地执行命令，并不拥有独立自主的学习意识，不过是看上去智能罢了。”[1]而一切和人工智能相关的产品都被称为“人工智能+”(AI+)，比如人工智能 + 化学、人工智能 + 医疗、人工智能 + 物流、人工智能 + 教育等。

未来人工智能还会发展到“通用人工智能”(AGI)阶段。通用人工智能就是指，“可完成认知任务，并且完成得至少和人类一样好的能力。”[2]换而言之，就是和人类一样有知觉、有自我意识，可以像人一样进行独立思考、分析、推理和解决问题的人工智能。而强人工智能(Strong AI)与通用人工智能的能力并无不同，所以二者经常混用。强人工智能可以被分为两大类：“类人的人工智能和非类人的人工智能。类人的人工智能，即机器的思考方式和思维水平同人是一样的；而非类人的人工智能，即机器产生了与人类大有出入的感知和认知水平，运用和人不一样的推理方法解决问题。”[1]

获得公民身份的索菲亚机器人就拥有和人类女性相似的外表，虽然她的表情还略显僵硬，但是不可否认的是，随着人类科技的发展，使机器人拥有和人类完全相似的动作和表情并不是不可能的；阿尔法围棋(Alpha Go)作为第一个击败人类职业围棋选手、第一个战胜围棋世界冠军的人工智能，就比较符合强人工智能非类人的特征。虽然这些人工智能还未完全达到强人工智能时代的要求，但是还是可以从中管窥到强人工智能时代的科技生活。

人类进步的脚步更加不会止步于强人工智能时代，接下来还会发展未知的赛博格(Cyborg)，这一时代的特点就是人机共存，即人与机器的混合体，不论是身体还是思想，都可以达到共享的程度。这一时代只能在科幻小说和科幻电影中被看到，比如在小说《机器人启示录》当中，威尔森就描述了一个人类肉体的一部分和机器零件相互融合的场景，甚至可以达到人类与机器人用意念沟通的效果[1]。

与此同时，人类社会的伦理规范、伦理思想并不能完全适应人工智能的快速发展，甚至还有些滞后。强人工智能时代的伦理困境在21世纪已经初见端倪了，为避免人工智能的伦理困境发展成为不可避免的后果，必须要早做预防。

2. 强人工智能时代的伦理困境

2.1. 强人工智能的发展伦理困境

1) 崇拜人工智能导致人的存在异化

随着科学技术的发展，人工智能已经逐渐融入人们的生产生活，但是随之而来的就是人类自身的异化。

首先是人的自然属性的异化。自然属性中又分为人的肉体性异化和精神性异化。在强人工智能时代，“类人机器人”将会普遍出现在人们的生活中，甚至人机结合也不会只出现在科幻电影中，人类机器化和机器人类化既冲击了人类的主体性地位，又使人们对自己的存在产生了质疑，即“我”到底是人还是机器？从而陷入“我是谁？”的窠臼中。

其次是人的社会属性异化。人的社会性就在于人是依靠社会关系存在的，但是如果强人工智能代替了人类大部分的社会关系，那么人与人之间的联系将会弱化，新的社会关系将会是人－机－人的人际关系，那时人们之间的联系与沟通就会是虚拟的，方便人类的同时也“将人际交往带入了虚拟沉迷、丧失自我、躲避责任、远离现实等异化窠臼中。”[3]

在人工智能智能化、信息化的趋势下，将会有大部分人类面临失业危机，尤其是普通体力劳动者，因为人工智能不仅拥有超越人类智力和体力，还可以代替人类从事一些机械、重复、3D (指Dirty, Difficult, Dangerous——肮脏、辛苦、危险)的工作，随着强人工智能时代的到来，一些“脑力”工作也会被人工智能替代，如画家、律师、诗人、医生等等需要大量学习的工作岗位，可以通过大数据处理方法被人工智能取代。就像2023年初火爆全网的ChatGPT，这个由人工智能技术驱动的聊天机器人，既可以回答许多专业问题，又可以和人类戏谑地聊天。这款聊天软件一经入市就被人类用来撰写学术论文，让多家学术期刊发表声明，完全禁止或严格限制使用ChatGPT等人工智能机器人撰写学术论文。从现实来看，这样的声明只能“防君子”，不能防“小人”，人们的学术道德又一次遭遇了挑战。

