科幻作品视角下的人工智能伦理

梁卫国*

（中央民族大学哲学与宗教学学院，北京 100081）

[摘 要] 人工智能快速发展的现实为科幻作品提供了创作素材和科技佐证，而科幻作品为人工智能提供了未来想象和发展可能，两者既相离又相交的关系构建起了丰富的精神实验和感觉体验。产生这些实验和体验的根本原因是，在符号主义、连接主义和行为主义指导下，以数据和算法为技术核心的人工智能逐渐突破了人类中心主义的伦理观。这种伦理观可能会导致人类丧失对机器的控制权，甚至导致人类的毁灭。要避免这种悲剧发生，需要构建起人类与人工智能相处的新伦理：重新审视人本身，保持对人造自然物——人工智能的敬畏，在人类中心主义的主体下，构建起人类与人工智能相处的准伙伴关系。

[关键词] 伦理　人工智能　科幻作品　主体性　人类中心主义

[中图分类号] TP18；B82-057　　[文献标识码] A　　[DOI] 10.19293/j.cnki.1673-8357.2020.03.005

在众多的科幻作品中，人类与人工智能相处的伦理问题常常让人揪心。在《终结者》(*The Terminator*) 中，最新型终结者 T-X 美丽俊秀、身手毒辣，以能量形式存在各种网络之中；她能随心所欲地变成各种形态，并能操控其他机器人，连最先进的液态金属 T-800 也无法逃脱她的操控。《黑客帝国》(*The Matrix*) 中，在 22 世纪，人类生活在机器所虚构的世界之中，肉体和精神都被超级机器控制。电影《机械姬》(*Ex Machina*) 中，人类造出的机器人具有独立意识，利用感情成功欺骗人类，逃离实验室，并试图毁灭人类。从逻辑上看，科幻片的构思似乎无懈可击。一方面，人工智能的确具有人类所没有的无限寿命、超强的计算能力及其他高

级智能。另一方面，在现实生活中，人类似乎无处可逃：街头车站涌现大量的手机“低头族”，网络社交评分无处不在，狂热青年忙着赚虚拟货币。导致人类无处可逃的是高速运转的人工智能，人工智能正把虚拟世界与物理世界的边界变得愈加模糊。这些巨变使得一些科学家认为，奇点来临后，科幻片中人类被追杀、毁灭的景象就会变成现实。那么，人工智能和科幻作品有何关系？如何看待人类和人工智能相处的伦理问题？面对这种伦理问题，人类应该何去何从？

1 人工智能和科幻作品的关系

1.1 科幻作品为人工智能伦理研究提供可能

如果从传统的学术谱系上看，人工智能

收稿日期：2019-10-27 基金项目：中央统战部宗教研究中心互联网宗教信息大数据监管体系研究（MT2009A）阶段性成果。

* 作者简介：梁卫国，中央民族大学哲学与宗教学学院博士研究生，高级编辑，《中国电力企业管理》杂志社责任编辑，研究方向：人工智能在宗教哲学领域的应用，E-mail：1064220365@qq.com。

主要指“ 一门研究如何构造智能机器（智能计算机）或智能系统，使它能模拟、延伸、扩展人类智能的学科”[1]，属于计算机科学的分支学科，是自然科学的范畴“；科幻作品是关注科技发展对人类社会自身发展及科学文化等更深层面的精神发展的影响的作品，是结合现代科技成就与文学意境的产物”[2]。科幻作品是在科学技术的基础上衍生出来的文学亚门类，是社会科学的范畴；那么，就人工智能和科幻作品纯粹学术研究的内部结构、运作机制、学术传统及代际传承方式来讲，两者相去甚远，并在各自独立的领域遵循着自身发展规律进行着生灭变化。目前，关于两者分别研究的自然科学界和社会科学界有着较多的学术成果，这些学术成果也为人工智能伦理研究奠定了重要基础。对普通公众来讲，人工智能伦理问题专业性较强等，因而理解起来有一定的难度。

