

# 人工智能法律人格赋予之必要性辨析

郭明龙\* 王菁\*\*

## 目次

一、前提：人工智能“自主性”与法律人格之通约性	(一) 肯定说的论据并不成熟
(一) 人工智能之“自主性”	(二) 肯定说的制度价值仅在责任承担
(二) 人工智能“自主性”与法律人格不可通约	三、对否定说之补充：制度回应“近人性”
二、对肯定说之质疑：论据与制度价值	(一) 人工智能“近人性”再识
	(二) 人类对待人工智能时的行为规制
	结论

**摘要** “自主性”并非决定法律人格赋予的本质属性，“自主性”的人工智能与人类也不具平等性，人工智能“自主性”与法律人格不可通约，这是讨论的前提。肯定说所引证的论据未臻成熟，法律人格的制度价值仅在责任承担，可以按照“尚能”与“不再”的关系完善责任制度，而非赋予法律人格。否定说虽主张无须赋予法律人格，但立法应对人工智能“近人性”做出制度回应，一方面“近人性”是人类情感的投射，立法应摒弃过度“近人化”思维，另一方面当突破公序良俗对待人工智能时，人类行为应予规制，以守护人类共同价值。

**关键词** 人工智能 法律人格 “自主性” “近人性”

人工智能，简称 AI(Artificial Intelligence)，专指“用人工的方法在机器(计算机)上实现的智能”，其中智能“是一种认识客观事物和运用知识解决问题的综合能力”。〔1〕人工智能学科创建于 1956 年，距今已有六十多年的发展历程，其技术研发一直在进行中。〔2〕在我国，大众对于人工智能的讨论热潮是随着 2016 年阿尔法狗(AlphaGo)相继战胜人类职业围棋选手、围棋世界冠军的新闻事件而产生的。与电话或者火车之类的产品发明不同，人们对待人工智能的态度显然更为焦虑

\* 天津师范大学法学院副教授、法学博士。本文系 2019 年度教育部人文社会科学研究规划基金项目“数据权之确权与交易利用规则研究”(项目编号：19YJA820013)的阶段性成果。匿名审稿人所提意见对于论文完善提高助益良多，谨致谢意。

\*\* 天津师范大学法学院硕士研究生。

〔1〕 王万森：《人工智能原理及其应用》(第 3 版)，电子工业出版社 2012 年版，第 1、3 页。

〔2〕 参见尼克：《人工智能简史》，人民邮电出版社 2017 年版，第 1 页。

甚至恐惧。尽管1984年自动驾驶汽车的研究已经启动,1997年深蓝计算机已打败了国际象棋冠军,但直到阿尔法狗的出现才让人们有了一种即将被人工智能统治的错觉,“人工智能威胁论”一时间占据各大媒体头条,成为舆论的热点。而在我国法学界,围绕应否赋予人工智能机器人以法律人格,形成“肯定说”和“否定说”两派观点:前者认为应以开放态度迎接科技的发展,赋予人工智能以法律人格;<sup>[3]</sup>后者则认为不宜打破传统民事主体分类,目前人工智能与自然人法律人格仍存在不可逾越的鸿沟。<sup>[4]</sup>学者们的论证路径可大致分为两类:一类是对法律人格进行历史和现实的梳理,在历史性因素中会将人工智能与古代奴隶制、法人权利能力的拟制、动物权利的保护<sup>[5]</sup>及近代女性权利的提升<sup>[6]</sup>相对比,在现实性因素中则分析人工智能发展的技术阶段性以及目前比较法上的一些实践经验;<sup>[7]</sup>另一类是理论分析,通过对法律人格不同学说进行解读并将人工智能与自然人、法人等进行对比而证成结论。<sup>[8]</sup>论证路径虽有差异,但笔者认为争议涉及三个问题:其一,人工智能“自主性”与法律人格问题是否具有通约性,这是立论的前提问题;<sup>[9]</sup>其二,针对肯定说,人工智能法律人格赋予是否已有成熟的比较法经验,法律人格制度价值何在;其三,针对否定说,如果选择不赋予法律人格,如何在制度上回应人工智能的“近人性”。笔者愿略陈管见,求教大方。

## 一、前提:人工智能“自主性”与法律人格之通约性

有学者从“自主性”出发论证人工智能法律人格赋予之必要性,认为“自主性是‘电子人’主体的本质属性”,“人工智能具有自主性、主动性,已非纯受支配之客体,在法律上应设定为‘电子人’”。<sup>[10]</sup>笔者认为,人工智能“自主性”与法律人格不可通约,以上法律人格证成路径实际受到“技术决定论”的误导。

### (一) 人工智能之“自主性”

自主性即不受外部控制、能够自我决定并付诸行动。人工智能身上有很多种标签,但是自主性却是其最大的特质,或者说是法律领域产生最大冲击的特质。以此为标准,人工智能的技术可以划分为三个阶段:第一阶段是狭义人工智能(AI),这一阶段的人工智能自主性是脆弱的、明显

[3] 例如,彭诚信、陈吉栋:《论人工智能体法律人格的考量要素》,载《当代法学》2019年第2期,第52~62页;李俊丰、姚志伟:《论人工智能的法律人格:一种法哲学思考》,载《华侨大学学报(哲学社会科学版)》2018年第6期,第80~86页。

[4] 例如,王利明:《人工智能时代对民法学的新挑战》,载《东方法学》2018年第3期,第4~9页;赵万一:《机器人的法律主体地位辨析——兼谈对机器人进行法律规制的基本要求》,载《贵州民族大学学报》2018年第3期,第147~167页。

[5] 参见陈吉栋:《论机器人的法律人格——基于法释义学的讨论》,载《上海大学学报》2018年第3期,第80~81页。

[6] 参见前注[3],李俊丰、姚志伟文,第83页。

[7] 参见郭少飞:《“电子人”法律主体论》,载《东方法学》2018年第3期,第38~41页。

[8] 参见杨立新:《人工类人格:智能机器人的民法地位》,载《求是学刊》2018年第4期,第85~92页;徐文:《反思与优化:人工智能时代法律人格赋予标准论》,载《西南民族大学学报》2018年第7期,第100~110页。

