

# 伦理人格与技术人格：人工智能法律主体地位的理论框架

梅夏英\*

**摘要** 目前人工智能法律主体地位的理论探讨遵循两种路径：一是将人工智能作为具有高级智能的类人“种群”来进行理论预判；一是探讨赋予当前弱人工智能某种法律拟制人格来承担责任和享有权利。对此有必要提出“伦理人格”和“技术人格”的区分理论，来界定人工智能主体地位的不同理论面向。人格区分现象在传统民法中已然存在，它呈现为以人格抽象程度为标志的人格递进序列，其中存在着两种人格的“渐变”和“断裂”现象。人工智能作为伦理主体遇到了“自我意识”的难题，目前的弱人工智能尚不能获得独立的伦理人格，规制技术开发者的科技伦理起主导作用，同时不排除人机交互伦理可能会赋予机器人某种道德性“权利”。就人工智能的技术性人格而言，如果只是将人工智能作为“个体”进行研究，并无自然人的意思表示机制和机器人自身财产的支撑，赋予机器人技术人格便不完全具备条件。未来人工智能的主体性将依照两种人格的路径各自发展，以技术人格的探索先行，逐渐进行伦理人格的塑造，人类或机器人的伦理人格最终成为技术人格的依归。

**关键词** 人工智能 伦理人格 技术人格 法律拟制 人机关系

## 引言

人工智能是否具有法律主体地位，或能否成为法律意义上的“人”，是近年来数字法学的一个重要理论问题。人工智能的出现，给法律领域带来了一个富有建设性和挑战性的事实，即它

\* 对外经济贸易大学法学院教授。本文系教育部重点研究基地重大项目“数字经济基础性私法制度研究”(项目编号:22JJD820016)及国家社科基金哲学社会科学领军人才项目“数据跨境战略博弈与全球数据治理研究”(项目编号:24VRC055)的阶段性研究成果。

创造和开启了一个前所未有的新型关系——“人机”关系。这种“人机”关系随着技术的进步快速演进,成为未来数字法学甚至整个法学必须面对的一种基础关系,并将很大程度上影响未来的立法进程。“人机”关系在法律上之所以具有历史性的、突破性的意义,主要原因在于它将深刻改变未来社会的社会关系、生活方式和社会形态,人与机器(主要指人工智能)将由传统的“主客体”支配关系转向一种深度“合作”关系,并遍布生活的每一个角落。在可预见的人机融合的未来社会场景中,法律如何看待人工智能以及赋予人工智能何种法律地位,成为数字法学必须回答的问题。当前法学界对于人工智能的法律主体地位问题已经进行了探讨,并主要局限在私法领域,这是由于在人工智能目前的发展状况下,私法领域的讨论已经具有相对确定的现实意义,因为它在解决诸如无人驾驶、医疗机器人、机器代理等领域的责任承担以及机器作品的创造者和权利享有者的界定等问题上均有所建树。但基于人机关系的普遍性和公共性,人工智能的法律地位问题在公法领域亦有充分的讨论空间。针对当前私法上关于人工智能法律地位问题的讨论,本文拟在传统私法的基础上,依据传统私法所隐含的(未被充分揭示的)两类主体内涵的差异,梳理人工智能在私法主体领域的两个面向或诉求:伦理性人格和技术性人格,并将之作为基础理论框架,分别探讨人工智能在两种人格类型上的意义和现实可行性,最后尝试性地提出理论展望,即人工智能的主体性是否将如传统民法主体制度一样,呈现出以“抽象度”为标志的两种人格融合场景。

## 一、人工智能主体性的两个理论面向:伦理性人格和技术性人格

当前关于人工智能主体地位的私法探讨,将人工智能纳入到传统民法的“人法”部分予以统一回答,并形成了两种截然相反的观点:否定说和肯定说。否定说目前成为大多数人支持的主流观点,该说认为人工智能目前成为民事主体并不现实,其主要理由在于:人工智能目前并不具有人类的理性和自我意识,亦不具有自身财产,且不赋予其主体地位并不影响法律解决与人工智能相关的法律问题;<sup>〔1〕</sup>肯定说则针对性地认为,对于一种新型民事主体的承认,法律并不必然要求主体具有人类理性,如法人(尤其是财团法人)就体现为财产的聚合;亦不要求责任财产的存在,如动物、胎儿和非法人团体等民法上的“有限主体”就与财产无直接关系。<sup>〔2〕</sup>显然,如果按照上述理论框架来论证人工智能的主体性,我们很可能无法在统一的理论基础上达成有效结论。其中被忽略的一个事实是,两种观点谈及的人格很可能不是一回事,故有必要对人工智能的人格类型进行区分,并回到传统民法找到这种区分的依据,以做出合理解释。

### (一)人工智能主体中伦理性人格和技术性人格的区分

人工智能能否成为法律上的主体,是数字技术为法学理论界提出的一个全新的问题。在将人工智能纳入传统法律主体制度予以定位时,除了要与原有制度形成体系性的契合外,也将

〔1〕 参见孙良国:“人工智能不应成为法律主体”,载《光明日报》2024年8月16日,理论版;袁康:“破除人工智能成为法律主体的臆想”,载《光明日报》2024年8月23日,理论版。

〔2〕 参见石冠斌:“人工智能成为法律主体不存在理论障碍”,载《光明日报》2024年8月16日,理论版。

加深我们对传统法律主体制度的整体性理解,甚至能在原有体系上有所丰富和发展。就上述关于人工智能主体地位的当前理论分歧而言,从表面上看,持不同观点的双方都是在对人工智能能否成为民法上的“人”这一命题进行论证,但实际上双方在人工智能成为民法上的何种“人”这一立场上却存在显著差别。具体而言,当前人工智能为我们提出的法律主体问题呈现两种不同的样貌:一是从智能或“自我意识”角度探讨人工智能是否具有“类人”智慧,或是否可能超过人类成为强人工智能或超级人工智能,并以此探讨人与机器的交互关系。<sup>〔3〕</sup>如果未来人工智能具有高度智能或“自我意识”,那么它就不能被仅仅局限于人机之间的支配关系中,而是带有“类人”的理性、种群性、伦理性和社会性,并会穿透民法进入到公法体系;二是从解决与目前初级智能机器人相关的侵权责任或权利归属的角度出发,探讨运用传统民法中财产性的“拟制”人格来解决上述问题,这个讨论角度带有浓厚的技术性、工具性和财产性的色彩,其针对的技术性人格形象与第一种“类人”人格形象具有显著的差别。上述人工智能的两种主体形象分别遵循不同的理论基础、规范路径和社会功能,基于上述两种形象分别立足于主体的“伦理性”和“拟制性”,本文在此正式提出“伦理性”人格和“技术性”人格的区分,这种区分似应成为研究人工智能法律主体问题的一个理论框架。