如果从科学普及的角度看，当公众在被《终结者》《黑客帝国》等这些科幻作品勾画出的未来而兴奋或担忧时，我们就会发现，从科幻作品的维度来研究人工智能伦理问题或许是个不错的视角。这是因为，人工智能作为科技的一种，常常也是科幻作品的内容。这些作品中反映的人工智能伦理问题为人工智能伦理的现实研究提供了某种可能。这种可能性也可以从人工智能伦理概念和科幻作品的概念中得以证明。虽然，目前人工智能伦理作为一个相对较新的事物在学术研究领 域并没有一个公认的统一的确定概念，但我 们不妨先从其上位概念科技伦理尝试进行定 义。科技伦理的定义是“，人们在从事科技创 新活动时对于社会、自然关系的思想与行为 准则，它规定了科学家及其共同体所应恪守 的价值观念”[3]。根据科技伦理这个定义，我 们可以将人工智能伦理定义为“ 人们在从事 机器人、语言识别、图像识别、自然语言处

理和专家系统等人工智能创新活动时对于人 类社会、自然关系的思想和行为准则”，一般 地，这个准则的核心问题是如何落实人工智 能的责任问题。这种“对于人类社会、自然 关系的思想和行为准则”与科幻作品是关注 科技发展“对人类社会自身发展及科学文化 等”影响的作品，存在一定重合关系。这个重合的主要因素就是，人工智能伦理和科幻 作品都关注科学技术对人类未来社会（生产、 生活、生存等）的影响。只不过，相对来说， 伦理问题关注的这个未来现实性更多些，而 科幻作品关注的这个未来幻想性更多些。如 果反映人工智能伦理和科幻作品两者本质属 性的概念有一定的融合因素，那么，从科幻 作品中去探寻人工智能伦理就存在可能。

科幻作品为人工智能伦理问题（关注如何 处理人类与自然的思想和行为的准则问题）研 究提供可能，还可以从具体的实例进行论证。

比如，在科幻电影《超验骇客》（*Transc- endence*）（2014）中，天才科学家威尔·卡斯 特死后，其妻子和好友将其意识与计算机网 络相连接，最终进化出了一个超级人工智能： 人工智能被赋予人的自我意识和独立思维； 计算机网络替代了威尔由碳原子组成的身体 而成为人工智能的硬件部分。这个超级人工智能给人类带来巨大的益处：挽救绝症患者， 净化空气和水源，将沙漠变为绿洲。这些益 处也是现实生活中人类生理数据库和医疗算 法、空气清洁器和水源治理系统、机器造林 技术等人工智能的逻辑推演。正是人工智能 这些现实为《超验骇客》提供了创作的素材， 这些现实也使得人们更容易相信这个科幻作品所塑造的事实。

《超验骇客》的科技特征并没有就人工智 能的合理逻辑推演止步，而是远远超出现实。 当威尔的独立意识侵入电脑成为人工智能的 一部分，弥补了电脑没有自我认知与缺乏价

值判断的缺陷时，科技就变成了“ 上帝”（威尔的意识 + 计算机网络)。这个全能全知的“ 上帝” 可以进入任何一台电脑系统获取信息，可以监控世界上任何一个角落；它洞烛先机，并且发明新的纳米技术——修复受损细胞、增强再生能力、治愈一切残疾人。但是，这个互利双赢的人工智能 (“ 上帝”）也导致一些政府首脑和科学家产生恐惧感：如果一些人可以永生，如果生态可以快速修复，那么人类生存的意义何在？正是在这种恐惧心理下，一些科技部队开始对这个超级人工智能进行围剿和屠杀，但诡异的是，被人类的炮火等武器破坏后的设备和人员经由纳米技术又迅速恢复原样——只要设备和人员处于无线联网状态（只要有空气、水、土等介质存在），人类就难以战胜这个人工智能。

剧情的发展似乎进入了一个公共悖论：人类的欲求创造出了人工智能，而人工智能又强大到了人类无法控制。何以解决？要么人类灭亡，要么机器灭亡；要么机器还是人的工具，要么人类沦为机器的奴隶。悖论的这四种答案，常常成为人工智能科幻创作的四个思路，如《迷失的一半》《杰克茜》《吾乃母亲》《黑客帝国》《复仇者联盟 2：奥创纪元》《地球停转之日》《天外魔花》《世界尽头》等。这些科幻创作与其说是贩卖人工智能技术下人类生存的焦虑，不如说是将人工智能和人类的冲突作为创作基础，并且在这种人类被人工智能科技灭亡的可能性和必然性中创造永恒话题，不断牵动观赏者的神经：生存还是毁灭？幸运的是，《超验骇客》中的威尔是为爱为生的，在听取了威尔妻子爱的谎言后选择了自我死亡。最终，影片以爱结束。

