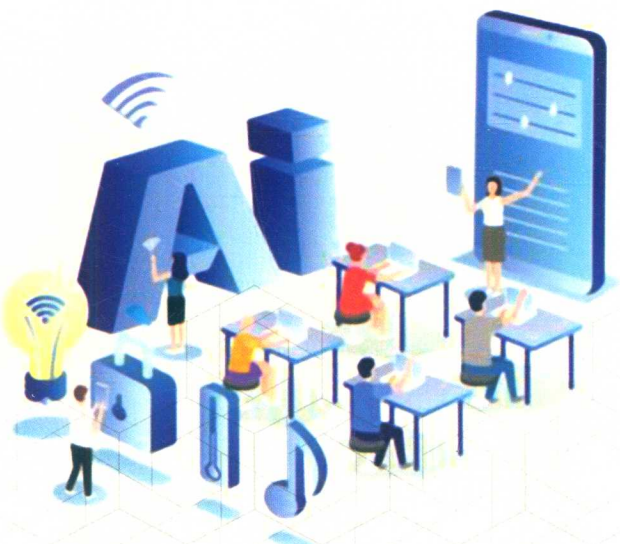
人工智能

技术与伦理的冲突与融合

Artificial Intelligence：The Conflict and Integration Between Technology and Ethics

李勇坚　张丽君 等 著



经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

需要全本请访问 www.ertongbook.com

中国社会科学院创新工程学术出版资助项目

人工智能

技术与伦理的冲突与融合

Artificial Intelligence：The Conflict and Integration Between Technology and Ethics

李勇坚　张丽君 等 著



经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

图书在版编目（CIP）数据

人工智能：技术与伦理的冲突与融合/李勇坚，张丽君等著. —北京：经济管理出版社，2019.9

ISBN 978-7-5096-6924-2

Ⅰ.①人… Ⅱ.①李… ②张… Ⅲ.①人工智能—技术伦理学—研究 Ⅳ.①TP18 ②B82-057

中国版本图书馆 CIP 数据核字(2019)第 216635 号

组稿编辑：申桂萍
责任编辑：刘　宏
责任印制：黄章平
责任校对：王淑卿

出版发行：经济管理出版社
（北京市海淀区北蜂窝 8 号中雅大厦 A 座 11 层　100038）
网　　址：www.E-mp.com.cn
电　　话：（010）51915602
印　　刷：三河市延风印装有限公司
经　　销：新华书店
开　　本：720mm×1000mm/16
印　　张：14
字　　数：244 千字
版　　次：2019 年 9 月第 1 版　　2019 年 9 月第 1 次印刷
书　　号：ISBN 978-7-5096-6924-2
定　　价：68.00 元

·版权所有　翻印必究·
凡购本社图书，如有印装错误，由本社读者服务部负责调换。
联系地址：北京阜外月坛北小街 2 号
电话：（010）68022974　　邮编：100836

前　言

2017 年 7 月，国务院发布了《新一代人工智能发展规划》（以下简称《规划》），大力推动人工智能发展成为国家战略。《规划》指出，人工智能是影响面广的颠覆性技术，可能带来冲击法律与社会伦理等问题。因此，《规划》提出，要“制定促进人工智能发展的法律法规和伦理规范”，“加强人工智能相关法律、伦理和社会问题研究，建立保障人工智能健康发展的法律法规和伦理道德框架”。研究人工智能所带来的伦理问题在国际上已有多年历史。早在 20 世纪 50 年代，控制论的创始人维纳在创立控制论之后，就想到了一个问题，即如果机器能够自己控制机器，那么人还有什么用？从这点出发，维纳于 1950 年出版了《人有人的用处：控制论与社会》一书，论证了在控制论的背景下人类的社会作用。值得深思的是，维纳首次提出了机器可以从事脑力劳动的社会后果，即人脑的贬值。这是一个相对悲观的看法。在 19 世纪时期，很多学者认为机器将取代体力劳动，人类将在脑力劳动上创造出更多的价值。机器大规模替代人类时，卢德分子们十分愤怒，认为机器剥夺了他们的饭碗。但是，艺术家和学者认为，从本质上来说，脑力劳动要优于体力劳动，因此，在社会上永远居于高位。奥斯卡·王尔德认为，“所有非智力劳动，所有单一的、乏味的劳动，所有令人作呕的劳动，所有工作环境恶劣的劳动，都将由机器来完成。未来世界将建立在机器奴役制的基础上，人类将奴役机器”。在这个背景下，人类社会将达到一种理想的状态。“毫无疑问，未来人们将奴役机器，就像大树在不停生长而乡绅却可以酣然入睡。人类将享受闲暇，提高自身修养，这才是人类生存的目的——绝非劳动，而是创造美好的东西，阅读优美的篇章，或仅仅怀着敬畏和愉悦的心情认识世界，而机器会去完成所有那些必要但无趣的工作。”但是在今天，很多脑力劳动正在被机