具体而言,伦理性人格立足于人格的智性、种群性、伦理性和社会性,以某种形式的复杂系统的“生命”活性为表征。当前理论界经常非正式地将人与人工智能并列为“碳基生命”和“硅基生命”,亦可视为某种生命伦理倾向。技术性人格在民法上亦可称为“工具性人格”或“财产性人格”,指的是立法通过某种拟制技术来创造享有权利和承担义务的主体,这种主体在现实社会的伦理生活中并不必然存在,它的目的是为了在技术上合理安排财产的法律状态和责任承担者。当然这两种人格的区分并不是绝对的,它只是为我们提供了一个观察人工智能在社会不同场景中所扮演角色的视角,且不同角色是可以依据场合而变化的。尽管如此,这种理念上的区分仍然具有基础性的意义和必要性,主要体现在如下方面:首先,这种区分可以有效避免在人工智能主体问题讨论中常见的范畴混淆。基于传统民法的主体制度并未严格区分伦理人格和技术人格,当前的理论探讨沿袭了传统民法上统一的人格理论,对人工智能的主体问题做出统一的肯定或否定的回答,但由此便会导致一个理论误区的出现,即学者常常会用一种人格形象的必要条件缺失,来否定另一种人格形象成立的可能性。如用人工智能没有理性和意识这一理由,来否定人工智能可能获得财产法上的“有限主体”资格;<sup>〔4〕</sup>又如以“未成年人”“胎儿”等伦理主体没有任何财产亦可构成民事主体为由,来降低对于人工智能责任财产的要求。<sup>〔5〕</sup>这种人格错位在传统民法上的自然人上亦有显现(下文将会述及),只是由于机器由“客体”成为“主体”的假设本身理论跨度过大,故在人工智能领域体现得更为明显。通过对伦理性人格和技术性人格进行区分,就可以有效避免上述范畴混淆的误区。

〔3〕 See Rawas Soha, “AI: The Future of Humanity,” *Discover Artificial Intelligence*, Vol. 4, No. 25, 2024, pp. 1-14.

〔4〕 参见朱艺浩:“人工智能法律人格论批判及理性应对”,《法学杂志》2020年第3期,第135页。

〔5〕 参见彭诚信、陈吉栋:“论人工智能体法律人格的考量要素”,《当代法学》2019年第2期,第55页。

其次,这种区分可以有效厘清两种人格的理论和立法路径。两种人格虽然可以毫不违和地并存于民法中,但由于机器人毕竟不具有自然人的天然身份,故两种人格从一开始就体现出不同的理论和立法路径。如2017年沙特阿拉伯授予美国汉森公司的人工智能“索菲亚”以法律公民身份、日本在个案中确立了陪护老人的宠物机器人“帕罗”的户籍地位等,就是将机器人视为伦理意义上的主体而在公法上做出的法律判断,并不首先体现为私法上的主体。除此之外,一旦“类人”机器智慧体被法律承认,其首先要考虑的是它在整个社会层面的地位,而不会首先考虑财产问题,如2016年美国国家公路安全交通管理局认定谷歌无人驾驶人工智能系统可以被认定为“司机”,亦是基于智能方面的判断。<sup>〔6〕</sup>人工智能伦理人格遵循由公法论证确认后下沉至私法领域的路径,它更多考虑的是“类人”智能的社会身份(如公民或户籍等)和社会权利(如选举权、自由权、财产权甚至家庭生活的权利)等,继而会在私法上得到体现。而传统私法对于人工智能的主体界定则只能局限于民法范围内,通过技术拟制人格来解决财产地位和责任承担等私法问题,这种拟制人格无法延展至公法体系。如果立法不明示人工智能人格的类型,我们就无法判断将其安置于何种法律范畴,如2016年欧盟委员会法律事务委员会曾主张赋予人工智能“电子人”的法律主体身份,这种抽象的主张并不能明确“电子人”属于何种人格,故无从将其纳入合适的法律领域予以讨论。<sup>〔7〕</sup>

最后,人工智能伦理人格的理论颠覆性远远大于技术人格。对当下尚未达到人类智能程度的机器人赋予某种工具性人格,只是民法的拟制技术以及法解释学的运用问题,只要形成逻辑闭环即可(尽管障碍重重),并不会对现实社会的公法秩序造成实质影响。而一旦人工智能具有“类人”的伦理人格,则会产生颠覆性的社会变革,甚至法律本身亦会受到巨大的冲击。比如,如果人工智能具有“类人”智慧或自我意识,那么人类的法律如何能有效约束他?<sup>〔8〕</sup>换言之,人类是否有能力为具有高度智能的另类智慧体立法,以及人工智能是否愿意听从人类的安排就成为一个问题。这个问题会转化为,人类如何与另一个“硅基”生命种群和平共处。比如美国科幻作家艾萨克·阿西莫夫(Isaac Asimov)为机器人设定的三个伦理定律并没有得到技术上的有效支持,当前的自动驾驶尚未能保证完全遵守上述定律。又如,由于人工智能缺少时间概念,无自然衰老死亡,亦无生物繁殖能力等人类典型特征,它是否能分享人类的律法规则,以及人和机器之间的伦理和法律规范如何形成和发展,目前完全无法想象和预知。故伦理性人格的意义更为宏大和深远,有必要与私法领域的财产性人格分开探讨。上述对于两种人格的理论假设必要性的论述,部分地超出了传统的理论背景,但通过一定的理论想象,有助于我们更好地认识上述人格区分的法律意义。

〔6〕 参见美国国家公路交通安全管理局(NHTSA)给谷歌公司的回函,载美国国家公路交通安全管理局网, [https://isearch.nhtsa.gov/files/Google%20---%20compiled%20response%20to%202012%20Nov%20%2015%20interp%20request%20---%204%20Feb%202016%20final.htm#\\_ftnref6](https://isearch.nhtsa.gov/files/Google%20---%20compiled%20response%20to%202012%20Nov%20%2015%20interp%20request%20---%204%20Feb%202016%20final.htm#_ftnref6),最后访问日期:2024年11月26日。

〔7〕 参见郭少飞:“电子人法律主体论”,《东方法学》2018年第3期,第41—42页。

〔8〕 See Genevieve Kaess, “Could Consciousness Emerge from a Machine Language?” *Macalester Journal of Philosophy*, Vol. 20, No. 8, 2011, pp. 144-161.

