分类号_____ 密级_____
UDC _____编号_____

云南师范大学
YUNNAN NORMAL UNIVERSITY

硕士学位研究生学位论文

论文题目：人工智能技术应用的伦理问题研究

英文题目：The research on the ethical issues of artificial intelligence applications

学　　院　马克思主义学院

专 业 名 称　科学技术哲学

研究生姓 名　杨 帆　学 号 14010108006

导 师 姓 名　杨 胜 荣　职 称 副教授

2017 年 5 月 18 日

独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：杨帆

2017年5月26日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权云南师范大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

学位论文作者签名：杨帆　　指导教师签名：杨胜荣

2017年5月26日　　2017年5月26日

摘要

人类进入20世纪后，一门只有60年历史的年轻科学----人工智能技术正以其特有的方式影响着世界。人工智能技术被称为本世纪三大科技成就之一，人工智能技术的应用有：机器学习、专家系统、模式识别、自然语言处理、人工神经网络、机器人等。这些人工智能技术的应用给自动驾驶、安检等领域带来了新的机遇。随着人工智能技术应用的快速发展，人工智能技术的应用也带来了许多伦理问题。

对此，论文主要分四个部分来阐述：第一个部分，介绍本文研究的意义和背景，阐述了前人所做的成果。第二个部分，介绍人工智能技术应用，它们分别是机器学习、模式识别、专家系统、自然语言等,以及人工智能技术应用的内容。第三个部分，首先考察人工智能伦理评价的义务论与功利论之争，进而分析人工智能技术异化的伦理风险、人工智能的智能和自主性进化对人类主体性的挑战。第四个部分，阐述人工智能技术应用的伦理原则与伦理对策。

任何科技都是一把双刃剑，有好的一面也有坏的一面。机器学习的应用使阿尔法狗在围棋比赛中赢了人类的世界冠军棋手，却给人类带来了机器智能超越人类智能的担忧；机器学习和图像识别技术让无人驾驶车给老年人、小孩带来方便的同时，也给人类在给机器设计道德上带来了困境。机器学习让人脸识别给安检带来了方便，但是却又带了侵犯隐私的伦理问题。无论是义务论还是功利论，亦或是美德伦理学，都没有一个是完美的理论，只有从多元的伦理理论视角来分析，才不至于片面，才相对全面。因此，需要认识人工智能技术应用的伦理问题，密切关注它的发展，更好地管控风险，让人工智能技术的应用更好地造福于人类。

关键词：人工智能技术；应用；伦理问题；义务论；功利论

Abstract

After entering 20th century, an only 60-year old young science, artificial intelligence technology is influencing the world in its unique way. Artificial intelligence technology is known as one of the three major science and technology in this century, the application of artificial intelligence: machine learning, pattern recognition and expert system, natural language processing, artificial neural network, robot and so on. The application of artificial intelligence technique for automated driving, security and other fields has brought new opportunities. With the rapid development of artificial intelligence technology, the application of artificial intelligence technology also brought a lot of ethical issues.

Concerning this, the thesis is mainly divided into five parts: the first part, introduced in this paper, we study the meaning and background, this paper expounds the achievements of predecessors did. The second part, this paper introduces artificial intelligence technology, they are machine learning, pattern recognition and expert system, such as natural language. And the content of the artificial intelligence technology. Including the concept of artificial intelligence technology, a brief history, development, as well as the specific domain. The third part deals, firstly, with the debate on ethical evaluation of AI between Deontology and Utilitarianism. Secondly, it analyses the ethical risks brought about by the alienation of AI technologies, and the challenges to subjectivity of human beings, brought about by the development of the intelligence and subjectivity of AI. The fourth part of this paper elaborates the ethical principles and strategies when AI technologies are applied.

Every coin has two sides, every sience has a good side and bad side. The application of machine learning to alpha dog won a chess game at the human world champion chess player, has brought human machine intelligence beyond the concerns of the human mind; Machine learning and image recognition technology for unmanned vehicles for the elderly, children bring convenient while, also brought human to machine a moral dilemma. Machine learning to face recognition brings security convenient, but also ethical issues with the invasion of privacy. Theory of

deontology and utilitarianism, or no ethics, no one is perfect theory, only according to the analysis of theory of multiple ethical perspective is not one-sided, only relatively comprehensive. Understanding the ethical problems of the application of artificial intelligence technology, pay close attention to the development of it, the better control risk, make the application of artificial intelligence technology to better benefit the mankind.

Keywords ： Artificial intelligence technology ； Application； Ethical issues； Deontology ； Utilitarianism

目录

第一章 绪论

第一节 问题的提出及研究意义

一、问题的提出

人工智能从达特茅斯会议第一次提出以来，像其他学科的发展一样遇到过冷冻期，现在又因为脑学科、认知学科的迅速发展而迅速发展。人工智能被称为本世纪三大科技成就之一。其中人工智能技术的应用，比如机器学习催生了很多新的岗位和产品应用，掌握人工智能技术应用的人才成为了各大 IT 公司争抢的“香饽饽”。陆琪入职百度，李飞飞入职谷歌都带来了许多社会的关注。人工智能技术的应用如机器学习、专家系统、人工神经网络、机器人等开辟了新市场，为人们的医疗诊断、交通运输、教育、刑事司法、金融、环境和能源等重要领域都带来了许多新的机遇。人工智能技术的应用毫无疑问给我们开辟了一条人类开发自身智力的道路，但是人工智能的技术应用也带了许多伦理问题，引发了大量的争议。本文在前人的基础上，将对人工智能技术的应用所产生的伦理问题进行比较系统的分析，这是本文的出发点。

二、研究意义

现在人工智能技术的应用如火如荼，可能会在未来持续为经济贡献力量，也是改善世界面貌的很宝贵的工具。机器学习、专家系统、模式识别等等，受到各界的追捧。因为人工智能技术的应用不断快速发展，所以问题也越来越多，本文从伦理学视角分析人工智能技术应用所带来的问题，通过阐述人工智能技术应用的概念，理清人工智能技术的应用领域，通过人工智能技术应用过程中的典型案例展开分析，以便为人工智能技术应用的伦理学研究努力贡献出自己的一份力量。需要了解人工智能技术应用中面临的各种伦理问题，才能有效地促进设计和开发符合人类需求的人工智能技术应用方法。这也是笔者想要达到的预期效果。

第二节 文献综述

（一）人工智能技术与伦理学是什么关系？

人工智能技术与伦理学有什么关系呢？1998 年胡增顺在开封大学学报发表的《人工智能的伦理思考》[①]一文中，从本体论角度出发，分析人工智能与伦理学之间的联系，得出“人工智能绝不单单是电脑技术问题，它与社会学、心理学、伦理学等诸多方面有着千丝万缕的联系。所有这些都将是人工智能研究过程中所必须思考的”[②]的结论。

如果说伦理不会凭空产生，而是在相互作用中才能体现，那么人工智能产生了哪些伦理问题呢？2010 年朱勤和王前在《欧美工程风险伦理评价研究述评》一文中通过对欧洲工程风险史的简述，指出将技术看作是对生活的威胁是一种悲观的看法，《欧美工程风险伦理评价研究述评》这篇文章不认为工程风险是价值中立的。朱勤和王前在文中认为，工程风险在多大程度上能够被接受，是一个涉及公正问题的伦理问题，与工程的伦理分析家们、公众、决策者的伦理价值观有关系。[③]2008 年张一男在《人工智能技术的伦理问题及其对策研究》[④]一文中，分析了人工智能技术的三个伦理问题即：1，人权伦理问题，也就是说人工智能的目的是模拟人类，当人工智能发展到一定程度后，是否应该对这些人造生命给予人权，还是它们应该继续被人类变相地奴役。2，道德伦理问题，张一男在文中提到如果人类造的这些“人工生命”一旦某天具有了人类的情感、意识后，如果继续让它们代替人类做危险的工作等，是不尊重 “人权”的，会让人类的道德体系崩溃的危险。3，代际伦理问题，张一男指出年轻人和老人有交流的代际问题，类比人工智能可能有类似伦理问题。该文作者假设了人工智能将来有一天有繁衍后代的能力，那么就有导致人类代际伦理出现危机。2016 年杜森在《人工智的法律与伦理意识形态问题研究》[⑤]一文中从法律的角度提出了这些伦理问题。他认为人工智能的发展一方面促进了法律的公平，同时也给法律和伦理的制定带来了混乱。这一挑战主要是人工智能主体的界定，因为人工智能或者机器人是否具有道德能力，或者什么样子的人工智能需要为自己的社会行

① 胡增顺. 人工智能的伦理问题[J]. 开封大学学报, 1998, 12(3): 57-59.
② 胡增顺. 人工智能的伦理问题[J]. 开封大学学报, 1998, 12(3): 57-59
③ 朱勤, 王前. 欧美工程风险伦理评价研究述评[J]. 哲学动态, 2010(9):41-47.
④ 张一南. 人工智能技术的伦理问题及其对策研究[J]. 吉林广播电视大学学报, 2006(11): 123-124.
⑤ 杜森, DuSen. 人工智能的法律与伦理意识形态问题研究[J]. 黄冈职业技术学院学报, 2016, 18(2):64-67.

为负责？如果人工智能的行为给社会或者他人造成损害，谁需要为此负责？是研发者还是运营者？还是使用者？武汉理工大学的李俊平在《人工智能技术的伦理问题及其对策研究》这篇硕士论文中，通过对人工智能技术的分析，论述了人工智能技术存在的主要伦理问题有:人权伦理问题、责任伦理问题、道德地位伦理问题、代际伦理问题、环境伦理问题；在此基础上，比较详细地做了成因分析，并且给出了解决的对策。《人工智能技术的伦理问题及其对策研究》这篇文章对于人工智能技术的伦理问题与其他文献相比，相对系统。 笔者的硕士论文与李俊平的论文的不同之处在于，李俊平写作切入的点是技术，而笔者写作的切入点是技术应用。李俊平是对人工智能技术的伦理问题进行研究，而笔者关注的是人工智能技术应用的伦理问题。

（二）对人机关系是应该悲观还是应该乐观？

对人机关系持有悲观态度的有，玛丽·雪莱在《弗兰肯斯坦》[①]一书里面讲述了这样一个故事，主人翁维克多·弗兰肯斯坦是一个科学家，他热衷于生命的起源，创造了一个无名无姓的“怪物”，然而等到他完成了，当他看着自己创造的丑陋的巨人，他感到厌恶，于是抛弃了相貌丑陋、身形高大的“怪物”。“怪物”被人们歧视，但是渴望得到友谊、伴侣的爱情等。于是“怪物”不断地找他的创造者弗兰肯斯坦，这过程中造成了弗兰肯斯坦的弟弟、朋友、妻子死亡等一系列悲剧，等到复仇的弗兰肯斯坦积劳成疾，含恨而死在北极冰川后，“怪物”也最终以火结束了自己的生命。这是一个悲伤的故事，有科幻的成分，但是作者向我们展示了如果人类创造了一个新的类似人的生命，有人类的智能，能够像人类一样思考，甚至像人类一样有情感，也像人类一样渴望得到尊严、友情、爱情，人类该如何去协调这类矛盾，如何做到协调与平衡。因为如不做到协调与平衡，就可能会给科技和人类带来类似的伤害。2016 年，魏鸿在《科幻电影中人机关系认同研究》硕士论文中，通过对阿西莫夫人机认同伦理准则进行分析，认为科幻电影中，人类对于机器人的身份认同的探索就是对人类自身的认同的探索，机器人身份探索的失败也预示人类自身认同的波折。文章作者借用吉登斯的话指出“转换的每一片段都会变成我们的认同危机，而认同的主体变得虚无就会导致危机。”[②]

对人机关系持有乐观态度的有，林德宏教授在《人与机器——高科技的本

① 玛丽·雪莱著，胡春兰、侯明古译. 弗兰肯斯坦[M]. 北京:人民文学出版社, 2004.
② 魏鸿. 科幻电影中人机关系认同研究[D].华东师范大学.2016.

质与人文精神的复兴》[①]一文中讨论了人机共生论这种观点。他认为人机共生论是一种双主体的观点，认为人的主体地位是不容挑战的。他认为机器只是人类的工具，机器的智慧是人类智慧的延伸和体现。林德宏院士认为人机可以共生。2016 年，陶悦宁在《论西方科幻电影中的新型伴侣关系》一文中，以人和人工智能体谈恋爱为切入视角，讨论了类似于由导演斯派克·琼斯(Spike Jonze)导演的电影《她》中萨曼莎与西奥多之间形成的新型伴侣关系。通过对新型伴侣关系的思考，提出了“人的生命究竟如何？人的边界如何确定？人类未来又将成为什么样子？”的问题。并认为，人工智能最后能否超过人类，关键在感性方面。接着提出了“机器或者程序一旦获得了情感学习，那么它们的身份该如何定位？拥有理性与感性双向能力的人工智能体，是否应该被纳入到人类的道德体系当中?造物者是否应该考虑被造物的幸福发展?人类将依然生活在人类中心主义的光圈下，还是被不断学习的人工智能体丢弃在这个旧世界中?”[②]2010 年，姚洪阳在《试论人机关系的历史发展及其文化考量》这篇硕士论文中，认为，人工智能与人类的关系，只是新阶段的机器与人类的关系的表现形式而已。该文作者认为在不同的历史条件下，机器与人类关系有不同的表现内容。[③]他认为人与机器的对抗的问题很突出，是因为人与机器的关系背后是人与人之间不平等的关系的体现。他也肯定了机器有时候是可以帮助人类做很多事，是人类的助手，人机关系存在着对抗的同时，机器与人类的关系也有着和谐的一面。机器的发展，从模仿人类的手到模仿人类的脑的发展方向发展。他还认为，机器是无心的，而工程设计人员、技术专家、科学家是有道德能力的。他认为只有人类才有义务和责任对人类、对社会负责。在文章的后面部分，写了机器人伦理的两个层次的内涵。第一种，认为机器人是人类社会的准他者，是从主体地位方面谈的，以人类为主体的伦理规范来看待他者机器人。第二种是从机器人的设计和制作方面谈的，机器人的设计和制作要考虑人类的伦理规范。 姚洪阳认为思考人类自身会变得怎么样比思考机器会怎么样更重要。他认为是人应该得到更多的关注而不是机器得到更多的关注。他还认为，机器人伦理是时代发展的产物。[④]

对人机关系提出警醒的有，2007 年法国的布律诺·雅科米在《PLIP 时代:技

① 林德宏. 人与机器:高科技的本质与人文精神的复兴[M]. 江苏教育出版社, 1999.
② 陶悦宁. 论西方科幻电影中的新型伴侣关系[J]. 影视长廊, 2016, (3): 39-44
③ 姚洪阳. 试论人机关系的历史发展及其文化考量[D]. 长沙理工大学, 2010.
④ 姚洪阳. 试论人机关系的历史发展及其文化考量[D]. 长沙理工大学, 2010.

术革新编年史》一书中从技术的发展讲了社会的发展史，对技术的革新进行了历史的回顾。对于机器的发展讲了三个阶段，第一个阶段是机器对于人类的手或者肢体的延伸，第二个阶段是机器对于人类的感官的延伸。比如，可以看到目前阶段，很多研发机构或者企业就有对于机器人视觉、听觉的模拟的技术。第三个阶段是对人类智慧的延伸。这个时候就有逼近人类智能并且机器控制人类的危险。书中从控制论的角度对人类与新阶段的机器进行了分析和论述。布律诺·雅科米在该书中对机器与人类的敌对或者竞争关系进行了思考，认为社会文化因素是很重要的。[①]计海庆在《'机器人'观念的形成及其影响的哲学考察——以分析技术的本质和人与机器(人)关系为视角》[②]这篇博士论文中，认为人机可以共生，不能把计算用在人身上，不然会造成人机关系的扭曲。王晓楠在《机器人技术在发展中的矛盾问题研究》一文中分为，"机器与机器人的矛盾"中提到，机器与机器人的矛盾；"人与机器人的矛盾"；"机器人自身的矛盾"，提出要处理好人和未来机器人之间的关系。王晓楠认为机器与机器人的区别是机器对人的模拟的程度决定的，即人工智能的发展程度。人和机器人的关系是人和物的关系。他认为机器人不能取代人，所以不用担心，因为机器人没有人类的社会性。[③]王晓楠认为机器人技术在应用的过程中产生的伦理问题，包括:"机器人影响了人的创造性的发挥;影响了对人的主体地位的界定;发展机器人技术忽视了人文精神的发展。"[④]他提出了"人类是否应该对自己创造出来的机器人负责?人应该怎样对待机器人?如何看待机器人犯下的错误?机器人的社会地位能否真正得到承认，即使得到承认了，社会地位会和人一样平等吗?"[⑤]这些问题。这些都些机器人技术引发的伦理思考。在此基础上提出了两条对机器人技术使用中的伦理问题的改进措施是：增强工程师和哲学家的对话；确立人的主体地位，增强人的道德责任感。最后呼吁，处理好人和未来机器人之间的关系。

社会也已经存在这样的一种共识，就是需要通过研究来创建人类和人工智能系统之间的有效交互。比如有的新闻中提到的手术机器人达芬奇，它在完成任务的过程中就是和人的相互协作。笔者认为，大部分人类开发出来的人工智能，带来的人机关系可以是一种相互协作的关系，不一定是一种取代关系。

① 布律诺·雅科米, 雅科米, Jacomy,等. PLIP时代:技术革新编年史[M]. 中国人民大学出版社, 2007.
② 计海庆. "机器人"观念的形成及其影响的哲学考察——以分析技术的本质和人与机器(人)关系为视角[D]. 复旦大学, 2005.
③ 王晓楠. 机器人技术发展中的矛盾问题研究[D]. 大连理工大学, 2011.
④ 王晓楠. 机器人技术发展中的矛盾问题研究[D]. 大连理工大学, 2011.
⑤ 王晓楠. 机器人技术发展中的矛盾问题研究[D]. 大连理工大学, 2011.

