

江晓原 季卫东 刘宪权 **鼎力推荐**

[美] 瑞恩·卡洛 迈克尔·弗鲁姆金 编  
[加] 伊恩·克尔 陈吉栋 董惠敏 杭颖颖 译



彭诚信 主编

# 人工智能ROBOT 与法律AI LAW 的对话

人类的法律体系能跟上人工智能的脚步吗？  
性机器人合法吗？

上海人民出版社

# 人工智能 与法律的对话

〔美〕瑞恩·卡洛

迈克尔·弗鲁姆金

〔加〕伊恩·克尔

陈吉栋 董惠敏 杭颖颖

译

编

上海人民出版社

## 图书在版编目(CIP)数据

人工智能与法律的对话/彭诚信主编;(美)瑞恩·卡洛(Ryan Calo),(美)迈克尔·弗鲁姆金(A. Michael Froomkin),(加)伊恩·克尔(Ian Kerr)编;陈吉栋,董惠敏,杭颖颖译.—上海:上海人民出版社,2018  
书名原文:Robot Law  
ISBN 978-7-208-15357-8

I. ①人… II. ①彭… ②瑞… ③迈… ④伊… ⑤陈… ⑥董… ⑦杭… III. ①人工智能-法律-研究 IV. ①D912.174

中国版本图书馆 CIP 数据核字(2018)第 163048 号

策 划 曹培雷 苏贻鸣  
责任编辑 秦 堃 夏红梅  
封面设计 田 松

## 人工智能与法律的对话

彭诚信 主编

[美]瑞恩·卡洛,[美]迈克尔·弗鲁姆金,[加]伊恩·克尔 编  
陈吉栋 董惠敏 杭颖颖 译

出 版 上海人民出版社  
(200001 上海福建中路 193 号)  
发 行 上海人民出版社发行中心  
印 刷 常熟市新骅印刷有限公司  
开 本 635×965 1/16  
印 张 27.25  
插 页 4  
字 数 356,000  
版 次 2018 年 8 月第 1 版  
印 次 2018 年 8 月第 1 次印刷  
ISBN 978-7-208-15357-8/D·3262  
定 价 88.00 元

# 主编序

彭诚信

—

无论生物学意义上的自然人类(以下简称人类)是否做好准备,人工智能时代正逐步走来,而这恰恰是由人类自身所引起。

初级的人工智能或许能为人类带来便捷,在我国,或许还能带来规则意识,甚至法治理念的真正普及。这是因为,人工智能的本质就是算法,任何算法必然建立在对某项事物认识的共性与常识之上。也正是在此意义上,人工智能能为人类服务,能代替自然人为人类服务。初级的人工智能,如果还没有深度学习能力,或者深度学习能力尚不充分,它就难以进行诸如自然人价值判断与情感判断的活动,比如包含爱的交流与体验,难以对疑难案件作出理性裁判,对案件的漏洞填补与价值补充等。在此意义上,人工智能产品还主要表现为人工智能物,仅在有限的意义上具有自然人的属性。但即便是初级的人工智能,在我国也具有非常重要的意义,主要表现为规则意识与诚信观念的建立。人工智能最核心的“大脑”就是算法,算法本身便是规则。初级人工智能对人类的服务就是规则服务;而人类要接受人工智能的服务,就必须接受算法设定的各种规则。人工智能,尤其是结合网络运用的人工智能,会促使与提升自然人的规则意识,因为无论自然人在线下是否遵守规则,也无论规则在线下如何难以推行与实现,只要自然人接受线上服务,就必须遵守线上规则;无论自然人在线下如何不守信,他在线上也必须诚实,否则