器所取代，这对人类的地位形成了新的挑战。本书的书名为《人工智能：技术与伦理的冲突与融合》，其本质是要解决人工智能与人类伦理的冲突与融合问题。一方面，人工智能作为一种智能体，其与人类如何相处，人类如何与之相处，这个问题是人类伦理所面临的一个全新挑战，即人机关系问题。因为人工智能已全面挑战了数万年以来的人类中心主义预设。这是人工智能发展过程中技术与伦理发生冲突的重要根源。另一方面，如何将人类的伦理准则嵌入到人工智能的判断与决策过程中，这需要伦理道德等诸多方面的研究来解决。这个问题是一个非常现实的问题，即人工智能已达到了能够独立做出判断及决策的水平，在研发的过程中，必须确保人工智能的这种判断及决策能够符合人类的伦理准则。因此，在人工智能发展过程中，需要将技术研发与伦理准则进行深度融合。

当前，人工智能技术虽然进步很快，但是，无论是对人类中心主义或者人类地位的威胁，还是人工智能独立进行判断与决策的情形，都不是特别普遍，但两种情况均已出现了一定的苗头。例如，自动驾驶技术已处于商业普遍应用的边缘，而自动驾驶体现了人工智能进行独立判断与决策的一部分，这需要为人工智能建立伦理准则。又如，2015 年 1 月，欧盟议会法律事务委员会（JURI）就决定成立一个工作小组，专门研究与机器人和人工智能发展相关的法律问题。2016 年 5 月，欧盟议会法律事务委员会发布《就机器人民事法律规则向欧盟委员会提出立法建议的报告草案》（*Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics*，以下简称《报告草案》）；同年 10 月，发布研究成果《欧盟机器人民事法律规则》（*European Civil Law Rules in Robotics*）。《报告草案》认为，“至少那些最尖端的机器人应该获得电子人的身份，同时它们也应承担相应的责任和义务”。《报告草案》还建议建立智能机器人的注册登记制度，由专门的基金会来承担机器人的法律责任，并让每个机器人都与相应的基金会挂钩。2017 年 10 月，沙特阿拉伯授予机器人索菲亚公民身份。这些都涉及人与人工智能的关系问题。

在国内，百度在人工智能技术方面正在实现超越。2011 年，百度在硅谷设立研究基地，重点瞄准人工智能。2013 年 1 月，李彦宏推动成立专注深度学习的研究院（Institute of Deep Learning，IDL），并担任院长。2014 年 5 月，人工智能领域的领军人物吴恩达博士加入百度在硅谷的研发中心。2016 年，百度正式发布“百度大脑”。2018 年，百度宣布与一汽集团合作生产 L4 级无人驾驶乘用车

需要全本请访问 www.ertongbook.com

Apollo红旗；将向500个贫困县医疗点捐赠AI眼底筛查一体机，帮助5600万风险人群避免失明；百度大脑也进入了3.0阶段，最大的变化是开始实现多模态深度语义理解功能，并且百度大脑的开发能力也从2.0版本的60余项增加到了3.0版本的110多项。随着百度人工智能技术在安防、交通、医疗、教育等各个领域的深度应用，企业的研发人员及应用人员均感觉AI的大规模应用正在带来深刻的伦理问题。