（三）人工智能能否超越人类智能？

对于人工智能能否超越人类智能，存在着不同的声音。有一部分学者认为人工智能不会超越人类智能。

中山大学的翟振明教授在《“强人工智能”将如何改变世界——人工智能的技术飞跃与应用伦理前瞻》[①]一文中认为，人工智能大爆发是一系列老问题的叠加而不是新问题，不是对人以前的人类经验的消灭，而是对以前的经验的颠覆。他认为，人工智能是否能赶上人类智能的设问的前提是人类心智得到最终解释。也就是说如果人类心智现象没有得到最终解释，那么人工智能超越人类智能的设问也就没有大前提。他认为人工智能在下棋方面，无论哪方赢了都是人类的胜利，因为都是人类的的智能。不这么认为的人，“不是神化了下棋技艺的智力本质，就是给下棋程序横加赋予了‘人性’特质。”[②]他认为，智能再强大也是机器，是人类的不会闹情绪的“秘书”。日益强大的智能，与超越人类或者威胁人类的后果并没有什么太大联系。他还认为，“关于人工智能的忧虑中，最为值得关切的是人工智能的应用伦理及其价值植入的技术限度。”[③]学者周昌乐在《无心的机器》[④]一书中认为机器的智慧不一定按照人类的概念的来定义，并且他认为人类不会被机器取代，因为人类能被机器取代的时候，人类不存在了，也就不存在人机关系了。在《人工智能“风口”，医疗与金融先起飞?》这篇报道中采访的专家认为中国人工智能的发展和欧美、德国、日本相差了 10 年左右。但是中国人口多，有数据优势，并且国家重视，有相关政策。这篇报道中还提到，目前阶段，人工智能有这么几件事目前还不能做到，“比如说第一件事情是环境的适应性，这个机器目前是很难做到的;第二个人的想象力、创造力，机器相当长的时间内是不可能替代的;还有人伦情感这种东西，是人这个物种本身所具备的。但在记忆、类别分类、推理等方面，智能计算的能力是太强了，这个方面已经显现出来很强的能力。”[⑤]

有一部分学者认为人工智能会超越人类智能。

雷·库兹韦尔被比尔·盖茨称为预测人工智能的最权威人士，在他的《奇点临

① 翟振明. “强人工智能”将如何改变世界——人工智能的技术飞跃与应用伦理前瞻[J]. 人民论坛·学术前沿, 2016(7):22-33.
② 翟振明. “强人工智能”将如何改变世界——人工智能的技术飞跃与应用伦理前瞻[J]. 人民论坛·学术前沿, 2016(7):22-33.
③ 翟振明. “强人工智能”将如何改变世界——人工智能的技术飞跃与应用伦理前瞻[J]. 人民论坛·学术前沿, 2016(7):22-33.
④ 周昌乐. 无心的机器[M]. 湖南科学技术出版社, 2000.
⑤ 徐豪, 邹锡兰. 人工智能“风口”,医疗与金融先起飞?[J]. 中国经济周刊, 2016(35):31-33.

近》[①]这本书中以一种崭新的视角对于未来人工智能科技发展做了分析，提到机器的智能将在 2045 年超越人类的智能。杜严勇在《人工智能安全问题及其解决进路》一文中认为，人工智能未来可能会超过人类，该文作者认为，人工智能超过人类不可怕，而是需要人类对人工智能拥有控制的能力和权利。该文作者在分析人工智能安全问题的解决路径时候指出，要限定人工智能的适用范围，还举了克隆人的例子来进行了论证，不能等到危险产生了再做伦理分析。指出要限定人工智能的智能水平，以及会有公众对人工智能的恐惧与担忧，所以要把公众对于人工智能的接受与否以及接受程度也要作为伦理问题进行考察。[②]最后文章指出了一条可参考的路径，“解决人工智能安全问题在一定程度上就转变成如何保证人工智能安全管理系统的可靠性问题。”[③]贺欣晔在《科幻文学中人工智能与人类智能的关系》一文中，从进化论的角度，对于人工智能将来会不会超越人类，论述了三种观点。第一种是，认为人工智能只是人类实践的物，是一种工具论，认为人工智能的智能是不会超越人类的智能的。第二种认为人工智能有学习能力，将来某天会超过人类。第三种是接近论，认为人工智能会进步但是只是无限接近人类的智能却不会超过人类的智能。文中还提到了，人工智能无道德论和人工智能可以拥有道德论。还提到了在科幻文学作品中存在着三中关于人工智能的自由的观点。第一种是认为人工智能随着意识还有情感的发展，人工智能会远远超越人类，人工智能会限制人类的自由。第二种认为人工智能随着情感、意识的发展会和人类争夺自由的权利。第三种认为在某一个阶段内人工智能会和人类存在着矛盾，但是由于人工智能来源于人类的智能，两者的矛盾最终会在一段时间后和解。[④]

笔者通过阅读文献发现，随着科学的发展，存在人工智能无限逼近人类的智能的可能性。但是由于人类的大脑还有认知过程是很复杂的，所以要让机器模拟出来，这个研究过程是需要时间的。

（四）怎么给人工智能设计伦理？

哥伦比亚大学的彼特·丹尼尔森教授(Peter Danielson)于 1992 年，在《Artificial Morality : Virtuous Robots for Virtual Games》[⑤]这本书里面第一次使用计算机技术

① RAYKURAWEIL. 奇点临近[M]. 机械工业出版社, 2015.
② 杜严勇. 人工智能安全问题及其解决进路[J]. 哲学动态, 2016(9):99-104.
③ 杜严勇. 人工智能安全问题及其解决进路[J]. 哲学动态, 2016(9):99-104.
④ 贺欣晔. 科幻文学中人工智能与人类智能的关系[J]. 沈阳师范大学学报(社会科学版), 2016, 40(2):111-115.
⑤ Danielson P. Artificial Morality : Virtuous Robots for Virtual Games[J]. Routledge, 1992.

对于建立了机器人道德模型进行了探索。这是开创性的，当时属于机器人伦理学领域的最高水平。[①]

王悠然在《预防人工智能伦理缺失》这篇报道中，提到了研发伦理机器人的两种方式，一种是机器学习，也就是不是预先写入代码而是需要机器自己通过学习积累经验，这样做会造成的难题是，一部分人为此感到困惑，因为不知道机器在做抉择的时候是按照什么伦理准则来的，他们需要按照指令做事才能感到安全。报道中提到的另一种是预先在机器程序中写入程序，告诉它应该怎么做。但是也有人提出困惑，因为伦理准则遇到冲突的时候需要分析情况分别对待。[②]2014 年，王东浩在博士论文《机器人伦理问题研究》[③]中讨论了机器人能否成为道德主体、人类能不能允许机器人成为道德主体等问题，从道义系统和功利主义的视角，对机器人伦理的设计提出了三种路进：1，一种是至上而下,2，一种是至下而上,3，还有一种是两种结合的。接着阐述了能不能让机器人有责任心？怎么样认识机器人的责任？最后得出了要对机器人伦理进行有效地预警的结论。王绍源在《论瓦拉赫与艾伦的 AMAs 的伦理设计思想——兼评<机器伦理:教导机器人区分善恶>》[④]一文中，先讲了，随着发展世界需要建设人工物道德行为体(artificial moral agents，以下简称 AMAs)，所以在人工智能领域兴起了人工物道德和机器人伦理的讨论。一些计算机科学家、工程师、哲学家也投入到该领域的研究中。然后引出这本书的作者瓦拉赫与艾伦把图灵测试这个论题，以一种新的道德维度引入机器智能。接着写了机器智能道德体的前提是“强人工智能”，瓦拉与艾伦认为实现强人工智能的过程是很不容易的，因此实现机器智能道德体的过程也是很艰难的。接着写了三重区分的理论，第一种是“自上而下式”的，是基于康德义务论的。第二种是“至下而上式”的。第三种，是一种混合型机制。最后认为瓦拉赫与艾伦建设机器道德体系统是一种开创性与全新的研究课题。[⑤]

基于案例设计的有，于雪和王前在《“机器伦理”思想的价值与局限性》[⑥]一文中认为，人与人的直接伦理通过机器转变为人与人的间接伦理。还指出了“机

① 何华灿，李太航等. 人工智能导论[M]. 西北工业大学出版社，1988.
② 王悠然. 预防人工智能伦理缺失[M]. 北京:中国社会科学报，2015. 1-2.
③ 王东浩. 机器人伦理问题研究[D]. 南开大学，2014.
④ 王绍源. 论瓦拉赫与艾伦的 AMAs 的伦理设计思想--兼评《机器伦理:教导机器人区分善恶》[J]. 洛阳师范学院学报，2014(1):30-33.
⑤ 王绍源. 论瓦拉赫与艾伦的 AMAs 的伦理设计思想--兼评《机器伦理:教导机器人区分善恶[J]. 洛阳师范学院学报，2014(1):30-33.
⑥ 于雪，王前. “机器伦理”思想的价值与局限性[J]. 伦理学研究，2016(4):109-114.

器伦理思想的局限性之一是其本质上是将人类的道德行为转换为可以计算的数字，这是对伦理可计算性的一种认可。”文中还提到，目前看来，将人类的自然语言表达成计算机的算法的形式，存储伦理程序现在存在着两种程序设计形式：“基于伦理原则”或者“基于案例”。[①]认为应该采取至下而上的路径的有，杜严勇在《机器伦理刍议》一文中认为，“选择何种伦理理论作为机器伦理的指导原则，如何解决哲学语言的模糊性与计算机程序的精确性之间的矛盾等问题是实现机器伦理可能遇到的主要障碍。”[②]他还认为，至下而上的伦理设计路进具有灵活性，只是对机器学习还有道德理解能力的要求比较高，需要长期努力才能在某种程度上达到。文章提到的康德义务论设计进路，笔者赞同文中引用专家观点表达认可康德的义务论具有形式的观点有可取之处。因为人类的语言模糊，机器算法程序需要明确的语言告诉机器人“应当做什么”，而康德义务论具有普遍的形式法则，而且可以明确地告诉机器人“应当”做什么。这是处于机器人还没有情感阶段，未来如果机器人具备情感那么可能就要继续讨论进一步的问题。因为有学者认为“义务论”侵害主体的自由、权益，是一种必要的以“恶”制恶。

（五）机器人有道德主体地位吗？

美国佐治亚理工学院“移动机器人实验室”负责人罗伯特·阿尔金(Ronald Arkin)教授是美国机器人伦理领域的代表。他在《Governing Lethal Behavior in Autonomous Robots》[③]这本书中认为机器人伦理的难点是，对于机器人是否有道德主体地位的界定。他在对机器人在医药和军事领域的应用作了分析。还讨论了机器人在社会角色进行了讨论。耶鲁大学生命伦理学跨学科研究中心的温德尔·瓦拉赫(Wendell wallach)和印第安纳大学认知科学工程中心的科林·艾伦(ColinAllen)合著的《机器伦理：教导机器人区分善恶》《Moral Machines: Teaching Robots Right From Wrong》[④]这本书从理论和实践两方面讨论了机器人伦理学，包括机器人是否能成为行为的主体，从而能不能成为道德主体，以及自治系统能不能实现等内容。

有一部分学者认为机器人有道德主体地位。

---

① 于雪，王前. “机器伦理”思想的价值与局限性[J]. 伦理学研究，2016(4):109-114.
② 杜严勇. 机器伦理刍议[J]. 科学技术哲学研究，2016，33(1):96-101.
③ Arkin R C. Governing Lethal Behavior in Autonomous Robots[J]. Crc Press，2009，37(2).
③ Lisa Damm. Moral Machines: Teaching Robots Right from Wrong[M]. Oxford University Press，Inc. 2008.
④ Lisa Damm. Moral Machines: Teaching Robots Right from Wrong[M]. Oxford University Press，Inc. 2008.

王东浩在《人工智能体引发的道德冲突和困境初探》[①]一文中认为，“理论层面上，我们需要意识到人工智能体也属于道德主体的一部分，其在权利与义务上与人类是等同的。”杜严勇在《机器人伦理研究方法论原则》一文从阿西莫夫的机器人三原则出发，指出有的学者认为阿西莫的三原则是一种人类中心主义，极度放大人类利益的存在，对其他生物物种的利益加以否定。杜严勇不赞同这种观点，并且对此进行了论述。他说非人类中心主义的从本质上讲还是为了保护人类的利益，而且人工智能是人类造出来的，这与大自然的生命还是有很大差异的。换句话说，就是阿西莫夫认为的机器人法则是有一定的缺陷，但机器人伦理是从人类利益这一理论层面延伸出来的。”[②]

王绍源在《应用伦理学的新兴领域:国外机器人伦理学研究述评》一文中认为“我们必须正视机器人的伦理问题和社会影响。”[③]王绍源在《国外机器人伦理学的兴起及其问题域分析》这篇文章中，通过机器人的应用产业着手进行梳理，分析了机器人伦理的问题域，但是对于人工智能技术应用的伦理问题讨论的不够具体、不够详细。笔者将在自己的硕士论文里面做到更具体、更详细。

还有一部分学者认为人工智能模拟人类的情感太困难了，人工智能不可能有人类的认知，也没有道德主体地位。

（六）军事机器人有哪些伦理问题？

2013 年卫华在《杀人机器人的伦理问题及辩护》一文中，论述了致命性自主机器人，也就是杀人机器人，在战场上是武器还是作战主体的伦理地位难以确定，因为，杀人机器人只有行动能力却没有责任能力。文中还提到，杀人机器人没有人的情感、不会疲劳、也不会感到饥饿，战争机器人不会区分军人和平民，一旦用上战场，将威胁到人类生命安全。如果不对杀人机器人加以限制，那么杀人机器人走向战场了，将会决定人类生死。作者从人道主义和功利主义视角讨论了杀人机器人的利弊，认为杀人机器人既不符合人道主义又不符合功利主义，该文作者认为是危害大于所带来的利益，促进和维护世界和平。该文作者反对杀人机器人的研发，呼吁限制杀人机器人研发。[④]2014 年杜严勇在《现代军用机器人的伦理困境》一文中通过分析军用机器人的研发现状与优势，论

① 王东浩. 人工智能体引发的道德冲突和困境初探[J]. 伦理学研究, 2014(2):68-73.
② 杜严勇. 机器人伦理研究方法论原则[J]. 中国社会科学报, 2015.1-2.
③ 王绍源. 应用伦理学的新兴领域:国外机器人伦理学研究述评[J]. 自然辩证法通讯, 2016, 38(4):147-151.
④ 卫华. 杀人机器人的伦理问题及辩护[C]// 全国军事技术哲学学术研讨会. 2013.

述了军用机器人与人性冲突，他认为人性是厌恶战争的，而军用机器人没有这类情感。他在文中提出了对于机器人恐怖的担忧，之后又写了机器人伦理设计的困难。提到了责任问题，他认为除了军用机器人的研发者、生产商以及使用者都应该为军用机器人造成的不良后果负责外，还有人工智能作为人工道德行为体也负有责任，他指出许多科学家只关注技术研发却不关心政治和道德责任的问题，在文章最后发出希望越来越多的科学家加入进人工智能伦理问题研究的愿望。[①]2015 年，陈升和孙雪在《国内外军用机器人的现状、伦理困境及研究方向》[②]一文中，指出了这样一种情形，就是机器人上战场，是机器人去作战，因此可以做到人类士兵的“零死亡率”，但是，就是因为不用考虑人员伤亡，所以有可能会被滥用。凭借军用机器人，军事强国对于军事较弱的国家就可以进行无所顾忌的打击，因为作战机器人是没有人性的，从世界和平的角度看，这样用没有人性化的作战机器人作战，哪怕胜利了也会遭受舆论强烈谴责。

（七）其他方面的研究

文献综述过程中发现，机器人伦理问题的文献在人工智能技术应用领域中写得最多。这其中机器人在军事领域的文献又比其他行业的要多。像同样属于人工智能技术的应用的决策系统、机器学习，没有机器人文献多。机器学习是最近人工智能技术应用目前最火的，伦理问题的讨论也应该跟上。由于目前正在大范围发展与应用人工智能技术，所以从伦理角度对人工智能的推广运用进行研究势在必行，就像我们不能等克隆人出来了再去做伦理研究一样。人工智能技术应用伦理问题研究相关的文献，目前要么是就单个的伦理问题讨论，不够系统；要么是系统却不够详细。我将在我的论文里面系统详细地论述人工智能技术应用的伦理问题。值得欣慰的一点是从伦理角度对人工智能体应用的探究正在吸引越来越多的哲学家、科学家、工程师的注意力，并且投身到这个领域的研究中来。在 1996 年的《电脑会永远正确吗？》一文中，提到以往人类命运的深刻变化，没有一次是只有绝对的益处却没有弊端的。该文作者认为电脑革命同样遵循这条规律，但是，这并不能成为人类拒绝进步和拒绝文明的理由。文章结尾在“达尔文进化论”视角下，呼唤伦理的变革、道德的变革，说只有在伦理、道德、理念的变革的阵痛中去实现人类的进步。[③]2013 年，胡增顺在《计算

① 杜严勇. 现代军用机器人的伦理困境[J]. 伦理学研究, 2014(5):98-102.
② 陈升, 孙雪. 国内外军用机器人的现状、伦理困境及研究方向[J]. 制造业自动化, 2015(11):27-28.
③ 佚名. 电脑会永远正确吗[J]. 中国技术监督, 1996(2):40-41.

机的伦理学困境与出路》一文中提出了，哪些地方的人工智能研发和使用需要加强?从什么角度出发的智能开发必须被削弱?另外，人工智能的可靠性如何?也就是说，我们能否把自己的身家性命托付给人工智能的专家系统?在人工智能开发问题上，我们必须慎之又慎，得出法律法规是解决伦理问题的出路的结论。[①] 单明在《科技发展要在乎伦理》在这篇文章中提出了这样一种忧思，就是人工智能的发展，那么很多岗位都可以用人工智能来做，那么劳动不是第一需要后，人类该如何安排自己？马克思的观点是按需分配。该文没有进一步讨论。但是单明觉得福利会很好的同时，还举例北欧来说明福利好的程度。但是他担忧将来的岗位没有人类，且不需要人类，那么人类行动能力或会退化，还有情感、思维也懈怠了。那时人类只有情感，那么情感将依附什么？最后提出人工智能的研究不只是科学技术的问题，还需要伦理学、心理学、社会学等的关切和研究。[②]张蕊、张佳帆、江灏在《可穿戴式柔性外骨骼人机智能系统可靠性及应用伦理问题研究》一文里，认为“科研工作人员的价值观和科研道德以及其所设计制造的外骨骼系统所应用的领域，将直接影响柔性外骨骼技术的发展趋势，所以需要建立相关的伦理准则，对外骨骼系统的设计和应用进行约束。”[③]前不久为了推动人工智能伦理学研究，麻省理工 MIT 和哈佛大学将会在人工智能基金的伦理和治理（Ethics and Governance of Artificial Intelligence）投入 2700 万美金，来推进人工智能伦理在公共利益领域的发展。2014 年谷歌成立了人工智能伦理委员会，监督人工智能机器人的发展趋势。

第三节 研究思路与研究方法

一、研究思路

研究综述之后，先介绍人工智能的技术应用情况，以及技术应用带来的的各种后果；接下来，分析这些后果中具有伦理性质的事件或者案例，从其所造成的伦理革新和伦理风险两方面展开分析。

二、研究方法

① 胡增顺. 计算机的伦理学困境与出路[J]. 开封大学学报, 2013, 27(4):86-88.
② 单明. 科技发展要在乎伦理[J]. 当代工人, 2015(9):19-19.
③ 张蕊, 张佳帆, 江灏. 可穿戴式柔性外骨骼人机智能系统可靠性及应用伦理问题研究[J]. 机电产品开发与创新, 2008, 21(5):19-21.