通过上述分析我们可以看出，一方面，人工智能的现实为科幻作品提供创作素材和 科技佐证，人工智能是科幻作品得以实现的 必然；另一方面，科幻作品为人工智能提供

未来想象和发展可能，科幻作品成为人工智能（主要是超级人工智能）得以实现的可能。两者这种复杂的关系为我们构建起了丰富的精神实验和感觉体验：相离的研究使得两者得以向纵深发展，相交的科幻作品使得二者创造出了色彩斑斓的精神世界。正是在这种复杂的矛盾统一中，人工智能伦理问题得以展开。

1.2 共时性中的人工智能伦理问题

除了从人工智能和科幻作品的关系角度来研究人工智能伦理问题外，我们也可以从人工智能本质属性的概念（共时性研究）和人工智能的发展路径（历时性研究）两个方面，将人工智能伦理问题的探讨推向深入。

从概念上看，目前人工智能的概念有很多。其中比较公认的是，人工智能指用机器（主要指计算机）实现人的计算能力、感知能力、记忆能力、逻辑思维能力等智能活动的技术。人工智能概念最早源于 1950 年英国科学家图灵提出的“机器能思考吗”这个知名的论题，在这篇《计算机器与智能》的论文中，图灵还提出了实现机器思考的心理实验（是否存在可想象的计算机能够通过一个混淆人类智能与机器智能的游戏）。图灵测试者指出，如果在 5 分钟内，一台智能机器不仅能够顺利、正确地给出人类测试者所要的答案，并且这些答案能够使超过 1/3 的人类测试者认为那台被测试的机器回答的答案就是人类回答的答案，那么，这台智能机器就算通过测试，即这台智能机器相当于拥有人类智能。按照这个逻辑，要制造一台拥有人类智能的机器，就转化为制造一个模拟人类童年的大脑机器，然后再对它进行学习英语、数学和下棋等教育训练，经过训练后的类儿童脑机器在经过场景实践学习后就可以发展为成人脑的机器。

如果，人工智能本质上是人造的智能，

是机器对人智能的模仿，反映在科幻创作中，人工智能的作品也应该是在人与机器的对立和统一中展开。如果“科学是作为支撑作品存在的、不可移除的核心线索存在的”[2]，科幻创作的两个基本特征是科学性和幻想性，那么，人工智能题材的科幻创作中作为科学的一种的人工智能也应该是不可移除的核心线索，且也应呈现人工智能的科学性和人工智能的幻想性两个基本特征。人工智能的伦理问题也应该是从科学性和幻想性出发，科学性构建起人工智能伦理求真和确定性的一面；人工智能不仅意味着前沿科技和高端产业，未来也能够广泛应用到解决人类社会面临的长期性挑战”[4]。而这些技术性问题可能是自然科学发挥作用的地方。幻想性构建起人工智能伦理中求变和创新性的一面。这些问题可能带来人文和社会问题。比如，如何看待人类与机器的关系问题；生物化学方法使人很快得到快乐，那么如何看待人类的痛苦问题；随着计算机的发展如何看待其带给人类的失业问题。

从共时性研究得出的人工智能伦理需要考虑的科学性和人文性问题，也可以从关于人工智能的两部经典小说来说明。人工智能小说《明天的两面》(*Two Faces of Tomorrow*)描述了一个世界，那里的复杂文明只有一个全球性的计算机网络才能控制。这个超级计算机集合了大量的逻辑程序，但它缺乏常识，并且它那些基于逻辑的决策开始导致太多致命的突发事 故发生。研究者担心超级计算机可能会脱离人 类的控制，所以他们决定到太空里测试这台计 算机，如果出现错误，就可以摧毁它。但是， 已经产生知觉的电脑很不喜欢这种测试，故事的矛盾就此展开。

被评为人工智能优秀小说伊恩·班克斯(Iain M.Banks)的《无限异象》(*Excession*)也是这样一部作品。小说中描写的“心智”

是超智能的人工智能生物，它们之间的交流 像是没有标题的电子邮件，它们也试图对人 类进行统治。很明显，这两部小说都是以人 类与人工智能的冲突展开叙事，且科学性都 是故事的决定性因素。在笔者搜集到的人工 智能作品中，也几乎没有只讲幻想性而不顾 科学性的作品。