## (二) 两种人格区分的民法学溯源解释

伦理性人格和技术性人格的区分在传统民法中已然存在,只是这种区分并未在“人法”部分得到彰显,并被正式地作为讨论民事主体的理论工具。其直接原因在于,民法上对自然人作为伦理主体和财产主体的双重身份并未进行区分,而是将自然人与作为财产主体的法人(还包括非法人团体)并列,以此形成人法的基本结构和内容。<sup>[9]</sup>但两种人格的区分是客观存在的,时常会造成民法理论上的困境甚至体系上的紊乱。如德国学者迪特尔·梅迪库斯(Dieter Medicus)认为,亲属法和继承法规定了相互之间具有联系的生活事实,而物权和债法规则体系不是基于生活事实的相似性,而是基于法律后果层面上的相似性。他揭示了,基于生活中的伦理共同体形成的人身法人格与基于财产占有和交易形成的形式化技术人格具有本质的不同,并进而提出:“民法典的人法部分仅仅是一件未完成的作品。人们几乎不可能从这些规定中推断出一般性的结论。”<sup>[10]</sup>另一当代德国学者罗尔夫·克尼佩尔(Rolf Knieper)则从财产的角度指出,法人的本质是对财产的法律调整,丝毫不涉及个人身体体格的调整。<sup>[11]</sup>这说明诸多学者已经注意到民法上两种人格的存在,只是未对民法典的人法体系做进一步的延伸论述。

伦理人格与技术人格的区别在民法体系中体现在如下方面:一是两者所依赖的人格原型不同。伦理性人格针对的人格原型是具有生物性、伦理性和社会性的自然人,并调整基于上述特征而形成的亲属关系和人格权关系。而法人针对的对象原型则是围绕特定财产形成的社会组织、财团甚至单一的独立财产,并在此上形成了抽象的符号化人格,从而完全脱离了人的生物性形象,进入了彻底非个人化的领域。亲属法上的自然人与法人以何种共性能同时成为民事主体,至今在理论上仍难以得到完整解释。二是两者在立法上被抽象的性质和结果不同。基于法律的普遍性,亲属法人格或人格要素会被适当地抽象予以分类,如亲权上的父子、人格权法上的身体、肖像诸要素等,但始终保留自然人的社会生物性特征。而法人则不以自然人为立法抽象化目标,而纯粹是对财产的形式化抽象,凭空创造出一个“符号”化的人格,从而实现了从“自然人”向“非人”、从“生物人”向“观念人”的转变,它与自然人分享的共同点仅在于“权利能力”,经过技术上的过滤和抽象,自然人与财产实际上处在同一地位。这种理论上的跨越在统一的人格理论下也很难得到充分解释。<sup>[12]</sup>三是两者的价值基础不同。民法上的主体概念具有强烈的价值导向,它遵从近代欧洲法律对于自然人的平等、理性和自由等基本价值的倡导和显扬,彰显自然人的“目的性”价值,在公私法上分享同样的价值意义。而法人则作为被形式化的抽象主体符号,纯然由实证法创造,表征的是相对空洞的“手段性”价值,这种技术“拟制”产物仅具有私法上的形式化意义,并不能适用于公法领域。究言之,这也是德国民法将人

[9] 参见李永军:“民法上的人及其理性基础”,《法学研究》2005年第5期,第16页。

[10] 参见(德)迪特尔·梅迪库斯:《德国民法总论》,邵建东译,法律出版社2000年版,第20、778页。

[11] 参见李永军,见前注[9],第23页。

[12] 参见梅夏英:“民事主体的抽象度问题及其对私法体系的影响”,《经贸法律评论》2019年第1期,第92页。

格、主体等价值性概念由技术性概念“权利能力”统一替代后的结果。因为“权利能力”是一个功能性的概念,一个实体只要能获得权利和承担义务,就可能成为民事主体;而非可以反过来理解,即先有主体再有权利享有问题,自然人的主体价值在此并未被优先考虑。

除了上述主要表现外,两种人格区分还体现在民法体系中人身法与财产法的显著分野上,即人身法并不能当然适用民法典抽象的形式理性和概念系统,如果将亲属法、人格权法和继承法从民法典中抽离,也丝毫不会影响民法典内外逻辑体系的完整。另外,以身份和人格为前提的人身关系也不能完全适用“权利能力”和“法律行为”的概念规则系统,这两个概念主要是对财产主体能力和财产行为的抽象。

在了解了两种人格区分在体系上的价值后,我们在理论上依然应该承认法人等财产主体在法律上的正当性和科学价值,只是应当将民法上的相关主体合理地归入伦理主体和财产主体两大系统中,并适用不同的规则体系。这种区分会给我们解决传统主体问题提供新的视角和契机,它的理论意义体现在如下方面:

首先,传统民法人格理论没有对自然人的两种人格进行区分,遮蔽了两种人格区分的体系价值。自然人在人身法和财产法上人格的显著差别,因在民法上被归入统一的“自然人”范畴而被强制熨平了,这极大地淡化了人格区分在理论和法律适用上的意义。实际上,自然人在亲属和人格关系中的人格属性,与自然人在与法人财产交往中的属性迥乎不同,按照主体对等原则,一为人伦主体,一为形式化的“符号”主体,两者完全无法被通约为一种身份。尽管两种主体都体现为自然人,但如果我们想象一下一个在家庭环境中的自然人与一个远在电商互联网另一端以卡通形象示现的自然人,就可想见两者的巨大区别,也可了解两种交往的民事规则系统的不同。

其次,两格人格区分可以在理论上明晰民法上多种“主体”的讨论范畴。除了自然人和法人外,民法存在多种“中间主体”或“第三主体”,如动物、胎儿(甚至胚胎)和非法人团体等,过去我们都将其纳入统一的主体理论予以定位,讨论其是否具有民法上的“主体”资格或“权利能力”,实则适用的是抽象的、形式化的财产法主体标准。实际上如果理论上接受两种人格的区分,则可以将上述主体归入不同范畴:如对于胎儿和动物而言,不存在“主体”或“权利能力”问题,法律保护他(它)们的目的是基于人伦关系或生物种群伦理,来善待同类或特定生物种群,与财产并无直接关联,并不需要适用财产法主体规则,直接适用由人伦关系或种群伦理决定的行为规则即可。<sup>〔13〕</sup>当前理论上对于胎儿(以及胚胎)、动物等主张的“主体说”“非主体说”和“有限主体说”等都如出一辙,难以自圆其说,均是用错了人格范畴的缘故。<sup>〔14〕</sup>对于第三主体(如合伙)或非法人团体(如筹建中的法人)等,因主要涉及财产关系,则应将其归入民法上占主导地位的抽象财产人格范畴进行讨论和定位。