（1）文献研究法。笔者通过阅读大量的人工智能技术应用的专著、伦理学专著、人工智能伦理研究等相关文献，主要从伦理意义角度对人工智能应用进行探究。本文在对人工智能进行探究的角度与其他文献有明显差别。国内外也有很多学者展开了对人工智能应用的思考，而从理论层面反思人工智能运用的却很少。而且有也不够系统。通过在前人的基础上系统的整理，力求形成比较清晰的关于人工智能技术应用的伦理问题的研究成果。

（2）对比分析法：笔者通过义务论的视角和功利论的视角的对比分析人工智能技术应用的伦理问题。两种视角各有所长，这样结合集中伦理视角，使人工智能技术应用的伦理问题不片面，从而使分析相对多元，相对全面。

三、论文结构安排

第一章，绪论。

第二章，介绍人工智能技术应用。

第三章，人工智能技术应用的伦理挑战。

第二章　人工智能技术的应用领域

第一节 人工智能技术概要

人工智能是一种具有巨大社会和经济效益的革新性技术。与此同时，人工智能是一门正在迅速发展的新兴综合性很强的学科。人工智能的英文原名是Artifical Intelligence,简记为 AI。国内也有人主张将 AI 译为智能模拟。在国外还有人主张用 Machine Intlligence(MI.机器智能)一词来称呼人工智能这一研究领域（例如在英国），但在国际上主要还是用人工智能这一术语。[①]科技的每次基础突破，都会激起创新的热潮。“有人将人工智能比作第四次工业革命，来显示人工智能度各行各业带来的深远影响。”[②]

一、人工智能的定义

人工智能就是在各种环境中模拟人的机器。人工智能如同许多新兴的学科一样，至今也没有一个统一的定义。“下一个一般性的定义几乎是不可能的，因为智能似乎是一个包含着许多的信息处理和信息表达技能的混合体。”[③]主要是因为主要是因为不同的学科从各自的角度看，不同的学科从各自的角度看，下的定义也不一样。另外，最根本的一点就是人工智能的（如听觉、视觉、知识的表达等等）本质或是机制是什么，人们目前还不是清楚。人工智能词典里面是这样定义的“使计算机系统模拟人类的智能活动，完成人用智能才能完成的任务，称为人工智能。准确地说这不能用来作为人工智能的定义，到目前为止还没有合适的方式对机器的职能参数进行测试（Turing test)].而一些学者正专注于研究人工智能的开发，并且在这方面也有一定的成果，并推广运用到很多领域。例如，人工智能技术在医学、生物学、地质、地球物理、航空、化学和电子学等等领域都获得了应用。人工智能的应用对这些领域发生了重大影响。对人工智能的探究涉及很多方面，其中包括机器学习、证明自动定理、专家系统、智能

① 蔡自兴，徐光祐. 人工智能及其应用[M]. 清华大学出版社，2010.10.
② 人工智能大小古. 最新人工智能行业研究报告[EB/OL]. http://mp.weixin.qq.com/s?__biz=MzIONDc2MDEzMQ==&mid=2247483771&idx=2&sn=e6c15b3049628b4ac1fde20710f426b7&chksm=e959aa57de2e2341fe4fa40031842e47a16c88700cf5290e07b450542a6d5d36f4ddd29f1daa&mpshare=1&scene=23&srcid=0312pC4ctlq3ZxHLJZ18Si97#rd, 2017-03-12
③ 蔡自兴，徐光祐. 人工智能及其应用[M]. 清华大学出版社，2010.10.

控制、理解自然语言、神经元模型、知识工程、机器人学、智能数据库、计算机辅助设计、自动编程以及模式识别等。”[①]

二、人工智能发展简史

通过了解人工智能的历史可以更全面地了解人工智能的内涵和发展。人工智能发展历程主要有下面几个时间点。

（一）孕育（1956 年以前）

从古至今，人们就不断期望人的脑力劳动能由机器所替代，进而人们能更容易征服自然。而在探究人工智能发展历程中也出现了很多有关键影响的成果：

（1）古希腊哲学家亚里士多德（Aristotle）发表的《工具论》中就指出了形式逻辑中存在一定的规律，现在演绎推理发展的根本依据也来源于当时的三段论。

（2）英国哲学家培根（F.Bacon）还对归纳法进行了系统性说明，他当时就强调知识的重要性。这些都深深地影响到了对人类思维活动的探究，以及后来以知识为核心对人工智能的不断探究。

（3）德国哲学家、数学家莱布尼茨（G.W.Leibniz）探究得到万能符号以及理论计算机，他指出能够构建万能的语言符号，而且运用这种符号可以完成推理运算。这种推理思想开启了数理逻辑的发展之路，还产生出现代机器思维设计这一概念。[②]

（4）英国逻辑学家布尔（G.Boole)专注于探究思维的固定形式，将思维机械化，还创建有布尔代数。布尔著作的《思维法则》中首次以符号语言对思维推理形式进行了说明。[③]

（5）1936 年英国数学家图灵从数学角度对计算机模型进行设计，图灵计算机从理论层面开启了电子数字计算机的先河。

（6）1943 年美国神经学家匹兹（W.Pitts）、麦克洛奇（W.McCulloch）成功完成首个 M-P 模型（神经网络模型）的构建，开始了从微观角度对人工智能进行探究，并为探究人工神经网络做铺垫。

---

① 吴胜，王书芹．人工智能基础与应用[M]．电子工业出版社，2007.
② 王万良．人工智能及其应用[M]．高等教育出版社，2016.
③ 蔡自兴．人工智能及其应用:研究生用书[M]．清华大学出版社，2004.

（7）1946 年美国数学家埃柯特（J.P.Eckert）、莫克利（J.W.Mauchly）成功创造了首台 ENIAC 电子计算机，这项史无前例的成果为探究人工智能提供了物质支持。

由于上面的发展过程可以看出，人工智能的产生和发展绝不是偶然的，它是科学技术发展的必然产物。[①]

（二）形成阶段（1956-1969 年）

人工智能发展到二十世纪中期已经从人类快速发展的科技领域正式产生。到 1956 年，斯坦福大学教授麦卡锡（McCarthy）、明斯基（Minsky）、朗彻斯特(Lochester)、香浓(Shannon)共同发起，邀请莫尔(More)、塞缪尔(Samuel)、纽尼尔(Newell)和西蒙(Simon)等，[②]在美国参加一个为期 2 个月的学术交流会，针对机器智能化问题进行探究。

这次会议之后，人工智能在实验研究上取得了两项重大突破： 一个是美国的纽厄尔、肖（J.Shaw）和西蒙合作编制了一个名为逻辑理论机（The Logic Theory Machine 简称 LT）的程序系统。[③]另一个是塞缪尔 1956 年研制成功的跳棋程序，这项程序不但可以与对手对战，在下棋的过程中还能够累积经验，在学习、组织与适应方面有很强的自发性。[④]

1956 年，另一个有深远影响的成就就是侨姆斯基（N.Chomsky)提出了一种文法的数学模型，开创了形式语言的研究。在模式识别方面，1959 年赛尔夫里奇推出了一个模式识别程序。同年，籍勒洛特发表了证明平面几何问题的程序。纽厄尔、肖和西蒙等人又通过心理学实验，发现人在解题时的思维过程大致可以分为：解题计划、解题过程、修订解题计划。[⑤]麦卡锡研制出表处理语言 LISP，在人工智能的二哥哥领域中都得到广泛的应用。因为它不仅能处理数值，而且可以更方便的处理符号，LISP 是研究人工智能的重要工具。LISP 语言武装了一代人工智能科学家。[⑥]

① 吴胜，王书芹．人工智能基础与应用[M]．电子工业出版社，2007.
② 何华灿，李太航等．人工智能导论[M]．西北工业大学出版社，1988.262.
③ 吴胜，王书芹．人工智能基础与应用[M]．电子工业出版社，2007.
④ 吴胜，王书芹．人工智能基础与应用[M]．电子工业出版社，2007.
⑤ 廉师友．人工智能技术导论[M]．西安电子科技大学出版社，2007.
⑥ 徐志敏，李栗．人工智能：梦想现实未来[M]．四川教育出版社，1992.133-134.

1969 年组建国际人工智能联合会议（International Joint Conferences On Artificial Intelligence,简称 IJCAI），它标志着人工智能这门新兴学科已经得到了世界的工人和肯定。[①]

（三）知识应用阶段（1970-至今）

进入 20 世纪 70 年代，这个时期主要确立了知识在人工智能中有着非常重要的地位。人工智能也从理论或者实验室走向了实际应用，在各个应用领域中获得了显著的效果。

在人工智能诞生的初期，在实验室中研究人工智能基本原理和方法的是一批科学家，他们认为人工智能科学家不应陷入到各个应用领域，人工智能技术推广运用到社会生产生活的各个方面，这理应是各领域工程师的任务。

美国斯坦福大学费根宝姆（E.Feigenbaum）教授坚持积极倡导将人工智能的原理和方法应用于解决实际领域中的问题。在他的领导小组内，第一个专家系统 DENDRAL 开始研究，并且在 1968 年开始投入使用。[②]“被誉为‘专家系统和知识工程之父’费根鲍姆发现：要使计算机工作得想专家那样出色，必须为它提供专家所具有的知识。在这一思想基础上他设计出首个专家系统 DENRAL,此专家系统与同领域化学家在解决问题水平方面基本无差异”。[③]专家系统进一步研发，对人工智能发展探究产生很大影响。肖特里菲(E.HShorliffe)等人在 1972 年开始研制专家系统 MYCIN 用于争端和治疗传染性疾病。以后经过不断地改进和完善成为了第一个功能比较晚上的专家系统，MYCIN 不仅能对传染性疾病做出专家水平的诊断和治疗选择。[④]

但是，由于最初的机器翻译会出现错误，好多国家都停止了对机器翻译的资助。不仅如此，机器学习，问题求解，神经网络等也都遇到了挑战。这些挑战让人工智能研究陷入困难之中。这其实与其他的一些科学一样，人工智能的发展也不是平坦的。

由于费根鲍姆在 1977 年提出了知识工程（Knwledge Engineering,KE）的概念，进一步推动了人工智能的发展。这个时期的特点是人们意识到了知识在人工智能中的重要地位，所以非常重视知识。人工智能系统的三个比较基本的问题是：知识利用、知识表示、知识获取。人工智能的各项研究开始复兴。

---

① 杨祥金. 人工智能[M]. 科学技术文献出版社重庆分社, 1988.404.
② 廉师友. 人工智能技术导论[M]. 西安电子科技大学出版社, 2007.
③ 廉师友. 人工智能技术导论[M]. 西安电子科技大学出版社, 2007.
④ 何华灿, 李太航等. 人工智能导论[M]. 西北工业大学出版社, 1988.

同一时期同一时期机器学习、机器人、自然语言理解，人工神经网络等也继续发展。

三、强人工智能与弱人工智能

（一）弱人工智能

弱人工智能(TOP-DOWN AI)指利用设计的程序对动物以及人类逻辑思维进行模拟，使智能体表现出的行为与人类相似，但智能体缺乏思想意识。持弱人工智能的观点的学者认为，不能制造出真正能推理和解决问题的智能机器。目前很多电子产品都具备一定的智能性，当外界数据输入发生变动就会运行相应的程序，得到不同结果，能代替人们完成重复简单的任务。弱人工智能随处可见，现在洗衣机、电视和微波炉都具备称重、感光、计时和感温等功能。但弱人工智能的运用也只限于模仿人类低等行为。[①]

（二）强人工智能

强人工智能(BOTTOM-UP AI)属于更高级的人工智能，强人工智能的认为有可能制造出真正具备思想意识、思考能力以及感情，能解决问题、能够推理的智能机器，对强人工智能来说，计算机不但能够对意识进行研究，换句话说，计算机经过相应的程序设计后就会具备一定的认知能力，从这个角度理解计算机也是有意识的。后文是基于强人工智能环境下进行的哲学思考。[②]翟振明:指出“和弱人工智能相比，强人工智能貌似要强一点，但事实上两者是完全不同的。这里的“强”是“弱”的一种超越，能有一定的思维能力、态度和情感。”[③]

第二节 人工智能技术的应用

一、专家系统

（一）专家系统的定义

在现阶段人工智能技术应用的研究中，专家系统可以算是最活跃的一个分支。由于越来越多的具体的专家系统的问世，使人工智能的应用越来越广泛。但是，从另一个方面看，专家系统的发展还远未成熟，甚至可以说还仅仅是开

① 龚园. 关于人工智能的哲学思考[D]. 武汉科技大学, 2010.
② 龚园. 关于人工智能的哲学思考[D]. 武汉科技大学, 2010.
③ 刘功虎. 中山大学人机互联实验室主任翟振明谈围棋人机大战:技术群体战胜了天赋个人[EB/OL]. http://www.whxc.org.cn/2016/0315/28966.shtml, 2016-03-15

始。[①]所谓专家系统其实就是一类程序系统，从功能上可以把它定义为“一个在某领域具有专家水平阶梯能力的程序系统”，它能像领域专家一样工作，能运用专家们多年来积累的工作经验与专门知识，在很短的时间内对问题得出高水平的解答。[②] 被誉为“专家系统和知识工程之父”的斯坦福大学的 Edward Feigenbaum 教授，对专家系统的定义为：一种智能计算机程序，它运用知识和推理来解决只有专家才能解决的问题。也就是说，专家系统是一种能模拟专家决策能力的系统。模拟的意思是做得跟专家一样。

（二）专家系统的类型

专家系统的类型有：

1、解释专家系统。它的任务是通过对一组信息和数据的分析与解释，确定他们的涵义。[③]

2、预测专家系统。它的任务是通过对有关于一个过去和现在的数据进行系统地分析，来预测未来该事物一段时间的发展趋势。[④]

3、诊断专家系统。诊断专家系统能根据取得的现象、数据或事实推断出系统是否有故障，并能找出产生故障的原因，给出排除方案。

4、设计专家系统。设计专家系统的任务是根据设计要求，求出满足设计问题约束的目标配置。[⑤]

5、规划专家系统。它的任务是根据约束条件，作出行动的安排或调度方案。[⑥]

6、教育专家系统。主要用于培训和教学，不但能够教学和辅导，而且能够指出学生错误并纠正错误。

7、控制专家系统。控制一个客体的全面行为，使之达到预期要求，更可以分析当前，预测未来。

8、调试专家系统。对于争端出现的故障，给出修改的补救措施。

9、监督专家系统。比较期望结果和观察结果。

---

① 郑丽敏主编. 人工智能与专家系统原理及其应用[M]. 中国农业大学出版社, 2004.
② 冯定. 神经网络专家系统[M]. 科学出版社, 2006.
③ 何华灿, 李太航等. 人工智能导论[M]. 西北工业大学出版社, 1988.
④ 郑丽敏主编. 人工智能与专家系统原理及其应用[M]. 中国农业大学出版社, 2004.
⑤ 何华灿, 李太航等. 人工智能导论[M]. 西北工业大学出版社, 1988.
⑥ 吴今培. 智能故障诊断与专家系统[M]. 科学出版社, 1997.

10、修理专家系统。该系统具有诊断、调试、计划和执行能力等功能，使故障客体恢复正常工作。

（三）专家系统的一般特点

启发性：该系统依据某一些条件选定一个假设，使推理进行。也就是该系统能够运用经验与知识进行推理、判断与决策。

透明性：专家系统一般都有较好的透明性，即会解释本身的推理过程。比如一个病人患了扁桃体炎，并且需要用一种药治疗，那么该系统会解释为什么用这种药。

灵活性：随着专家家系统中知识的不断增加，只要推理方式不变，推理及部分可以不变。因此，专家系统具有十分广泛的应用领域。

二、机器学习

（一）学习是什么？

从心理学角度来看,学习是一种新的行为模式的形成，比如一个西方人以前不懂中文，通过学习会说中文了，这就是行为模式的改变。按照人工智能专家西蒙的观点，系统在不断的工作中对本身能力的增强或者改进，使得系统在下一次执行同样任务的时候或者相类似的任务的时候，会比以前效率高。

（二）机器学习是什么？

机器学习又称知识的获取，就是机器模仿人类学习的智能行为。按照H.Simon 的观点，学习是系统改进自身性能的过程，通过自动地学习获取知识和不断地自我完善。一个没有学习能力的计算机系统，就难以称为智能系统。有的人可能会认为，一部机器过了一段时间还是那部机器，但是这是针对没有学习能力的机器而言的，对于一部有学习能力的机器而言，可能过了一段时间，就不是以前的那部机器了，因为通过学习，机器里面的系统知识早已经更新换代了。现在国内外的 IT 巨头正在深入研究机器学习，目标是模拟人类的大脑，试图研究出拥有人类智慧的大脑。[①]

（三）机器学习分类

1、机械式学习（Rote learning），也叫死记硬背式学习，是一种最简单的学习方法，靠记忆存储库来存储知识，其成功与否取决于机器有没有一个好的知

① 麦好. 机器学习实践指南:案例应用解析[M]. 机械工业出版社, 2014.6.

识库，而不需要做假设和判断。在机械学习中，知识是以较为直接和稳定的方式进行的。[①]

2、讲授式学习(learning from instruction)，也叫教授学习或者指点学习。例如，TEIRESIAS，由来自于斯坦福大学的 Randall Davis 建立，它能够对于用户给的信息有一定的选择能力，并能把用户的指示或者建议形式化、具体化。[②]

3、类比学习(learning by analogy)。类比是只能更深层次的知识行为，一般类比分为两个步骤：先要找到两个对象的相似特征，在认为两个对象具有相似性之后，将比较对象的知识与求解方法转换到求解的目标对象中去。这相当于是人类用已有或者以前的经验来求解新的问题。[③]

4、归纳学习（Learning from induction），也称实例（example）学习，其特点是从个别实例得到规则。例子学习是一种高级的学习能力。比如，“青苹果是苹果”，“红苹果是苹果”等等，机器通过推理得到这些实例的一些知识，比如“苹果”的概念。因此专家们一般称这种方法为“从实例中学习”。我们要求学习的个例中没有错误的案例，比如“橘子是苹果”，也不能有歧义或者模棱两可的例子，否则会造成机器判断失误。[④]

5、观察发现式学习（Learning by observation &discovery）是归纳学习的更高一个层次，是通过概念的聚类发现规律或规则。[⑤]例如幼儿刚学会走路的时候，是跌跌撞撞的，反反复复练习才学会了走路。幼儿再大一点，上幼儿园了，学会写字（这就是机遇行动的学习或者是称为自主学习），从幼儿园老师给的卡片小猫小狗获得了关于小猫小狗的概念（通过例子的归纳学习）。再长大一点，上小学了，背书，刚开始是死记硬背，这是属于（机械式学习或者死记硬背式学习）；后来遇到挫折了，懂得总结经验，下次处理类似的事情效率更高，这就是归纳学习或者实例学习。

三、自然语言理解

（一）自然言语理解是什么？

① 何华灿，李太航等. 人工智能导论[M]. 西北工业大学出版社，1988.262.
② 蔡自兴，徐光祐. 人工智能及其应用[M]. 清华大学出版社，2010.10.
③ 蔡自兴，徐光祐. 人工智能及其应用[M]. 清华大学出版社，2010.10
④ 徐志敏，李栗. 人工智能：梦想现实未来[M]. 四川教育出版社，1992.133-134.
⑤ 吴胜，王书芹. 人工智能基础与应用[M]. 电子工业出版社，2007.129.