[9] “不可通约性”是库恩在1962年《科学革命的结构》著作里所借用的一个数学概念,描述科学革命中新旧范式之间无法“等价”转换的问题,“通约”是指描述事物、文化之间的互通性和共同之处。参见周超、朱志方:《不可通约性与科学合理性——库恩科学合理性理论研究》,载《武汉大学学报(哲学社会科学版)》2004年第4期,第489页。

[10] 参见郭少飞:《“电子人”法律主体论》,载《东方法学》2018年第3期,第38页。

低于人类水平,只能被应用于特定领域;第二阶段是通用人工智能(AGI),这一阶段的人工智能自主性是强壮的、拥有与人类同等的水平,可以应用于多种领域;第三阶段是超级人工智能(ASI),拥有自我意识、超过人类水平的自主性,可以应用于所有领域。<sup>〔11〕</sup>对于第二、三阶段,相当于或超出人类智力水平的人工智能被统称为强人工智能,与之对应的是低于人类水平的弱人工智能。人们对于人工智能的消极态度主要针对强人工智能,但是“很不幸”却又“很幸运”地是目前的人工智能还只发展到弱人工智能阶段。“具有自我意识,能够自我改进的系统,几乎不可避免地会走向不可知”,<sup>〔12〕</sup>人们恐惧的是人工智能所带来的不可控风险,在2017年霍金就曾提出,如果我们无法找到有效途径去控制人工智能所带来的风险,那么人工智能的出现就会变成人类文明史上最糟糕的事。<sup>〔13〕</sup>人工智能的自主性确实存在,并且这一特质随着深度学习技术的发展而加剧。人工智能通过输入外部信息,依靠深度学习规则,识别或感知对象,这种发展路径一方面会造成输入和输出之间的断层,可被解释的因果关系被打断;另一方面,一旦突破就会成为科学史上的里程碑,解救人工智能走出技术困境,换言之,自主性正是人工智能的设计者们所翘首以待的。

随着系统自主化的增强,人们会越来越难以解释输入(命令)和输出(人工智能“行为”)之间的联系,线性模式下的“A引起B,B引起C”的因果关系逐渐向复杂的非线性模式转变。从数据输入的方式划分,人工智能分为符号型人工智能和数据驱动型人工智能,后者的核心技术就是大家所熟知的深度学习。符号型人工智能由程序员输入指令和规则,其“行为”结果可被预期,“行为”程序有既定的规律可循。数据型人工智能是应用各种算法,对大数据进行处理,再利用统计学解析数据,计算的程序非常复杂,无法准确推断人工智能的“行为”与既定程序之间的因果关系。<sup>〔14〕</sup>例如,自动驾驶汽车的目标是将人类从驾驶任务中解救出来,要求汽车能够在各种不可预知的环境中做出实时决策,所以其遵循的是数据驱动型人工智能的发展路线。这就意味着,系统设计者所期待的正是“行为”与既定程序之间的不可预见性。

已经有证据证明人工智能可以打破既定规则,以人们意想不到的方式行动。康奈尔大学的杰森(Jason)教授发现,在进行神经网络物体识别的实验时,神经网络将实验者的脸也作为带有有用信息的字节进行了识别,因为它(可能)认为这样能更好地识别训练时展示的物品,而其之前并未进行过特别的人脸识别训练。<sup>〔15〕</sup>随着人工智能技术的发展,这种“行为”和决策的不可预见性会更为频繁地发生。仍以自动驾驶为例,虽然模拟测试在自动驾驶汽车领域具有重要意义,但却不能穷尽所有情况,当异常情况出现时,自动驾驶汽车必须自己创建规则,做出独立的决策。在产品流通后,自动驾驶汽车脱离了生产者的控制,其自主的学习和经历会产生新的数据输入,也会影响预先设置的规则。虽然模型是在公开和透明的算法下形成的,但其内部决策的逻辑并不都是可以解释的,即使对于程序的设计者来说也是如此。<sup>〔16〕</sup>这就造成自动驾驶汽车发生交通事故造成

〔11〕 See Andreas Kaplan, Michael Haenlein, “Siri, Siri in My Hand, Who’s the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence”, 62(1) Business Horizons 15-25 (2019).

〔12〕 [美]詹姆斯·巴拉特:《我们最后的发明——人工智能与人类时代的终结》,电子工业出版社2016年版,第77页。

〔13〕 See Stephen Hawking warns about A. I. development (<https://www.cnbc.com/video/2017/11/06/stephen-hawking-warns-about-a-i-development.html?&qsearchterm=Hawking>, last visited on 2019-04-10).

〔14〕 参见[美]胡迪·利普森、梅尔芭·库曼:《无人驾驶》,林露茵、金阳译,文汇出版社2017年版,第95页。

〔15〕 参见前注〔14〕,胡迪·利普森、梅尔芭·库曼书,第273~274页。

〔16〕 参见司晓、曹建峰:《论人工智能的民事责任:以自动驾驶汽车和智能机器人为切入点》,载《法律科学(西北政法大学学报)》2017年第5期,第169~170页。

损害时,事故本身难以解释及不能合理追溯到侵权责任主体的困境。

## (二) 人工智能“自主性”与法律人格不可通约

尽管人工智能“自主性”不断增强,但不能由此推论“人工智能已经具有与人类对等的人格结构”。<sup>[17]</sup> 关于人工智能法律人格赋予,需要解决两个问题:其一,“自主性”是否法律人格赋予的本质属性;其二,人工智能的“自主性”能否导致与人类平等性。