但是，本来人工智能的发展和应用是为了解放人类，是为了实现人类自由而全面的发展的。如果反过来，人们崇拜人工智能，被人工智能所奴役，未来的人工智能就会成为人们心灵上的枷锁，甚至各国之间会为了争夺人工智能核心技术而进行战争，对人类自身产生威胁，那么强人工智能不仅达不到全人类解放的目的，反而会造成全人类的禁锢。

如同金银货币的创造是为了方便人类，解放人类一样，人工智能也是为了实现这一目的而得到人们的广泛关注，但同时不得不警惕未来的人工智能会走向和拜金主义一样的老路——技术崇拜，它们归根到底都是犯了物本主义的错误。技术崇拜的错误需要得到及时地遏制。

如果说，人类对人工智能的技术崇拜还可以及时发现，及时遏制，那么，强人工智能有可能引发的发展伦理困境——人类自身的片面性发展，人工智能就是在潜移默化地代替人类。

2) 人类自身的片面性发展

人工智能是“转化为人的意志驾驭自然界的器官”[4]，如果过度依赖机器，则有可能成为“单向度的人”[5]，即人类的片面性发展。这并不是危言耸听，人工智能通过大数据的分析和推理确实能够快速准确地做出最有利人类全体的决定。马尔库塞(Herbert Marcuse)[1]认为，“技术的异化通过使人依附于机

[1]赫伯特·马尔库塞(Herbert Marcuse, 1898~1979)，德裔美籍哲学家和社会理论家，法兰克福学派左翼主要代表，被西方誉为“新左派哲学家”。他一生在美国从事社会研究与教学工作，代表作品有《理性与革命》《爱欲与文明》《单向度的人》等。

器而实现单向度社会的构建，使得社会成为一种‘没有反对派’的社会，也使得生活在这个社会中的人成为批判停顿的单向度人，人们失去了否定、批判和超越的能力而忠心耿耿地臣服于‘技术的进步’”[6]。

如果过度依赖人工智能，则极有可能重现工业社会时代的“技术异化”，和拜金主义一样，人类极易形成技术崇拜，因为“技术的权威的增长，使人崇拜技术，技术竟成了人的宗教”[7]。在强人工智能时代，智能机器人会超过人类的智力水平和能力水平，如阿尔法围棋(Alpha Go)从还会有一两次的失败到和世界顶级围棋高手对战毫无败绩，只用了一年左右的时间。人工智能的发展如此快速，如果人工智能发展到“智能爆炸”[2]的时候，人工智能的发展恐怕会呈指数型增长。那时候强人工智能的各种能力完全超越人类，同时又不会有所谓的死亡威胁，不死不灭的神何尝不就是人工智能，那么，对神的信任和崇拜转向人工智能是显而易见的。同时对人工智能的信任和崇拜会使人们丧失掉基本的思考能力，因为只需要询问人工智能就可以有行之有效的解决方案了，如 ChatGPT 的应用。

如果人类失去了批判性思维，那就只是一个行尸走肉，这是对人类自身威胁最大的一点，当每一个人的个性都成为同一个个性；每一个人的观点都成为同一个观点，表面看起来一派和谐，社会意见达到了高度的统一，实际在人工智能面前，每个人都丧失了独立思考的能力，被同化成为了一个被人工智能掌管的“机器人”。“人的鲜活的生命力变成了抽象的、空洞的符号，人们虽然自由但再无个性，如同被资本宰制一样被人工智能钳制着。”[8]那么，人类引以为傲的理性和创造力实际上也就不复存在了，那么强人工智能对于人类社会来说，就不是在进步，而是退步到卢梭所说的自然状态下的人类，在这一状态下，人类作为自由主体和动物的最大区别就在于有选择“做与不做”的自由，而人类的自由仅限于此罢了。

与上文中人工智能的“威胁论”不同，另一部分人却对人工智能的发展充满了信心，不担心或忽略掉强人工智能时代的伦理困境。

3) 发展人工智能技术的不确定性

目前的程序师、设计人员等研究人工智能技术的相关人员仍旧没有认识到人工智能的伦理问题有多么的重要，一些高级知识分子尤其是理工科人员，仍然将程序计算作为重点研究的方向(虽然这是本职工作)，更多关注的是突破人工智能的技术难题，很少甚至完全不关注人工智能有可能带来的伦理问题。除了如今的弱人工智能时代的机器人确实无法威胁人类以外，研究人工智能不同领域的科技人员还是带有一种技术乐观主义，认为人工智能尚处于自身知识的控制之下，并不会反过来“谋害主人”，在笔者看来弱人工智能确实无法从根本上威胁人类，但是仍然要防止强人工智能时代人类利用智能机器人自相残杀，乃至发动全球战争。比如，火药的发明最开始也只是用于炼制丹药，并没有被应用于战争中去，火药也可以用于制作烟花而不是作为战争工具而存在，强人工智能的应用还是需要人们谨慎对待。