1.3 历时性中的人工智能伦理问题

从历时性的角度看，人工智能概念被提 出100 多年来，主导人工智能发展的指导思想 主要有逻辑主义（或称符号主义、计算机主 义或心理主义 ），连接主义（生理学派）和行 为主义（或称进化主义或控制学派）三个。

逻辑主义（符号主义）的逻辑起点是， 符号（如数字、字母甚至服饰颜色等）是人 类认知的基本元素，用符号表示的系列运算 就是人类认知事物的过程。物理符号系统 假设（所有智能行为都等价于物理符号系 统）和有限合理性原理（如谷歌、百度等 使用关键词进行模糊搜索来逐渐得到问题正 确答案的计算检索过程）是其核心。纽厄尔（Newell）和尼尔逊（Nilsson） 等符号主义 者认为，符号是人类思维的单元，思维是符 号程序化、算法的结果；人工智能的实现的 路径就是，在遵守逻辑系统规则的前提下， 给机器输入大部分程序，使得机器通过 0、1 二进制符号实现人类的智能。目前，符号主 义是人工智能的主流观点，并在知识认知、 知识表示等方面取得了重要进展。但符号主 义者的局限性在于其线性关系和排中律的预 设，即对智能的模仿主要依据其代数学和数学定理的机器实现。这为机器功能划定了界限，而连接主义的出现一定程度上可以克服此局限。

连接主义的逻辑起点是，思维的基本元素是神经元，思维过程是这些大量并行连接的神经元的运动或活动。连接主义认为，仿

照人类神经网络的运行规则和连接机制就能形成学习算法，按照这些算法制造的机器就能实现人工智能。与符号主义的线性处理相比，分布式存储和并行协同处理的实现方式使得神经网络理论发展较快。

行为主义的逻辑起点是控制论和感知 - 动作系统。行为主义认为，智能不一定必须用符号表示，也不一定必须使用仿生学结构来模仿，智能行为源于主体与环境的互动和变化，即智能主要取决于感知和行动。既然，现实世界是智能行为形成的基础，那么，人工智能的实现方式就是制造出一个模拟人类儿童脑的机器，之后，让这个类儿童脑的机器像人类的儿童一样在现实中给予其教育培训或让其自我学习，从而得到一个类似人类成年脑的机器。该观点主要认为，是机器的自我学习造就了人工智能。

既然逻辑主义（符号主义）、连接主义和行为主义是主导人工智能发展的主要思想，那么人工智能的科幻创作也应该是逻辑主义、连接主义和行为主义三者所决定的人工智能科幻创作。换句话说，人工智能题材的科幻创作可以分为这三种形式的科幻创作。人工智能的伦理也应该从这三者中进行分析，在符号主义下，人是符号的动物，人人就会被标签化，与这种标签相关的信息就被密集推送给了被标签化的每个具体的人，这就会造成信息渠道狭窄的风险（如刻板印象、沉默的螺旋效应）——人类创造了工具，工具也创造了人。在连接主义下，人工智能伦理问题可能是对人类大脑的过度开发，人与机器的界限会日渐模糊，这很可能导致心与芯的竞争，最终机器的智慧可能会代替人类的智慧。行为主义带来的伦理可能是超级人工智能的到来，这会导致超级人工智能取代人类的那一天早日到来。

2 人工智能伦理与人类中心主义

2.1 人工智能发展中人类中心主义的悖论

如果说人工智能为科幻创作提供了必然性的话，那么，科幻创作为人工智能提供了可能性。这种必然性和可能性在人工智能的科幻创作中为人工智能伦理提供了一个技术实现路径。从人工智能的发展现状看，人工智能的伦理路径问题已经成为决定人工智能如何发展甚至生存的问题。人工智能的伦理问题如此重要，是因为这个伦理打破了农耕文明后人类社会中逐渐成为主流的伦理价值观——人类中心主义。通常来说，这个人类中心主义伦理价值观强调“一切以人为中心，或一切以人为尺度，为人的利益服务，一切从人的利益出发”[5]。