他就进入不了虚拟世界，便也无从获得特定人工智能的服务。在初级的人工智能时代，人类仍是核心，是世界的主宰，毕竟自然人仍是规则的制定者，是人工智能的服务对象。

而到了高级人工智能时代，即，当人工智能能够进入深度学习与感情交流，可以进行团体合作与共同行动时，换句话说，当人工智能可以改变甚至完全脱离自然人为其设计好的初始算法而创制新的算法时，那时的人工智能物便实实在在地变成了人工智能人。人工智能人如何改变自然社会，甚至如何引导与影响整个自然社会走向，已非自然人所能完全掌控与想象，恐怕也为人工智能人本身所不知。尤其是，当人工智能人可以在虚拟世界制定规则(创制新的算法)，而这种规则又必然会影响到自然世界时，那时自然世界的主宰到底是人工智能人，还是自然人，或许现在的我们(人类)已经难以给出确定答案。那时的人类在自然世界或虚拟世界中处于何种主体地位，现在的我们也不得而知。当人工智能人有了情感交流能力并具有生物生成功能后，在自然人与自然人、人工智能人与人工智能人以及自然人与人工智能人之间的多元关系中，谁来制定规则，为谁制定规则，谁是自然世界或者虚拟世界的主宰或规则主体，以及各种形态主体之间具体的生活样态如何等问题，可能都远远超出了我们当下的想象，或许那时的社会状态本身就不可想象！

正是为了认真面对这些问题，警惕与体味这些问题，以便未来更好地深入研究或应对这些问题，上海人民出版社曹培雷副总编辑、法律与社会读物编辑中心苏贻鸣总监、秦堃编辑等及本人一起探讨决定编译人工智能丛书，帮助我国读者了解既有的人工智能研究，并以此为切入口对人工智能进行深度了解与学习。我们筛选并翻译了国外有关人工智能研究的较有影响力的三部经典著作，推荐给中国读者。这三部著作便是意大利学者乌戈·帕加罗所著的《谁为机器人的行为负责?》、美国律师约翰·弗兰克·韦弗所著的《机器人是人吗?》以及美国学者瑞恩·卡洛、迈克尔·弗鲁姆金和加拿大学者伊恩·克尔编辑的文集《人工智能与法律的对话》。

二

《谁为机器人的行为负责?》一书,由张卉林、王黎黎和笔者共同翻译。该书通篇都在试图回答一个问题:“谁来承担责任(Who Pays)”。作者建构了一种分析法律责任模型。他在刑法、合同法和侵权法的框架下讨论了27种假设情况,例如刑法中的机器人士兵、合同法中的外科手术机器人以及侵权法中的人工智能雇员等,目的是分析在不同的情况下设计者、生产者、使用者和机器人之间应当如何分配责任。作者还讨论了机器人对现代法学体系中的若干重要内容带来的挑战,比如刑法中的正义战争理论、合同法中的代理资格以及侵权法中的责任承担。上述问题的讨论建立在作者对法律责任和义务的概念的分析基础上,讨论法律基础是否会受到机器人技术的影响。最后,作者讨论了“作为元技术的法律”,即如何通过法律实现对技术发展的控制。

《机器人是人吗?》一书由刘海安、徐铁英和向秦翻译。该书认为,人工智能可以达到如同与真人一样进行语音交流的程度,并自主学习知识和判断问题。作者讨论了人工智能的知识产权享有和责任承担问题。作者认为,面对人工智能承担法律责任,可以通过人工智能保险或储备基金支付赔偿费用。如何规范人工智能?作者以美国各州对自动驾驶汽车的相关立法为例,对未来人工智能统一立法作出合理预测:(1)当产品制造商、开发商和人类都没有过错时,不同体系的机构将会为涉及人工智能的事故受损者建立赔偿或补偿基金;(2)至少在初期,很多形式的人工智能产品的使用将被要求获得执照许可背书;(3)在初期,往往需要对人工智能进行人为监督,但是最终,只有那些主要用于改善人类表现的人工智能才需要人为监督;(4)尽管最初的立法将会经常把人类作为操作者(行为人),即使这种标签不适用于人工智能的类型,但最终立法会在以确定操作者责任为目的时变得更加细分;(5)立法将始终区分用于测

试目的的人工智能和向消费者提供的人工智能；(6)立法将始终要求这样一个机制，允许人类脱离人工智能但很容易重新控制人工智能；(7)立法将始终要求在自动化技术失败时，人工智能产品能向周围的人发出警告；(9)对采集个人信息的担忧将会迫使法律要求披露人工智能运作时所收集的信息。