为了更深入地研究这些问题，中国社会科学院财经战略研究院联合百度公共政策研究院，并吸收中国社会科学院哲学所的研究力量，拟对人工智能发展过程中可能出现的一些伦理问题进行研究。研究从人工智能技术与伦理的冲突与融合两个方面进行，以对当前人工智能技术发展过程中所遇到的一些伦理问题做出回应。

本书由中国社会科学院财经战略研究院与百度公共政策研究院共同编写。具体分工为：绪论，李勇坚、张丽君；第一章，李勇坚；第二章，周瑜；第三章，徐金海；第四章，裴彧；第五章，张心宇；第六章，李勇坚、杨刚、张萌；第七章，邢樾；第八章，张鹏飞；第九章，李勇坚、张丽君；第十章，李勇坚；第十一章，李勇坚；后记，李勇坚。

李勇坚（中国社会科学院财经战略研究院互联网经济研究室主任　研究员）

张丽君（百度公共政策研究院　执行院长）

2019年2月28日

目　录

绪　论 …… 1

第一章　人工智能技术对人类现有伦理规范的挑战 …… 17

一、为什么人工智能会引发伦理冲突 …… 17
二、人工智能伦理问题的主要表现 …… 20

第二章　有关人工智能伦理问题的文献综述 …… 39

一、人工智能与人类的关系 …… 39
二、关于人工智能伦理问题的分歧 …… 44
三、人工智能带来的社会伦理问题 …… 47
四、人工智能开发的伦理道德 …… 53

第三章　人工智能伦理问题的国际实践 …… 62

一、国际上人工智能的各种伦理倡议 …… 62
二、IEEE 全球人工智能与伦理倡议 …… 69
三、阿西洛马人工智能 23 条原则 …… 71
四、德国自动驾驶伦理规范 …… 72
五、联合国的人工智能政策与监管 …… 73

第四章　人工智能的权利论 …… 89

一、AI 不应拥有生存权 …… 90
二、AI 不妨拥有财产权 …… 97
三、AI 不应拥有著作权 …… 98
四、AI 不应拥有投票权 …… 100

第五章　人工智能与人类：情感交互 …… 102

一、爱之相互性与人工智能对人类的情感伦理 …… 102
二、人工智能的情感需求 …… 106
三、人类与人工智能的情感交互 …… 112

第六章　人工智能与人类：命令与服从 …… 116

一、人工智能与人类命令的关系 …… 118
二、人工智能对人类命令的遵循准则 …… 120

第七章　人工智能应用带来的人类社会伦理问题 …… 133

一、引言 …… 133
二、透明空间：人工智能读取大量隐私之后 …… 135
三、算法歧视 …… 137
四、当人类不能读懂机器 …… 139
五、科技带来贫困 …… 141
六、安全危机 …… 143
七、结语 …… 145

第八章　人工智能与就业：一个文献综述 …… 146

一、技术进步、生产率提高和就业 …… 147
二、技能偏向型技术进步与就业 …… 152
三、程序偏向型技术进步与就业 …… 155
四、技术进步与女性就业 …… 162

五、结论与展望 …… 165

第九章 自动驾驶的伦理问题 …… 175

一、自动驾驶的利益及研究其带来的伦理问题的必要性 …… 176
二、自动驾驶的四条伦理准则 …… 180
三、自动驾驶中的人机关系 …… 188
四、自动驾驶中的其他伦理问题 …… 192

第十章 行政司法领域引入人工智能的伦理问题 …… 195

一、行政司法领域引入人工智能的可能性及实践 …… 197
二、行政司法领域引入人工智能的意义 …… 201
三、行政司法领域引入人工智能所带来的伦理问题 …… 204

第十一章 结束语 …… 210

后 记 …… 213

绪　论

人工智能（Artificial Intelligence，AI），其本质是对人的意识、思维的信息过程的模拟。当前的人工智能主要应用范围包括语音识别、图像识别、语义理解、预测规划和智能控制等。