除此以外，人类在人工智能研究中的伦理困境还包括有“交互偏见”(用户由于自己与算法的交互方式而使算法产生的偏见)、“潜意识偏见”(算法错误地把观念与种族和性别等因素联系起来)、“选择偏见”(用于训练算法的数据被倾向性地用于表示某个群体或者分组)和“确认偏见”(数据驱动偏向于那些先入为主的信息)等[9]。这些偏见并不是显而易见的，都是在潜移默化地伤害人类自身，最终导致矛盾的激化、爆发冲突乃至战争。

每一次科技的进步，其发明本意都是善的。但随着科技的进步，我们就会发现，越是强大的技术，容错率就越低，即使只发生一次事故，就有可能造成巨大的破坏，足以抹杀所有的裨益。所以，与其亡羊补牢，不如防患于未然，强人工智能的伦理问题在设计之初就应该得到全面而详尽的考虑，以免成为伤害人类自身的一个武器。

[2]智能爆炸(Intelligence Explosion)，能迅速导致超级智能的迭代式自我改进的过程。

2.2. 强人工智能的责任伦理困境与道德主体地位问题

人工智能的道德主体地位问题并不是只有在未来才会出现，反而是一直以来学术界讨论的重点内容，并且这个问题的讨论会一直持续到从现实中得出结论的那一刻才会停止。人工智能是否需要承担道德责任，这一问题究其根本就是人工智能是否拥有道德主体地位的问题或者说道德意识的问题。

人工智能是否会和人一样拥有一颗“七窍玲珑心”，拥有一颗和人类良心一样的“机心”，是人工智能获得道德主体地位的第一步。但是关于道德是什么？善是什么？如何才是善的？这一系列问题，人类社会还处于众说纷纭的阶段，并不能达到世界统一的标准，就拿正义的标准来看，关于什么是正义的讨论至今还没有统一的标准，是使用功利主义者边沁和密尔“最大多数人的最大幸福”的正义原则为标准，还是使用罗尔斯《正义论》的两个正义原则[3]的“合乎最少受惠者的最大利益”原则为标准，并没有得到明确的认定，有时候还需要按照情况而随时变换正义原则。

在弱人工智能时代，人工智能的道德主体地位当然不能被认同，因为这些机器人都没有发展到独立思考的地步，更加不会拥有“机心”，生物本质主义者们也认为：“意识是生物有机体的属性；由硅和铜制成的电线和晶片不会有主观感受，也不会有自我意识。”[10]硅基生物，只能是冷冰冰的机器，不能和碳基生物一样有血有肉，拥有“恻隐之心”。人工智能的判断永远是出于程序设定，而做出相应的行为反应，就算未来的智能机器人可以完全模拟人类的种种表情、感受，但是诸如疼痛、欢愉、幸福等感受，人工智能永远不可能真正感同身受。连最基本的肉体感受力都没有，“机心”从何而生？又何谈道德主体地位？

与此相反的是，技术乐观主义者认为，“意识完全可以通过计算机产生，所以足够复杂的人工智能必定会有意识。”[10]随着科技的进步和时间的推移，人工智能机器人能够通过模拟人类的生物大脑，通过自下而上的学习，逐渐“养成”人类的肉体感受力，并产生出“机心”；或者通过自上而下的办法——计算机工程师输入人类的感情程序而产生“机心”。

另一种认同人工智能道德主体地位的观点认为，按照通用人工智能(即强人工智能)的定义来看，人工智能是有道德意识的，如果否定了人工智能的道德主体地位，就是否定了这一定义，这是相互矛盾的。直接从定义上认可了人工智能的道德主体地位。

无论是否认同人工智能的道德主体地位，每种理论都各有自己的逻辑自通性。无论人工智能机器人能否获得“机心”，都是一个未知数，但可以明确的一点是，强人工智能时代的最好的结果是构建人机合作、人机共存或者说人机交互的智能化社会。相对来说，人机交互的社会可能不是一个最好的社会，但却是对人类伤害最小的社会。