最后,人格区分理论对于界定特定主体的人格序列及立法模式亦具有重要的意义。民事

〔13〕 参见许可、梅夏英:“一般人格权:观念转型与制度重构”,《法制与社会发展》2014年第4期,第96页。

〔14〕 同上注,第96—99页。

主体存在一个明显的序列性规律,即从亲权到股份有限公司法人之间存在一个伦理性递减和财产性递增的趋势,其中伦理人格和财产人格的转换是在自然人身处不同的民事关系场景中不经意间完成的,如民间所言的“亲兄弟,明算账”即为适切场景。由此我们可以根据主体所处序列位置(以“抽象度”为表征)来确定特定主体适用的私法规则领域。如对于人格权而言,其主体人格仍属于伦理性人格的范畴,只是其伦理性较亲权稍弱,但与财产性人格有本质的区别,故人格保护无法在传统“人法”中解决,因为民法中以“权利能力”为基础的核心人格理论主要针对的是财产法人格,无法调整伦理性的人格内容。<sup>[15]</sup> 同时亦不能通过侵权法来获得完全的保护,因为侵权法解决不了人格利益的法律内涵和人格要素法定的问题,故通过人格权法来调整是较为妥当的方法。其他如环境法学界一直有学者倡导的“环境法人”等,由于其针对的主要是财产责任和赔偿受益对象问题,故应归入财产人格序列讨论。<sup>[16]</sup> 关于不同民事主体的序列定位,下文将会进一步涉及。

上述关于伦理人格和财产人格相区分的讨论,对于人工智能法律主体地位的探讨亦具有重要意义,并将构成未来研判人工智能法律地位问题的理论框架。值得注意的是,人工智能在主体领域正在同时沿着两种人格路线展开,并展现出不同的理论意趣和走向,对此下文将分别进行阐述。

## 二、伦理性人格:人工智能作为新的智能“种群”?

当前人工智能的快速发展,客观上激发了当代社会对于人工智能具有类人智慧,甚至越过“奇点”成为超级人工智能的想象。<sup>[17]</sup> 过去的几十年里伦理学和社会学对此进行了大量的前瞻性研究,而法学界则相对保守,对于人工智能能否成为具有人类意识的独立主体一直持审慎态度,其中,人工智能是否具有人类理性或自我意识又是重要关切点。<sup>[18]</sup>

### (一) 理性抑或自我意识?

人类具有与生俱来的理性能力是近代法律建立的重要理论前提,它决定了法律主体固有的理性基础和道德实践能力,近代私法和经济学均将理性人假设作为社会主体的理论基础。当前人工智能是否具有人类的理性与能否成为伦理主体具有直接关系。从目前的智能程度上看,人工智能已表现出类似人类的“理性”能力,如自 2016 年谷歌开发的“AlphaGo”战胜韩国选手李世石之后,ChatGPT 生成式大模型系列在应对人类重要考试中陆续超过人类,据称新

[15] 参见刘银良:“从三维人格同一性探究人格权的客体”,《中外法学》2024 年第 5 期,第 1220 页。

[16] 参见李光禄、郑德光、陈国申:“环境法人的构建”,载国家环境保护总局武汉大学环境法研究所、福州大学法学院主编:《探索·创新·发展·收获——2001 年环境资源法学国际研讨会论文集(上册)》,山东科技大学应用法学研究所 2001 年版,第 129—131 页。

[17] See Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, New York: Viking Penguin, 2005, p. 203.

[18] 参见杜严勇:《人工智能伦理引论》,上海交通大学出版社 2020 年版,第 28—37 页。

近发布的 ChatGPT-4o1 在“推理”方面已经达到博士能力水平。<sup>[19]</sup> 这种计算智能还将在算力、算法和数据的增强和优化下继续发展进化。理性能力在传统理论中是相对于非理性行为（如情绪、欲望等）而言的，它体现在对客观事物的感知、辨识以及推理和选择能力。就此而言，计算智能似乎符合人类理性的表面特征，它们均以冷峻的、一丝不苟的程序执行者著称，但以此认定人工智能具有人类理性还远远不够。其原因在于，人类的理性能力可以自然延伸至道德能力上，对此人工智能尚不具备；另外，人类的理性能力意味着人类能够内观反省自身，认识到自己的存在，并通过调整来完善自身，这一点当前的人工智能亦不具备。由于人工智能算法由人类设定，也是人类理性的延伸，故“理性”标准在衡量人工智能主体性上并不是一个好用的概念，实际上按照笛卡尔的“我思故我在”标准，主体只有在意识到自己思考时，才能成为理性人，法律真正应该关注的其实是人工智能是否具有“自我意识”。

关于人工智能可否具有自我意识，目前理论界和产业界对此分歧甚大。如西方人士霍金、马斯克以及加拿大学者杰弗里·辛顿(Geoffrey Hinton)等对人工智能可能进化为超过人类的自主意识体表示深切的担忧，认为人类对此应高度警惕，提前谋划应对。英美学者如罗杰·彭罗斯(Roger Penrose)、诺姆·乔姆斯基(Noam Chomsky)以及燕乐存(Yann LeCun)等则认为人工智能离具有自我意识距离尚远，甚至在理解、想象、直觉和创造力上永远达不到人类意识的水平。<sup>[20]</sup> 否定的立场大致体现如下：一是人类意识并不能通过计算获得。基于库尔特·哥德尔(Kurt Gödel)的自指理论以及“理发师悖论”，图灵认为计算机并不能完全判定自身是否会“停机”，这就构成算法的根本逻辑局限，类似于人不能反观自身，而反观是自我意识的基础。彭罗斯则认为意识不是一个机械的计算过程，而是源于大脑微管中的量子过程，量子状态在量子引力的作用下坍塌，这一瞬间的变化就是我们意识体验的来源。<sup>[21]</sup> 二是意识并不完全来源于大脑。有学者认为，意识没有存在于大脑中的生理基础，甚至记忆也不存储于大脑中。<sup>[22]</sup> 乔姆斯基则认为语言能力是人先天的能力，人类意识只是处于语言思维的表层，并不完全由基因进化决定，当代的机器学习依赖于暴力计算，并没有真正的思维和意识。又如彭罗斯认为意识不仅仅是大脑的产物，它还于宇宙的基本结构有深刻的联系，意识在某种程度上影响着现实的结构，并意味着主观现实和客观现实之间的界限变得模糊了。<sup>[23]</sup> 三是机器学习只是软件的暴力计算，目前看不到它会发展到人类意识的迹象。如燕乐存认为，目前的生成

[19] See OpenAI, “Learning to Reason with LLMs,” <https://openai.com/index/learning-to-reason-with-llms/>, last visited on 25 November 2024.

[20] See Roger Penrose, *The Emperor’s New Mind*, Oxford: Oxford University Press, 1989; Noam Chomsky, “The False Promise of ChatGPT,” <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>, last visited on 25 November 2024.

[21] See Penrose, *supra* note 20, pp. 552-562.

[22] See Penfield Wilder, *The Mystery of the Mind*, New Jersey: Princeton University Press, 1975, p. 114.