根据人类中心主义伦理价值观而发展起来的人工智能，可能会导致人类的毁灭并使得人类中心主义最终丧失。理由是，按照人类利益为中心的逻辑，发展满足人类记忆、认知和运算的人工智能自然是满足人类利益的；人工智能认为人类的活动只是数据和算法，那么由碳原子组成的智慧（人）比由硅原子（芯片）组成的智慧并无高低贵贱之别。如果人类的伦理与人工智能的伦理没有高低贵贱之分，那么，人工智能取代人类也是自然的事情。如果人类被人工智能取代，那么人类将不会存在。如果人类不存在了，那么人类的生命也就不再具有意义。而这一点，则是人工智能发展的又一悖论：本来是以服务人为目的的人工智能最终却将人类终结了！

如何避免人类被毁灭又能享受人工智能带来的富足和快乐，这是人工智能的伦理挑战，也是科幻创作矛盾冲突的一个逻辑起点。从某种程度上，正是人类中心主义的内在矛盾才使观赏者将人工智能和自己的生命联系起来：一方面自己离不开人工智能，另一方

面人工智能逐渐使自己丧失主体性。这种冲突再加上亲情的渲染（如《星际穿越》），加上大胆想象（《超验骇客》中威尔妻子希望的一个人造的水、空气和土壤的世界），加上恐怖威胁（如《黑客帝国》中机器对人类的征服和毁灭）等这些合理幻想，就使得科幻作品产生了惊心动魄、引人至深的传播效果。

2.2 人工智能伦理应随着时代发展而变化

“一切宗教、艺术和科学都是同一株树下的各个分枝。所有这些志向都是为着使人类的生活趋于高尚，把它从单纯的生理上的生存境界提高”[6]。爱因斯坦认为，一切技术上奋斗的主要目标就是关心人类本身，因此，人工智能的伦理问题就其一般性来讲属于科技伦理问题，就其特殊性来讲，应该在于其智能性如何服务于人又不伤害人的问题。那么，当前人们如何看待这个伦理问题呢？

当前世界各地的人们谈论人工智能伦理路径问题有两个特点。一是人工智能概念被泛化。各群体在参与人工智能伦理讨论时，其讨论的对象实际是含生命科学、基因编辑等数据驱动技术的泛指，远超出科技界对人工智能概念的外延。这一现象在非专业人群中尤为明显，缺乏专业背景的公众对数字技术更多的是直觉认知和判断，而这也是公众中出现人工智能“无用论”“万能论”等截然不同答案的原因[7]。二是政界、学界、实业界等主体的诉求各有侧重。虽然三者的伦理规则大多以号召或软性原则为主，普遍关注人工智能增进人类福祉、技术的包容公平、维护人类尊严和自主性、保障安全和隐私等内容，但是，这三类的关注点略有不同。从对文本的统计分析来看，学术界较多关注人类价值观和责任；产业界更关注协作，而较少提及安全和隐私[8]。

概念泛化和自说自话的原因，从现象上看，是各利益主体受先前认识局限而形成了

“ 前理解”，实质上则是，因为人工智能技术 是个不断发展的新兴技术，它的风险伴随着 创新进步而不断显现。

3 构建人工智能伦理的思辨

3.1 人工智能伦理在哲学上难以构建

当我们从人工智能伦理的背景回到人工 智能伦理内部的时候，就会发现，人工智能 伦理问题本身是人与智能机器的关系问题。 而要构建两者的和谐关系，无法避免且必须 要回答的问题是，是否存在“自由个人”。生 命科学者认为，所谓“ 自由个人” 并不是真 实的存在，人类不过是生化算法的组合而已。在生命存续期间，各种各样的生命体验都被 大脑的生化机制制造出来。但是，这些体验 不是一直存在的，而是像电影的镜头一样都 是短暂停留后立马就消弭于无形，之后，更 多的体验被大脑再次创造出来……生命的过 程就是这些体验不间断地闪现又不间断地消 失，不间断地出现又不断失去的过程。在闪 现和消失之间宛若影像镜头般快速相连而使 人认为这些体验就一种不动的存在。《未来简 史》的作者赫拉利等还把这种体验分为体验 的自我（experiencing self，主要是自我所经历 的生化反应）和叙事自我（narrating self，对 这些生化反应尝试编织各种故事的自我）。生 命学者的这种看法与 2000 多年前哲学界的观 点不谋而合。佛教释迦牟尼（人是五蕴和合 的混合）、道教（人人都能化生为仙）、古希 腊德谟克利特（Democritus）（万物皆由原子 组成的）等也都认为个人概念是一种虚妄。 数据主义者的观点也支持生命学者的看法， 认为宇宙由数据组成，任何现象或实体的价 值就在于对数据的处理的贡献[8]。全人类可被 看成一个巨型的数据处理系统，每个人都是 这个系统中的一个芯片，我们对一切事物的 认知及所做出的反应都是在执行自身生化算