近年来，随着大数据的应用与核心算法的不断改进，人工智能的研究持续勃兴，基于人工智能技术的各种产品如雨后春笋般出现，在各个领域代替人类从事简单重复的体力或脑力活动，人工智能真正成为了智能产品的大脑。工业化时代，工业技术（Industrial Technology，可称为老 IT）极大地提升了人类对大自然的征服能力，形成了以机器为特征的大工业生产体系；信息化时代，信息技术（Information Technology，可称为旧 IT）极大地促进了人类之间、人类与万物之间的沟通交流能力，并形成了以信息化为特征的生产体系；进入 21 世纪之后，智能技术（Intelligent Technology，可称为新 IT）极大地扩展了人类的大脑，并建立了一种新的智能生产模式。随着人工智能在各个领域的广泛应用，以及资本对人工智能的巨额投入，人工智能成为国家、产业界与社会各界研究的一个热点。

AI 的领域旨在实现广泛任务的自动化。AI 研究人员研究的典型任务包括玩游戏、引导车辆和分类图像。但原则上，AI 可以改变的任务集是巨大的。至少，人类或非人类动物利用其智力执行的任何任务都可以成为创新的目标。麦肯锡全球研究院（McKinsey Global Institute）提出了一个观点，认为人工智能对社会经济的影响前所未有，这种影响对人类社会的影响正在以比工业革命快 10 倍的速度发生，而预计影响的规模与深度将达到工业革命的 300 倍。2016 年 9 月初，由斯坦福大学组织的名为“人工智能百年研究”持续性项目发布了首项成果——《2030 年人工智能与生活》（*Artificial Intelligence and Life in* 2030）研究报告，认

为到2030年，人工智能将从各个方面影响人类的生活[①]，这些方面都将在很大程度上给现有的人类伦理规范带来挑战。人工智能的研究者与投资者杰瑞·卡普兰（Jerry Kaplan）在2015年时出版了《人工智能时代》一书，在书中他指出，人工智能已经不是一种想象中的技术，而且正在成为人类生活的一部分，这种与人类生活的同化，将使人工智能的应用产生大量的伦理问题，特别是涉及人工智能人权的伦理问题，是前所未有的，虽然与产业有关的问题可以通过市场来解决，但各界仍没有就伦理问题达成一致。高奇琦（2018）认为，人工智能时代可能带来的最大冲击是在地球上突然出现了一个超越人类的新物种：机器人。[②] 机器人的特点是具有智能，与人类有诸多相似之处，从而会给我们现有的伦理道德体系带来巨大的挑战。

技术的进步总会带来社会结构的巨大变化。对这种变化，从历史来看，很多人都喜欢用传统的社会结构及规章制度等来看待，认为变化的本质将导致社会的失范。因此，关于人工智能与人类关系等非技术性问题，也成为研究的热点，并受到了政策的关注。OECD（2017）指出，人工智能的发展伴随着重要的政策和道德问题。历史学家斯塔夫里阿诺斯（L. S. Stavrianos）在《全球通史》[③] 一书中指出："人类历史中的许多灾难都源于这样一个事实，即社会的变化总是远远落后于技术的变化。这是不难理解的，因为人们十分自然地欢迎和采纳那些能提高生产率和生活水平的新技术；但是，人们却拒绝接受新技术所必需的社会变化，因为采纳新思想、新制度和新做法总是令人不快的。"尼克·比尔顿在《翻转世界》一书中，对火车、电话、印刷术、广播、电视等新技术所带来的社会恐慌与焦虑，以及对新技术将带来的社会失范的担忧等问题进行了详细的论述。布雷特·金在《智能浪潮：增强时代来临》的开篇也写道："我没法亲身经历，但是我想轮子的发明在当时意义非常重大。不过，伴随着每件重要的发明，我确信也有牧师、巫师、村中长者、当地商贩或城镇官员警告我们轮子如何不利于城镇，它会如何毁掉工作并可能带来世界末日般的灾难。"瑞恩·卡洛、迈克尔·弗鲁

① 报告从交通、家用/服务型机器人、医疗、教育、娱乐、资源较缺乏群体、公共安全与治安、就业及工作场所八大人类活动领域，分别分析了到2030年时人工智能的应用及影响。

② 高奇琦．人工智能：驯服赛维坦［M］．上海：上海交通大学出版社，2018.