人机交互并不等于完全解决了强人工智能时代的伦理困境，人机交互过于频繁依然会产生相应的伦理问题。

2.3. 强人工智能与人类产生的情感伦理困境

人与人工智能产生情感伦理问题多来源于服务型机器人这一领域，这一伦理问题是人机交互频繁而造成的。人机情感的产生，可以是人类与类人机器人，也可以是人类与虚拟人物产生恋爱情感、依赖情感等。如虚拟学生华智冰[4]、虚拟偶像洛天依[5]等，但不论是在现实生活中还是在虚拟世界中，我们现如

[3]两个正义原则：一是平等自由原则，二是差别原则和机会的公平平等原则。

[4]“华智冰”由三方合作诞生：北京智源人工智能研究院领衔开发超大规模智能模型“悟道 2.0”；智谱 AI 团队作为骨干参与开发“悟道 2.0”，并主要开发平台应用生态；小冰公司提供全球领先的人工智能完备框架，同时负责声音、形象的开发应用。是具有一定推理和情感交互能力的机器人，会创作音乐、诗词和绘画作品。

[5]洛天依，2012 年 7 月 12 日诞生(诞生即正式出道)，中国内地女虚拟歌手、虚拟偶像，上海禾念信息科技有限公司旗下虚拟艺人。2022 年 2 月 2 日，洛天依参加在北京举办的“相约北京”奥林匹克文化节开幕式上，并演唱歌曲《Time to shine》受到广泛关注，它成为第一位登上奥运舞台的中国虚拟歌手。

今的技术都达不到完全人性化的人工智能，无论这类机器人与人类有多么相像的身体外观和生理特征，人类都能自然而然地区分类人机器人和人类。但是当人工智能机器人和人类达到完全一样的程度，且人工智能机器人在各个方面比人类自身还完美，可以满足大部分人的各种情感需求时，人类对“完美恋人”、“完美朋友”产生感情是控制不住的。那么人类和机器人产生情感是否道德？人们又该如何对待像“朋友”、“家人”、“恋人”一样的机器人？机器人应该获得道德主体地位吗？等等伦理问题。

1970 年日本学者森政弘提出“恐怖谷理论”[6]，指出了一种现象，当机器人与人类在动作上、外表相似，人类会产生正面的情感；但是当这一相似度超过一定程度时，人类就会变成负面情感和反感；但是如果这种相似度继续上升并达到和普通人之间的相似度的时候，就会再次回到正面的情感反应，产生人类与人类的移情作用。这三个阶段的情感反应实际上都是随着人工智能技术的发展而产生的，人工智能本身并没有任何的错处可以被指摘。归根到底都是人类自身的情感变化随着人工智能技术的发展而发生了变化。

无论是主动还是被动，未来的社会必然会大面积使用服务型机器人，比如养老型机器人、家居型机器人、甚至性爱机器人也会越来越被接受。为了更好的服务人类，这些机器人被不断完善直至能够完美模拟人类的情感反应，预测人类的情感需求，并源源不断地为人类提供情绪价值。如果抛开机器人的“非人类性”，可以说，服务型机器人简直是一个完美的朋友、恋人甚至亲人。但人类的情感本来是通过人与人之间的交往而产生的，而人工智能开启了交往方式的新模式，那么如果人类个体不与现实的人类个体进行交流、交往，就很容易产生一些伦理问题，如：

1) 机器人能否取代人的道德责任？例如，陪伴老人的机器人能否代替儿女尽孝？

2) 研发用于其他目的的伴侣机器人，诸如酒伴侣、性伴侣等，有无道德问题？

3) 是否会产生对机器人的情感依赖？倘若我们的情感为机器人所左右，是好是坏？[11]

那么在强人工智能时代，如何面对情感伦理困境呢？这需要具体问题具体分析，但是可以确定的一点是，对待机器人还是应该是“善”的，友好的，不应该随意虐待机器人，人之所以是高级动物就在于对动植物的态度是友善，对待机器人也是这样的，即使机器人不会获得和人类同等的社会地位，也不能因为机器人没有痛感就随意虐待，这是不符合人“善”的本性的。

3. 强人工智能时代伦理困境的伦理治理

面对强人工智能时代可能发生的种种伦理困境，纵观古今，我们或许可以从中得到一些解决这些伦理问题的启发。

3.1. 开展中国传统“以道驭术”的伦理思想教育

先秦时期，儒家、道家、以及墨家就提出了“以道驭术”的科技伦理思想。所谓“以道驭术”，“就是技术行为和技术应用要受到伦理道德规范的驾驭和制约。”[12]如果没有伦理道德的约束，科技的发展只会为了功利目的而不择手段，这一伦理规范思想在强人工智能的发展中同样适用。