法而已。

如果，个体的自我从物质到精神都是变动不居的碎片（原子、电子）和算法，那么仿照人脑而制造的智能机器也只能是碎片化的反映或算法的反映。从这个意义上来看，人与人工智能的伦理关系可能不是人类与自然物，或人类与智能机器的关系，而是社会中那种一台机器操控者与另一台机器操控者之间的关系，即社会关系。

在语言、文字、金钱、机构等社会性存在的前提下，人类构建了一套人类（虽然没有主体性或只有主体间性）与客体（人化的自然）的伦理规则。如“保存自我的努力是德性的首先的唯一的基础”；又如“以理性为指导，而行动、生活、保存自我的存在”；等等[9]。《未来简史》作者赫拉利等甚至认为，在21世纪，人类需要的伦理可能是获得永生、幸福快乐、化身为神。

按照以上理解，哲学意义上的人工智能伦理构建是艰难的，而社会学意义上的构建应该是必要和可能的。这种构建至少可以提供一个社会秩序稳定的预期和消除一些人对未来不确定的恐慌情绪。

3.2 人工智能伦理在社会学上态势复杂

关于社会学意义上的人工智能伦理是目前学界讨论的重点，其主要观点有以下几种。欧洲政策研究中心认为，现在没有迹象表明，人工智能将发展出类似人类的感知（perception）和意识（awareness）。当前人工智能主要被应用于优化（optimization）、搜索/推荐（search/recommendation）和诊断/预测（diagnosis/prediction）三个领域[7]。按照他们的观点，人与人工智能的伦理关系只是人与自己创造的劳动工具的关系。

当代人机接口技术的主要开创者费尔森斯丁则认为“，今天的人工智能技术越来越倾

向于以人类为中心的傀儡学”，费尔森斯丁还强调，人类与人工智能的关系“，是一种共生性的伙伴关系”[10]。

英国学者塔迪欧（Mariarosaria Taddeo）和弗洛里迪（Luciano Floridi）建立了一个以 数据本身、数据算法，以及与数据和算法相 应的实践过程所组成的三元数据伦理框架。 人工智能的伦理问题可近似模拟为“ 数据伦 理学、算法伦理学和实践伦理学的三个轴”。 在同一个人工智能的概念层面“，人工智能的 伦理问题是以三种值来区分的点。”[11]

张浩、黄克同在《迈向制度化的人工智 能伦理》一文中认为“，探索使人工智能治理 原则落地的规则和机制，将是未来的重点和 方向”。他们提出了技术安全、非歧视性、隐 私保护、可问责性的伦理观[7]。

笔者认为，人工智能具有类主体的性质，随着人工智能技术的发展，人类具有排他性（排除其他动物、植物甚至自然不动物）主体地位的伦理时代很可能随着奇点的到来而完成历史使命。尤其是在认知科学、神经语言学、互联网技术快速发展的背景下，人的思维、记忆、情绪如果能够被芯片化，能够被外部软件测度的话，人首先应该确定的是如何与人工智能相处的问题。目前，各国科技竞争的态势决定了这种可能，且这种可能正日益渐进性地成为现实。新的伦理观应该是，人类不仅要对自然存在敬畏感，也要对人造的自然物——人工智能存在敬畏感，随着技术进步和社会发展，人类与人工智能的共处应该成为常态，人类与人工智能应该成为准伙伴关系。新的伦理规则应该是以人类中心主义为主体的情况下，增加对人工智能这个准伙伴的一些约定。