[23] See Stuart Hameroff and Roger Penrose, “Consciousness in the Universe: A review of the ‘Orch OR’ theory,” *Physics of Life Reviews*, Vol. 11, No. 1, 2014, pp. 39-78.

式大模型尚缺乏人类基本的推理、想象和长久记忆能力。亦有一些学者认为,大型神经网络算法只是面对过去的事实数据,无法提供面向未来的想象力和创造力,算法仍无法脱离人类意志的支配。与此相反,肯定的立场则更多地集中在工程学上,以探讨意识与智能、体验、情感等的区别和上述事物的生成机制为核心,并形成了丰富的实践性理论和流派,在此不一一赘述。

从认知科学、脑科学和量子科学等角度探讨人类意识问题,属于纯粹科学探究,但对于法学研究人工智能主体问题却是必要的。因为人工智能“意识”问题是当前人工智能领域的新关口,它直接决定了人工智能未来的发展趋势和角色定位。虽然目前关于人工智能是否具有自我意识尚未有定论,但随着科学的发展并非完全不可能,只是生物意识和硅基意识有可能分享不同的技术路线。目前人工智能因快速发展而呈现出来的高度“自主性”和“智能性”,让人们们对大型数字复杂系统涌现出类似“意识”的事物充满期待。当然在目前人工智能未具备人类意识的情形下,法律还无法将其直接赋予伦理人格,但以深度学习为标志的具身机器人已然出现在社会生活的各个领域,并为我们提出了有关科技伦理和人机伦理的现实问题。

## (二)科技伦理和人机伦理

在人工智能是否具有类人意识尚不确定的情形下,当今人工智能发展的伦理治理已经被提上日程。人工智能伦理相对于法律而言,更多地是鼓励科技向善,促使业界制造出可信的、安全的以及负责任的人工智能。无论是人际间的伦理规范,还是种群间、人机间的伦理规范,其核心要义均在于践行对人类、生物种群以及特定智能体的尊重,大多数伦理规范无法直接上升到强制性的法律层面。就人工智能伦理而言,目前存在着两个维度的伦理要求:一是约束人类开发行为的科技伦理,一是关于人类如何对待机器人,并与机器人的和谐相处的人机伦理,以下简称之。

人工智能的科技伦理目前仍以人类中心主义为依托,旨在避免人工智能对人类和社会造成伤害。在谷歌公司将“别作恶”(Don't be evil)作为公司的座右铭之后,当今较有影响力的人工智能伦理为一群美国科学家倡导的《阿西洛玛人工智能准则》(2015—2017),该准则共 23 条,最后 5 条明确体现为对超级智能的预知和应对。<sup>[24]</sup> 近几年国际组织如欧盟、美国电气工程协会(IEEE)以及各大公司等提出了自身的人工智能道德框架,中国也于 2019 年发布了《人工智能北京共识》。上述倡议存在的一个共同点是可操作性甚弱,与技术实施并不同步。如美国科幻作家阿西莫夫在 1939 年即提出了“机器人三定律”,成为当时对人类和机器人保护的基本理念,但事实上人工算法在技术层面上并不能达到上述要求,如仅仅就判断是否对人类造成“伤害”而言,人工智能就不能在编码中考虑到所有的相关因素,并选择唯一的行为来达到目的。“电车难题”则又是另一个伦理命题,它要求人工智能具备一定的道德实践能力,但法律对人类通过考试获得驾照都不做如此要求,要求人工智能解决上述问题离现实生活过远,且这一伦理难题并没有统一答案,发生的机率甚微。实际上基于对不确定未来的设想,科技伦理目前承载了太多口号性的内容,在发展阶段对人工智能要求过高,如要求人工智能具有道德实践

[24] See Future of Life Institute, “Asilomar AI Principles,” <https://futureoflife.org/open-letter/ai-principles/>, last visited on 25 November 2024.

能力等,这是不符合目前行业发展实际的。<sup>[25]</sup>

针对上述问题,一个合理的做法是先确定人工智能可能对人类带来的诸种负面影响,再通过技术上的要求来对科技伦理予以具体化。当前人工智能对人类社会带来的负面影响体现在如下方面:①人工智能带来的失业问题;②算法异化问题,即算法改变人的工作性质,使人成为算法的奴隶;③人的基本权利受侵害的风险,这主要体现在人类通过各种风险评估工具(如危害评估和预测警务工具等),在决定是否对人实施法律手段时代替人做出判断,侵害基本人权;④机器人杀手问题,即用当代无人机等作为自主杀人武器,用于战场和犯罪活动;⑤算法偏见问题,即由于数据集的偏差以及设计算法的人本身的偏见,导致算法输出的结果对特定人群造成不公平的伤害;⑥人工智能造假,当前的算法通过自动化推荐为用户构造了一个与现实不同的“信息茧房”,为用户制造了一个虚假的世界,同时目前也存在一些软件通过“深度合成”技术制造出假视频、假新闻,制造社会舆情混乱。上述负面影响是目前人工智能应用带来的实际后果,除了失业问题外,这些后果都是通过算法来实现的,而人是算法的设计者,故有效的科技伦理应主要侧重于算法设计者的伦理约束上。具体表现在:①解释义务,即算法设计者需要对算法针对他人的决策和行动做出合理解释,但这在机器深度学习(尤其是无监督学习)时代存在挑战;②设计人责任,即人工智能目前尚不能成为道德智能体,软件本身并不能被追究责任,与算法相关的责任只能由设计者承担;<sup>[26]</sup>③透明度原则,即算法所使用的数据来源是明确的、可获取的,算法本身对用户而言也是清楚明晰的。<sup>[27]</sup>总之,通过对算法的伦理约束,机器人造成的风险和危害能得到最大限度的抑制。

人机伦理则是关于人机如何共处的伦理问题,是人类历史上前所未有的人与机器之间的良性互动和信任达成问题。基于深度学习技术,具身机器人成为未来人工智能发展的主流方向,亦是人机伦理着重关注的对象。伦理学界较早地涉及到“机器人权利”这一领域,他们重点关注机器人是否可能享有某种被尊重的权利,不被人类滥用、奴役或虐待。<sup>[28]</sup>具体而言,在情感机器人、助老机器人和电子宠物的使用情境中,涉及到人的情感依赖问题,甚至有可能动摇传统社会的两性关系和婚姻关系(如情侣机器人)。虽然机器人是否享有“有限”的道德权利目前尚缺乏确定的理论基础,但在主客支配关系前提下,我们还是能够提供两个正当性理由:一是由动物享有的“道德”权利带来的启示,一是机器人所具有的“高度”智能,这两者分别在“生命”和“智能”上与人类相似。在目前机器人智能尚未达到强人工智能的阶段,针对机器人伦理的研究实际上仍处于人类中心主义的话语体系中,其本质仍然是在规范人类自身,引导人

[25] See Lorin Brennan, “AI Ethical Compliance Is Undecidable,” *Hastings Science and Technology Law Journal*, Vol. 14, No. 2, 2023, pp. 311-338.