《人工智能与法律的对话》由陈吉栋、董惠敏和杭颖颖翻译。本书共分讨论起点、责任、社会和道德意义、执法和机器人战争 5 个部分，共 14 篇论文。其中，大部分是首次在“*We Robot*”这一跨学科会议上发布的最新论文。这些论文探讨了机器人的日益复杂化以及它们在各个领域的广泛部署，重新思考了它所带来的各种哲学和公共政策问题、与现有法律制度不兼容之处，以及因此可能引发的政策和法律上的变化。整本书为我们生动地展现了一场内容广泛、启发深远的对话，如本书第二部分有关机器人行为责任的讲述：F.帕特里克·哈伯德教授《精密机器人所致人身损害的风险分配》一文对普通法应对技术变革的能力提供了一种乐观的评估：“普通法系统包含了内部机制，能够为应对机器人化的世界作出必要的相对较小的变化”；而柯蒂斯·E.A.卡诺法官在《运用传统侵权法理论“迎接”机器人智能》一文中则提出了截然相反的观点：传统的过失和严格责任理论不足以应对真正自主性机器人的挑战。

需要说明的一点是，我们从 2017 年 9 月确定翻译书目，10 月组建翻译团队，到 12 月后陆续落实版权并着手翻译，翻译时间可谓十分紧张。丛书译者多为高校或者研究机构的青年科研教学人员，需要克服繁重的教学和科研压力；加之，所译著作内容涉及法律、计算机和伦理等多元且交叉的学科知识，远远超出了多数译者所在的法学学科领域，翻译不当甚至错误恐在所难免，在此我们衷心恳请并接受各位读者、专家批评指正。

### 三

2017年7月中华人民共和国国务院发布《新一代人工智能发展规划》，强调建立保障人工智能健康发展的法律法规，妥善应对人工智能可能带来的挑战，形成适应人工智能发展的制度安排。《规划》为此要求“开展与人工智能应用相关的民事与刑事责任确认、隐私和产权保护、信息安全利用等法律问题研究，建立追溯和问责制度，明确人工智能法律主体以及相关权利、义务和责任等”。但正如弗鲁姆金(Froomkin)指出的，也可能是本译丛三本书的作者们皆认可的：“(1)对于机器人和监管问题，现在还为时尚早；(2)技术问题远比律师想象的复杂，法律、伦理和哲学问题比工程师想象的更有争议(有时也更复杂)；(3)我们要彻底解决这些问题的唯一办法就是扩大和深化我们跨学科的努力。让世界为机器人做好准备的同时，使机器人也为世界做好准备，这必须是一个团队项目——否则它可能会变得很糟糕。”由此揭示出，对于人工智能的探讨与研究，即便是对于人工智能的规范性研究，并非法学一个学科所能胜任。人工智能本身就是一个具有综合性、复杂性、前沿性的知识、智识与科学，它需要几乎所有的理工与人文社会科学学科进行交叉性研究，也需要研究者、实体技术者与产业者等各个领域的人配合与对话。法律人在人工智能的研究、开发、规则制定等各个环节中是不可缺少的一环，但也仅仅是一个环节，他只有加入人工智能的整体研究与发展中去，才会发挥更大的价值。我们期待这套译丛的出版有助于人工智能在法学及其他领域展开深入讨论，为跨学科的对话甚至团队合作提供一定程度的助益。

无论未来人工智能时代的社会生活样态如何，无论人工智能时代的社会主体如何多元，多元的主体依然会形成他们自己的存在哲学，也许依然需要他们自己的情感系统。无论未来的人工智能时代多么不可预



测，问题的关键还是在于人类的自我与社会认知。就像苹果公司首席执行官蒂姆·库克(Tim Cook)在麻省理工学院(MIT)2017 届毕业典礼演讲中指出的，“我并不担心人工智能能够像人一样思考，我更关心的是人们像计算机一样思考，没有价值观，没有同情心，没有对结果的敬畏之心。这就是为什么我们需要你们这样的毕业生，来帮助我们控制技术”。是的，我们或许不知未来的人工智能是否会产生包含同情与敬畏的情感，但我们能够确信的是，即便在人工智能时代，我们最需要的依然是人类饱含同情与敬畏的“爱”！未来的人工智能时代无论是初级样态还是高级学习样态，能够让多元的主体存在并和谐相处的，能够让多元主体维系在一起的或许唯有“爱”。这个从古至今在自然世界难以找到确定含义的概念，在虚拟与现实共处的世界中更是难以获得其固定内涵，但我们唯一知道并可以确信的是，如果没有“爱”，那么未来的人工智能时代就真的进入了一个混沌而混乱的世界！