3.3 人工智能伦理的困境是人自身矛盾的展开

一直以来，人类生存的原子物理空间的

现实是由可能性所创造的，而人类的科幻创作扩大了人类的可能性，并将不可能变成了可能（如用影视方式模拟精神实验）。在这个不可能的可能性的创造过程中，人类精神不断超越物理时空而飞跃到一个由语言、文字甚至思想所建构的世界之中（比如读科幻小说的读者可以在作者用文字创造的意境世界中畅游）。如果说，随着互联网技术的出现创造出了一个更大的虚拟空间（这个空间因为非具身性等特征而将人类的虚拟空间无限扩大，把物理空间无限压缩）的话，人工智能则是将这个虚拟空间变成物理现实的重要手段。人工智能的创造与历史上人类其他工具不同，其他工具更多的是人的眼、耳、鼻、舌、四肢的延伸，而人工智能是对人类最高思维的模仿。在这个模仿中，如果仅仅是模仿的话，可能人工智能的这个机器只是人的工具，但如果人没有在伦理约束下无限扩大这种模仿的话，这种工具一定会超越人，伤害人，还有可能成为人类的统治者。

从某种意义上，人工智能的发展是人自身欲望与克制矛盾的双重展开。比如为了便利而以失去隐私为代价。在这个展开过程中，科幻小说、科幻影片等具有独特的作用：一方面，它的故事展开以现实科技的发展为思维基石；另一方面，它又依靠创作者的猜测和想象来构建自己的意义空间。在这个独特的空间中充满了贪嗔痴慢疑的人性与真善美的人性冲突，这种冲突往往围绕着科技是人 类的工具还是人的本身不断展开。从工具角 度看，无论《黑客帝国》《机器姬》《人工智 能》等科幻片中人类面临多么大的威胁，现 实世界的人类还是逐渐将主导生活的权利交 给机器来决策判断：离开导航系统，很多人 不知道如何开车行路；离开机器诊疗，很多 医生难以判断病情；离开实时监控心跳、步

数的穿戴设备，很多人会怀疑自身的健康； 离开智能手机，很多“ 低头族” 不知道如何 消费大把时间；离开虚拟货币、虚拟客服， 很多公司（物流、金融、电信等）的竞争力 会大打折扣。从人本身的角度看，《黑客帝国》《机器姬》《人工智能》等科幻创造的影像中 的现实（这个比特形式实现的电子事实正模 糊着物理的和虚拟的边界，使得一些网络成 瘾的青年人往往分不清是现实还是虚拟），如 果现实物理世界提供的产生这些影像中现实 实现的条件足够的话，那么影片中的警醒将 会变成真正的事实，人类可能会真的面临被 奴役、被屠杀的局面。

如果说人类能取代恐龙，那么机器为什 么就不能取代人类呢？当人类真的只是芯片 形式存在，并没有物质身体的时候，有谁能 说人类是获得了幸福还是失去了幸福（正如 庄子所讲的是人化了蝶还是蝶化作了人）呢？ 或许，科幻创作的价值就在于，人类自主创 造出一个世界，同时又毁灭另一个世界。生 灭之间，是现实和未来的展开，是可能与不 可能的交织，是即生即死、复生复死。

4 结语

本研究从科幻作品的视角对人工智能伦 理问题进行了探讨。得出的基本结论是，按 照主导人工智能发展思想的标准，可以把科 幻作品分为符号主义的人工智能作品、连接 主义的人工智能作品、行动主义的人工智能 作品；这些作品往往是以反思人工智能伦理 问题展开的，在人类中心主义的主导下，人 工智能得以快速发展，而这种发展内在地会 导致人类的灭亡，从而彻底埋葬人类中心主 义。要走出这一困境，我们从哲学和社会学 两个维度对人工智能伦理观进行了考察，从 哲学维度看，在人类没有独立自我的前提下，

人工智能伦理问题即人类与智能机器相处的关系问题是没有终极答案的；从社会学维度梳理几种常见的伦理观：人工智能对人类生活不会影响、不同利益主体对人工智能伦理问题判断不同，人机接口的傀儡主义等。本文的最终结论是，人工智能伦理问题的构建需要对机器保持敬畏，人类与人工智能应该是准伙伴关系。