自然语言理解也称为自然语言处理，语言是互通信息的重要方式，自然语言理解主要是研究怎么样让机器理解人类的语言，实现人与机器之间的自然语言交互[①]。设定的自然语言机理被人工智能的计算机程序表达出来，实际上自然语言理解的过程是一种映射，一种表达被转化成另一种表达的过程。自然语言理解包括文字理解和语音理解，[②]自然语言的生成比书面语言难，自然语言理解的难点是知识的表达和应用。 机器要听得懂人发出的语音信息，比人听得懂机器发出的自然语言要难得多。语言的编码是一个很复杂的编码和解码的过程。不过一些语音合成器也早已投入使用，比如，由菲利普斯兄弟公司推出的 Copilor，它会即时提醒司机“请注意油量”。自然语言理解意味着可以通过“人机对话”，机器可以被人类的语言而不是人类的手所控制。许多研究者确信，当知识结构和表达理论被建立起来，自然语言理解就会取得成功[③]，并且许多研究表明语言和思维有着密切联系，通过研究自然语言理解也有助于研究人类思维的谜底。

（二）自然语言理解的层次

1、语音分析。语音分析是根据音位规则，区分出独立的音素。每个词汇都有它的语音形式，音素构成了音节，音节组成了一个词的发音。语音分析是一种高层次的语音识别。

2、语法分析。语法就是语言还有句子表达的组织规则[④]。如果让一个没学习过也没听过意大利语的人听懂意大利语是不可能的，要想让语言被计算机理解，也离不开语法。一般有两种不同的方法，第一种只使用模式匹配，以便信息被从句子中得出来。第二种采用被简化的语言子集[⑤]。

3、语义分析。对输入的语言进行分析，检查其语法结构的正确性，从而加以分类，这种过程称为语义分析。[⑥]对于语言所表达的意义，这涉及到结构和词的歧义问题，如英语词 come 可能有多种意义。但是一个词的意义再多，结合上下文，在词组中其表达的意思也是唯一的[⑦]。

① 王万森. 人工智能原理及其应用:(第 2 版)[M]. 电子工业出版社, 2007.230.
② 廉师友. 人工智能技术导论[M]. 西安电子科技大学出版社, 2007.235.
③ 杨祥金. 人工智能[M]. 科学技术文献出版社重庆分社, 1988.404.
④ 蔡自兴. 人工智能辞典[M]. 化学工业出版社, 2008.191.
⑤ 杨祥金. 人工智能[M]. 科学技术文献出版社重庆分社, 1988.405-406.
⑥ 蔡自兴. 人工智能辞典[M]. 化学工业出版社, 2008.193.
⑦ 蔡自兴, 徐光祐. 人工智能及其应用.第 4 版[M]. 清华大学出版社, 2010.310.

4、语用学。语用学的因素与知识、推理、上下文有关。语用是研究形式语言的操作含义，即如何实现的问题[①]。语法、语义、语用之间是相互作用和相互联系的[②]。

四、机器人

（一）机器人的定义

“Robot”一词最早出现在科幻作家笔下。美国机器人研究院下的定义是，机器人是一种可再编程序的多功能的操作装置。其实也可以广义地理解，就是模拟人的视觉、听觉以及行走的控制装置。机器人的视觉、听觉的研究是难点。很多小孩子都能解决的比如绕过一个障碍物椅子，但是机器人却十分困难，因为机器人的视觉的模仿很困难。[③]

1、第一代机器人（first generation robots）：分为没有工作程序和有工作程序。没有工作程序的机器人，比如遥控机器人，对外界没有感知能力，且不能独立完成工作，要人遥控它才能完成工作。有程序的第一代机器人把程序储存起来，记住人交给它的一连串动作后，机器人自动重复完成人交给它的动作，如 UNIMTE[④]已经进入应用。

2、第二代机器人(second generation robots)：相比于第一代机器人，多了独立性，能够又触觉等感应外界的能力，能够获取操作对象的简单信息。能够通过计算做出简单的推理和对动作进行反馈[⑤]，只需要人在关键的地方给与提示和指点。目前还在继续研发，也有的机器人已经进入商业应用[⑥]。

3、第三代机器人(third generation robots)：第三代机器人又叫高级智能机器人，能用语言与人对话，具有丰富的传感器，它能进行识别、逻辑推理、判断环境以及自己的状态，自己决定自身的行为，还能听从语言命令动作，并且根据自己的任务制定计划，完成工作。[⑦]

机器人已经在国防、宇宙探索、工业、海洋开发、商业、医疗、农业、空中、教育、旅游业等领域应用。[⑧]

---

① 蔡自兴. 人工智能辞典[M]. 化学工业出版社, 2008.196.
② 杨祥金. 人工智能[M]. 科学技术文献出版社重庆分社, 1988.409.
③ 蔡自兴, 徐光祐. 人工智能及其应用[M]. 清华大学出版社, 2010. 10.
④ 徐志敏, 李栗. 人工智能: 梦想现实未来[M]. 四川教育出版社, 1992.187.
⑤ 芮延年. 机器人技术及其应用[M]. 化学工业出版社, 2008.4.
⑥ 徐志敏, 李栗. 人工智能: 梦想现实未来[M]. 四川教育出版社, 1992.188.
⑦ 杨祥金. 人工智能[M]. 科学技术文献出版社重庆分社, 1988. 448.
⑧ 杨宪泽. 人工智能与机器翻译[M]. 西南交通大学出版社, 2006.16.

五、人工神经网络

（一）神经元（neuron）

神经元是脑细胞的基本单位，包括树突和一条很长的轴突。神经元的主体部分是细胞体，由细胞体向外延伸的最长的一条分支称为轴突即神经纤维。大脑由大量的神经细胞组成，每条神经又会连接几千条其他的神经。大脑是人和动物的智能的来源，因此许多科学家、学者都在研究神经元的生理机制，并且对此进行模拟出神经元模型。因为神经元是很复杂的，人们至今对神经元的认识还很很浅。[①]轴突是用来传递和输出信息的，其端部的许多轴突末梢为信号输出端子，将神经冲动传给其他神经元。由细胞体向外深处的其他许多较短的分支称为树突。树突相当于细胞的输入端，树突的全长各点都能接受其他神经元的冲动。神经冲动只能由前一级神经元的轴突末梢传向下一级神经元的树突或细胞体，不能反向的传递。[②]

（二）人工神经网络

人工神经网络（Artificial Neural Network,简称 ANN），是由麦克丘洛奇（Mc Culloch）和皮茨（Pitts）在 1943 年首先提出的神经网络模型。[③]从那时候开始，就进入了人工神经科学理论研究的时代。人工神经网络是一种模仿生物神经网络的数学模型，它可以用来模拟人类神经系统的结构和功能。人工神经网络是拥有自己学习和自己组织的智能机构，它用大量的人工神经元来计算，每个神经元代表一个特定的输出函数，然后由大量的“神经元”连接起来组成了网络[④]。

（三）人工神经网络的工作原理

大脑是由神经元组成的，因此大脑就是一个神经网络，这是一个不争的事实。[⑤]《人工智能的未来》的作者认为：“如果一个科学狂人将你的每一根神经元用同样的微型电脑复制品替代，替代的过程完成之后，你会感到自己还是原来的自己，没有任何变化。”[⑥]

对神经网络模型、算法、理论分析和硬件实现的大量研究，为神经网络计算机走向应用提供了物质基础。[⑦]每个神经元能够收发能量，然后神经网络并不

① 王树林. 人工智能辞典[M]. 人民邮电出版社, 1992.135.
② 王万良. 人工智能及其应用.第 2 版[M]. 高等教育出版社, 2008.218.
③ 王树林. 人工智能辞典[M]. 人民邮电出版社, 1992.134.
④ 麦好. 机器学习实践指南:案例应用解析[M]. 机械工业出版社, 2014.158.
⑤ 霍金斯. 人工智能的未来[M]. 陕西科学技术出版社, 2006.20.
⑥ 霍金斯. 人工智能的未来[M]. 陕西科学技术出版社, 2006.20.
⑦ 蔡自兴, 徐光祐.人工智能及其应用[M]. 清华大学出版社, 2010.11.

是立即工作，而是通过接受各种能量，达到了一定的能量，然后通过节点传输给其他的节点。

由于计算机不能够很好地处理非数值的形象思维信息，所以用神经网络来处理形象和直觉思维信息的时候，作用比用传统方式处理的效果要好。研究神经元的构成、实现、连接方式、连接权值、输入和输出特性、网络、网络形态、信号的传输和处理等是研究神经网络的主要任务。1982 年霍普菲特（Hopfield）研究了用能量函数实现神经网络，解决了一批 NP 完全问题。这一思想已渗透到语音研究、文字识别、计算机视觉、认知科学等领域。神经计算机理论和方法已经提出，有人希望以神经网为总体结构基础，建立第六代计算机。[①]神经网络的记忆和知识都分散在所有的连接上面，这和真正的大脑很像。但是，神经网络不能对信息进行中央存储，因为它与计算机不同，它没有 CPU。[②]

（四）神经网络学习方式

（1）、监督学习（有教师学习）,有教师学习需要有个“导师”或者“教师”来提供目标输出信号或者期望。因为有教师学习需要“导师”对给的输入提供正确的输出，再根据期望来调整系统参数。

（2）、非监督学习（无教师学习），是没有外界的“导师”，学习系统也不需要知道期望输出，学习系统根据输入模式能够自己调节自己的参数，自动地适应连接权，来表示外部输入的某种固有特性（如聚类），无教师学习的例子有 Kohonen.

（3）、再鼓励学习（强化学习），是教师学习的一种特殊的例子。这种学习介于上面所描述的两种情况之间。强化学习采用一个评论员来评价信息（奖或惩），并不是给出正确的答案，学习系统根据受惩罚或者奖励的情况来完善、改进。强化学习算法的例子是遗传算法（GA）。

如果说一个人每隔几年他或者她的细胞会换一遍，但是从最重要的部分来看，那个人还是那个人。同样的道理也适用于大脑，如果一个精通技术的人用具有同样功能的电脑复制品去替代一个人的神经元，替代后，那个人依旧感觉还是原来的自己，没有什么变化。同样的道理，可以设想，如果一个人工系统拥有了与人类的大脑结构相同的机器智能结构，那么它也可以向人类的大脑一

① 王树林. 人工智能辞典[M]. 人民邮电出版社, 1992.134.
② 霍金斯. 人工智能的未来[M]. 陕西科学技术出版社, 2006.20.

样聪明，[①]并且具有“智能”。

（五）神经网络学习种类

迄今为止有 30 多种人工神经网络被开发出来了，并且开始应用。比较有代表性的是：反（BP）向传播网络、Hopfield 网。

（1）、BP 神经网络（Backpropugation Neural Network）。BP 网络是一种误差反向传播的网络，就是多层向前网络。它的学习的算法称为 BP 学习算法，是一种有“教师”的算法。BP 算法包括第一阶段的计算实际输出，第二阶段修改网络的联结权值和阀值。它的优点是算法的推导比较清楚，学习的精度很高。经过训练后的 BP 网络，运行速度非常迅速，可以用来实时处理。从理论上讲，有的神经网络可以学习学会任何可学习的东西。[②]只是标准的 BP 算法也存在一些缺陷，容易形成局部最小，得不到全局最优；训练次数多，使学习效率不高，收敛的次数也慢。[③]

（2）、Hopfield 神经网络（HNN） 是由 Hopfield 提出的，是全互联反馈神经网络模型，它的每一个神经元都和其他神经元相连接。Hopfield 网络的学习过程是在系统向稳定性转化的过程中逐渐完成的。[④]是典型的全连接网络。Hopfield 网络用于联想记忆的时候包括存储操作和提取操作两个阶段。

① 霍金斯. 人工智能的未来[M]. 陕西科学技术出版社, 2006.32.
② 王万森. 人工智能原理及其应用.第 3 版[M]. 电子工业出版社, 2012.206.
③ 冯定. 神经网络专家系统[M]. 科学出版社, 2006.31.
④ 王万森. 人工智能原理及其应用[M]. 电子工业出版社, 2000.206.

第三章 人工智能技术应用的伦理挑战

人工智能技术正在兴起，并且具有广阔的应用空间，但同时也带来一系列充满风险的社会后果，从伦理学的角度看，主要有道德判断的困境、技术异化的伦理风险、人工智能的智能和自主性进化对人类主体性的挑战。

第一节　人工智能伦理评价的义务论与功利论之争

智能计算机系统以往一直是科幻小说的主题。现在，我们正在进入一个人工智能技术应用对于我们的日常生活产生深远影响的时代。从交通到安检再到医学等等，智能计算机系统的应用已展现了广阔的前景。在分析这些影响的伦理挑战之前，我们先考察涉及到人工智能伦理评价的义务论与功利论之争，以确定本文的理论视角和价值立场。

一、义务论

也叫“道义论”、“道义学”、“本务论”或者“非结果论”。义者，宜也，应当也。务者，事也，任务也。义务同职责和责任有相同的含义，是人们出于义务心的为完善品德而完善品德，它是人们主观上的自觉意识和内心的自愿要求。义务论认为判断人们道德是看人们的义务心，也就是说只看动机是否符合道德规则。义务论是把道义作为道德的最终标准，把增减每个人品德的完善程度做为道德的最终目标，而不以获得额外的报偿或者权利为前提。

义务论反对功利主义。有学者解释说，义务论所反对的利、功利，仅仅是反对私利而不是反对公利，反对的是目的利己，而不是反对目的利他。冯友兰说：“求自己的利，可以说是出于人的动物的倾向，与人所以为人者无干。为实现人之所以为人者，我们不能说，人应该求自己的利。但求别人的利，则于人之所以为人者有干。为实现人之所以为人者，我们可以说，人应该求别人的利。”[①] 马克思主义则认为，义务是人的社会性和社会本质的必然产物，凡是在人与人有关系的地方，就一定会产生义务的要求。[②]道德义务的根源在于社会物质生活条件和社会经济关系。义务论者们往往认为，道德起源与目的是自律的，因而

① 王海明. 伦理学原理[M]. 北京大学出版社, 2009.128-129.
② 王泽应. 伦理学[M]. 北京师范大学出版集团. 2012.183.

义务论者同时也是道德起源与目的自律论者。康德、罗尔斯的义务论是义务论的主要代表。

康德伦理学是近代伦理学中最重要的伦理学说，也是伦理学史上义务论伦理学的典型代表。康德的伦理学著作有《实践理性批判》、《道德形而上学基础》、《道德形而上学》等。

康德对伦理学的影响或者说贡献是把伦理学中的幸福论清除了。康德反对幸福主义，因为早前西方的伦理学是既讲至善又讲结果幸福的，如果结果不幸福那么也算不上是道德的。康德反对把幸福等同于道德，他认为这样子不利于建立普遍的原则。他在《道德形而上学基础》的开头就指出："在世界之中，甚至在世界之外，除了善的意志，没有什么能被称作无条件善的东西。"[①]也就是说，康德认为世界上除了"善良意志"外，不可能预想一种无条件的善。"善良意志"之所以善良是因为它本身"善良"，而不是因为它产生的结果。同样的，他认为判断一个人的行为是否善良，不是看这个人行为的结果，而是看这个人行为的初衷或者动机是否善良。如果初衷或者动机是善良的，那么这个行为就是善良的。如果有一个好的结果，但是行为的初衷或者动机不是善良的，那么这个行为也算不上是道德的。

康德反对功利主义，反对为了幸福、利益、欲望、成功等而行使道德。他认为这样子会破坏道德的纯粹性。也就是说，道德不是为了达到目的之手段，而是目的本身。康德把有条件的行为原则叫"假言命令"，把无条件的原则叫"绝对命令"，认为只有从"绝对命令"出发的行为才是道德的。康德把"绝对命令"作为道德的最高原则。邓晓芒教授指出："康德认为纯粹的实践就是道德。纯粹实践理性就是道德原则（道德律），在道德领域，人为自己立法，用自由意志来限制自己的规律，用自由意志的规律来作为自己的自由意志所遵循的规律。"善良意志不是因为它所期望的事物而善，也不是因为达到目的而善，而是因为意愿而善，因其自身而善。康德还认为，道德不会因为外界的因素而改变，康德把这个叫做道德韧性。如果一个人遵守道德，没有变得富贵，但是他拥有的道德本身也像宝石一样闪闪发光，熠熠生辉。

关于"绝对命令"，康德在《道德形而上学基础》一书中说："不但道德规律及其原则本质上不同于任何含有经验成分的实践知识，而且全部道德的哲学都

① 伊曼努尔·康德，康德，孙少伟．道德形而上学基础:全新译本[M]．中国社会科学出版社，2009.1.

完全以其纯粹部分为基础。”[①]“通过分析，我们就会发现，道德原则必定是一个定然律令，而且这个律令所命令的不多不少正好就是这个自律性。”[②]康德把责任看作是绝对命令，但是康德认为前提条件是我们是有着自主性的理性行为者。“康德反对功利主义的的义务观，认为人必须为尽义务而尽义务，而不能考虑任何利益、快乐、成功等外在因素；只有出于善良意志即义务心，对道德规则即绝对命令无条件遵守的行为，才是真正道德的行为。”[③]

黑格尔批判康德的义务论，认为“伦理学的义务论，如果是指一只种客观学说，就不应该包括在道德主观性的空洞原则中，因为这个原则不规定任何东西”。叔本华认为：“如果一个人真的把义务与责任当作伦理学的基本概念，这当然是不可避免的；因为这些观念本质上是相对的，而且它们的重要意义全靠可能的惩罚和允诺的奖惩而定。”[④]“换句话说，它是幸福论，它已经被康德视为一个擅入者郑重其事地从他的体系前门推出去了，反而以最高善的名义又让它从后门爬进来。”[⑤]叔本华指出，康德把人类行为之有无价值看作是有没有尽义务，但是，康德用“绝对的”这种词去代替与这样的本质密不可分的外在条件是拙劣的。[⑥]约翰·罗尔斯在评论康德伦理学时说，“日常伦理道德行为，如果不是受到‘善良意志’的限制，而仅仅凭借一般意志的驱使，那么，这些行为就有可能变成不道德的行为。”[⑦]“绝对命令作为普遍的客观法则，只能在一定条件下转化为主观准则才其作用。”[⑧]

意志的自律性与他律性是康德伦理与其他伦理主要区别标志。但是也有学者批判康德的绝对命令的表述自动排除了对人类以外的对象的尊重。有学者认为，“康德的道德的超功利和自律在现阶段不可能完全实现。”[⑨]

笔者认为，康德伦理学有他的美好的可取之处，但是有准则主义的片面性。

二、功利主义

功利主义（utilitarianism）也叫“功利论”、“功用主义”，是一种把实际效用

① 伊曼努尔·康德，康德，孙少伟. 道德形而上学基础:全新译本[M]. 中国社会科学出版社，2009.4.
② 伊曼努尔·康德，康德，孙少伟. 道德形而上学基础:全新译本[M]. 中国社会科学出版社，2009.78.
③ 朱贻庭. 应用伦理学[M].上海辞书出版社.2013.34.
④ 叔本华，孟庆时，任立. 伦理学的两个基本问题[M]. 商务印书馆，2011.165.
⑤ 叔本华，孟庆时，任立. 伦理学的两个基本问题[M]. 商务印书馆，2011.165.
⑥ 叔本华，孟庆时，任立. 伦理学的两个基本问题[M]. 商务印书馆，2011.166.
⑦ 李小科. 正义女神的新传人[M]. 河北大学出版社，2005.44.
⑧ 李小科. 正义女神的新传人[M]. 河北大学出版社，2005.44.
⑨ 王淑芹，伦理秩序与道德研究，中央编译出版社，2015，23.