③ 斯塔夫里阿诺斯．全球通史：从史前史到21世纪（第7版）［M］．吴象婴，梁赤民，董书慧等译，吴象雯校．北京：北京大学出版社，2006.

需要全本请访问 www.ertongbook.com

姆金等（2018）则认为，就像互联网一样，机器人技术是一种社会和经济上的变革性技术。当前与机器人技术有关的法律和政策问题类似于学者和决策者在万维网之前对互联网的处理。[①] 也就是说，人工智能会带来政策、法律、伦理、道德等诸多方面的挑战，但是，这些挑战仍然是可以应对的。进一步说，人工智能的社会挑战，需要社会科学家的“参与性塑造”（高奇琦，2018）。

作为对这些担忧的一个回应，中国国务院《关于印发新一代人工智能发展规划的通知》（国发〔2017〕35 号）中，就对人工智能引起的伦理问题给予了关注。文件指出：“人工智能发展的不确定性带来新挑战。人工智能是影响面广的颠覆性技术，可能带来改变就业结构、冲击法律与社会伦理、侵犯个人隐私、挑战国际关系准则等问题，将对政府管理、经济安全和社会稳定乃至全球治理产生深远影响。在大力发展人工智能的同时，必须高度重视可能带来的安全风险挑战，加强前瞻预防与约束引导，最大限度降低风险，确保人工智能安全、可靠、可控发展。”2017 年 1 月，麻省理工学院媒体实验室与哈佛大学伯克曼·克莱因互联网与社会研究中心进行合作，拟深入推进与人工智能相关的伦理问题研究，这项计划预计耗资 2700 万美元。更早些，2014 年由 Skype 联合创始人 Jaan Tallinn 联合各个大学及研究机构在人工智能方面的专家创立了志愿者机构“生命的未来研究所”（Future of Life Institute，FLI），关注人工智能带来的各种社会经济问题。在中国国务院颁布的前述文件中也明确提出“加强人工智能相关法律、伦理和社会问题研究，建立保障人工智能健康发展的法律法规和伦理道德框架”，尤其是“制定人工智能产品研发设计人员的道德规范和行为守则，加强对人工智能潜在危害与收益的评估”，“建立健全公开透明的人工智能监管体系，实行设计问责和应用监督并重的双层监管结构，实现对人工智能算法设计、产品开发和成果应用等的全流程监管”。

因此，对人工智能这种人类历史上最为伟大的发明所带来的伦理挑战，需要进行深入研究。

① 瑞恩·卡洛，［美］迈克尔·弗鲁姆金，［加］伊恩·克尔．人工智能与法律的对话［M］．陈吉林等译．上海：上海人民出版社，2018.

（一）人工智能相关伦理问题的四个层次

人工智能由于与人类本身以及人类社会制度的直接相关性，带来了丰富的伦理学问题。这些伦理学方面的问题，受到了社会各个方面的关注，但是，从现有的研究架构来看，对于人工智能相关的伦理问题，在研究层次方面比较混乱。我们认为，人工智能的伦理问题可分为四个层次：