先秦各家“以道驭术”的主要伦理思想各有特色。首先，儒家的伦理思想讲究现实性，提倡经世致用。尤其是儒家主张“三事”，所谓“三事”，指的是“正德、利用、厚生”，“‘三事’，要求技术发展目标既对国计民生有利，又有道德教化功能。”[13]第二，道家的“以道驭术”观念中主张技术活动各要素之间的和谐，包括技术应用中人际关系的和谐、技术活动与社会的和谐、技术活动与自然的和谐。第三，墨家的“以道驭术”思想则主要关注于工匠个人的道德修养。“注重以技术道德规范约束群体或个体工匠的技术活动，要求门徒学习大禹治水吃苦耐劳、栉风沐雨的精神，毫无功名利禄之心，勤生薄

[6]“恐怖谷”一词由 Ernst Jentsch 于 1906 年的论文《恐怖谷心理学》中提出，而他的观点被弗洛伊德在 1919 年的论文《恐怖谷》中阐述，因而成为著名理论。

死，以赴天下之急。”[14]

总体上看，先秦时期的“以道驭术”，是观念和行动的有机结合，但“思想家的伦理道德观念并不能直接变成普通民众特别是工匠的道德意识和行为，这里要经历一个不断教化熏陶的过程。”[15]在人工智能伦理学中就可以借鉴这一方法，对德行优良的人工智能科研工作者要加以赞誉，而对在研究人工智能活动中的歪风邪气要加以鞭笞。久而久之，德行良好的科研人员就会在研究活动中占据有利地位，由他们给人工智能机器人进行道德嵌入，自然不会有“性恶”的人工智能，研究人工智能的不良风气也会逐渐被摒弃。

3.2. 确立强人工智能“以人为本”的伦理原则

无论将来的人工智能有多么先进，机器人和人类多么相像，首先要明确一点：人只能是目的，而不能是手段。一切人类创造的产品，不论是否拥有和人类一样的“肉体感受力”[7]，都不能忘记“以人为本”这一基本原则。不论是国际组织还是各国政府、企业，各类主体发布的人工智能伦理原则都不能离开这一基本原则。

早在 1942 年，阿西莫夫(Isaac Asimov)在《我，机器人》中就给出了三条机器人定律：“第一条，机器人不得伤害人类个体，不能目睹人类受到伤害而不干预；第二条，机器人必须服从人类的命令，命令与第一条冲突除外；第三条，机器人在不违反第一条、第二条原则的情况下，要保护自身安全。”[16]鉴于以上三条原则还不够成熟，阿西莫夫后续又增加了一条根本性原则，即机器人不得伤害或者侵犯人类。以阿西莫夫的机器人定律为基础，各类主体提出了大同小异的人工智能伦理原则。

2017 年，阿西洛马人工智能 23 条原则，提及“安全性、透明性、责任、人类价值观、隐私、利益分享、人类控制”等伦理原则；2018 年 6 月在加拿大召开的 G7 峰会上，七国集团领导人通过了一项《人工智能未来的共同愿景》文件，提及“以人为本、增加信任、保障安全、保护隐私”等若干伦理原则[17]；2019 年 4 月，欧盟人工智能高级专家组发布《可信任的人工智能伦理指南》，其中提出了“尊重人类自主、防止伤害、公正与可解释性”四大伦理原则；2022 年，中共中央办公厅国务院办公厅印发了《关于加强科技伦理治理的意见》，其中明确了科技伦理原则：增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明[18]。

总之，人工智能的伦理原则都不能离开“以人为本”这一基本原则，但是要把握好尺度，避免从以人为本的良好出发点走向人类中心主义的错误路线。

3.3. 强人工智能时代的多元协同伦理治理

面对人工智能技术的伦理困境而制定的伦理原则和伦理思想，主要依靠政府的推动实施，但是人工智能技术的发展涉及到多元利益主体，仅靠政府单一主体，其推行效果不会太理想，所以要构建强人工智能的多元协同治理模式，形成强人工智能的伦理治理网络。

首先，“成立统一技术风险治理共同体，以克服现有分散的、单项的风险治理机制，实现部门之间、区域之间的风险协调机制。”[19]避免因为信息的不对等而重复同一个人工智能伦理错误。

二是构建全球联动治理机制。由于各国之间技术标准、准入制度等差异，带来了伦理风险治理的制度性差异，所以要在人类命运共同体的思想指导下，构建全球联动治理机制，制定统一的标准才更有利于各国人工智能的发展和进步。