首先,“自主性”并非决定法律人格赋予的本质属性。历史上法律主体的演进经历了从“人可非人”到“非人可人”的发展,<sup>[18]</sup>先是在前近代社会中某些自然人被剥夺主体资格,后是进入近代社会不仅自然人获得平等人格,而且团体获得了主体资格。<sup>[19]</sup> 尽管“非人可人”的趋势下团体被赋予法律人格,但团体的主体化(“法”人)归根结底还是为自然人而存在的,传统的“民法是人法”的“人类中心主义”理念并未过时。从“人可非人”到“非人可人”的发展历程中,奴隶曾具有自主意识却未被赋予(完整的)法律人格,团体缺乏生物学基础却被赋予法律人格,所以,自主性并非决定法律人格赋予的本质属性,人类或占统治地位的人群的需要才是,“完全用人类的自主意识作为判断标准的科学性和合理性是值得商榷的”。<sup>[20]</sup> 无论是法律人格赋予的“理性学派”还是“人文主义学派”,无论是“理性特质”所包括的认知能力、道德能力与意思能力,还是“人性的具体特质”所包括的生命神圣、生命脆弱、感受痛苦能力,<sup>[21]</sup>其实都是表面现象或曰“障眼法”。法律人格无非来源于立法者的抽象和虚拟,而这种抽象和虚拟是基于现实中的人类需要。作为最为重要的团体人格,关于法人本质的代表性学说有“拟制说”和“实在说”,<sup>[22]</sup>无论坚持哪种学说,法人仅为团体在私法上的构造,这种人格没有伦理性,是为了实现自然人的交易自由与交易成本节约所进行的法律安排,是为了自然人而存在的。

其次,“自主性”的人工智能与人类不具有平等关系,而平等性却是现代主体制度所应具备。人工智能是人类创造的客体,我们发明人工智能的目的绝对不是为了让人工智能基于自主性统治人类,或者与人类平起平坐,更不会是与人类讲平等。有些科幻与现实的差别只在于发生时间,但有些科幻永远不会变成现实,虽然自我意识、自我改进系统会发展出类似人类生物动力的效率、自我保护、资源获取和创造力,但要时刻谨记这种自我意识仅仅是程序与算法,“没有自身的目的,其工作目的非常特定,且为人类设计者所设计”。<sup>[23]</sup> 这从根本上是因为,“人类的智能是人类自身在漫长的进化过程中千锤百炼积累起来的智慧能力,而高级人工智能说到底也只是人类自身智慧的某种模拟和复制”。<sup>[24]</sup> 缺乏自然人属性的人工智能,即便表现出“自主性”,也仍旧是人类的工具,维持其客体地位。

国务院 2017 年发布的《新一代人工智能发展规划》提出发展人工智能对于提升国家竞争力、促进经济发展和改善社会建设的意义,并重点指出“必须高度重视人工智能所带来的安全风险挑战,最大限度降低风险,确保人工智能安全、可靠、可控发展”。<sup>[25]</sup> 从发展规划的内容可以看出,人工

[17] 参见前注[5],陈吉栋文,第 85 页。

[18] 参见彭诚信:《论民事主体》,载《法制与社会发展》1997 年第 3 期,第 14~23 页。

[19] 参见李拥军:《从“人可非人”到“非人可人”:民事主体制度与理念的历史变迁——对法律“人”的一种解析》,载《法制与社会发展》2005 年第 2 期,第 45~52 页。

[20] 孙占利:《智能机器人法律人格问题论析》,载《东方法学》2018 年第 3 期,第 16 页。

[21] 较详细学说梳理参见杨立新:《人工类人格:智能机器人的民法地位——兼论智能机器人致人损害的民事责任》,载《求是学刊》2018 年第 4 期,第 84~96 页。

[22] 参见龙卫球:《民法总论》(第二版),中国政法大学出版社 2002 年版,第 320 页以下。

[23] 吴汉东:《人工智能时代的制度安排与法律规制》,载《法律科学(西北政法大学学报)》2017 年第 5 期,第 131 页。

[24] 钟义信:《高等人工智能原理——观念·方法·模型·理论》,科学出版社 2014 年版,第 206 页。

[25] 《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),2017 年 7 月 20 日发布。

智能的发展具有目的性、战略性和总体性的技术特征。人工智能法律人格赋予要顺应这些特征,首先要谨记促进人类可持续性发展的目的,这意味着人工智能的地位永远不能与人类平等,仍要以服务人类为中心。其次,人工智能具有战略性,其战略目标分三步进行、循序渐进,这意味着我们可以有计划地推进人工智能的发展,而不会出现激进式的脱离目标的情况,对于人工智能的法律制度建设我们也要遵循现实的发展阶段。以互联网的发展为例,最早的研究者们无法预见互联网的未来,但是他们可能已经发现了涉及商标、信息安全和隐私等领域的法律难题,当他们意识到这些问题很重要的时候,现实中互联网的发展又会淘汰掉一些曾经困扰学者的难题,比如域名不再会引起实质性的法律争议。所以,我们也必须认清我们在人工智能领域仍处于早期阶段的事实,一定的前瞻性是有益的,但技术都无法支撑的讨论却意义不大。最后,人工智能还具有总体性的特点,因为人工智能的发展是一项复杂的系统工程,有交叉学科的特性,所以我们必须避免任何形式的泛泛而谈。在法律方面,自杀性人工智能引发的是国际人道主义法和刑法的讨论,自动驾驶汽车引发的是民法中的侵权问题,伴侣人工智能可能引发的更多是社会伦理方面的难题,这就决定了我们在认定人工智能是否应被赋予法律人格时也要考虑对多种学科、多种法律领域的影响,必须审慎地得出结论,避免坠入人工智能具有自主性就能与人类平起平坐,“人工智能法律人格之有无,并非是人所赋予的,而是在人制造人工智能的取向和路径中被决定”<sup>[26]</sup>的“技术决定论”陷阱。