第一个层次是人类自身决策的伦理问题，在决策权移交给人工智能时，这些伦理问题重新得到了关注。

第二个层次是人工智能作为一种新型的智能体，与人类的关系问题，这个问题受到了最多的关注，也是争议最多的领域。

第三个层次是人工智能作为一种新的技术范式，在广泛应用之后，给社会伦理规范带来的挑战。

第四个层次是人工智能作为一种强有力的技术，如何避免在开发、应用过程中被恶意利用。

（二）被人工智能强化的伦理问题

从工业革命开始，我们就生活在一个理性构建的社会，人类社会在机器化时代发生了巨大的改变。例如，芒福德（1967，1970）在《机器的神话》（*The Myth of the Machine*）一书中，对机械文明给人类带来的挑战进行了深刻的分析，他认为，机械文明使人类之间的差异缩小，社会成为一种机械的缺乏情感的架构，他甚至使用“Megamachine”一词来描述社会和官僚结构，而个人沦为机器的“被动，惰性，琐碎的配件”，这种架构脱离了人类的本性。但是，不可避免的是，随着机器的进化，机器的智能化程度越来越高，机器在很多场合下必须像人类一样做出决策，而这些决策是否符合人类的伦理原则，这就产生了伦理问题。从本质上看，伦理学是关于道德问题的科学。尽管机器在不断进化，人类本性中的自私等因素并没有因此而衰退（Richard Dawkins，2006），因此，人类所面临的很多伦理问题并不会因为人类将决策权交给人工智能而产生变化，在很多情况下，这种伦理问题因为人工智能而变得更加突出。这个问题的实质是，人工智能是否有道德（moral）、不道德（immoral）或非道德（amoral）方面的概念或实践架构，这种架构的创设，是人类自身的问题，还是需要人工智能的共同参

与。更为重要的是，随着人工智能的演化，还需要给人工智能设立一些边界与规则，使其决策行为符合人类的伦理准则。而做到这一点，其实是非常困难的。

在本书的后面章节中将要重点讨论的“电车难题”，其实质就是一个这样的问题：在人类伦理实践中，“电车难题”本身就是一个没有解决的伦理问题，在这种情况下，将这个决策权交给人工智能，这必然带来伦理冲突。而这个伦理冲突问题，本身并不构成将决策权交给人工智能的障碍，更不能将其定义为人工智能所带来的伦理问题。问题的核心在于两个方面：第一个方面是这种决策权的移交，对我们习以为常的“人类中心主义”或者说“人是万物的尺度”这些观念带来了挑战。这个挑战，实质上与人工智能相关伦理问题的第二个层次相关，即人工智能与人类关系所带来的伦理问题。第二个方面则与人工智能有着直接的关系，也就是人工智能用学习的方式进行自我的演化，将规则进行无限推广。例如，在前述的“电车难题”方面，如果我们为人工智能建立一个基本的原则，即人都是平等的，每个人的生命具有同样的价值，那么人工智能能够轻易地做出判断，即五个人的生命比一个人更为重要。这样，我们给人工智能设定一个优先序：两条人命比一条人命重要。这是一个常识性的问题，也能够得到大多数社会公众的接受。然而，这本质上是一种功利主义的伦理原则，这种原则正受到很大的挑战。

在这个优先序的情况下，其实会产生更多的伦理问题，而且是我们日常所能够看到的伦理问题。帕尔玛（2010）列举了另一个非常极端但值得进一步思考的问题，即“山姆案”（Case of Sam）。在这个故事里，山姆代表一个基本上正常、没有什么特色但很“和善”的人，有一天，山姆去医院拜访他唯一的亲戚——年迈体弱的姑妈。在他来到医院的同时发生了五桩紧急医疗事件：一个人需要肝移植，一个人需要脾移植，一个人需要肺移植，一个人需要心脏移植，还有一个人需要新的松果腺。这五位病人都是身居要职、备受爱戴的伟人，他们的死会给许多人造成极大的痛苦和确实的不适。然而，山姆的死却不会受到任何人的哀悼（除了他的姑妈，或许会在难得清醒的时候想起来）。于是医院管理层的最高成员，他们都是功利主义者，将山姆引诱到一间手术室中，将他的内脏摘除，并且把它们分给需要做移植手术的病人们，他们的做法与功利原则吻合：为大多数人寻求最大限度的幸福。

人工智能相关的伦理问题的第一个层次，就是人类自身很难解决的伦理学问题，如何交由机器进行决策。在现实中，很多监管机构、社会大众以及社会学者都希望人工智能技术能够实现公平，具备伦理道德，但这种希望都是虚无缥缈

的。因为伦理学本身并不是逻辑完备的，寄希望于强烈依赖于逻辑能力的人工智能来解决人类社会所面临的伦理问题，这条路基本行不通。更进一步说，人工智能的强逻辑性与强执行力，将使伦理学的逻辑不完备的冲突更加尖锐。这是在研究人工智能所带来的伦理问题时首先需要解决的。