或者利益作为行为的评价标准的伦理学说。功利主义用行动后果的价值来衡量行为的善恶。避苦求乐、趋利避害是人的本性。使人不开心、不快乐的就是恶的，使人快乐、幸福的就是善的。这种价值观把行动的结果作为伦理考虑的主要因素。也就是说，功利主义者把增减每个人的利益总量作为评价一切行为的善恶的标准。如果能够增加每个人的利益总量，那么行为就是善的；如果一个行为减少每个人的利益总量，那行为就是恶的。一个行为是增加还是减少社会利益总量，是评价道德的终极标准。

功利主义的诘难有两个非常著名的例证："惩罚无辜"和"奴隶制度"。

"惩罚无辜"是说在一个小镇上，出现了一系列的恐怖事件，镇上的人们跑到警察局跟警察局长说，如果在规定时间内不找到坏人，那么就要发生百余人的暴乱，警察局长经验丰富，是一位功利主义者，马上判断暴乱一旦发生后果不堪设想。警察局长看到村镇里面有一位无辜的流浪者，于是把他抓起来，告诉小镇的人们说，这个就是凶手，然后宣判乐流浪者死刑。镇上的暴乱被拦住了。按照功利主义者的看法，这位警察局长是道德的，是善的。但是这位警察局长明明知道流浪汉是无辜的，却还是判了他死刑，这位警察局长是非正义的，因为惩罚无辜者是非正义的。

关于"奴隶制度"，功利主义者认为如果一个社会的奴隶制度能够增进每个人的利益总和，那么奴隶值得就是道德的，是应该的。

康德和冯友兰批判功利主义是巧于算账的行为。中国当代学者王海明对功利主义则有同情的理解和独到的阐释。王海明认为，上面的两个例子不能用是非正义来说，他认为不能两全的情况下，是为了避免更大的恶，也就是，两恶相权取其轻。王海明认为，功利主义是在人们的利益不发生冲突的时候，表现为"不损害任何一个人的利益，增进每个人的利益"，在利益发生冲突的时候，不能两全的时候表现为"最大利益尽余额"[①]。比如前面的的两个例子，他认为功利主义是在惩罚无辜与百人暴乱相冲突，奴隶制度与社会总利益相冲突的情况下才表现为"社会最大利益"，当两者利益不冲突的时候，功利主义不伤害任何一个人的利益。因为不发生冲突，却伤害个人的利益，这是违背增进社会总利益的，也就违背了功利主义。王海明认为，功利主义只赞同前者反对后者。王海明还指出，大家只注意到功利主义可能会导致非正义，而忽略了"不损害个人的利益，增进每个人的利益之和"这条原则，然后又把利益发生冲突的时候，不能两全的

① 王海明. 伦理学原理[M]. 北京:北京大学出版社, 2001. 127.

时候选择的“社会利益最大化”，夸大成“最大利益”，说功利主义即“最大利益”。而在事实上，只有利益冲突的时候，功利主义才赞同“社会利益最大化”。没有冲突的时候，惩罚无辜和奴隶制度是不对的，是不适用的。[①]所以，功利主义不会导致非正义。王海明还认为，有时候社会的个体的人也需要有牺牲精神，需要舍己为利人，需要牺牲个人保全社会的整体利益，如果不这样，每个人可能会面对社会更大的恶。人的生存和发展依赖于社会，如果没有社会，一个人单枪匹马是不适合生存的。如果这个时候没有牺牲精神，人人只为自己，那么社会就会遭受更大的恶，那么也就没有个人的利益可谈。[②]在王海明看来，道德是他律的，这是真理，而义务论是道德目的论，是谬论。道德是为了限制每个人的欲望和自由的“必要的恶”，也就是说，不是看行为对于增进社会的道德总量，而是看对于欲望还有自由侵犯最少，看是否带给个人的经济利益、快乐等总量最高。功利主义者认为，社会总利益是由个人利益之和组成。如果能够增进个人利益之和，那么便是优良的道德，如果减少，那么就是恶劣的道德。

边沁是功利主义的代表，他的伦理学主要集中在《道德与立法原理》这本书中，该书系统地阐述了他的伦理学思想。边沁是一元论的价值观。他认为通过满足个人的利益来满足最大多数的人的利益是功利主义的基本准则。道德是对人性的客观的表现，而不是道德学家一厢情愿的的表达。[③]他认为一个行为增加了人们的快乐就是善，增加了人们的痛苦就是恶。让人快乐幸福的行为就是善，让人痛苦的不信服的行为就是恶。边沁认为个人幸福是社会幸福的基础，为了在追求个人幸福时候，不侵犯别人的幸福，那么得借助于法律。[④]人的行为目的就是追求幸福，所以把幸福作为判断一切行为的标准。[⑤]边沁的学生密尔改进了这种学说，他说除了身体的幸福，还有精神的幸福。他用高级幸福和低级幸福来做了区分，避免把功利主义变成一种低俗的享乐主义。

有学者认为，边沁把“道德公益理解为社会利益总和，但是没有提出怎样保证社会的和谐。”[⑥]也有学者认为边沁的幸福价值在换算的时候会遇到困难，因为认得幸福或者难过都是一种抽象的感受，并且不是恒定不变的，要受到外界的影响。例如，一个人在喝一杯牛奶的时候是快乐的，按照功利主义，那么喝这

① 王海明. 伦理学导论[M]. 上海:复旦大学出版社, 2009. 57.
② 王海明. 伦理学原理[M]. 北京:北京大学出版社, 2001. 118-121.
③ 刘琼豪. 密尔对公里原则的道德哲学辩护[M]. 北京:中国社会科学出版社, 2014. 48.
④ 罗国杰主编. 伦理学名词解释[Z]. 北京: 人民出版社,1984.
⑤ 程炼. 伦理学导论[M]. 北京:北京大学出版社, 2008. 152.
⑥ 王润生. 西方功利主义伦理学[M]. 北京:中国社会科学产版社, 1986. 17.

杯牛奶是道德的。不过，如果在喝牛奶的时候吃了一种特殊的药丸，导致这个喝牛奶的人在喝这杯牛奶的时候感到痛苦而心情不好，那么喝牛奶这一行为此时是道德的吗？也有学者批判说，功利主义虽然是说社会的最大多数的幸福，但是落脚点还是个人，所以不免落入个人主义。功利主义最重要也最普遍的不足和遗憾是："几乎所有功利主义者都不懂得功利原则乃是由若干原则构成的道德原则体系；却以为功利原则只是一条原则，从而将其完全等同于'最大利益净余额'或'最大多数人的最幸福'原则。"[①]

笔者认为，在进行伦理分析和评价的时候，我们需要把义务论和功利主义结合起来看，既看行为效果又看行为动机。以下关于人工智能技术应用中出现的伦理问题的分析，便试图贯彻这一看法。

第二节　道德判断的困境

一、无人驾驶带来的的伦理问题

每次科技的应用都会给传统道德带来挑战，人工智能技术应用的发展也在社会的方方面面带来了许多挑战，让人们陷入道德判断的困境中。让我们来看看下面两个情境带来的伦理问题。

情境（1）：从伦理学视角来看，前面横着一辆拖挂车，如果来不及刹车，无人驾驶汽车如果一边是一只狗，一边是一个人，那么无人驾驶汽车会撞向那一边呢？如果一边是一百只狗，一边是一个人呢？如果写程序时候写成一个人约等于一百只狗，那么会引起很多动物保护者的不满，甚至鄙视。

情境（2）：如果迎面过来的不是一辆拖挂车，而是活生生的十个人，车里是一个人。刹车失灵的情况下，车是选择撞向十个人，车内的人幸存；还是选择撞向旁边的建筑物，车内人因为猛烈撞击而死亡？这是一个伦理困境。

无人驾驶汽车是什么呢？无人驾驶汽车又叫轮式机器人或者自动驾驶汽车。无人驾驶汽车所使用的是一种人工智能技术----机器学习。以车载传感器传递路况或者周围信息，并且通过感知反馈过来的路况、道路上的障碍物、所处的位置等信息，依据车内的自动驾驶仪来实现车辆的行驶。机器上路之前通过机器学习，积累驾驶经验，并且有的无人驾驶车能够把自己积累的驾驶经验上传，然后能够分享给其他的无人驾驶车。

① 王海明. 伦理学原理[M]. 北京:北京大学出版社, 2001. 126.

在当今社会，确实有一部分人能够从中受益，比如一个年龄大的人可能学习驾驶汽车很吃力或者由于体力不便长途驾驶，这时候如果有不需要自己亲自驾驶的无人车来驾驶，确实可以给人们带来诸多方便。无人驾驶汽车不会疲劳、不会注意力分散，可以在多种路况如高速公路、野外、山地等上面驾驶。

特斯拉无人驾驶汽车比较有名，但是却出现了事故。2016 年 5 月，特斯拉在美国佛罗里达州发生了第一起死亡车祸，导致车上的人死亡。事故发生过程是这样的：特斯拉行驶在有双向车道的路上，旁边都是草坪，而且特斯拉无人驾驶汽车能够识别任何障碍物；当时的特斯拉无人驾驶汽车前面有一辆白色左转的拖挂卡车，由于光线强，特斯拉无人驾驶汽车没有识别对面白色的拖挂卡车，由于智能系统搜集的感知信息失误，做出了失误的“控制”，可能是强光导致特斯拉的视觉----图像识别功能丧失，变盲，直接从前面的白色拖挂车下面穿过，挡风玻璃破碎，里面的人员也死亡，造成了悲剧。

让我们回到开头提到的两个情境之中的伦理问题（2），义务论者认为应该保护人的生命安全，不应该选择撞建筑物，因为车上的人是无辜的。不论种族背景、学历、男女老少、社会贡献多少，都应该一视同仁地救助人，这是国际红十字会的救助准则。这种伦理准则是按照义务论的“应当”这条行为原则，这个行为的初衷是保住车上的人的生命。如果初衷是道德的，那么就应当是善的、是道德的。善的道德不会因为对面的人数是十个人或者百个人而性质改变。保住车上人的生命，这一善的行为本身具有道德价值。康德认为“道德完善就是出于义务（即法则不仅是支配他行动的规则，而且是他行动的动机而履行义务）”[①]。

功利主义者则认为上面的选择是片面的。功利主义者认为不仅仅要谋利，而且要看看是为谁、为哪一边谋利。在功利主义者看来，应该选择撞旁边的建筑物，牺牲车内的无辜者，保全前面十个人的生命。功利主义者是判断行为的好坏，是从行为是否能增加社会的整体利益做判断的。十个人的生命价值利益之和比一个人的生命利益总价值要大，因此功利主义者选择保十个人的生命，牺牲车内一个无辜的人。虽然撞十个人与一个无辜的人的死亡都是恶的，功利主义者认为两恶相权取其轻，牺牲车上的一个无辜的人能够避免十个人丧生这个更大的恶。但是义务论者会反问这样子伤害无辜的选择是正义的吗？因为“惩罚”无辜是非正义的。

这就导致了伦理的冲突。这其实是一个现代版的“电车难题”。虽然以前一个

① 康德，郑保华.康德文集[M]. 改革出版社.1997.

人很难遇到这种类似的两难的境地，但是无人驾驶汽车的到来，就让人们不得不面对这样的困境，并且不得不做出抉择。可以想见，无人驾驶汽车的到来将人们带入比以往还要复杂的道德判断的困境。

二、人脸识别的伦理问题

随着社会的发展，人们对于隐私的保护观念也越来越强，人工智能技术应用的发展也带来了隐私保护方面的伦理问。

什么是隐私？隐私权又名为宁居权、私生活秘密权、生活安宁权等，它指公民个人信息、私人住宅以及个人身体隐私等受法律保护，不受非法侵犯。也就是说，个人隐私和公众没有关系，公民有意识地隐藏的信息都属于隐私权保护的范围。对隐私权的保护主要包括：他人不能以任何形式侵犯隐私权；当隐私权被侵犯时可利用法律武器进行保护。[①]从安全方面而言，隐私的核心是安全，试想当一个人住哪里，去什么地方都可以像透明一样被知道，那么就会给人带来焦虑甚至安全隐患。

随着社会的进步，对于人脸识别这项用于身份识别的技术也在不断地得到应用。2016 年春节，火车站进站检测时就用到了人脸识别技术。人的行为特征或者生理特征只要这样几个要求：普遍性、独特性、稳定性、可采集性，就可以作为生物特征来识别。[②]人脸识别因为采集的手续简单、方便，比基于虹膜、指纹、视网膜的身份识别更友好、自然、对用户干扰少、更容易被用户接受、而且能够快速有效地进行身份验证，因此具有非常广阔的应用前景。[③]

人脸识别技术应用到了人工智能技术中的人工神经网络、模式识别、计算机视觉等技术。那么这些人工智能技术的应用与人们的隐私安全有什么关系呢？让我们来看看人脸识别这项技术的工作内容。在人脸识别中，定位人的眼睛往往是识别的第一步。通过人工神经系统把观察到的五官形状特点和之间分布关系传送到模糊神经元系统，然后按照其所储存信息来识别人脸，基于此再分别出对应的五官分布。[④]在进行识别过程中，要综合多方面因素所带来的人脸细小变化，且在此细微变化基础上保证脸部识别的准确性。[⑤]光照和姿态变化

① 冯继宣. 计算机伦理学[M]. 清华大学出版社, 2011.146.
② 景英娟, 董育宁. 生物特征识别技术综述[J]. 桂林电子科技大学学报, 2005, 25(2):27-32.
③ 中国智能制造网. 人脸识别技术应用广泛 “刷脸”时代即将来临[EB/OL]. http://news.dichan.sina.com.cn/2016/10/12/1216729.html, 2017-10-12.
④ 梁路宏, 艾海舟, 徐光档,等. 人脸检测研究综述[J]. 计算机学报, 2002, 25(5):449-458.
⑤ 张翠平, 苏光大. 人脸识别技术综述[J]. 中国图象图形学报, 2000, 5(11):893.

会对人脸识别带来挑战。人脸识别已成为众多重要研究领域的前沿课题。[①]

2016 年发生了这样一件事，一个叫 Nimesh Patel 的人状告 FACEBOOK 公司侵犯了他的隐私，因为 FACEBOOK 公司开始搜集他的眉眼之间的距离、嘴唇信息、面部轮廓等脸部信息，这一举措违反了美国某某州的某某法律。在 Nimesh Patel 状告 FACEBOOK 公司之前，Nimesh Patel 就知道 FACEBOOK 公司在搜集他的信息，他是在 FACEBOOK 公司更进一步搜集到他的面部具体信息时，才将 FACEBOOK 公司状上法庭的。

人脸识别能提取人脸的几何特征，能通过模板匹配来实现识别功能，人的面部信息通过人脸识别记录并且存储下来。当需要找到这个人的时候，可以在茫茫人海中一下子识别出来。但是，这就有可能侵犯人们的隐私权。

从单个的主体而言，尊重隐私也就是尊重人格的尊严。在许多侵犯隐私的法律案件中可以看到，如果不尊重隐私，那么很可能伤害到主体内在的认同和内在的自我。[②]正如吕耀怀在《信息隐私问题的伦理考量》这篇文章中所表达的一样，“尊重他人隐私也即为尊重他人的人格；另一方面，对隐私尊重的忽视也即为忽视尊重他人。”[③]按照康德的伦理学，在主体隐私受到侵犯时，主体自由随之被干涉，而作为人生存的基本权利就是人的尊严和人身自由，因此侵犯他人隐私就是侵犯他人的基本生存权利。[④]

一个功利主义论者会认为，单个主体的隐私应当被保护，因为这样有利于人增进主体的安全感。隐私的核心是安全，安全感具有内在价值。增进每个人的安全感是有利于增进社会利益之和的，虽然安全感不等于幸福感，但是失去安全感可能会消减主体的辛福感。增加主体的安全感可能让主体体会到纯粹的幸福感，而不需要消减幸福感。所以保护主体的隐私是有利于增进社会每个人利益之和的，是善的，是道德的。边沁的情感计算虽然没有明确地指出安全感可以用多少数字来表征，但是如果人的感觉是可以计算的话，相对安全是善的，用正数算，失去安全感是恶的，用负数算，正数显然比负数大，那么增加安全感的情感计算的数值显然是大于消减安全感这个负数值的，因此保护隐私是善的、是道德的。案例中的 Nimesh Patel 不仅失去隐私，而且被侵犯了隐私权。因

① 李武军，王崇骏，张炜,等. 人脸识别研究综述[J]. 模式识别与人工智能，2006，19(1):58-66.
② Brown W S. Ontologcal Security， Existential Anxiety and Workplace Privacy. Journal of Business Ethics，2000 (23)；61 一 65.
③ 毛磊. 保护隐私权:个人尊严和价值的体现[J]. 紫光阁，2003(5).
④ 唐凯麟，李诗悦. 大数据隐私伦理问题研究[J]. 伦理学研究，2016(6):103.

为，由于 FACEBOOK 公司在未经允许的情况下采集他的脸部信息，脸部信息又不能像其他信息一样能够马上轻易地换一个，使得 Nimesh Patel 对于自己隐私权的掌握大大减弱，也就是 Nimesh Patel 被 FACEBOOK 公司侵犯了隐私权。按照功利论者的观点，这减少了 Nimesh Patel 的利益，这是恶的，是不道德的。

如果是个体的隐私与社会整体的利益需要冲突的时候，在做选择的时候会选择社会整体的利益。比如之前提到的春运火车站，因为是为了保护社会整体的安全，所以应用了人脸识别技术来安检。因为这个社会整体的安全利益之和大于个体的安全利益。按照功利论者的观点，个人隐私保护与社会整体的利益相冲突的时候，功利主义者会选择社会整体的安全。从功利论者角度看，社会整体的利益比个体的隐私要大，功利主义者以增进最大利益之和为善。

这样看来，功利主义论者主张保护个体的隐私，当保护个体隐私与社会整体利益相冲突的时候，选择社会整体的利益。

从义务论者的角度看，隐私权属于人的权利，人的隐私权利不应当被侵犯，而且在无论何种情况下都不应当被侵犯。正如康德所言，人是目的本身，而不是手段。人脸识别技术需要大量的数据来进行训练，此时人就被当作手段。FACEBOOK 拥有社交网络，社交网络往往拥有更容易获取用户数据的优势。为了获取人脸的信息数据来训练机器，也是为了更好地服务于社会，还可以获取更多的商业利益。基于此，那么 N 对于 FACEBOOK 公司具有工具价值。此时，案例中的 N 没有被当作目的本身，而是被 FACEBOOK 公司用作获取商业利益的工具，这在康德主义者眼中是不道德的。安检中的人脸识别应用，一个功利论者会认为是为了社会安全的整体利益，是善的，但在康德义务论里却是相反的，因为康德义务论认为，保护隐私是义务，那么在何种情况下都应当保护隐私。这是根据行动的规则来作出评价，不会因为社会利益的后果就不尊重个体的隐私。义务论者认为不应该因为其他的原因就改变这条道德义务。比如有的人在整容前被搜集了人脸识别的信息，私人信息如果一旦泄露，就会给该主体带来舆论压力或者生活中的困扰，成为这项技术的无辜受害者。

两种观点各有其长处，但又都只看到了问题的某个方面。不过，正是因为社会中存在着这两种观点，才有利于为隐私保护的立法提供比较全面的理论参考。美德伦理学认为尊重隐私就是尊重人，因为隐私权是一种具体的人格权，所体现和保护的是人作为一个人的尊严。尊重人的人格尊严，本身就是一种善的品德，所以保护隐私是一种美德。在 FACEBOOK 侵害 Nimesh Patel 的案例中，

显然 FACEBOOK 在侵犯隐私权方面没有这种保护隐私的美德。也就是说，在美德论者眼中，FACEBOOK 公司的管理者没有保护隐私权的品格。

三、谷歌伦理委员会透明度的伦理问题

谷歌应用人工智能的机器学习技术，做了 Google Play、翻译软件 Google Translate、无人驾驶汽车。此外还收购了几家机器人的公司。科技在应用过程中有好的一面，也有坏的一面。显然谷歌这一世界科技巨头公司也意识到了人工智能技术应用带来的伦理问题。比如，像物理学家斯蒂芬·霍金（Stephen Hawking）和企业家埃隆·马斯克（Elon Musk）都对人工智能的进步表示过担忧，霍金甚至认为人工智能是人类最大的敌人。谷歌成立了“谷歌伦理委员会”，来研究与监管人工智能的伦理问题，以避免人工智能技术被滥用。

成立谷歌伦理委员会本来是一件很好的事情，但是由于不透明，引起了不少质疑。对于业内认识的质疑，谷歌也没有给予回应，而且谷歌委员会的成员也没有对外公布。这就引发了麻省理工麻省理工学院媒体实验室主任伊藤穰一(Joi Ito) 的忧虑。

谷歌在几年前把一家名为 Deepmind 的创业公司收购了。当时这家公司向谷歌提出的收购条件是创立一个人工智能委员会。当时 Deepmind 这家公司也意识到了人工智能技术的应用是把双刃剑，因为不乏这样的科幻电影，由于心态不健康的个别科研狂人在某个不知名的小岛，研制出一种杀伤力极大的机器人，结果给人类社会带来杀戮，使人类社会损失巨大。有人指出，对于人类社会真正的威胁还不是类人型机器人，而是基于社交网络和搜索引擎的应用产品。

谷歌伦理委员会成立后，并没有对外公布成员和伦理委员会的工作，这就导致了媒体的质疑。企业以盈利为目的；另一方面，企业也有社会责任，这就需要企业具有康德所说的“自律”。康德说：“意志自律（Autonomie）是道德律与相应义务相统一的原则；对应的，所有的他律（Heteronomie）是没有任何责任约束的，并且还和责任与道德相违背。由于德行的准则就是完全独立于法则的，另一方面还依据准则可约束的立法程序对其他进行限制。而上述所说的独立却不是积极方面的自由，而完全并且能够自身践行的自己立法的就是积极理解的自由。”[①]这就需要企业在信息对外公布与保护自己企业的商业秘密之间保持一个适当的度。

① 康德 著，杨祖陶 绘，邓晓芒. 实践理性批判[M]. 人民出版社，2004.