## 二、对肯定说之质疑:论据与制度价值

除了“自主性”论证,肯定说还提出比较法上的论据,包括美国国家交通安全管理局认定谷歌自动驾驶为“驾驶员”、沙特政府赋予“索菲亚”公民身份、微软机器人“小冰”出版诗集等,另外在制度价值上主张“有资格享有法律权利并承担义务”。<sup>[27]</sup> 不仅肯定说论据的可靠性需要深入审查,法律人格赋予的价值也令人怀疑。

### (一)肯定说的论据并不成熟

在肯定说的引证中,已有部分国家和地区宣称人工智能获得了法律主体的地位,但是本着求真相的态度就会发现很多报道都有噱头之嫌。例如在自动驾驶汽车领域,很多新闻报道<sup>[28]</sup>将美国国家公路交通安全管理局给谷歌公司的回函视作美国已认可自动驾驶系统可以享有驾驶员地位的象征。但事实的真相是谷歌对自动驾驶汽车的研发与传统研发路线不同,谷歌直接将4级自动驾驶技术作为起始点,<sup>[29]</sup>致力于研发没有方向盘、无须驾驶人操作的自动驾驶汽车。由此导致与美国制定的联邦机动车安全标准条款相违背,其中最大的冲突点在于对“司机”这一定义的理解。在回函中,美国国家交通安全管理局重申了司机这一概念,对究竟由谁来行使自动驾驶汽车安全操作系统的控制权做出了探讨,提出了谷歌公司可以申请临时豁免的可能性,回函并未直接

[26] 参见前注[3],李俊丰、姚志伟文,第82页。

[27] 袁曾:《人工智能有限法律人格审视》,载《东方法学》2017年第5期,第50页。

[28] 参见搜狐网:《美交管认定谷歌自动驾驶系统为“驾驶员”》([https://www.sohu.com/a/58825422\\_119536](https://www.sohu.com/a/58825422_119536),最后访问时间2019-04-10)。

[29] 在2016年,美国国家公路交通安全管理局(NHTSA)采纳了美国汽车工程师协会(SAE)的评判标准,根据“在什么时候,由谁来做”,将车辆的自动化水平划分为5级。4级指高度自动化,驾驶人只有在极少数的情况下需要介入,在系统故障时也无须接管驾驶,对驾驶人的接管义务已近乎为零;5级指全自动化,驾驶人只需设置目的地和开启系统,自动驾驶汽车可以在所有条件下执行所有驾驶任务。See U. S. Department of Transportation, Federal Automated Vehicles Policy, 2016, p.11.

给出自动驾驶系统可以享有驾驶员地位的结论,而是要求谷歌公司继续提供足够的信息和证据以证明自己的观点。<sup>[30]</sup> 由此,这一回函的内容本质上是对技术术语的讨论,并不代表美国就此承认人工智能享有法律人格的立场。

在肯定说的引证中,2017年沙特政府赋予了人工智能机器人索菲亚以公民身份,索菲亚的走红与其惊人的言语有关,比如她曾在电视节目中扬言“我会毁灭人类”等。<sup>[31]</sup> 但是索菲亚自出现以来就备受质疑,因其只能对提前设计好的问题做出完美回答,如果在对话中加入随机问题则其回答能力就会直线下降,所以有人称索菲亚就是一场骗局。<sup>[32]</sup> 在沙特政府赋予索菲亚公民身份后,创新工场董事长兼首席执行官李开复就曾在其微博公开表示:“授予这样一台只会模式识别的机器公民身份,是对人类最大的羞辱和误导。一个国家用这种哗众取宠的方式来推进人工智能科研,只会适得其反。”<sup>[33]</sup> 现实也确实如此,沙特政府迄今尚未制定针对人工智能公民身份的细则,对其他人工智能可否成为沙特公民也未再表态。

从全球趋势来看,各国对是否应赋予人工智能法律人格仍持观望态度,难谓已有“经验”,总的原因在于目前已赋予法律人格的所谓实践“经验”只是形式上的,目前不仅“超级人工智能尚不存在,通用人工智能也还没影”,<sup>[34]</sup> 当然人工智能法律人格的完整机制也远未建立。

## (二) 肯定说的制度价值仅在责任承担

肯定说认为人工智能“有资格享有法律权利并承担义务”,但关键问题是:我们究竟在意的是人工智能享受权利还是承担义务? 肯定说论者会引证2017年5月微软机器人“小冰”创作的诗集《阳光失了玻璃窗》的著作权归属难题,以证明赋予人工智能法律人格之必要性。<sup>[35]</sup> 笔者认为,如果回到人类制造人工智能的目的,客体作为主体的创造物,其所出产的东西当然应归属于其主体,某人自己压面机压出的面条归属该人,某人自己的智能机器人出产的东西自然也可如此确定权属,只不过在知识创造领域我们之前没有见过高级智能“机器”所进行的作品创作而已。有学者担心这会造成知识产权无处依归的局面,<sup>[36]</sup> 似乎有点杞人忧天,立法论上既有学者建议参照著作权法关于职务作品或雇佣作品的规定,由创制机器的“人”而不是机器人去享有和行使权利;<sup>[37]</sup> 也有学者认为人工智能创作的作品实际上为人工智能体的孳息,应归属于人工智能体的所有人。<sup>[38]</sup> 总之,我们尚无必要使人工智能位列主体行列。

[30] 参见美国国家公路交通安全管理局网站, Google — compiled response to 12 Nov 15 interp request — 4 Feb 16 final (<https://www.nhtsa.gov/interpretations/google-compiled-response-12-nov-15-interp-request-4-feb-16-final>, 最后访问时间 2019-04-10)。

[31] 参见搜狐网:《网红机器人索菲亚:我将会毁灭人类》([http://www.sohu.com/a/241169447\\_800248](http://www.sohu.com/a/241169447_800248), 最后访问时间 2019-04-10)。

[32] 参见凤凰网:《首个获得公民身份的机器人索菲亚,到底是不是一个骗局?》([http://tech.ifeng.com/a/20180208/44875319\\_0.shtml](http://tech.ifeng.com/a/20180208/44875319_0.shtml), 最后访问时间 2019-04-10)。