在人工智能是人类的福音还是噩耗的问题上，科幻作品给了我们无限的想象和创造空间。对科学家来讲，还有许多现实任务有待完成。这个现实是，我们当前的人工智能时代是人机共存的时代。这个时代人工智能的任务可能是，如何让人类的感官与机能的功能范围依托机器而更加广阔灵活，如何让人类的信息交流、学习方式乃至生命形式产生质的变化。从上述变化的角度来看，本研究还有很多不足之处或遗留问题。这些问题主要有：人的最高本质如何实现？人工智能的自我意识何以可能？如何回答人工智能的生成机制问题？人工智能如何与宗教融合发展？如何避免人工智能把人的高级思维活动变为低级的物理电子活动？如何评价人类的存在形态、价值观念、思想意识的变化？人工智能与人的心灵体验有何关系？而对上述问题的回答需要自然科学学者与人文学科的学者共同来解决。“ 我们只能看到眼前很短的距离，但人类有很多事情要做”（We can see only a short distance ahead, but we can see that much remains to be done），这是人工智能鼻祖 图灵在其论文《计算机器与智能》末尾的一 句名言。做好现在，才能创造更好未来。

参考文献

[1] 王永庆．人工智能原理和方法 [M]，西安：西安交通大学出版社，2015：4.
[2] 尹霖．关于我国科幻发展状况的调研报告——“ 科幻创作与青少年想象力培养” 研讨会综述 [J]. 科普创作通讯，2011(2)：3-11.
[3] 王忠伟．战略视野下的科技伦理 [J]. 四川经济管理学院学报，2008(2)：21-
23. [4] 袁勇 . 人工智能伦理三问 [N]. 经济日报，2018-07-12（12）.
[5] 余谋昌．走出人类中心主义 [J]. 自然辩证法研究，1994(7)：8-14，47.
[6] 爱因斯坦 . 爱因斯坦文集（卷 3）[M]. 许良英，译．北京：商务印书馆，1979：149.
[7] 张浩，黄克同．迈向制度化的人工智能伦理 [J]. 人工智能，2019(4)：32-38.
[8] 尤瓦尔·赫拉利 . 未来简史：从智人到智神 [M]. 林俊宏，译，北京：中信出版社，2017：359.
[9] 陈远，于首奎，梅良模，等．世界百科名著大辞典：社会和人文科学 [M]. 济南：山东教育出版社，1992：

172. [10] 蓝江 . 人工智能的伦理挑战 [N]. 光明日报，2019-04-01(15).
[11] 杨庆峰 . 数据偏见是否可以消除 ?[J]. 自然辩证法研究，2019，35(8)：109-113.

（编辑 姚利芬）

Research on Artificial Intelligence Ethics from the Perspective of Science Fiction Creation

Liang Weiguo

(School of Philosophy and Religion , Minzu University of China , Beijing 100081)

Abstract : The rapid development of artificial intelligence (AI) provides creative materials and scientific evidences for science fiction creation , while science fiction creation provides AI research with future imaginations and developing possibilities. Such separating and interacting relationship forms rich spiritual and sensory experiences. The fundamental reason for these experiences is that AI (which takes data and algorithms as its core technology) has gradually broken through the ethics of anthropocentrism under the guidance of symbolism , connectionism , and behaviorism. This change of anthropocentric ethics may lead to the loss of human control on machines , or even the destruction of humankind. To avoid this predicament , we need to build a new ethic in which human and artificial intelligence could coexist , namely , a quasi-partnership between humans and artificial intelligence based on anthropocentrism.

Keywords : ethics ; artificial intelligence ; science fiction creation ; subjectivity ; anthropocentrism

CLC Numbers : TP18 ; B82-057 Document Code : A DOI : 10.19293/j.cnki.1673-8357.2020.03.005

Analysis on Public Attitude to GMF Risk and Its Cause of Formation: A Study Based on Virtual Ethnography

Cui Bo[1] Lin Fangyu[2]

(Communication University of Zhejiang , Hangzhou 310018) [1]

(Hainan Hinews Co. Ltd , Haikou 570206) [2]

Abstract : The paper investigates how the supporters and opponents of genetically modified food (GMF) in two different QQ groups defend their own stands , and finds that both sides take advantages of existing local knowledge to construct their understanding of GMF risk.

To some extent , traditional Chinese notions of nature and food act as a social amplifier of the risk.

Keywords : GMF ; supporters of GMF ; opponents of GMF ; local knowledge
CLC Numbers : G206.3 Document Code : A DOI : 10.19293/j.cnki.1673-8357.2020.03.006

Empirical Study on Network Communication Power of the National Research Institutions of China: A Case Study on Institutes under Beijing Branch of CAS

Yuan Zhibin[1] Li Meng[2]