一个狭隘的功利主义者会认为，谷歌自己研究的知识，产权属于该公司，企业从创办之初就是为了获利，能够为企业获得利益，就是善的。如果谷歌自己研究的知识公布给大众，那么谷歌从自己的公司研究出的知识成果获利就减少了，这是对于谷歌公司不利的。

而一个极端的义务论者会认为，企业就应该对外公布一切信息，因为人工智能技术的发展关系着人类的发展。霍金教授就曾经严重警高人类，人工智能可能导致人类灭绝。因此，该领域的研究应该毫无保留地告知社会，促进决策民主，虽然这可能会让企业受损失。在极端的义务论者眼里，人的行为的目的是为了追求义务和道德，而不是为了获利而去行使道德。

如果谷歌一味地全部对外公开信息，那么就会把自己的知识研究成果全部奉献给别人，自己从中获益就少了，长期让企业吃亏是不现实的。另一方面，企业如果一味地把研究成果为自己的商业服务，却不承担社会责任，就会引起社会大众对它的反感，相反达不到企业想达到的商业目的。过尤不及，亚里士多德意义上的美德伦理论者会认为，企业可以一方面从自己的研究成果中获得应有的合法的利益，一方面又要勇于承担起社会赋予企业的责任。企业需要帮助提高公众对于人工智能技术应用的风险的认识，舒缓大众对于人工智能技术应用的伦理焦虑。适当地对社会公布人工智能技术应用的伦理问题，促进社会社会对人工智能技术应用的决策民主化，从而在企业的利益和公众知情权中保持一个合理的平衡。义务论与功利论之间持久的紧张在人工智能技术应用上带来的道德判断的困境，或许可以通过美德伦理实现一定程度的缓解。

第三节 技术异化的伦理风险

一、无人机关系伦理问题

最近，有一种叫胡椒的机器人，不仅具有情感、文化而且有认知能力，正在被研发。随着人们生活质量的不断提高，人类寿命的延长，老龄人口的数目的增加，“胡椒”可以减轻医院、医疗机构、养老院的压力。胡椒机器人可以缓解老龄人的空虚，陪伴老龄人，给予老龄人精神上的安慰，甚至能够叮嘱老龄人按时服药。这样可以改善家庭护理的质量，减轻社会各方面的压力。[①]

① 智详传感器. 机器人也可以照顾老年人了[EB/OL]. http://info.ec.hc360.com/2017/02/031655893746.shtml, 2017-02-13.

2017 年年初，在中山大学附属一院心脏外科，有一位患有先天性心脏病的男性患者，通过达芬奇机器人成功实施了心脏房间隔缺损（多孔型）修补手术。据了解，这是华南地区首例通过达芬奇机器人手术成功开展的心脏手术。[①]

未来社会，有扫地机器人、聊天机器人、娱乐机器人、护理机器人等，那么必然会产生一种新型的社会关系即人-机器人关系。可以大胆地想象，未来很有可能是一个人与机器人共存的世界。甚至你的同事都可能是机器人。

机器人为人类做了这么多事情，那么机器人有人权吗？现在还是弱人工智能时代，当机器人拥有了类似人类理性的能力和自我意识的强人工智能时代到来后，当机器人照顾小孩、老人，到海底勘探矿产，协助医生做手术，履行了照顾人类、为人类服务尽义务后，它能够拥有权利吗？机器人与人类有交互关系，在契约论者看来，道德不是单独存在的，而是存在于处理关系的过程之中。契约论者认为：“道德并不是某种先在的事物，而是人类为了更好地生存，为了维护自身的利益并且对利益之间的冲突进行调节所发明的东西。所谓道德就体现在拥有利益与理性能力的人，与同样拥有利益与理性能力的人一起签订的有关权利的相互保障与义务的各自履行的契约上，道德就是为了保障权利而履行义务的规则系统。”[②]于是有人提出了这样的道德考量，当机器人和人类一样拥有了智慧、情感等，当它也要求和人类同等的权利，人类会赋予它人权吗？

“所谓人权，是作为个体的人在他人或某一主管面前，对于自己的基本利益的要求与主张，这种要求或主张是通过某种方式得以保障的，它独立于当事人的国家归属、社会地位、行为能力与努力程度，为所有的人平等享有。”[③]谈到人权，首先会谈到生命权，因为生命权是人权的基础。那么因为要消灭恐怖分子，该不该鼓励为了更多的人的生命而选择自我牺牲？无辜的平民有没有权利拒绝这种牺牲？有的人在没有轮到自己的时候会说，要消灭恐怖分子，这样子的误伤也是难免。但是当这个误伤是自己的时候，这个时候个体又会反悔。[④]

可以将机器人伦理的实在论进路的基本思想转化为了亚里士多德式的三段论推理模式:(1)任何拥有特征 P 的实体都具有道德地位 S;(2)实体 X 拥有特征 P; (3)实体 X 具有道德地位。在此，P 表示与道德相关的各种特征如意识、理性能力、

①中山大学. 华南首例达芬奇机器人心脏手术在附属一院成功实施[EB/OL]. http://news.sina.com.cn/o/2017-02-17/doc-ifyarrcc7563027.shtml, 2017-02-17.
② 甘绍平. 人权伦理学[M]. 中国发展出版社, 2009.196.
③ 甘绍平. 人权伦理学[M]. 中国发展出版社, 2009.196.
④ 甘绍平. 人权伦理学[M]. 中国发展出版社, 2009.196.

情感等;S 代表具体的道德地位如道德主体、道德病人等;X 则代表各式各样的机器人。[①]机器人与人类是不同的，但是这种区别并不能够阻碍机器人成为道德主体。[②]如果机器人能够成为道德主体，应该要满足的前提条件是什么，或者说，要满足道德主体地位需要哪些道德特征？契约论认为是有理性和自我意识。“John P.Sullins 将一定程度的自主性、意向性和责任性三者作为道德主体性的条件，他认为，只要机器人能够在某一抽象水平上显示这些特征，那么，它们就是道德主体。”[③]

有不少学者认为，人工智能成为道德主体的前提条件是具备一定的意识能力，能够进行思考。如果机器人没有足够的智能或者理性，那么就不能称为道德主体。也就是说，判断人工智能是否能成为道德主体的核心是是否有认知功能，其次是意识和感知。[④]但是，在强人工智能时代，当具备理性和自我意识的机器人履行了很多为人类服务的义务后，和人类要求得到相应的权利的时候，那么是不是也会采用契约规定机器人拥有哪些权利？当机器人对某个人类做了许多事，这个人对机器人有什么回报和责任？

如果机器人有了情感，那么它也有可能和人类谈判。契约是一种自主的意愿。道德是为了人类而存在的，而不是相反。契约论者认为，这对于要求自保的每个个体而言绝对是有益处的。这样一来，人与机器人的关系中就出现了一种对等的权利与义务的构造。每个拥有为人类服务的义务的机器人，也就拥有了权利；每个拥有权力的机器人，也就拥有了为人类服务的义务。这是一种机器人的权利与义务对等的秩序。而在一个契约论者眼中，道德正是一种对等的秩序，所以让机器人拥有权利与义务的对等，这是道德的。

“robot”一词，来源于捷克语“robota”，意思是“苦工”或“奴役”。有的学者提出，机器人会不会助长人类的奴役意识；多年来，人类适应了人类是万物之主的角色，在主客体对立的情况下，有的人认为，技术是人类创造的，技术的应用理所当然也是为了人类服务。所以，机器人是人类创造的，理所当然也是为了人类服务。也有学者认为，机器人让人类万物之灵的中心地位受到挑战和威胁；但是有的学者认为这是大可不必担心的，该部分学者认为，机器人没有人

① The Moral Standing of Machines :Towards a Relutaonal and Non 一 Cartesian Mora Hermeneutics
② 王东浩. 机器人伦理问题研究[D]. 南开大学, 2014.
③ John P.Sullins.When Is a Robot a Moral Agent? (J).International Review of Information Ethics ,2006, 6(6):23-30.
④ 王绍源. 机器(人)伦理学的勃兴及其伦理地位的探讨[J]. 科学技术哲学研究, 2015, v.32;No.180(3):103-107.

类的社会性，因此不必要担心；有的学者认为，对于机器人的认同的探索即是对人类自身认同的探索；阿西莫夫的机器人三定律建立在人类中心主义的基础上。有学者指出机器人三定律，这是人类对自己创造力的认同；人类中心主义认为，机器人只是为人类实现人类的目的而产生的工具，就算谈到道德也是应该围绕符合人类利益的人类的道德规范。而非人类中心主义者所关注的是机器人本身的尊重，对机器人持有一种非工具论的关切。[①]李德毅教授认为，人机可以实现人机共生。那么会出现一种人机共生的社会。笔者认为，人机共生的社会或许是一种比较合理的设想。

二、军事机器人伦理问题

最近，笔者看到一则新闻《美军军事机器人或将推出陆空混用无人机》[②]。二十一世纪，许多地区仍然战争不断，军用机器人也被用上了战场，比如，阿富汗战争中应用的“packbot”，还有波斯顿公司研发的军用机器人“BigDog”。根据有关数据表明，“在美国对巴基斯坦附反恐战争中，在 2006 年到 2013 年期间，军用机器人共杀死 2514 名武装份子，但与此同时也杀害了 153 名无辜的公民，截至 2014 年，这个数据还在上升。”[③]

如果从功利主义的角度出发，打击恐怖分子是为了避免更大的牺牲。所以应当利用军事机器人打击恐怖分子，尽管军事机器人会导致一定数量的无辜贫民的牺牲。从功利主义角度出发去计算，社会总的利益大所以就应该选择利益大的。但是军事机器人伤害的是人的生命，生命与钱财等“身外之物”的利益有本质的区别。功利论不能作为对于生命的考量，生命只有一次，是独特的、唯一的，失去便不再复有。就像一种极端的观点，认为要求生病了的人，不要浪费社会资源，实施安乐死，这是不可接受的。所以功利主义的观点在此处是站不住脚的。从义务论的观点出发，伤害人的生命，无论因为何种原因、无论在何种条件下都是不道德的。

亚里士多德的德性伦理认为道德的结果取决于品行，而不是某种原理的知识。在道德实践中，德性伦理是根据行动者的品德来判断的。百度首席科学家吴恩达说：“人工智能也可能具备意识”。假如进入强人工智能时代后，军事机器

① 王绍源. 机器(人)伦理学的勃兴及其伦理地位的探讨[J]. 科学技术哲学研究, 2015, v.32;No.180(3):103-107.
② 非常在线. 军事机器人或将推出陆空混用无人机[EB/OL]. http://www.veryol.com/article/150433.html, 2017-03-04
③ 杨笔锋，詹艳军.基于射频识别的智能车辆管理系统设计[J].计算机测量与控制.2010.18(1):97-99.

人具有了类似人类的认知能力、人类的意识、理解能力，理性等。那么军事机器人可以拥有道德主体地位，军事机器人所做的行为就是体现军事机器人的德性。德性并不是遵守某种普遍有效的原则，更不在于理性的利益计算，而是强调行为者本身善的认知。也就是说，军事机器人如果有善的认知，那么在德性论者眼中便是有道德的。军事机器人不应该误伤平民的生命，要在执行消灭恐怖分子的过程中保护平民的生命。这样的军事机器人是有德性的军事机器人。

当出现了上述案例中军事机器人伤害平民生命的这种情况时，谁人该为此负责呢？是设计者？工程师？还是机器人自己？有学者认为，军事机器人的使用者、生产商、研发者都应该为军事机器人造成的不好的后果负责。那么机器人能否承担道德责任？如果机器人是道德主体，它能不能承担道德责任？ 如果机器人能够承担道德责任，那么他应该遵守怎样的道德原则呢？以往的文献只是指出人工智能、研发者、使用者需要负责，但是具体的分配却没有详细的解答，为何要负责也没有说。

责任分配的依据，即角色分配的依据，只能是行为主体的认知能力和责任能力。[①]人工智能成为道德主体的前提条件是具备一定的意识能力，能够进行思考。研发机器人的工程师是有认知能力的，因此具有责任。使用机器人的应用者也有认知能力，那么也有责任。强人工智能时代具备理性且有类似人类的自我意识的机器人，也有责任。那么这三者之间的责任如何划分？根据认识能力划分吗？工程师认知能力最高，责任最大；普通的使用者的认知能力次之，责任次之；当机器人的有认知能力但是还不及人类的时候，责任也比人类小。那么当机器人的认知能力逼近甚至超越人类的时候，责任也就逼近人类或者超越人类。也就是说，当机器人的认知能力超越人类的时候，一个机器人对于行为所需要付出或者需要承担的的责任甚至超越人类。当机器人没有足够的智能时，不能对人类的生死作出判断，当机器人具备足够的智能还有理性的时候，机器人能够对人类的生死作出判断。具有类似人的情感，拥有做事的动机和行为的过程，那么也需要为自己所做的行为负责，那么也就有理由让机器人接受惩罚。机器人的自住性越高，风险越大，有的学者指出，此时不应该让指挥官为机器人的错误负责。

根据罗尔斯的观点，军事机器人消灭恐怖分子，打一场正义的战争是必要的，但是军事机器人伤及无辜的平民的生命却不能算在此列。这是一种责任伦

① 程东峰. 责任伦理导论[M]. 人民出版社, 2010.

理的困境。

爱因斯坦曾经就说过，许多科学家只是对科学感兴趣，但是科学所带来的社会后果却很少考虑。他们“为了科学而科学”，这样很容易被社会的不良分子所利用。当科技触目惊心的副作用显示出来的时候，我们不得不对科技是否符合造福于人类这一原则开始反思。不仅要义务论的初衷善，也要符合功利论的结果的善。科学家不要把人工智能技术应用所带来的副作用看作是与自己无关的事情。也就是应当把在军事机器人追求科技真理与军事机器人造福于人类的价值判断相结合。

从接受军事命令而言，军事机器人既是道德行为的接受者，从军事行动而言，军事机器人也是道德行为的输出者。当军事机器人作为道德主体实施军事行为的时候，那么军事机器人具有行为动机，完成任务；军事机器人在完成任务的过程中，也有行为的后果。当产生伤害了无辜平民的后果时，军事机器人也应该为此而负责，因为军事机器人也是道德主体，道德主体有意愿的自由，也就是说，军事机器人可以选择不做不符合道德的事情。如果军事机器人对于人类的生死有判断的话，军事机器人就应该知道伤害无辜的人的生命是不符合道德的。如果军事机器人做了选择执行不符合道德行为，那么就应该为不良后果负责。

军事机器人的研发者、生产者、使用者，不应该为了科学而科学，认为军事机器人所造成的不良后果与自己无关。因为科技的研发应该符合造福于人类的原则，而伤害无辜的平民的生命这一不良后果，显然是不符合造福人类的原则的。因此，技术上的成功，不等于道德上的应当，军事机器人的研发者、设计者、使用者也应该为无军事机器人造成的不良后果负责。机器人的研发者应该具有人文关怀，人文关怀不会直接对于科技产生影响，但是会成为科技的研发者、生产者、使用者的内驱动力，会在设计机器人的时候“嵌入”人文关怀的理念，从而更好地为人类服务，造福于人类，造福于世界和平。军事机器人涉及到人类的安全问题，应当引起政府、科学家们、哲学家、社会大众的充分重视，这并不是杞人忧天，而是一个紧迫的社会问题。[①]

总之，人工智能技术原本是为了服务于人而创造的，但是在应用的过程中，有时候却反过来给人带来伤害，比如上述案例中军事机器人原本是为了消灭恐怖分子、保护人们的安全，但是却伤害了无辜平民的生命。人工智能技术的进

① 杜严勇. 论机器人权利[J]. 哲学动态, 2015(8):83-89.