上海交通大学凯原法学院

2018年7月10日

## 前 言

A.迈克尔·弗鲁姆金\*

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。接下来几个章节将探讨机器人的日益复杂化以及它们在各个领域的广泛部署,如从家庭到医院、公共场所以及战场,同时需要重新思考它们所带来的各种哲学和公共政策问题、与现有法律制度不兼容之处,以及因此可能引发的政策和法律上的变化。

这些讨论是必要且及时的。当今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。想象一下,如果在互联网开始发展的1980年(第一个互联网标准文件RFC 1出现的十多年后),或者在第一个网站上线的1992年,一些跨学科的学者试图研究互联网有关的政策和法律问题。毫无疑问,研究者无法预见今天我们享受到的互联网的所有产物,无论其是好是坏。然而,他们很可能会发现与域名、商标、信息安全、残疾人使用以及隐私等相关的重大困难。事实上,当意识到这些问题很重要的时候,机器人的应用基础已经强大到足以改变相关协议和应用中存在的争议(例如IPSEC),包括在某些情况下(例如域名)是不切实际的争议。

正如后面的章节所描述的那样,尽管在考虑机器人的社会和法律效应方面我们仍然处于早期阶段,但我们已经可以确定机器人在社会中广泛应用所带来的一系列问题。机器人引发的问题涉及广泛的学科领域,可

---

\* A.迈克尔·弗鲁姆金,美国迈阿密大学法学院法学杰出教授。

能不仅影响日常生活和商业生活的各个方面,而且还会影响战争。除此之外,正是问题识别的这一过程揭示了还有多少定义和应用工作仍有待完成。以一种有序的方式处理这些问题是一项重大的任务,它需要一种延伸的对话,必须跨越学科,才能使其真实有效。

xi

关于什么应该算作“机器人”,目前仍然没有定论,这说明我们处于一个十分早期的阶段,大多数人包括本书的所有参与者都会同意,能够对外部刺激作出反应并且不需要直接的或持续的人类控制就可以采取行动的人造物体就是一个机器人,尽管有些人可能会争论并要求一个更广泛的定义。这种相对狭窄的、可能涵盖不足的工作定义的三个关键要素是:(1)某种传感器或输入机制,没有这种机制就不会有刺激反应;(2)一些控制算法或其他系统将管理对感测数据的响应;(3)对外部世界产生显著影响的能力(我最喜欢的定义是一个范围更窄的版本,因其尖锐性而非准确性受到重视,它说一个机器人是“带着电锯的iPhone”)。

这个工作定义允许很大的兼容性。机器人的尺寸可以是任意大小,从纳米机器人到巨人机器人,但除了纳米机械可能引发的特殊问题之外,并不是尺寸让机器人变得有趣和存在潜在问题。

机器人实例化及其独立性有更大的变量,也更为重要。机器人可以是有形的,也可以是纯软件产品,只要它们从外部世界接收信息输入,能够以对主机程序之外的世界进行响应;软件当然是由算法控制的。机器人的功能范围比较灵活;它们的行动和决策自由范围可能非常狭窄,也可能非常宽泛,甚至可能是无限的。在极端的情况下,人们可以在“机器人装配线”上找到自动机器。在某些情况下,这些机器重复着同样的事情,并不算作机器人。但在其他情况下,即使是只能小范围运动或识别范围较窄的机器,在技术上也可能符合上述工作定义:例如,机器人手臂可能设计为检测某一部件流下传送带的不同位置,并且可以抓住并固定它。根据我们的说法,这可能足以将其定性为机器人,尽管这不是一件很有趣的事情。