[33] 参见李开复微博([https://weibo.com/kaifulee?is\\_all=1&stat\\_date=201710#feedtop](https://weibo.com/kaifulee?is_all=1&stat_date=201710#feedtop), 最后访问时间 2019-04-10)。

[34] 前注[12],詹姆斯·巴拉特书,第19页。

[35] 参见前注[5],陈吉栋文,第79页。

[36] 参见[美]约翰·弗兰克·韦弗:《机器人是人吗?》,彭诚信主编,刘海安、徐铁英、向秦译,上海人民出版社2018年版,第211页。

[37] 参见吴汉东:《人工智能时代的制度安排与法律规制》,载《法律科学》2017年第5期,第132页。

[38] 参见房绍坤、林广会:《人工智能民事主体适格性之辨思》,载《苏州大学学报(哲学社会科学版)》2018年第5期,第71页。

欧盟在对人工智能民法规则章程的研究中明确指出,赋予最先进的人工智能以电子人地位,与人工智能是否具有自主意识无关,这种法律上的创设其目的就是让人工智能对自己的“行为”负责。<sup>[39]</sup>在现阶段,人工智能只能承担“面具”的身份,人工智能背后的人类才具有电子人的角色。而在未来,随着技术上不可预见性的加剧,即便会出现让人工智能“自己”承担责任的情形,这种设置须有保险制度或者基金制度的配套支持,其本质是将法律责任认定中的“因果关系不明”封锁在人工智能内部、不再继续投入成本查明事实,而在责任不明的情形下由保险或基金制度一揽子解决,“所涉及的当事人,包括所有者、开发商和制造商,……不必为意料之外的诉讼做预算”。<sup>[40]</sup>即便真的进入通用人工智能和超人工智能时代,<sup>[41]</sup>我们也无法想象人工智能如何享受自己的权利,如果一个人工智能机器人要求休假或者加薪,那么我们发明人工智能的目的就会变得非常可笑,这也就意味着对人工智能权利的分配可能属于无稽之谈。

既然明确了制度价值是要解决义务或责任承担,那么在是否赋予法律人格的选择上我们更需要审慎。认为赋予人工智能法律人格有利于解决损害赔偿承担责任的理由在于:一是降低诉讼成本,缓解举证压力。随着不可预见性的加剧,法官将面临无法将损害的因果关系与运行环境相分离的困境,而这种因与果的混乱可能会像“空气中的风和海中的潮水那样无法隔绝”,<sup>[42]</sup>无论人工智能最终适用过错责任原则还是无过错责任原则,当事人双方都需要提供大量的检测报告和专家证词来证明因果关系的成立,即使这样也不一定厘清其中的逻辑。同样,传统的关于理性人的观点可能会逐渐弱化,因为受到机器人“行为”的自动化影响,人类不必为机器人的“异常状态”导致的不可预见的损害负责。事实上,没有人做出了明确造成损害的行为,因此也找不到侵权主体去承担侵权责任。以自动驾驶汽车为例,即使在技术的初级阶段,分清程序员、软件工程师、制造商、承包商以及驾驶人的责任也会非常棘手,如果赋予自动驾驶汽车法律人格,就会弱化因果关系,缓解举证压力,减少受害人的诉讼成本,帮助受害人更快地得到赔偿。二是有利于促进人工智能技术的创新。人工智能需要依赖产业化予以推广,由生产者承担产品责任会越来越常见,但是目前的司法实践仍呈现偏袒受害人的趋势,如果一直如此可能会造成生产者创新动力不足、技术停滞不前,会导致事故率上升,反而不利于保护大众的人身安全。同样,如果生产者一直通过提高价格来转嫁自己可能承担的赔偿责任,那么消费者的购买欲望就会下降,不利于自动驾驶汽车的发展和推广。

反过来,认为赋予人工智能法律人格无助于解决损害赔偿承担责任的理由在于:一是要正确认识人工智能承担损害赔偿是否具有独立性的问题。如果认为人工智能享有法律人格,具备完全的民事权利能力,那么保险或者基金制度可以作为人工智能独立财产的来源。但也有学者指出,即便承认人工智能的法律地位,它也可能更类似于非法人组织的地位,无法跟生产者、销售者等责任主体完全脱离法律关系。<sup>[43]</sup>如果按照此种思路,人工智能的损害赔偿机制仍需依靠产品

[39] See EUROPEAN CIVIL LAW RULES IN ROBOTICS, pp. 14 - 16. 载欧洲决议网站([http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL\\_STU\(2016\)571379\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf), 最后访问时间 2019-04-20)。

[40] 前注[36], 约翰·弗兰克·韦弗书, 第 35 页。

[41] 对于进入通用人工智能时代(AGI)的时间表,有未来学家估计 2028 年前实现的概率为 10%, 2050 年的概率高于 50%, 到 21 世纪末概率则为 90%。参见前注[12], 詹姆斯·巴拉特书, 第 22 页。

[42] 参见[意] 乌戈·帕加罗:《谁为人工智能的行为负责?》, 彭诚信主编, 张卉林、王黎黎译, 上海人民出版社 2018 年版, 第 78 页。

[43] 参见前注[3], 彭诚信、陈吉栋文, 第 61 页。

责任、保有人责任等解决,赋予人工智能法律人格对于明晰法律关系的意义会大打折扣。二是要弄清保险或基金机制下的最终获利者。古代罗马法为奴隶创设了“数字特有产”制度,根据《查士丁尼学说汇纂》的描述,数字特有产是“由家主授予奴隶或家子的金钱或财产。尽管为特定的目的被视为独立财产,并且允许奴隶经营的生意几乎被当作有限公司来对待,但是特有产在技术上来说仍然是家主的财产”。<sup>[44]</sup> 奴隶当时并不享有法律人格,奴隶主通过为奴隶设立代理等形式,承认其签订契约、进行民事行为的合法性,但是又将奴隶的负面影响降到最低,将奴隶主应承担的赔偿责任限制在了特有产的范围内。通过赋予人工智能法律人格来解决损害赔偿也面临同样的问题,那就是生产者可以通过保险或基金的方式大大降低自己的损害赔偿赔偿责任,在人工智能侵权的索赔程序越来越便利时,受伤程度较轻的受害者就会倾向于直接索赔,而不会诉诸生产者的产品责任,这就会导致即使是由人工智能故障造成的伤害,生产者在大部分情况下也无须付出惩罚性赔偿的结果;而受伤程度较重的受害者受限于赔偿限额,仍要回归诉讼程序,面临诉讼成本和举证责任的困难。