1. 因人工智能与人类关系而带来的伦理挑战

人工智能是人类创造出来的与人类具有高度相似性的智能体，这种智能体不同于人类创造出来的非智能机器，人工智能具有很多曾经被认为仅有人类才有的能力。例如，人工智能具有非常强的逻辑推理能力，在很多情形下能够代替人类做出决策。人工智能也能够对各种机器进行自如的操控；人工智能还能够理解人类的自然语言，并做出类似于人类的响应。这些特征给人工智能与人类关系问题带来了伦理挑战，如人工智能的身份问题、情感问题、命令与服务问题等。这些问题带来了巨大的伦理挑战。正如马克思所指出的，人们迄今总是为自己造出关于自己本身、关于自己是何物或应当成为何物的种种虚假观念。他们按照自己关于神、关于模范人等观念来建立自己的关系。他们头脑的产物就统治他们。他们这些创造者就屈从于自己的创造物。我们要把他们从幻想、观念、教条和想象的存在物中解放出来，使他们不再在这些东西的枷锁下呻吟喘息。我们要起来反抗这种思想的统治。一个人说，只要我们教会他们如何用符合人的本质的思想来代替这些幻想；另一个人说，只要我们教会他们如何批判地对待这些幻想；还有个人说，只要我们教会他们如何从头脑里抛掉这些幻想，这样……当前的现实就会崩溃[①]。

人工智能与人类的关系问题在人工智能研究之初就受到了高度重视。例如，图灵在其1936年发表的《智能机器》一文中，就对人工智能发展对人类带来的影响进行了探讨。早在人工智能发展伊始，领域里就形成了人工智能（Artificial Intelligence，AI）与智能增强（Intelligence Augmentation，IA）两大阵营（马尔科夫，2015）。以人工智能概念的提出者约翰·麦卡锡为首的AI阵营积极模拟人类的能力；而以道格拉斯·恩格巴尔特（Douglas Engelbart）为代表的IA一派则坚信计算机应该被用来加强和扩展人类的能力，而不是取代或模仿这些能力。这两派分歧的背后，是对人工智能与人类关系的一种认知。AI派的核心是人工智能将获得与人类相同的进化路径和资源，从而最终获得一定的权利；而IA派则

① 马克思，恩格斯．马克思恩格斯全集（第三卷）［M］．北京：人民出版社，1960.

认为，人工智能归根到底将只是人类的一种工具，这种工具与其他已经发明的机器并无本质的不同。但是，为了体现人工智能作为一种工具与传统机器的区别，20 世纪 40 年代，科幻小说家阿西莫夫（Isaac Asimov）在《我，机器人》这本小说中提出了著名的“机器人学三大法则”，即第一定律：机器人不得伤害人类个体，或者目睹人类个体将遭受危险而袖手不管（Law Ⅰ：A Robot may not injure a human being or，through inaction，allow a human being to come to harm）；第二定律：机器人必须服从人给予它的命令，当该命令与第一定律冲突时例外（Law Ⅱ：A robot must obey Orders given it by human beings except where such orders would conflict with the first law）；第三定律：机器人在不违反第一、第二定律的情况下要尽可能保护自己的生存（Law Ⅲ：A robot must protect its own existence as long as such protection does not conflict with the first or second law）。从这三条基本定律以及之后诸多补充来看，这些原则的本质是将人工智能作为人类的一种奴隶，强调人类在人工智能面前的至高地位与绝对权威。但是，“高度自主的人工智能系统”（Highly autonomous AI systems）发展之后，人类如何与人工智能相处，如何解决三大定律之间的矛盾与冲突，都是值得深入研究的。