在物理性和不可测性方面的另一个极端是米勒(Millar)和克尔(Kerr)所著章节中所描述的一个运行在超级计算机上的医疗诊断机器人程序。它可能收集患者身上的传感器信息,或者它可能根本没有传感器,只是对诊断人员输入的数据作出响应。在任何一种情况下,其反应都只是文字和数字,除了建议之外,它没做任何其他的事情,所以它对世界的直接影响尽管广泛,但是有限的。但它非常独立,并且它在巨大数据集中发现相关性的方法非常难懂,我们可能永远无法理解为什么它会作出某种选择。相反,我们只能衡量它的成功率,将它与人类医生进行比较,并开始认真思考如果——或者更确切地说当——机器人医生开始表现出更高成功率时意味着什么。

xii

尽管如此,人们通常所称的“机器人”超出了上面的工作定义,也超出了这个项目的范围。被排除在这个工作定义之外的是许多在大众意识中大量出现的自动机器。机器完全由人类远程监控操作——如沃尔多斯(Waldos)——包括我们所知的目前部署在战场的所有军用无人机(与绘图板上的以及可能正在测试的那些不同),它们在某种意义上不是机器人,因为它们不会对刺激作出反应。它们无非是没有机器人驾驶的传统机动车。它们也不是那种能产生我们关心的有趣问题的机器人。相反,所谓的“半机器人”,即那些通过使用或植入某些技术而增强身体的人,并不是“机器人”,因为他们仍然是人,因为同样的原因,驾驶传统的汽车不能使驾驶员成为机器人。

尽管如此,即使在最狭隘的意义上,这些设备可能不被称为机器人,但其中一些部分的增强(如机械义肢),也会引发与机器人类似的问题。举个例子,考虑一下,一条机械腿被设计来响应甚至是预见与机械相连接的人类神经的命令。在理想情况下,从用户的角度来看,行走所需要的有意识的意向对于机械肢体和正常的肢体来说是没有区别的。然而,人们只需要想象一下,如果假体没有完全调试好,会发生什么:肢体的主人被困在一个鸡尾酒会上,听着一次特别无聊和不愉快谈话,想象着有人踢他一脚。无意识的冲动沿着神经传递,遇到机器人控制器,突然,聚会变得更有意思了。

玩笑归玩笑,这种半机器人意外的结局突出了由具有任何自主度的(至少是非软件的)机器人引发的法律和社会问题与互联网所带来的不同:因为它们存在于现实世界中,机器人可以对世界造成直接的影响。通过设计,无论是意外还是军事机器人,都可能造成实质性损害,伤害甚至杀害人类。机器人有很大的潜力来承担不愉快、困难或无聊的任务,但它们也存在真正的风险,需要设计者和决策者仔细规划。

关于新技术在社会、政治、道德和法律方面作用的研究,往往会在该技术尚处于萌芽阶段时遭遇不同寻常的障碍。部署尖端技术的工业科学家倾向于关注使机器人工作、将它们推向市场并获利。他们认为,社会政策及法律问题可以忽略或克服。学术科学家可能考虑更广泛的问题,但他们并不总是接受有关学科的训练,无论是社会科学、哲学或法律。相反,在这些领域接受过训练的学者往往缺乏相关技术的必要背景,并且往往对相关文献不熟悉;可能会发现自己无法理解。随着技术的推广和普及,这些问题往往会得到解决;当该技术成为主流,会吸引更多人。更重要的是,随着技术的发展,新技术带来的新问题变得明显——但到那时围绕问题改变技术设计已经太迟了。同样,潜在的赢家和输家可能会使政策或技术的调整实施变得比以前更困难。

本书汇集了不同学者的努力,他们每个人都努力克服这些早期障碍,进入并试图解决机器人普及所带来的后果。本书中的大部分贡献是首次在“我们机器人(We Robot)”上发布的最新论文,“我们机器人”始于2012年,是一年一度的跨学科会议,曾在迈阿密大学(两次)、斯坦福大学和华盛顿大学举办。就像下面章节的作者一样,“We Robot”的参与者们相信,如果要制造出能够适应一切环境的机器人,我们还有很多工作要做。如果我们要建造这样一种法律环境:一方面,正确平衡需要鼓励实验技术和投资的目的,保护公共安全,另一方面,平衡机器人行业的竞争经济参与者的需求——更不用说考虑工业机器人承诺的利益增强或取代,我们面临着更加艰巨的任务。