[26] 参见王乐兵:“自动驾驶汽车的缺陷及其产品责任”,《清华法学》2020年第2期,第97页。

[27] See Joshua A. Kroll, “Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems,” *In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, New York: Association for Computing Machinery, pp. 758-771.

[28] 参见杜严勇,见前注[18],第28—30页。

类善待机器人,并给机器人足够的发展空间。如果未来真有具有“自我意识”的超级人工智能出现,那么人机伦理的叙事方式就会改变,届时人与机器的支配关系将会逆转为两个智能“种群”的关系,新的“社会”契约将会以无法预知的方式上演。

### 三、技术性人格:责任主体和权利主体

技术性人格指的是人工智能在财产领域的人格表现,它主要是为了在技术上解决人工智能带来的责任承担和权利归属问题。这些问题目前已经大量发生,并引起了学界和业界的普遍关注,主要集中在智能助手、机器人侵权领域以及机器人创作领域。目前学界对于机器人主体地位的研究主要体现在财产法领域,鲜少涉及对于伦理性人格的诉求。倡导机器人享有某种法律主体地位的法律主张是基于一个共同的、特定的技术事实,即因机器人显著的“自主性”而产生的“自动化风险”,它的行为后果已经不能被自然人的“过错”规则所涵盖。<sup>[29]</sup> 基于此,国外有专家提出了赋予机器人主体地位的两难困境:一方面,将机器人当作“主体”缺少事实基础,因为目前机器人确实不能算是“人”;另一方面,一旦把机器人当作主体,又会面临机器人是否享有自然人的公法权利、亲权和人格权等问题,显然又失之过宽。<sup>[30]</sup> 这种“两难”困境实际上并不确定存在。如果我们能够按照上文所提出的两种人格区分理论进行理解,其并不会成为理论难题,因为在财产法上通过拟制赋予某种实体以“人格”,并不一定要求该实体是自然人;同时该实体具有某种形式化的财产人格以后,也不一定享有以生命形式存在的自然人的伦理人格权利。将这种主体资格定位于纯粹技术性财产人格领域就可以顺利解决上述问题,这也是本文反复强调两种人格区分的重要原因。人工智能的技术性人格目前体现在两个场景,一是对于责任分担主体的诉求,一是对于人工智能作品的权利主体地位的探讨,两者在理论逻辑上并不重合,亦不属于同一技术人格体。

#### (一) 责任主体

在侵权责任认定上,如果依照传统法律能够追究设计人、销售者或使用人的责任,就不需要机器人成为责任主体。如上文所述,机器人独立承担责任乃源于不可控制的“自动化风险”,它是在与机器人相关的民事主体不存在过错的情形下产生的。这里的“责任主体”的在民法上应属于主体的“行为能力”范畴,“行为能力”在民法上既包括从事合法行为的能力,又包括对不法后果承担责任的能力,责任能力属于后者。但机器人是否同时享有“权利能力”,按照传统理论似乎应当做出肯定的回答,因为在民法上权利能力是行为能力的前提,但在人工智能上却不一定如此,这也反映出高度自动化智能给法律带来的理论挑战。以下以智能合同和机器人侵

[29] See Kylie Porter, “Shaping the Future of Sustainable AI and Automation: Why Human Rights Still Matter,” *Human Rights Defender*, Vol. 28, No. 1, 2019, pp. 33-35.

[30] 参见(德)扬·埃里克·施默:“人工智能与法律人格:部分权利能力的引入——德国法上的部分法律地位”,载(德)托马斯·威斯迈耶、蒂莫·拉德马赫编:《人工智能与法律的对话 2》,韩旭至、李辉等译,陈吉栋等校,上海人民出版社 2020 年版,第 736—756 页。

权为例来做简要分析。

智能助手既包括如亚马逊的“Echo”或苹果的“Siri”等虚拟助手，也包括机器人管家等具身机器人，它们的共同特点是能够为主人从事诸如网络预订、购物等民事行为。智能助手在实际工作中带来的问题是，它常常会给使用人带来不必要或“错误”的订单，这就涉及到合同的效力问题。一般来说，机器人只是听从人类的指令行事，所以在法律上没有必要将其作为合同主体，但交易却是由机器人操作完成的，那么在民法上就存在是否可以适用代理关系的问题。<sup>[31]</sup> 在代理关系中，意思表示的传达和接受都要求双方具有法律主体必备的“意思能力”，但智能助手除了按指令和程序行事外，并没有超出特定事项的“意思能力”，另外即使智能助手能成为代理人，也无法承担无效代理的法律后果，故赋予其主体地位显然用力过猛了。对于自动化和程式化交易，我们很难以“错误”或“误解”为由来对交易相对人主张合同效力存在瑕疵，除非明显存在其他的指令人如“黑客”或“幼儿”等。故依合同主体和代理人来界定智能助手的地位，都不能与传统民法理论很好地融合。实则这里面涉及到的根本问题在于，用户对于因智能助手不合预期的交易造成的损失由谁承担？对此传统的法律无法自行解决，从而形成了新的“责任漏洞”。<sup>[32]</sup> 有德国学者反复提出通过“部分权利能力”来解决机器人代理问题，但主体的此种界定亦无法解决责任承担的问题。<sup>[33]</sup> 实际上由智能助手来承担合同责任目前看来是不可能的，在针对机器人的新的责任模式被创造出来以前，只能由用户自己承担。

侵权责任则是学界和业界更为关注的问题，因为机器人造成的损害在可预期的范围内日益增多，如机器看护有可能对家庭成员造成人身或财产损害，医疗机器人会造成自主范围内的医疗事故，自动驾驶亦会因交通事故对乘客或行人造成伤害等，都会产生新的侵权责任问题。<sup>[34]</sup> 这方面的案例和讨论颇多，本文不作具体分析。总的说来，基于“自动化风险”带来的侵权风险，国内外学界倾向于通过“替代责任”和“危险责任”来追究侵权责任。<sup>[35]</sup> 前者主要是类比监护责任和雇主责任，后者主要是类比动物责任和高度危险责任，基本上都是通过严格责任来防范自动化风险。但替代责任面临的障碍是，人工智能并不能像未成年人或雇工一样，

[31] See Roman Dremluiga, Pavel Kuznetsov and Alexey Mamychev, “Criteria for Recognition of AI as a Legal Person,” *Journal of Politics and Law*, Vol. 12, No. 3, 2019, pp. 105-112.