步带来的风险有待我们去评估，当机器人的研发者申请专利的时候必须要通过伦理委员会的审核，以确保我们的社会以及我们的后代不会因为此项科技遭受危险。人工智能技术在某些方面不能无限地使用，而是应当有限制地使用。还有，人类必须对自己的行为负责，因为军事机器人均有不确定性，技术研发者和科学家作为技术研究开发的主体，也应该对军事机器人负有道德责任。在研发的过程中遵守相应的伦理规范，对公众负责。

第四节 人工智能的智能和自主性进化对人类主体性的挑战

一、AlphaGo 的伦理问题

最近引起社会各界高度关注的是这样一件事，在 2016 年 3 月 15 日，AlphaGo 与九段棋手李世石长达两个小时的比赛，AlphaGo 在前三局和第五局取胜，李世石在第四局取胜。本来五局三胜结局胜负已经有了结果，但是为了满足人们对于 AlphaGo 与人类比赛的好奇，商定好了要把五局比完。比赛以下快棋的形式进行。最终 AlphaGo 以 4:1 战胜了李世石。

AlphaGo 是由谷歌旗下的 DeepMind 公司开发的，运用的是深度学习技术和蒙地卡罗搜树技术。它不是依据具体的知识，而是通过数目巨大的人类的棋谱图像进行训练学习。训练他的是一位普通段位的棋手。AlphaGo 拥有类似人类的“直觉”，其实这是应用蒙地卡罗搜树技术搜索出类似的棋谱，是一个“布局 vs 布局”的映射过程，然后 AlphaGo 通过计算机强大的硬件计算速度，计算出走哪一步更有利于胜利。

事实上，谷歌的 AlphaGo 的机器学习技术并不是突然出现，而是之前就有，只是谷歌将多种机器学习技术进行了结合。比赛结束后，有人指出，下快棋有利于 AlphaGo，因为机器里面的棋谱已经有那么多，搜素速度本来就很快，所以下快棋或者下慢棋对机器没有什么影响。而对于人类而言却是不利的，因为下快棋，就得需要快速的思考，而人类思考需要时间。在《神奇的数字 7±2 》这篇论文中，就提到我们人的工作记忆的容量大概是在 7 加减 2 的范畴下面，所以我们存的内容可想而知。不管 7 是怎样大小的一个数字，意味着我们的工作记忆很有限。这就是我们为什么不能一心二用的原因，除非是任务被我们高度智能化了，占用我们 CPU 的资源已经被我们降得很低了，才可以一心二用。比如，有的人骑自行车的同时做其他一件事，那是那个人骑自行车的技巧已经很

熟练了。[①]有学者认为，如果 AlphaGo 在第前三局输给李世石，那么说明 AlphaGo 也有破绽。如果李世石在第五局赢了 AlphaGo，那么还可以说明人类通过找到 AlphaGo 的破绽可以找到战胜它的方法。可是李世石是在第四局赢的 AlphaGo，所以这些就无从推论了。

机器与人类弈棋在历史上就有。第一次是 1989 年 10 月，在美国纽约艺术馆进行的比赛。比赛双方分别是国际象棋世界冠军卡斯帕罗夫对弈由卡内基一梅隆大学研发的计算机“深思”。最后的结果是人类代表卡斯帕罗夫以 4:2 赢了“深蓝”取得胜利。第二次是 1996 年 2 月，在美国费城，比赛的双方分别是卡斯帕罗夫对弈“深蓝”，“深蓝”以 3.5：2.5 赢了卡斯帕罗夫。第三次是 1997 年，参赛的双方是卡斯帕罗夫和美国 IBM 公司的超级计算机“深蓝”。比赛结果是“深蓝”以二胜一负三平战胜了当时的世界冠军卡斯帕罗夫，成为首个在标准比赛时限内击败国际象棋世界冠军的电脑系统。

人们在关注人机大战的时候，不仅仅是关注下棋的胜负，而是关注人工智能有有智慧吗？人工智能会超过人类智能吗？这些在哲学和科学领域都有长期的争辩。

英国计算机专家图灵在他的 1950 年发表的《机器能够思维吗？》这篇文章里早就说过，他希望“在一切纯智力领域内，机器将是最终和人相竞争。但是，最好从哪一个领域开始呢?.....许多人以为，像奕棋这种很抽象的活动也许是最好的领域”。[②]回过头来，我们也不难看出确实如此。拿 AlphaGo 与李世石下围棋来说，机器与人下棋的特点是：1，下棋的规则很明确，一个棋子怎么走，它的规则是很明确的。2，棋局的定义也非常清晰。这对计算机非常重要。3，棋盘的空间是非常有限的，就这么一个大的空间，国际象棋 64 个格子。这三点极其重要，意味着下棋这件事是可以精确地用计算机算法来描述或者表征的。写过程序的人都喜欢规范的数据，而这些特点正好产生很规范的数据。哪些是不规范的数据呢？我们的自然语言。危辉认为，下棋是人工智能界的“软柿子”，是比较容易做的。语言、图像是人工智能界比较不容易做的，机器翻译，图像识别，现在也做得不是很好。

那么机器到底有没有智慧？“智慧”和“智能”的定义的区别是，从狭义上讲智

① G. A. Miller, 陆冰章, 陆丙甫. 神奇的数字 7±2:人类信息加工能力的某些局限[J]. 心理科学进展, 1983, 1(4):53-65.

② 自然辩证法研究通讯编辑部. 控制论哲学问题译文集[M]. 商务印书馆, 1965.

慧比智能的范围要大一些。从广义上讲，两者没有区别。有的学者认为，AlphaGo在下棋的过程中确实有些棋步是没看过的，这是它的智慧。但是复旦大学的教授危辉认为，这只是 AlphaGo 搜索人类棋谱后走的，有的棋步是比较少见，但也是已经有人类走过的。还有另一种是根据现有的走步方法进行改良，他甚至举出炒菜要放三个辣椒，都是也可以只放一个。他否认这是一种创新。

其次，人工智能是否会超越人类智能呢？复旦大学的危辉教授比较乐观，认为 AlphaGo 与李世石的对弈不过是下棋而已，并不能说明人工智能就超越了人类。因为谷歌 Deepmind 公司研发的 AlphaGo 应用的人工智能技术的机器学习学习技术很早之前就有，并不是什么高深神秘莫测的技术，并不是什么断裂式的技术突破。中山大学人机交互实验室主任翟振明教授认为，与其说是 AlphaGo 战胜了围棋选手李世石，倒不如说是在单一的抽象博弈方面，AlphaGo 里面的技术群体战胜了天赋极高的自然个体九段棋手李世石。危辉教授也认为，“用蛮力”谁不会呢？他认为人机大战输赢都是人类的胜利，短期内不会出现类似电影《终结者》里面机器人统治人类的景象。翟振明教授认为，人工智能还没有达到“强人工智能”的境界，即还没有自我意识和自由意志，因此 AlphaGo 没有独立的“自己”，也就没有所谓的人机“大战”。因此，这场比赛的输赢，都是人类自我的一种投射。另外，人类有情绪，而机器没有，当然也不受情绪的影响，这也是机器赢的原因之一。AlphaGo 没有人类的情感、意向，不能像人类一样称为主体，这也是只能称呼它为工具机器且是没有第一人称的机器的原因。如果真有一天，机器达到了强人工智能阶段，也只是人类被新人类征服而已。[①]有的学者认为，AlphaGo 在比赛中取得胜利只能说明它下围棋好，不能说明 AlphaGo 智商高，就像如果有一天，人工智能做 IQ 智力测试得了很高的分数，只能说明它做 IQ 测试做得好，不能说明它智商高。“‘智能’是人类最引以为豪的能力，也是已知世界中最复杂的现象之一，背后一定隐藏着深奥的客观规律。”[②]更有极端持有机器智能不会超越人类智能的学者认为，AlphaGo 不过是一堆算法罢了，说人工智能威胁人类未免是一种杞人忧天。

那么到底是不是这样？也有许多人持有不同的观点。 很早以前，卡斯帕罗夫输棋后深有体会地说:“几年前，对于计算机是否拥有智慧，我总是抱之一笑。这种很容易用智谋战胜的机器固然算得快且多，但是它能有智慧吗?计算机本身

① 刘虎功. 中山大学人机互联实验室主任翟振明谈围棋人机大战[N]. 长江日报, 3-15(16).
② 黄铂钧. AlphaGo 来了![J]. 科学世界, 2016(4):4-11.

分辨不出显而易见的相同或不同，而是通过千万次运算来区分，这根本不能算是智慧。然而现在我不再那么肯定了。1996 年我在费城同‘深蓝’的比赛迫使我重新考虑这个问题。有时我真的会有这台机器偶尔也会有智慧的感觉”。[①]

中国科学院自动化研究所复杂系统管理与控制国家重点实验室的王飞跃认为，人工智能威胁论就像玛雅预言一样无可辩驳，除了等待。王跃飞认为，人工智能认知科学家和思想家明斯基(Minsky)所说的“是什么让我们又智慧？其中的奥秘是并不存在奥秘。智慧的力量来自于我们的多样性，而不是来自于任何单一、完美的理论”[②]这句话中的“我们”是指人类而非机器，“一旦离开了人类的智能，那么机器也将是没有力量的机器”[③]。另外，怎么区别看两个主体的能力强[④]，主要是看实践结果。如果说一场战争是一场敌我的博弈。那么，围棋只是对于战争的模拟，也是类似战争的博弈，只是这种战争是用围棋模拟的，是比战争小的博弈，而 AlphaGo 却在这场战争中赢得了这场对弈的战争。如果将来人工智能上战场了，和人类真正在战争中博弈，那么人工智能难免也会胜利。这个时候不能说人类有情感，所以人类厉害。说不通，因为实践结果在那里。“不论是外行还是专家，行为是智力的特征。”如果认为人工智能继续发展，哪怕超越了人类，或者代替人类接管地球，也是一种自然的类似达尔文进化论现象，那么这种观点是把自己摆在“上帝之眼”的位置引起的“造世伦理学问题”。

AlphaGo 的设计者说:“这将是一种完全不同的方式。你需要从头开发，让机器人学习新东西，处理不可预期的事件。"人类智能（Aritificial Intelligence）直接触及人类对于宇宙的思考。比如持有“非人类中心主义”的人，认为就如达尔文的进化论一样，优胜劣汰，如果有一天人工智能超越了人类，那么人类也就像恐龙的灭绝一样是一种自然而然的现象。持有这种观点的人认为，人工智能因为更高的智能，而代替人类接管地球，是一种更进步的物种取代一种落后的物种。 基于此观点，笔者想提出的是，如果按照这种“非人类中心主义”的逻辑，如果宇宙中有一种比人类更高智慧的的物种进入地球，那么人类就应该理所应当地被淘汰、被奴役、被服从、被灭绝吗？笔者认为，答案在绝大部分人的心里是“不”且“坚决不”。

“人类中心主义”强调的是服务于人类的根本利益，从功利主义角度看，是满

① 童天湘. 从“人机大战”到人机共生[J]. 自然辩证法研究, 1997(9).
② Minsky M. The society of mind[C]// Simon & Schuster, Inc. 1986:371－396.
③ 王飞跃. 从 AlphaGo 到平行智能:启示与展望[J]. 科技导报, 2016(7):72-74.
④ 霍金斯. 人工智能的未来[M]. 陕西科学技术出版社, 2006.27.

足社会每个人的利益之和的。一个功利主义者会认为，机器的智能不能用于伤害人类。必要的领域应当限制机器的智能发展程超越人类的智能，哪怕允许机器的智能超越了人类，但是掌握权还是得在人类手中。

在一定程度上，伦理是一种认知的追求，一个超级智慧也许可以比人类思想家做得更好。这意味着，关于伦理的问题，只要他们有正确的答案，可以通过推理和证据加权来达到，超级智能可以比人类更准确地回答。这同样适用于政策和长期规划问题，当谈到哪些政策将导致哪些结果，以及哪种手段将能最有效地实现既定目标的问题，超级智慧在这些方面可能将超过人类。因此，有很多问题，如果我们已经或即将获得超级智能，我们不需要自己回答；我们可以将许多调查和决定委托给超级智能。例如，如果我们仔细考虑了很长时间，如果我们不确定如何评估可能的结果，我们可以要求超级智能估计我们如何评估这些结果，因为他们有更多的记忆和更好的智能。当超级智能制定目标时，并不总是需要给出这个目标的详细、明确的定义，我们可以利用超级智慧来帮助我们确定我们的请求的真正目的，从而减少由于不恰当的措辞或误解导致对我们在回顾中看到不想要的结果的风险。[①]

在最近由著名的物理学家霍金、特斯拉 CEO 马斯克等参与讨论并且刚刚公布的《艾斯罗马人工智能 23 定律》 中，在伦理和价值观这一块就有：

第六条，安全。AI 系统在其整个使用寿命期间内应安全可靠，并在可行的情况下可验证。

第七条，故障透明度。如果一个 AI 系统造成损害，应该能够确定原因。

第八条司法透明度，任何自助系统参与司法决策都应提供令人满意的解释，并有人类权力机构审核。

第九条，责任。设计师和建设者，是先进人工智能系统的使用、滥用和行动道德意义上的利益相关者，他们有责任也有机会去塑造这些意义。

第十条，价值取向。设计高度自治的人工智能系统时，应该使它们的目标和行为，在整个运营过程中与人类价值保持一致。

第十一条，人类价值观。人工智能系统的设计和操作应符合人类的尊严、权利、自由和文化多样性的理想。

第十二条第十二条，个人隐私。人们应该有权访问，管理和控制他们产生的数据，因为 AI 系统有能力分析和利用这些数据。

① Nick, Bostrom. Ethical Issues in Advanced Artificial Intelligence[J]. 2003, (12): 1-5.

第十三条，自由和隐私。AI 对个人数据的应用不得不合理地限制人们的真实或感知的自由。

第十四条，共享利益。AI 技术应该使尽可能多的人受益。

第十五条，共享繁荣。AI 创造的经济繁荣应该广泛分享，以惠及全人类。

第十六条，人类控制。人类应该选择如何以及是否将决定权委派给人工智能系统，以完成人类选择的目标。

第十七条，非颠覆。控制高度现金的人工智能系统所赋予的权利，它应尊重和改善而不是颠覆社会健康所依赖的社会和公民进程。

第十八条，AI 武器竞赛。应该避免进行致命自主武器的军备竞赛。[①]

社会已有这样的共识，就是并不是要研发出取代人类的人工智能，而是人类开发出来的大多数人工智能将与人类合作，以实现最佳性能。

二、人工智能的道德设计

无人驾驶汽车里的案例，也引发了如何给人工智能设计伦理的问题。在美国 2016 年的《为人工智能未来做准备》这份政府报告里，显示有一多半的受访者认为，未来会出现与人的智能相当的智能机器人。它们的道德观价值观是否能够符合人类的价值观、道德观？那么给智能机器设计道德的路径就很重要了。

笔者在本文中曾阐述了三种道德设计路径：

第一种是“至上而下”的设计，这种理论从伦理学角度讲是康德义务论。就像康德认为诚实是永远应当的，杀人、作恶是永远不应当的。康德义务论的优势在于康德义务论的形式，可以很明确地通过编程，明确地告诉人工智能主体应当做什么，什么是永远不应当做的。“应当做什么”这种形式很精确，适合把人类的自然语言转化成计算机语言。但是在需要具体情况具体分析的情况下，这种伦理的设计也存在着缺陷。在弱人工智能时代，机器没有情感，可能会能够“服从”地“接受输入”人类价值观的算法，帮助人类做一些很危险的工作，比如矿洞里工作、野外作业等。但是，随着强人工智能的发展（“所谓‘强’，指的是超越工具型智能，达到第一人称主体世界，意向爱恨情感乃至自由意志统统发生。”[②]）如果机器获得了情感之后，它们还会心甘情愿地“服从”人类告诉它们的“应当”做什么的伦理规则吗？而且有时候对于同一个伦理问题的解决本身也存在着各

① AI 世代. 人工智能的 23 条“军规”，马斯克、霍金等联合背书[EB/OL]. http://tech.qq.com/a/20170207/031641.htm, 2017-02-07.

② 刘虎功. 中山大学人机互联实验室主任翟振明谈围棋人机大战[N]. 长江日报, 3-15(16).

个伦理观点的争议。如果一个被输入了“应当”做什么的机器在照顾年迈的老人的时候不懂得变通，依旧是按照先前输入的“应当的”伦理准则行事，可能会对被照顾的老人造成伤害。比如，康德伦理认为“撒谎”在任何时候都是不应当的。如果这个时候一个人在被追杀，那么这个人躲在了你的屋子里面，此时追杀他的人群到了，问那个人是否在这里，从康德义务论角度认为“撒谎”在何种情况下都不应当做的，此时康德义务论者会认为应当讲真话，那么这个人就没命了。同理，被人工智能照顾这位老人，也会在不懂得变通时候，而伤害到被照顾的老人。从功利主义论者角度看，生命的价值比说真话的善的价值高，此时会选择为了保护住那个人的生命而选择“撒谎”。功利主义论者会在价值产生冲突的时候比较两者的利益大小，选择价值高的。单从这方面来讲，功利主义者有时候比义务论者要懂得变通。

第二种设计是一种“至下而上”的路径。美国达特茅斯学院哲学系教授摩尔(James H.Moor)把内在地按照某种伦理规则运行的机器，依据其伦理判断与行为能力从低到高把机器区分为隐性道德行为体(Implicit Ethical Agent)、显性道德行为体(Explicit Ethical Agents)以及完全道德行为体(Full Ethical Agents)[①]。让显性道德的机器通过训练不断积累知识和经验，不断通过实践学习，然后从失败的实践经验中总结教训再改进，从成功的经验中积累经验，最终具有道德判断能力。一般通过足够的训练，机器也可以有很快的进步，甚至在很多领域超过了人类，比如下棋。通常机器学习的训练由机器学习专家完成。我们在获取信息的时候，不如显性的人工智能那样快速地作出选择。但是也有人提出了担忧，人工智能的情感、道德意识前提是强人工智能的实现。那么在强人工智能阶段，机器是否能作为道德主体？如果我们假设人工智能能够作为道德主体，那么我们人类在应用人工智能技术的时候，在做伦理问题选择的时候，机器是按照人类世界的伦理道德规则来行动吗？有人提出这样一种担忧，不知道机器在行动的时候，机器的自治性会使人工智能机器依据何种伦理道德准则。有的学者认为，机器是为人类的根本利益服务的，这是一种人类中心主义的看法。但是，如果智能的机器在拥有认知能力、感情、意识能力后它有自己的“判断了”，它还会认可基于“人类中心主义”为人类服务的伦理道德准则吗？也有一部分人认为，这是多虑，因为机器的智能无限逼近人类智能甚至是超越人类的智能，是基于人类对

① Moor J H. The Nature, Importance, and Difficulty of Machine Ethics[J]. Intelligent Systems IEEE, 2006, 21(4):18-21.

于人类智能现象的解释，而现阶段人类还没有完全把人类智能现象解释清楚，所以，距离人工智能超越人类智能还差得远。另一方面，有的人并不需要有太多自己判断的机器，他们需要的是能够执行行动的机器。

第三种“混合型解决方法(hybridresolution) ”是一种基于亚里士多德的美德伦理学建构的“混合型方法”，在一定意义上可以称之为“混合型解决方法”。亚里士多德伦理学重要的思想是中道学说，也就是类似中国的“中庸”，主张不极端，懂得变通。这或许能克服几种方法的僵硬不懂变通，特别是基于康德义务论的第一种路进。

澳大利亚的企业在“全球人工智能成熟度”这项调查中的排名中就不高，是因为在人工智能技术应用中没有涉及伦理问题从而导致了该国的人工智能的发展。[①]在人工智能技术应用的过程中需要要重视伦理问题，每一种伦理设计的路径都各有优势，又有局限性，我们可以根据具体的人工智能技术应用的产品的场景来选择哪一种道德设计路径。

科学技术是人类创造的，目的是为了人类服务，造福于人类。笔者在本文中将这些人工智能技术应用的伦理问题一一分析出来，以期有助于更好的去寻求解决之道。让人工智能技术更好的服务于人类的发展。路慢慢其修远兮，吾将上下而求索。

① 环球科技. 调查 ：伦理阻碍 AI 技术在澳大利亚发展[EB/OL]. http://www.techweb.com.cn/column/2017-01-18/2475399.shtml, 2017-01-18.