## 第一部分：起点

本书第一部分是律师尼尔·M.理查兹和机器人专家威廉·D.斯马特合作的《法律如何看待机器人?》，面对越来越多的机器人在公共空间、家庭、办公室和医院的使用以及面临的挑战，法律体系将需要决定如何处理相关问题。目前，我们的法律倾向于认为机器人的行为是由设计决定的，并把机器视为使用者的合法延伸。当机器发生故障时，我们尝试将责任分配给适当的制造商、设计者、维护者或修改者。这些方法都不容易转移到复杂的机器人世界，特别是具有任何自动性的机器人。当遇到能够采取突现行为的机器人时，对错误和产品责任界定的传统法律方法也会摇摆不定。制造商、程序员、所有者、用户和其他可能有意或以其他方式训练机器人的人之间复杂的相互关系，带来了错综复杂的证据和理论问题。理查兹教授和斯马特教授认为，我们如何将机器人概念化，即我们用来理解并为之关联的隐喻将会塑造出工程师设计机器人的方式，以及社会（尤其是律师）将如何管理机器人。作者认为，问题在于想象力特别是设计师和监管者的想象力，以及如何更好地将这两种截然不同的思维方式融合在一起。无论把机器人看作工具、宠物、仆人还是家庭成员，都会影响对它们的开发方式以及法律如何认识它们。

## 第二部分：责任

第二部分讲述的主题是机器人行为的责任问题。从两位律师开始，关于侵权体系能够有效应对由机器人引起的法律和社会干扰问题，他们观点截然相反。一方面，F.帕特里克·哈伯德的《精密机器人所致人身损害的风险分配：效率、公平与创新》对普通法应对技术变革的能力提供了一

种乐观的评估。他首先展示了当前的学说如何应用于机器人汽车等各种机器人的实用情形,并且表明根据当前学说,结果不会偏离我们想要达到的目标。此外,哈伯德教授提醒我们,今天的侵权法本身就是不断演变的产物,他认为,普通法系统包含了内部机制,能够为应对机器人化的世界作出必要的相对较小的变化。因此,他对大多数旨在促进复杂机器人创新的当代法律改革建议持怀疑态度。他总结说,要对一个在企业家和新设备支持者以及有权从伤害中获得赔偿的普通民众之间的冲突利益中已经取得平衡的系统做出重大改变,我们必须非常谨慎。

柯蒂斯·E.A.卡诺法官在《运用传统侵权法理论“迎接”机器人智能》中提出了截然相反的观点,部分原因是他专注于最难的案例——真正的自动机器人。在卡诺法官看来,传统的过失和严格责任理论和严格赔偿责任不足以应对它们的挑战。卡诺法官首先提出了一个关于机器人“自动”意味着什么的工作定义——这正是本章最重要的特征。根据这一定义以及美国侵权法对“可预见性”的定义,卡诺法官认定法律制度在假定线性因果关系方面存在不足之处,可能会导致线性因果关系,但会遇到更复杂的与更大环境相互作用的机器。他认为,机器人的行为将是非线性的,实际上是不可预测的,侵权制度被用来将大部分责任分配给那些被认为应当承担责任的人,这是一个旨在将大部分责任推给那些由于存在合理机会来控制或预见风险而被认为应该承担责任的人的侵权制度,但却无法胜任这项任务。侵权制度需要改变,我们的期望需要改变,并且根据该章在第一部分引入的一个主题的结论,机器人设计将需要找到一种方法来构建具有“常识”的机器人,以使他们的行为不会偏离我们的期望。

布莱恩特·沃克史密斯的《律师与工程师应当使用相同的机器人语言》也解决了有关法律与机器人相遇的概念性主题,这次伴随着从为无人驾驶汽车制定规则和标准的新尝试中得到的经验。概念部分认为,对于工程师或律师而言,将机器人与其所属的较大系统隔离开来是错误的。许多机器人正在或将要被设计为与其他机器、云以及最主要的人类进行交互。设计和规则必须考虑到这种嵌入性。然而,认识到机器人是系统的一部