[32] 参见杨旭：“法律与法的评价互动——克劳斯·威廉·卡纳里斯《法律漏洞的确定》述评”，《经贸法律评论》2023年第6期，第140页。

[33] 参见施默，见前注[30]。

[34] See Daria Onitiu, “The Limits of Explainability & Human Oversight in The EU Commission’s Proposal For The Regulation on AI—A Critical Approach Focusing on Medical Diagnostic Systems,” *Information & Communications Technology Law*, Vol. 32, No. 2, 2023, pp. 170-188.

[35] 国内学界就替代责任所做的论述，可参见吴汉东：“人工智能时代的制度安排与法律规制”，《法律科学》2017年第5期，第138页；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”，《法律科学》2017年第5期，第170页。就危险责任，可参见张建文、贾章范：“《侵权责任法》视野下无人驾驶汽车的法律挑战与规则完善”，《南京邮电大学学报（社会科学版）》2018年第4期，第28页。

他们的行为可被充分预见并可以被有效控制,它涉及的主体和环节过多,很难通过界定自然人的“注意”义务来确定合适的责任主体。另外,这里还要首先解决一个人工智能自身是否已经具有某种“人格”这一前置判断问题。“危险责任”面临的困难则在于,人工智能不似“动物”具有独立生命且不可控制,而是由人类设计创造并服务于人类,实际上它在应用中并不“危险”,其安全性甚至普遍高于人类(如无人驾驶),采用危险责任会对技术创新产生负面影响,并且面临风险转嫁、保险费用高昂等问题。<sup>[36]</sup> 严格责任对于人工智能来说并非灵丹妙药,其原因在于,作为当代智能工具,人工智能总体上仍然由人类设计和支配,且处于技术进化过程中,它产生的一些不可预知的风险,也只能通过技术进步来解决,责任追究应朝着促进技术进步的方向发展;通过强制保险和赔偿基金池等完全由社会分摊风险的做法,无助于促进开发者通过技术进步来减少甚至消除风险的发生。基于此,有学者提出人工智能侵权责任仍应坚持传统侵权法的原理,通过产品责任规则来追究开发者的侵权责任,通过过错规则来追究使用者的侵权责任。这种理念在目前情况下仍然具有一定的合理性。<sup>[37]</sup>

现在回到人工智能的财产性人格问题。赋予人工智能财产性人格完全是另外一个思路,它意图通过机器人的独立责任使其他主体从侵权责任中解脱出来。这种思路仍然是通过类比“自然人”或“法人”的财产性人格来确立一个新型的机器人“人格”。在技术上拟制一个法律人格并没有理论上的障碍,但跟伦理人格不同,技术人格具有强烈的目标导向,即法律通过某种程度上不存在的“人格”来解决财产问题。显然机器人的财产人格并不能做到这一点,具体原因在于:一是独立的机器人并没有社会交往的独立意思表示能力。机器人不似法人,法人虽然亦没有“法人”自身的意思表示能力,但它可以通过内部治理机制中自然人的组合产生出意思表示能力。当前的弱人工智能尚不能脱离人类指令自由交往,并形成自主财产关系。二是单个的机器人没有独立财产和责任能力。由于机器人缺乏伦理人格,故不能获得属于自身的财产,法律上设计的最低资产要求或者强制责任保险都是由相关参与者(生产者、所有者或使用者)贡献的,机器人并没有实际承担任何责任。三是机器人的财产人格无助于促进技术进步来预防损害的发生。即通过机器人独立承担责任将开发者从责任中解放出来,不仅无助于技术开发者通过技术进步来预防损害的发生,而且对机器人的未来重复侵权行为毫无作为。四是除了财产责任外,法律无从对机器人直接追责和惩罚,比如包括法律上的停止侵害、消除危险、赔礼道歉等救济手段,以及刑事责任的追究等,都无法实现。<sup>[38]</sup> 总而言之,目前探讨赋予机器人财产性人格尚为时过早,机器人给我们制造了一个同时兼具伦理人的形象和工具人的功

[36] See Buiten Miriam, de Streef Alexandre and Peitz Martin, “The Law and Economics of AI Liability,” *Computer Law & Security Review*, Vol. 48, 2023, pp. 1-20.

[37] See Karni A. Chagal-Feferkorn, “Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers,” *Stanford Law & Policy Review*, Vol. 30, No. 1, 2019, pp. 61-114.

[38] See Kathryn Bosman Cote, “Outsmarting Smart Devices: Preparing for AI Liability Risks and Regulations,” *San Diego International Law Journal*, Vol. 25, No. 1, 2024, pp. 101-140.

能的综合性幻象,使我们时常在两者之间来回跳跃,但可以肯定的是,在具有自我意识的人工智能出现之前,为一个单独的、没有“自然人”和“自身财产”依托的人工智能机器赋予技术性人格,可能性甚微。

## (二)权利主体

上文主要是从承担侵权责任的角度,论述了机器人作为责任主体的理论困境。与此同时,当前在生成式大模型领域却存在着人工智能的权利主体问题,即 ChatGPT 系列可否对其生成出来的文字、视频等作品享有知识产权。此处的“权利”主体与上文中的“责任”主体并不属于同一语境,前者蕴含的“创作者”带有较强的主体性人格色彩,且以权利能力为基础;后者则完全以责任承担为导向,仅作为侵权责任风险化解的工具存在,不以权利能力为基础,但在人格区分中都属于技术性人格范畴。目前关于人工智能著作权的理论探讨主要集中在两个方面,即作品的“可作品性”和人工智能的“可主体性”,这两个方面紧密联系且相互影响,实际上可归结为同一个理论基础,即是由主体性决定作品性,还是由作品性决定主体性。限于篇幅,下文简述之。

就第一个问题即人工智能生成的作品是否构成著作权法上的“作品”而言,目前在理论上赞成者居多。其主要理由在于,人工智能生成的作品具有独创性和艺术审美价值,符合法律上对作品的基本要求,通过著作权保护可以达到鼓励创新的目的;<sup>[39]</sup>反对的理论较少,主要还是认为创作是专属于人的活动,给机器生成作品赋予著作权会陷入“有成果必有作者”之谬误。<sup>[40]</sup>从客观形式上看,机器生成的作品无疑具有与人类作品相似的稀缺性和社会价值,甚至“艺术”水准还要高于人类,故问题的焦点不在于“作品”本身,而在于是否可纳入著作权范畴予以保护。显然,从外在作品形式角度就认定机器生成的作品属于著作权法上的“作品”,有些过于简单了,它给我们带来的挑战应在“作品”之外,体现如下:

一是对作品予以著作权保护是近代印刷术和工业革命兴起后的产物,它具有特定的历史和技术背景,在历史上的绝大多数时期人类的创作作品并没有受到法律保护,而是由人类成员自由分享。二是近代以来法律对“作品”的保护主要是激励创新,促进艺术文化的繁荣,但事实上著作权法并没有做到这一点。尽管法律强调“作品”的独创性不在于客观的和历史的创造性,而在于基于人类心理和思维个性的“心理创造性”,但很显然当今著作权几乎对所有个人创造的作品(包括正式和非正式、生活和职业、有价值或无价值等)都予以保护,没有任何客观价值的“门槛”,其激励创新的功能让人怀疑。<sup>[41]</sup>三是如果认为“非人”创造的作品亦成为著作