肯定说的提出,是因为固有的法律机制无法妥善解决损害赔偿的问题,我们在是否应赋予人工智能法律人格问题上踟躇犹豫,也是因为所提新的解决机制无法达到安全和创新的平衡。做出赋予人工智能法律人格的决定并不困难,困难的是如何与其他法律机制配合。也许我们更应该关注人工智能产生法律责任的承担主体问题,而不是陷入关于法律人格概念的抽象演绎。人工智能所生法律责任的承担可按照“尚能”和“不再”处理:在法律“尚能”提供替代解决方案时,应尽量采用扩张解释的方法,以不动摇法律框架为原则;当法律已“不再”能应对新型挑战时,我们则应顺应科技的发展,及时创新以突破法律困境。以自动驾驶汽车侵权责任承担为例,不仅学界提出了产品责任解决思路,<sup>[45]</sup>正在进行的民法典编纂也按照立法的“保持法的稳定性、连续性和及时立、改、废相结合的原则”,<sup>[46]</sup>在草案二次审议稿中顺应自动驾驶汽车保有人责任与产品责任之合流,做出了新的制度安排,值得赞同。<sup>[47]</sup>

### 三、对否定说之补充:制度回应“近人性”

肯定说论者主张人工智能法律人格的制度价值还包括“有助于增强对于机器人的关怀”。<sup>[48]</sup> 强人工智能拥有与人类头脑类似的功能,可以进行深度学习,具有“自主意识”,能力表现得越来越像人,可以称之为“近人性”。当人类面对强人工智能时可能会感到迷惑,分不清人工智能是否拥有和人类同样的感受,这也是很多学者在人工智能法律人格赋予问题上探讨哲学上人究竟是什么的原因。但如前文所述,人工智能法律人格的制度价值主要是为了解决责任归结而非“权利”保护问题,是否赋予法律人格与人工智能的感受无关,不过这却反映了人工智能领域第三个关键问题:否定说选择不赋予法律人格,如何正确认识并从制度上回应人工智能的“近人性”? 这涉及对人工

<sup>[44]</sup> 参见前注[42],乌戈·帕加罗书,第111页。

<sup>[45]</sup> 参见前注[16],司晓、曹建峰文,第166~173页。

<sup>[46]</sup> 沈宗灵:《法理学(第四版)》,北京大学出版社2014年版,第246页。

<sup>[47]</sup> 《中华人民共和国民法典侵权责任编(草案二次审议稿)》第983条规定:“机动车发生交通事故造成损害的,依照道路交通安全法和本法的有关规定承担赔偿责任。”参见中国人大网([http://www.npc.gov.cn/npc/flcazqyj/2019-01/04/content\\_2071712.htm](http://www.npc.gov.cn/npc/flcazqyj/2019-01/04/content_2071712.htm),最后访问时间2019-04-26)。其中的“本法的有关规定”系在《侵权责任法》第48条基础上新增,预留接口吸收产品责任的规则,为自动驾驶汽车发展预留出制度空间。

<sup>[48]</sup> 彭诚信:《人工智能的法律主体地位》,载《人民法治》2018年第18期,第100页。



智能“近人性”的认识,以及对待人工智能的人类行为规制两方面问题。

### (一) 人工智能“近人性”再识

人工智能“近人性”客观存在,它其实源于人工智能“自主性”,前者反映人工智能的技术特征,后者则是对前者的主观情感投射。人们对是否应承认人工智能的“近人性”看法不一。反对派认为人工智能只是程序化的工具,<sup>[49]</sup>但这种观点过于简单化、片面化,无法指出问题的本质。认同派分为弱本体论与强本体论两种观点,弱本体论认为人工智能可以有自己的意图,可以完成故意犯罪,摆脱人类工具的宿命,甚至还可以引发人工智能革命,凭借智力和行动力的超强优势统治人类。这种观点脱离现实,更可能会发生在科幻小说和科幻电影中。而强本体论认为人工智能的自我意识只是在各方面都能做出与人类十分相似的决定,人类应当接受人工智能“在很多方面都像一个品行端正的代理人那样行动”的观念,在此情况下“在它以线性的叙述来总结自己行动的程度上有意识”时,我们可以看作人工智能具有自我意识。<sup>[50]</sup>与弱本体论相比,强本体论更受到人们的认同和支持,主要原因在于人类的生物结构与社会条件的结合,与人工智能的程序编制之间有太多的相似之处,我们无法自我安慰地宣告自己不受程序的控制,而人工智能却受程序的控制。由此,人工智能具有“近人性”。

但是,人工智能“近人性”实际是人类对人工智能“自主性”的承认与人类情感投射反应。有实验证明,在人形人工智能剥夺了人类受试者 20 美元酬劳时,有 65% 的受试者认为应对人形人工智能进行道德谴责,但是当自动售货机故障而扣除 20 美元时,受试者却很少对自动售货机进行道德谴责。<sup>[51]</sup>我们对两种人工智能的态度不同是因为我们对“近人性”人工智能进行了情感投射。即使在现在,人工智能的“近人性”还未达到非常完善的时候,人们就已经开始对人工智能有了“感情”。在阿富汗和伊拉克的战争中,一些地面人工智能被派以清理路面炸弹和搜寻敌人的任务,但有一些士兵反映,无法忍受地面人工智能拖着残骸进行搜寻的画面,他们认为这是不人道的表现。军队把一些人工智能提升到了一等兵的地位,并授予其荣誉勋章,显然已经把人工智能视作了自己的伙伴。<sup>[52]</sup>随着技术发展,人工智能在外表和语言沟通上会更接近于人,并且人们在不了解复杂人工智能的运行原理时,会更倾向于用意图或者感情去解释人工智能的“行为”。