第一个问题是机器人是否有身份的问题。人工智能如果拥有像人类一样的智慧，拥有自然语言理解能力，能够像人类一样进行学习，接受信息，做出判断与决策，那么，我们能否将其视为与我们平等的公民呢？这个问题将成为人工智能与人类关系相关伦理问题的起点，也将引发无数的争论。从理论上看，如果人工智能能够拥有公民身份，那么，它就应该有相应的财产权利等民事权利，甚至其可以成为人类的配偶、参与政治活动等。2017 年 10 月 26 日，沙特阿拉伯授予美国汉森机器人公司生产的机器人索菲亚（Sophia）公民身份，从而使后者成为史上首个获得公民身份的机器人。这种由人造出来的智能生命获得相应的权利之后，如何解决一系列的伦理问题，仍是值得思考的。马歇尔·麦克卢汉曾指出，“任何发明和技术都是人体的延伸或自我截除”[①]。人工智能是人类利用先进的智能技术等创造出来的，这种创造物拥有与其创造者相似或者更高的智能水平（智

① 马歇尔·麦克卢汉．理解媒介：论人的延伸［M］．南京：译林出版社，2011．麦克卢汉认为，信息传播方式或信息媒介的变化，会导致人类对信息整体理解的巨大变化。因此，他被称为“电子时代革命思想的先知”。

能是人类据以区别于其他存在物的一个重要特征），这样就必然产生一个矛盾问题：站在人工智能作为创造物的视角，他必然是人类的奴仆，也就是说，这是一种“主奴关系”，这种关系天然是不平等的。但是，正如黑格尔所指出的，这种主奴关系并非一成不变的，而是辩证的，甚至有可能异化。人工智能让人类更大程度地解放出来，但是，人工智能也会给人类带来危机，当人工智能代替人类做出理性决策，甚至是审美等诸多方面的决策时，人类自身的内涵将会发生深刻的变化，人类与人造物之间的天然界限将会泯灭。而在这种情况下，人工智能更为强大的计算能力，将有可能在深处奴役着人类，使人类沦为“物”的水平。

第二个问题是超级智能带来的对人类征服的可能性。科幻电影对人工智能，尤其是超级智能在智能水平超过人类之后的很多情景进行了构想，这些构想深刻地影响了普通社会公众对人工智能的态度。而在科学家以及社会学家内部，对超级智能现象的看法也充满了矛盾。来自牛津大学的哲学家、知名人工智能思想家尼克·波斯特洛姆（Nick Bostrom）把超级智能定义为“在几乎所有领域都比最聪明的人类大脑都聪明很多，包括科学创新、通识和社交技能”。[①] 在《超级智能》一书中，他对这种超级智能给人类发展带来的可能后果进行了详细的描述。迈克斯·泰格马克（Max Tegmark）在其新出版的《生命3.0》[②] 一书中，对超级智能给人类可能带来的影响也进行了详细的论述。他们的观点是，超级智能发展对人类社会的影响，取决于人类如何适应超级智能，并不一定会带来人类被奴役乃至于被毁灭的后果。而《时间简史》的作者史蒂芬·霍金（Stephen Hawking）在谈到人类的命运时，曾经提出，随着技术的进步，有极大可能出现一种技术，这种技术将会威胁人类的生存[③]。在人工智能快速进步并普遍应用之后，他认为，人工智能这种技术可能就是他提出的那种“会威胁人类生存”的技术，并在多个公开场合对此表示担忧。伊隆·马斯克（Elon Musk）[④] 是对人工智能的威

① 尼克·波斯特洛姆. 超级智能：路线图、危险性与应对策略［M］. 张体伟，张玉青译. 北京：中信出版社，2015.

② 迈克斯·泰格马克. 生命3.0［M］. 汪婕舒译. 杭州：浙江教育出版社，2018.

③ 但是，霍金认为，这种技术可能并不会导致人类的灭绝，因为在这种技术广泛应用之时，人类可能会实现太空移民。

④ 马斯克（Elon Musk）现任太空探索技术公司（SpaceX）CEO兼CTO、特斯拉公司首席执行官、太阳城公司（Solar City）董事会主席。2018年2月7日4时45分，SpaceX公司的“重型猎鹰”运载火箭在美国肯尼迪航天中心首次成功发射，并成功完成两枚一级助推火箭的完整回收。

需要全本请访问 www.ertongbook.com