分,反过来又产生了一个新问题,即定义相关系统边界。使问题适当复杂化后,史密斯教授向我们展示了实际行业标准以及与自主机动车有关的拟议或实际规范性文件,特别是从系统分析的角度来看,发现了严重的概念和术语不一致性。他总结说,避免这种混淆及对驾驶者和第三方造成的实际风险的最好方法是工程师和律师更多地交流,以便他们能够建立关于机器人的共同语言。

xvi

我们对责任问题的讨论以杰森·米勒和伊恩·克尔的《委托、让渡与法律责任:专家机器人的前景》作为结束。这里的问题是什么时候依赖机器人专家才是恰当合理的。作者首先关注 IBM 的沃森[这是“深度问题解答”机器人,击败了两位历史上的《危机边缘》(Jeopardy)冠军],这个系统现在正在作为机器人诊断专家接受测试。如果研究表明这些机器人基于大数据相关性诊断的方法比面对同一患者基于逻辑的人类诊断表现更好,那么会发生什么?机器人可能会成为首选专家,这迫使我们质疑人类是否有责任将诸如医疗诊断和驾驶等任务授权给专家机器人。通过探索专业知识和基于证据的决策的观点,即认为产生最有利结果的行动是最合理的观点,作者提出了各种法律和道德方法来确定这些先进系统出现错误时,以及当人们不能像他们想象的那样依赖机器人时的责任。

### 第三部分：社会伦理意义

当机器人比人类更安全时,关于依赖机器人在道德上是否合理甚至必需的讨论将我们转向涉及机器人的更广泛的道德问题。阿荣·月等人的《开放式机器人伦理倡议与乘坐电梯的机器人》描述了一种开放源代码倡议,旨在加速机器人伦理学原则实施的自下而上设计。该倡议既反映了当前机器人伦理学举措的分散性,也反映了现有方法由专家开展的趋势;也就是说,他们的方向是自上而下的。自下而上的方法允许更多的公众参与,作者认为这对于基于伦理的规则的发展尤其适用,而这些规则本质上



必须与情境相关。为了说明这些规则如何依赖于具体情境，作者使用“乘坐电梯的机器人”为例提供了一个概念验证，出于安全和其他原因，它必须是建筑物内电梯上的唯一乘客。它应该如何应对已有乘坐人员的电梯？它应该要求他们走出去吗？当机器人在等电梯的时候，如果有人也想要使用电梯，那该怎么办呢？无论答案如何，作者都认为他们的开放式机器人伦理倡议(ORi)会产生一个结果，不一定是哲学家所说的最道德的，也绝不保证适用于不同的文化，却可能会反映出给定文化中的大众情绪和期望的行为。这种共识的精华是设计师在短期内需要的东西，从长远来看，它可以作为讨论道德上正确行为的起点。事实上，就“乘坐电梯的机器人”而言，作者设计了一项调查，该调查根据给定的变量例如机器人任务的紧急程度，人员的位置（电梯内或等待时间）以及人的特征（例如坐轮椅，携带重物）等来确定期待值。非常小的概念验证调查( $n=8$ )允许构建决策矩阵，机器人可以从中选择让步、坚持或与其他乘客对话，作者进一步讨论如何在机器人中实现这样的矩阵——以及如何设计它，以便根据需要进行简单的修改，作者建议的结果表明其开放式机器人伦理模型的实用性。

另一种开放性是戴安娜·玛丽娜·库珀所撰写的《以“充分的和具有选择性的开放式许可”限制与开放式机器人相关的责任和道德问题的应用》中的主题。开放式许可促进了机器人行业为实现“每个家庭中都有机器人”的目标所需的快速创新，但同样的开放性带来了各种各样的法律责任风险。开放式许可也意味着一个人的发明可能会被认为是道德上可疑的下游使用。因此，我们面临的挑战是找到一个既不会太开放而致风险，也不会太封闭而阻碍创新的许可证，一个“充分开放的许可”。本章探讨采用“充分开放”模式的障碍，然后讨论克服这些障碍的手段。最后提出了一个具体建议：道德机器人许可(ERL)的初稿。

辛赞娜·古图的《同意的机器人化》论述了机器人在道德上可质疑的使用，考虑了现实中性机器人与人类相互作用的可能后果。古图首先指出，由市场驱动的设计选择会带来女性机器人，即使它们不是性机器人，