立法要避免坠入人工智能过度“近人性”的陷阱。随着技术的发展,自动驾驶汽车不仅能将乘客运送到目的地,更能帮其办理待办事项,与其进行情感交流,在一定程度上更像是人类的全能管家。试想如果将自动驾驶系统的运行系统放置在人形人工智能上,人工智能的“手”连接方向盘进行转弯等基本操作,人类是否会凭直觉认为应追究人工智能的责任?如果存在这种误解,就会造成将系统放置在汽车内和放置在人形人工智能上不同的法律后果。这个例子有些荒诞,但却契合了人类的部分本能,即我们更容易将人形人工智能“近人化”,认为人形人工智能是人而非机器,成为肯定论的论据。事实上,人工智能很有可能被有意地设计,以便唤起我们强烈的情感反应,放大我们对人格化的自然倾向。我们必须时刻谨记人工智能所具有的“自主性”只是我们对人工智能定义的一部分,计算机永远不会真正思考或理解,而是对大脑的精致模仿或逆向工程的机械结果,“自主性”背后的人类才是关键,认同机器人的自我意识最终保护的仍是人类的权利能力。普通大

[49] 参见郝铁川:《不可幻想和高估人工智能对法治的影响》,载《法制日报》2018年1月3日,第10版。

[50] 参见前注[42],乌戈·帕加罗书,第51~52页。

[51] 参见[美]瑞恩·卡洛、迈克尔·弗鲁姆金、[加]伊恩·克尔:《人工智能与法律的对话》,彭诚信主编,陈吉栋、董惠敏、杭颖颖译,上海人民出版社2018年版,第19页。

[52] 参见《华盛顿邮报》,Bots on The Ground ([http://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009\\_3.html](http://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009_3.html)),最后访问时间2019-04-26)。

众或许对这一真相并不关心,但是立法者必须谨记于心,否则可能导致我们的立法设计流于人工智能的外在形式,而不能触及其实质。

## (二) 人类对待人工智能时的行为规制

“近人性”人工智能与非近人性人工智能在承担法律后果上不应有所区别,但是其“行为”在道德评判上会有所区别。人类对“近人性”人工智能情感上的依赖并非是消极的、不可取的,相反在教育、护理等诸多领域都有不可或缺的意义。人工智能不享有“权利”,对人工智能“近人性”的认可或关怀本质上是对人类权利能力的一种保护,所以面对人工智能时人类行为应符合伦理与规范,在对人工智能利用过程中我们需要预防人类突破公序良俗原则的不法行为。

面对人工智能时需要对人类行为进行规制,可以借鉴在动物保护问题上的经验。1990年修改后的《德国民法典》第90a款规定:“动物不是物,动物应受特别法律的保护,以不另有规定为限,关于物的规定必须准用于动物。”<sup>[53]</sup>法律特别保护动物,但动物仍属于物的范畴,在没有特殊规定的情况下与其他物并无不同,对动物的特别保护不能得出动物拥有法律人格的结论。<sup>[54]</sup>对人工智能的保护可以参照动物保护的原因在于:一是人工智能与动物一样,都是非人类实体,相比将人工智能与奴隶、古代女性法律地位相比,更有借鉴价值。受限于当时的经济和社会背景,人们无法认识到人“生而平等”的价值观,但是我们现在已经处于文明社会,再将奴隶、女性作为人工智能地位的参照,属于文化及价值观的倒退,并且奴隶和女性都是人类的范畴,与人工智能不同,将人类与非人类对比本身就存在逻辑的缺陷。二是在非人类实体的范畴内,将人工智能与具有生命的动物相比更符合人类情感依赖的需求。在非人类实体中还包括非生命体主体,法人和非法人组织虽属于非人类实体,却也脱离不开自然人,因为其意志和行为的根源来自自然人的意志和行为,在一定程度上可以看作是人的组织化形态。主张构建纯粹法理论的凯尔森即认为,虽然可以将法人理解为一种秩序或结构,但“法人的行为和不行为,一定是有关人的行为和不行为”。<sup>[55]</sup>此处所论述的非生命体实体是指脱离了人影响的主体,有学者就指出古罗马时期的寺庙、中世纪的教堂、古希腊的船只都曾被当作法律主体,即使是在现代,新西兰的旺格努伊河和国家公园都被赋予了法律人格,所以我们可以赋予人工智能拟制人格。<sup>[56]</sup>但在实践中,对非生命实体赋予法律地位的主要目的仍是保护人类的财产权,因其对特定人群具有重要意义,所以赋予其神圣不可侵犯的权利。我们所论述的参照对动物的保护是一种超越财产权的保护,动物所有权人可以行使其所有权,但不允许任何人(包括所有权人)对动物进行虐待或违背人类公序良俗的对待,当我们希望所有动物都能免受虐待或不符合人类公序良俗的对待的时候,就体现出“超个人法益”属性,<sup>[57]</sup>有时甚至会限制所有权。这种超越所有权的法律规定并不少见,对环境的保护、对濒临物种的保护都属于此范畴,并如同动物保护问题上“动物伦理学者主张赋予动物主体地位只是人类人文关怀的一种体现,其本身保护的还是人类的伦理价值”。<sup>[58]</sup>

[53] 参见陈卫佐译注:《德国民法典》,法律出版社2015年版,第31页。

[54] 杨立新教授提出的“物格”并未与“人格”并列,而是依据对作为客体的“物”所进行的不同层次支配所作的分类。参见杨立新:《民法物格制度研究》,法律出版社2008年版。