参考文献

一 著作类

[1] RAYKURAWEIL. 奇点临近[M]. 机械工业出版社, 2015.
[2] 布律诺·雅科米, 雅科米, Jacomy,等. PLIP 时代:技术革新编年史[M]. 中国人民大学出版社, 2007.
[3] 玛丽·雪莱著，胡春兰、侯明古译. 弗兰肯斯坦[M]. 北京:人民文学出版社, 2004.
[4] 王悠然. 预防人工智能伦理缺失[M]. 北京:中国社会科学报, 2015. 1-2.
[5] 周昌乐. 无心的机器[M]. 湖南科学技术出版社, 2000.
[6] 林德宏. 人与机器:高科技的本质与人文精神的复兴[M]. 江苏教育出版社, 1999.
[7] 麦好. 机器学习实践指南:案例应用解析[M]. 机械工业出版社, 2014.
[8] 何华灿, 李太航等. 人工智能导论[M]. 西北工业大学出版社, 1988.
[10] 徐志敏, 李栗. 人工智能: 梦想现实未来[M]. 四川教育出版社, 1992.
[11] 吴胜, 王书芹. 人工智能基础与应用[M]. 电子工业出版社, 2007.
[12] 王万森. 人工智能原理及其应用:(第 2 版)[M]. 电子工业出版社, 2007.
[13] 廉师友. 人工智能技术导论[M]. 西安电子科技大学出版社, 2007.
[14] 杨祥金. 人工智能[M]. 科学技术文献出版社重庆分社, 1988.
[15] 蔡自兴. 人工智能辞典[M]. 化学工业出版社, 2008.
[16] 杨祥金. 人工智能[M]. 科学技术文献出版社重庆分社, 1988.
[17] 蔡自兴, 徐光祐. 人工智能及其应用.第 4 版[M]. 清华大学出版社, 2010.
[18] 徐志敏, 李栗. 人工智能: 梦想现实未来[M]. 四川教育出版社, 1992.
[19] 芮延年. 机器人技术及其应用[M]. 化学工业出版社, 2008.
[20] 杨宪泽. 人工智能与机器翻译[M]. 西南交通大学出版社, 2006.
[21] 王树林. 人工智能辞典[M]. 人民邮电出版社, 1992.
[22] 王万良. 人工智能及其应用.第 2 版[M]. 高等教育出版社, 2008.
[23] 麦好. 机器学习实践指南:案例应用解析[M]. 机械工业出版社, 2014.
[24] 霍金斯. 人工智能的未来[M]. 陕西科学技术出版社, 2006.
[25] 王万森. 人工智能原理及其应用.第 3 版[M]. 电子工业出版社, 2012.
[26] 冯定. 神经网络专家系统[M]. 科学出版社, 2006.
[27] 王万森. 人工智能原理及其应用[M]. 电子工业出版社, 2000.
[28] 程炼. 伦理学导论[M]. 北京大学出版社, 2008.
[29] 蔡元培. 中国伦理学史, 蔡元培[M]. 江苏文艺出版社, 2007.
[30] 伊曼努尔·康德, 康德, 孙少伟. 道德形而上学基础:全新译本[M]. 中国社会科学出版社, 2009.
[31] 梯利. 伦理学概论[M]. 中国人民大学出版社, 1987.
[32] 王润生. 西方功利主义伦理学[M]. 中国社会科学出版社, 1986.
[33] 斯密何丽君. 道德情操论[M]. 北京出版社, 2008.

[34] 徐建龙. 伦理学理论与应用[M]. 合肥工业大学出版社, 2009.
[35] 罗国杰主编.马克思主义伦理学[M].北京:人民出版社.1982.
[36] 王泽应主编.伦理学.北京:北京师范大学出版集团[M].2012.
[37] 王海明. 伦理学原理[M]. 北京大学出版社, 2009.
[38] 王泽应. 伦理学[M]. 北京师范大学出版集团. 2012.
[39] 朱贻庭. 应用伦理学[M].上海辞书出版社.2013.
[40] 叔本华, 孟庆时, 任立. 伦理学的两个基本问题[M]. 商务印书馆, 2011.
[41] 李小科. 正义女神的新传人[M]. 河北大学出版社, 2005.
[42] 王淑芹，伦理秩序与道德研究[M].中央编译出版社，2015.
[43] 王海明. 伦理学导论[M]. 复旦大学出版社, 2009.
[44] 王海明. 伦理学原理[M]. 北京:北京大学出版社, 2001.
[45] 刘琼豪. 密尔对公里原则的道德哲学辩护[M]. 北京:中国社会科学出版社, 2014.
[46] 王润生. 西方功利主义伦理学[M]. 北京:中国社会科学产版社, 1986.
[47] 程炼. 伦理学导论[M]. 北京:北京大学出版社, 2008.
[48] 朱伯崑. 先秦伦理学概论[M]. 北京:北京大学出版社, 1982.
[49] 廖申白，伦理学概论[M].北京师范大学出版集团，2009.
[50] 冯继宣. 计算机伦理学[M]. 清华大学出版社, 2011.
[51] 康德, 郑保华.康德文集[M]. 改革出版社.1997.
[52] 康德 著, 杨祖陶 绘, 邓晓芒. 实践理性批判[M]. 人民出版社, 2004.
[53] 自然辩证法研究通讯编辑部. 控制论哲学问题译文集[M]. 商务印书馆, 1965.
[54] 霍金斯. 人工智能的未来[M]. 陕西科学技术出版社, 2006.
[55] 冯继宣. 计算机伦理学[M]. 清华大学出版社, 2011.
[56] 王前，杨慧明. 科技伦理案例分析[M]. 北京:高等教育出版社, 2009.

词典

[1] 朱贻庭. 应用伦理学[Z]. 上海辞书出版社,2013.
[2] 罗国杰主编. 伦理学名词解释[Z]. 北京: 人民出版社,1984.

二 学位论文

[1] 李俊平. 人工智能技术的伦理问题及其对策研究[D]. 武汉理工大学, 2013.
[2] 计海庆. “机器人”观念的形成及其影响的哲学考察——以分析技术的本质和人与机器(人)关系为视角[D]. 复旦大学, 2005.
[3] 魏鸿. 科幻电影中人机关系认同研究[D].华东师范大学.2016.
[4] 王晓楠. 机器人技术发展中的矛盾问题研究[D]. 大连理工大学, 2011.
[5] 王东浩. 机器人伦理问题研究[D]. 南开大学, 2014.
[6] 姚洪阳. 试论人机关系的历史发展及其文化考量[D]. 长沙理工大学, 2010.
[7] 龚园. 关于人工智能的哲学思考[D]. 武汉科技大学, 2010.

三 期刊

[1] 胡增顺. 人工智能的伦理问题[J]. 开封大学学报, 1998, 12(3): 57-59.

[2] 陶悦宁. 论西方科幻电影中的新型伴侣关系[J]. 影视长廊, 2016, (3): 39-44.
[3] 朱勤, 王前. 欧美工程风险伦理评价研究述评[J]. 哲学动态, 2010(9):41-47.
[4] 杜严勇. 人工智能安全问题及其解决进路[J]. 哲学动态, 2016(9):99-104.
[5] 杜严勇. 现代军用机器人的伦理困境[J]. 伦理学研究, 2014(5):98-102.
[6] 杜严勇. 机器人伦理研究方法论原则[J]. 中国社会科学报, 2015.1-2.
[7] 于雪, 王前. “机器伦理”思想的价值与局限性[J]. 伦理学研究, 2016(4):109-114.
[8] 杜严勇. 机器伦理刍议[J]. 科学技术哲学研究, 2016, 33(1):96-101.
[9] 张一南. 人工智能技术的伦理问题及其对策研究[J]. 吉林广播电视大学学报, 2006(11): 123-124.
[10] 陈升, 孙雪. 国内外军用机器人的现状、伦理困境及研究方向[J]. 制造业自动化, 2015(11):27-28.
[11] 翟振明. “强人工智能”将如何改变世界——人工智能的技术飞跃与应用伦理前瞻[J]. 人民论坛·学术前沿, 2016(7):22-33.
[12] 贺欣晔. 科幻文学中人工智能与人类智能的关系[J]. 沈阳师范大学学报(社会科学版), 2016, 40(2):111-115.
[13] 杜森, DuSen. 人工智能的法律与伦理意识形态问题研究[J]. 黄冈职业技术学院学报, 2016, 18(2):64-67.
[14] 徐豪, 邹锡兰. 人工智能“风口”,医疗与金融先起飞?[J]. 中国经济周刊, 2016(35):31-33.
[15] 单明. 科技发展要在乎伦理[J]. 当代工人, 2015(9):19-19.
[16] 张蕊, 张佳帆, 江灏. 可穿戴式柔性外骨骼人机智能系统可靠性及应用伦理问题研究[J]. 机电产品开发与创新, 2008, 21(5):19-21.
[17] 佚名. 电脑会永远正确吗[J]. 中国技术监督, 1996(2):40-41.
[18] 胡增顺. 计算机的伦理学困境与出路[J]. 开封大学学报, 2013, 27(4):86-88.
[19] 王东浩. 人工智能体引发的道德冲突和困境初探[J]. 伦理学研究, 2014(2):68-73.
[20] 王绍源. 论瓦拉赫与艾伦的 AMAs 的伦理设计思想--兼评《机器伦理:教导机器人区分善恶》[J]. 洛阳师范学院学报, 2014(1):30-33.
[21] 王绍源. 应用伦理学的新兴领域:国外机器人伦理学研究述评[J]. 自然辩证法通讯, 2016, 38(4):147-151.
[22] 王飞跃. 从 AlphaGo 到平行智能:启示与展望[J]. 科技导报, 2016(7):72-74.
[23] G. A. Miller, 陆冰章, 陆丙甫. 神奇的数字 7±2:人类信息加工能力的某些局限[J]. 心理科学进展, 1983, 1(4):53-65.
[24] 景英娟, 董育宁. 生物特征识别技术综述[J]. 桂林电子科技大学学报, 2005, 25(2):27-32.
[25] 梁路宏, 艾海舟, 徐光档,等. 人脸检测研究综述[J]. 计算机学报, 2002, 25(5):449-458.
[26] 张翠平, 苏光大. 人脸识别技术综述[J]. 中国图象图形学报, 2000, 5(11):885-894.
[27] 李武军, 王崇骏, 张炜,等. 人脸识别研究综述[J]. 模式识别与人工智能, 2006, 19(1):58-66.
[28] 毛磊. 保护隐私权:个人尊严和价值的体现[J]. 紫光阁, 2003(5).
[29] 唐凯麟, 李诗悦. 大数据隐私伦理问题研究[J]. 伦理学研究, 2016(6):103.

四 外文文献

[1] Danielson P. Artificial Morality : Virtuous Robots for Virtual Games[J]. Routledge, 1992.

[2] 3Arkin R C. Governing Lethal Behavior in Autonomous Robots[J]. Crc Press, 2009, 37(2).
[3] Lisa Damm. Moral Machines: Teaching Robots Right from Wrong[M]. Oxford University Press, Inc. 2008.
[4] Moor J H. The Nature, Importance, and Difficulty of Machine Ethics[J]. Intelligent Systems IEEE, 2006, 21(4):18-21.
[5] 黄铂钧. AlphaGo 来了![J]. 科学世界, 2016(4):4-11.
[6] 童天湘. 从“人机大战”到人机共生[J]. 自然辩证法研究, 1997(9).
[7] Minsky M. The society of mind[C]// Simon & Schuster, Inc. 1986:371–396.
[8] Nick, Bostrom. Ethical Issues in Advanced Artificial Intelligence[J]. 2003, (12): 1-5.
[9] Brown W S. Ontologcal Security, Existential Anxiety and Workplace Privacy. Journal of Business Ethics, 2000 (23); 61 一 65
[10] R, V, Yampolskiy. Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach[J]. Roman V. Yampolskiy: 389-393

五 电子文献

[1] 大数据实验室. 俄罗斯最大银行推出律师机器人，3000 名专家将被炒鱿鱼[EB/OL]. http://mp.weixin.qq.com/s?__biz=MzA3MDI3ODQxOA==&mid=2651245363&idx=2&sn=fd97601e11d586dcd7c48129dfa77cb9&chksm=84cd274eb3baae581bc3b0ad388cbb99ee9f3ec20bc681434619557e4d62caec25350c4a547f&mpshare=1&scene=23&srcid=01305BwcO1l7wXh32ZE0O6g9#rd, 2017-01-30。
[2] AI 世代. 人工智能的 23 条“军规”，马斯克、霍金等联合背书[EB/OL]. http://tech.qq.com/a/20170207/031641.htm, 2017-02-07.
[3] 中国智能制造网. 人脸识别技术应用广泛 “刷脸”时代即将来临[EB/OL]. http://news.dichan.sina.com.cn/2016/10/12/1216729.html, 2017-10-12.
[4] 环球科技. 调查 ：伦理阻碍 AI 技术在澳大利亚发展 [EB/OL].http://www.techweb.com.cn/column/2017-01-18/2475399.shtml, 2017-01-18.
[5] 腾讯科技. 人工智能的 23 条“军规”，马斯克、霍金等联合背书[EB/OL]. http://www.hao123.com/mid?key=pZwYTjCEQLw-mv68TgD8mvqVQvDEnW0kP10znjTEnjnkPWfkQh9YUf&from=tuijian, 2017-02-07.

六 其他文献

[1] 刘虎功. 中山大学人机互联实验室主任翟振明谈围棋人机大战[N]. 长江日报, 3-15(16).
[2] 卫华. 杀人机器人的伦理问题及辩护[C]// 全国军事技术哲学学术研讨会. 2013.

攻读学位期间发表的学术论文和研究成果

一、杨帆. 人工智能发展简述[J]. 西江文艺, 2016, (24): 228-228.

二、杨帆.人工智能会超越人类智能吗？[J].西江文艺, 2017, (9).236-236.

.

致谢

研究生的学习生涯即将结束了，回首三年的学习经历还有为了毕业论文熬夜奋战的日子，感慨万千，读研读得相当辛苦，写的论文内容太前沿，可搜集的前人的研究成果太少，而且还面临各种压力，很辛苦，收获也很多，值此论文完稿之际，感激之情油然而生。

首先感谢我的导师杨胜荣老师，由于人工智能技术应用属于前沿科技，我的导师和我之前都没有接触过，但是我的导师杨胜荣老师同意了我的选题，顶住压力同意了我写人工智能。尊重了我自己内心的想法和兴趣。杨胜荣老师，在哲学领域学识渊博，他完全可以让我选择一个他自己熟悉的领域，不用冒险在未知的领域，但是杨胜荣老师并没有这样做。因为人工智能是前沿科技，是创新，创新就会面临着失败和挫折，在写的过程中遇到过是否能够成功的质疑声，但是我的导师杨胜荣老师顶住了压力。每一种质疑声不但不会减少创新精神的价值，而且会增加创新精神的价值。正是因为创新之难，才让创新精神才更加难能可贵，熠熠生辉。

感谢导师论文写作的过程中，从论文的选题，到写作，杨老师给了很多宝贵的指导给了我极大地帮助。他严谨的治学态度、独立的学术品格，开拓进取的精神给我留下了很深的印象，给我很多积极的影响。从论文写作时框架的思路不清，期间遇到的困难超乎想象。本人才疏学浅，自觉对问题的理论深度还可以进行深入的挖掘，这是小小的遗憾遗憾。研究过程中很多时候，当写到没有思路的时候，还是以导师的治学态度为学习的榜样，一步一步地完成论文。坚信人生路慢慢，没有什么翻不过去的山，只有轻易言败的心。写论文克服难关的经历将使我终身受益。

感谢姐姐，论文的写作过程是坚辛苦和富有挑战的，姐姐让我知道了论文写作的时间管理的重要性。

在查阅文献，收集资料过程中也遇到过许多人的帮助。

感谢斯坦福访问学者王博士，在我写论文过程中由于搜集文献有限，他分享了人工智能领域相关的一些国外文献，开阔了视野，发散了思维。并且提了写人工智能技术方面的专业建议。感谢中山大学的朋友，帮助我借了中山大学的专著，并且把专著帮我邮寄到学校，给我提供了帮助，在我论文迷茫的时候给我帮助和鼓励；感谢林肯大学留学的朋友，帮我提供了珍贵的文献资料，丰富文献来源。感谢计算机专业同学的同学们在人工智能技术方面给我很多帮助。

感谢张天祥老师，在写作过程中遇到困难的时候给了无条件的支持和鼓励。学生铭记在心。

感谢家人对我的支持和帮助，在我写论文的过程中给了我鼓励与支持，每

当遇到困难，父亲常常讲长征精神、延安精神我听。做过老师的父亲说：“你是学哲学的肯定懂，道路是曲折迂回的，方向是前进上升的嘛。”父亲用马克思主义辩证唯物主义理论告诉我，凡事有两面，在遇到困难的时候要看到希望。”爸爸会心一笑，让我倍受鼓舞。每个时代都有自己的长征路，我也要走好自己的“长征”路。家人是我强大的内心的支柱，是力量之源。

在借阅书籍，查找文献的时候得到过学妹、图书馆老师的帮助，在得知我图书卡借满了，又需要看书的时候，图书馆的老师毫不犹豫的用自己的图书卡帮我借阅书籍，缓解了寒假图书馆不开的带来的查阅书籍的困难；在食堂得到过打菜的云南姑娘帮助过；在存放书籍的时候；得到过我们的图书馆管理员的帮助，书籍很多每天搬来搬去不现实，图书馆的管理员帮我找到放书的书柜给我放书，每天看书没有后顾之忧。这些一点一点，都铭记在心中，正是因为这些热情善良的人们，让我对于师大是有着自己很深的热爱的。

感谢马克思主义学院的领导。感谢在研究生生涯过程中，前任学院党委书记，杨纪武老师，在不仅在学习知识上给予了很大的帮助，而且在生活工作方面，给予了我很多的支持和关怀。感谢刘华军老师、陈林老师的支持和鼓励，他们尽自己所能在生活方面给我很多建议和鼓励。感谢，当我要考试的时候，陈林老师给予鼓励。

感谢马克思主义学院的诸位老师们。感谢你们传道授业，尊尊教诲。从你们身上学习到很多。尤其感谢张天翔老师，从他身上学习到了做人做事应有的原则。也很感谢张天翔老师在我的研究生生涯中的支持与鼓励。

感谢校园里面的同学们，学习到很多。感谢那些青春岁月，感谢校园里那些温暖的人。

感谢母校，提供了良好的学习平台，母校学习氛围浓厚、科研精神求真务实、校园风景怡人，治学态度严谨，这些都深深的影响着我。让我成为更好的自己。

本文在撰写的过程中得到过很多的人的帮助，在此一并谢过。你们让我感到自己是一个幸运的人。

几经彷徨求索，论文终于完成。每个人生命中都有段时光、这段时光让你思考，让你奔跑。硕士三年是我人生宝贵的阶段，这三年在美丽的云南师范大学度过。校园时光匆匆，倍感珍惜、倍感感恩。一路走来，说不尽的感谢。而我将带着思考奔赴新的行程。