三是构建、完善相关法律规范。为了预防强人工智能时代可能引发的矛盾与冲突，需要提前着手建立专门针对人工智能的法律法规。

[7]有的学者认为当机器人拥有和人类一样的肉体感受力时才能开始拥有和人类一样的“机心”，否则，无论机器人和人类多么相似，都只能是一个冷冰冰的机器人。在 2.2 关于“强人工智能的责任伦理困境与道德主体地位问题”的讨论中已经说明，不再赘述。

四是成立人工智能伦理监督委员会。伦理监督者的存在必不可少，未来人机共存的社会中必然需要有一个公正的“裁判”来处理各种人机伦理问题，这样才不至于走上人机对立的危险境地。

4. 结语

第四次科技革命的到来使科技快速发展的同时，人们忽略了科技伦理的同步跟进，历史的经验告诉我们，科技越是发达，一旦出错，对人类的伤害就越加不可挽回。面对强人工智能时代的伦理困境，诸如强人工智能的发展伦理困境；强人工智能的责任伦理困境与道德主体地位问题；强人工智能与人类产生的情感伦理困境等，我们不得不提前采取措施，为人机共存的发展前景打下基础。其方法有：开展中国传统“以道驭术”的伦理思想教育，通过教化、学习，摒弃歪门邪道，创造一个由内而外的“善”的人工智能；确立强人工智能“以人为本”的伦理原则，制定统一的伦理标准，人工智能才能更好地服务人类；构建强人工智能时代的多元协同伦理治理，带动全社会、全球、全人类的协同治理才能真正地处理好强人工智能时代的伦理困境。

参考文献

[1] 刘子玥. 关于弱人工智能与强人工智能的思考[J]. 电子元器件与信息技术, 2021, 5(7): 7-8.

[2] [美]迈克斯·泰格马克(Max Tegmar). 生命3.0: 人工智能时代 人类的进化与重生[M]. 汪婕舒, 译. 杭州: 浙江教育出版社, 2018: 51.

[3] 周玄, 赵建超. 人工智能的伦理困境与正向规约[J]. 江西社会科学, 2022, 42(10): 37-43.

[4] 马克思, 恩格斯. 马克思恩格斯全集: 第三十一卷[M]. 北京: 人民出版社, 1990: 102.

[5] 赫伯特·马尔库塞. 单向度的人: 发达工业社会意识形态研究[M]. 刘继, 译. 上海: 上海译文出版社, 2016.

[6] 吴永源. 人工智能与单向度的人[J]. 高等理科教育, 2020(2): 13-15.

[7] 肖雪慧, 等. 主体的沉沦与觉醒——伦理学的一个新构想[M]. 贵阳: 贵州人民出版社, 1998: 114-115.

[8] 娄延强. 人工智能的伦理困境与正解[J]. 道德与文明, 2022(1): 131-139.

[9] 于雪, 段伟文. 人工智能的伦理建构[J]. 理论探索, 2019(6): 43-49.

[10] 詹姆斯·道斯, 张伟, 李冰清. 推理性人权: 人工智能与人类未来[J]. 中国政法大学学报, 2022(1): 289-304.

[11] 尚新建, 杜丽燕. 谈机器人伦理学[J]. 外国哲学, 2020(1): 222.

[12] 王前, 等. 中国科技伦理史纲[M]. 北京: 人民出版社, 2006: 7.

[13] 孙宏安. 中国古代科学教育史略[M]. 沈阳: 辽宁教育出版社, 1996: 291-292.

[14] 汝莉, 李生荣. 中国科技教育史[M]. 长沙: 湖南教育出版社: 1992: 88-89, 285.

[15] 王前. “以道驭术”——我国先秦时期的技术伦理及其现代意义[J]. 自然辩证法通讯, 2008(1): 8-14+110.

[16] 阿西莫夫. 银河帝国 8: 我, 机器人[M]. 叶李华, 译. 南京: 江苏文艺出版社, 2012: 2.

[17] 吴红, 杜严勇. 人工智能伦理治理: 从原则到行动[J]. 自然辩证法研究, 2021, 37(4): 49-54.

[18] 中共中央办公厅 国务院办公厅印发《关于加强科技伦理治理的意见》[J]. 中华人民共和国国务院公报, 2022(10): 5-8.

[19] 谭九生, 杨建武. 人工智能技术的伦理风险及其协同治理[J]. 中国行政管理, 2019(10): 44-50.