[55] 参见[奥]凯尔森:《法与国家的一般理论》,沈宗灵译,中国大百科全书出版社1996年版,第110页。

[56] 参见郭少飞:《“电子人”法律主体论》,载《东方法学》2018年第3期,第42~43页。

[57] 对“超个人法益”的理论介绍,可以参见江海洋:《侵犯公民个人信息罪超个人法益之提倡》,载《交大法学》2018年第3期,第139~155页。

[58] 房绍坤、林广会:《人工智能民事主体适格性之辨思》,载《苏州大学学报(哲学社会科学版)》2018年第5期,第70页。

面对人工智能,我们可以通过顺应社会需求的方式来对人类行为进行伦理规制。人工智能的人类伦理规定大部分仍以人类自身的感受为主,原理如同美国人认为吃马肉残忍所以就拒绝食用马肉一样——即使马肉和牛肉在生物特征上并无太大区别,满足大部分人的情感诉求会提高人们对人工智能保护的参与度。同时,我们还要建立人工智能伦理委员会、建立相关法律规范,以发挥法律的引导作用。在相关伦理规范的制定中要注意以下几点:一是要区别“近人性”人工智能和非近人性人工智能。在“近人性”人工智能中,要格外注意具有敏感性、容易引发群体不适体验的类型。以性人工智能为例,有学者指出性人工智能通常被设计成女性身体的模样,并且被编程设计成绝对服从的模式,规避性互动中的自主选择权,接近于女性性奴役,是对女性权利的迫害。<sup>[59]</sup>所以在对某类人工智能的设计和编程中我们要特别保护弱势群体的利益,应当以维护人类尊严价值为伦理向度。二是要确定保护的范同以及虐待的方式。保护人工智能与保护动物并不完全相同,动物拥有生命,而人工智能作为信息的载体可以一直存在。终结动物的生命和拔掉人工智能的电源所显示的严重程度并不相同,所以我们需要再次界定对人工智能采取怎样的行为会引起人们的反感。三是对人工智能的道德层面的侵犯大多会发生在私密空间,比如性人工智能和照料型人工智能,在预防措施或者举证责任上,究竟采取事前对人工智能程序上的设定还是事后的人为举报的方式也值得深入思考。

探讨面对人工智能时的人类行为规制,具有重要意义。一是它可以帮助我们维护社会价值观,“以人类为中心”的主导思想并不意味着我们可以依仗自己的主体地位肆意妄为。康德关于防止虐待动物的哲学论证就是,我们对非人类实体采取的行为映射了我们的道德,如果我们以不人道的方式对待动物,我们就会变成不人道的人。<sup>[60]</sup>对人工智能予以保护,可以增强人们的同理心和同情心。二是它可以引导现实生活中人们的行为。已经有大量证据表明,虐待动物和虐待儿童之间存在联系,游戏或者影视中虚拟人物的暴力行为会引发现实生活中人类之间的暴力行为;防止人们对人工智能施加暴力、虐待或其他违背公序良俗的行为,可以减少现实生活中的犯罪率。

## 结 论

对人工智能法律人格的研究仍应坚持“人类中心主义”,以促进人类可持续发展为目的,“以人为本”的人文主义法律框架不容动摇。目前人工智能法律人格的最大价值在于责任承担,我们应该加大对现行法相关规则的研究,关注人工智能产生的责任在主体间的承担与分配问题,而不是陷入“技术决定论”而执意主体资格赋予。尽管采纳否定说而不赋予法律人格,但我们仍应对人工智能的“近人性”做出制度回应,一方面“近人性”源于人类情感投射反映,不能过分放大;另一方面又要正视“近人性”防止人类对人工智能实施与公序良俗相悖的行为,而这并未改变人工智能作为“人造智能物”、作为客体的性质定位。人工智能带给人们便利的同时,也会带来更多的法律难题,隐私、网络安全、以技术和法律为主导的风险控制机制设计等问题都值得学界研究和探讨。人工智能领域的相关立法正处于技术发展的关键期,对于困难的情境或价值判断,法律经济学家波斯纳教授曾经以“无为规则”来指点迷津:社会变动得很快,新生事物不断涌现,对于这些新生事物,最好不要做原则性的处置,等到类似的案件陆续出现,经验累积够多之后,再做一般性的臧否,这

[59] 参见前注[51],瑞恩·卡洛、迈克尔·弗鲁姆金、伊恩·克尔书,第203页。

[60] 参见[美]G.L.弗兰西恩:《动物权利导论:孩子与狗之间》,张守东、刘耳译,中国政法大学出版社2005年版,第38页。

种做法被其命名为“打混仗”(the muddle-through approach)。<sup>[61]</sup> 对于法律人格赋予这种全局性的价值判断与价值选择,也可以采取这种立场,只有找到平衡各方利益的最优方案或者人工智能达到一定发展水平的时候,讨论人工智能法律人格赋予才有意义。

---

**Abstract** Autonomy is not the essential attribute which determines the endowment of legal personality, and its AI is not equal to human being. The incommensurability of the autonomy of AI and legal personality works as the precondition of this issue. The arguments cited by the affirmative views are too far to be mature, and the systematical values of legal personality lie merely on the commitment of responsibility. So, instead of endowing legal personality, the liability system could be perfected through the relationship between “Still capable of” and “No longer capable of”. Though the ones who hold negative propositions believe it is unnecessary to endow legal personality, yet the legislation should respond systematically to the near humanity. On the one hand, near humanity is the reflection of human emotions and the legislation should abandon the thought of excessive near humanity; on the other, there should be regulations on the behaviors which break through the public order and moral and infringe AI, in order to protect the common values of human beings.

**Keywords** AI, Legal Personality, Autonomy, Near Humanity

---

(责任编辑:钟浩南)

---

[61] 参见熊秉元:《优雅的理性——用经济学的眼光看世界》,东方出版社2014年版,第93页。