[39] 参见易继明:“人工智能创作物是作品吗”,《法律科学》2017年第5期,第137—147页;吴汉东:“论人工智能生成内容的可版权性:实务、法理与制度”,《中国法律评论》2024年第3期,第118—120页;曾田:“AI时代著作权独创性标准适用问题研究”,《经贸法律评论》2024年第4期,第142页。

[40] 参见王迁:“三论人工智能生成的内容在著作权法中的定位”,《法商研究》2024年第3期,第196—197页。

[41] 参见丁晓东:“著作权的解构与重构:人工智能作品法律保护的法理反思”,《法制与社会发展》2023年第5期,第116—118页。

权法上的作品,那么可以想象,大自然为人类创造了无与伦比的自然作品,动物亦为人类创造了丰富的生态作品,并成为人类在观光旅游中欣赏的对象,是否亦应为之设定一个“作品”和“作者”呢? 四是通过“作品”保护机器人生成物,并不能有效维持目前的人类作品保护。它给我们带来的启示是,机器人以大规模的“创作”取代了人类通常水准的创作,并使人类只有通过高水准的创作才能获得社会认同和经济利益,这就如同传统社会的艺术家和工匠,近代机器化生产取代了大部分工匠的工作,艺术家只有专注于独特性和高水准的创作才能获得社会的青睐。除此之外,尚有作品法律保护可能性的问题,如传统版权是通过出版环节的版税来实现经济利益的,互联网时代这一环节很大程度上被电子化了,这意味着作品通过类似纸质时代“锁定”作品利益的方式会被无差别的网络分享所冲击,对于人工智能作品而言,它在系统里随时大规模地生成、复制和流动,作品的确权和利益实现缺少稳定可靠的“锁定”方式,作品会惯性地溢出到公共领域。基于以上考虑,通过所谓的“作品”来决定“主体”的方式并不完全符合社会和生活常识,甚至有些本末倒置。

就第二个问题即人工智能是否可以成为作品主体而言,由于上文已对机器作品的“可作品”性进行了质疑,故主体性问题相应地也不会存在。但为了使问题更清晰,不妨也从主体角度来做一番设想。设若我们将人工智能作为创作主体,它会面临以下困难:一是人工智能在目前的技术条件下无法自主创作,它是人类设计出来的智能工具,并接受人类的指令自动生成作品。二是人工智能目前缺少自我意识、想象力和直觉,它生成的作品可能具有历史创新性,但绝对缺少心理创造性,而后者是“作品”的灵魂。<sup>[42]</sup> 三是人工智能作为创作人,与著作权服务于人际域的初衷相违。著作权是人类的创造并服务于人类,它调整的是人际关系中的创新激励和利益分配,也是私法在近代的权宜性创造,它不似财产和人格具有历史纵深感,而是可供取舍的技术性的权利。故著作权理论不能改变自身小口径之外的大口径问题,有关机器人人格和财产的相关问题应当在基础法理层面上予以深入探讨。四是人工智能作为创作人,并不能实实在在地享有和行使权利,亦不能因此承担责任。

基于此,学界将目标转移至人工智能的用户上,以论证用户取得作品著作权的合理性。关于此问题的讨论甚多,在此仅简要陈列几项主要理论障碍:一是用户的创作行为并没有被有效坐实。将通过指令、提示以及参数设定等方式由机器人生成的作品,归之为用户的创作,失之勉强,为众多案例和学者所不认同。<sup>[43]</sup> 二是人工智能生成作品,取决于投资人、开发人、所有人和用户等多个主体的共同贡献,其中开发者的贡献应占主导地位,故人为设定用户为主体失之武断。人工智能是复杂的技术综合体,我们无法在错综复杂的“主体”丛林中任意指定一个

[42] 参见欧阳友权:“人工智能创作的艺术隐忧和伦理边界”,载《光明日报》2024年7月20日,第9版。

[43] 参见菲林律师事务所、百度网讯科技公司著作权侵权案,北京市互联网法院民事判决书,(2018)京0491民初239号。该案法院认为原告利用计算机软件智能生成的内容不是通过自己的智力劳动创造获得,原告不享有著作权。参见王迁:“论人工智能生成的内容在著作权法中的定性”,《法律科学》2017年第5期,第155页。

创作人。三是如果指定用户作为著作权人,那么他应同时承担与作品相关的责任。如果机器人作品中涉及到侵害他人著作权的行为,用户作为创作人就应当应承担相关责任,试问这种情况该如何处理?其他不一一赘述。

基于上述分析,当今关于机器人作品及主体问题的探讨,很大程度上受传统思维惯性的影响,即只要存在“作品”,就会有“主体”存在,反之亦然。人工智能的“作品”和“主体”问题实则为同一问题,一项不成立应当随之导致另一项失效。故人工智能创作是一个全新的领域,不一定受传统规则的辖制。我们忽视了除了受现代著作权保护的作品外,大量的天然和人工作品自始以来即由人类共同自由分享,时代的发展亦促使人类在人工智能时代创造出新的创作激励机制,而非将有益于人类的机器人作品人为地纳入现有权利框架,以迎合传统社会对所有的所谓人类“作品”予以平面化和普惠化保护的最低水平需求。这一部分主要探讨的是人工智能能否成为“权利”主体的问题,此处以著作权为例所做的分析,亦大部分适用于其他“权利”场合。总结本部分的论述,一言以蔽之,将人工智能作为技术人格中的责任主体和权利主体,目前并没有任何成功案例和成功征兆,除非具有自我意识的强人工智能出现。

#### 四、未来人工智能两种人格融合的“抽象度”分析和展望

通过上文的分析,我们大致了解了目前赋予人工智能两种人格面临的理论和实际困难,但这并不意味着未来一直如此。随着智能科学的发展以及人类对意识大一统理论的探索,人工智能成为新兴的、具有自我意识的智能体并非完全不可能,故我们应以开放的眼光看待人工智能的未来。基于这种愿景,在此我们可以尝试对强人工智能时代的机器人人格作一个前瞻式的预想,即假设人工智能具有真正类人的智能,那么它将具有何种法律上的人格形态,以及两种人格将以何种方式融合并呈现。这就涉及一个传统民法上的人格融合问题,主要体现在自然人领域。通过第一部分的分析,我们了解到民法上实际存在着泾渭分明的两类人格,但由于民法理论和结构并未对此进行区分,加之亲属法和人格法体量较小,两种人格被人为地统一于具有财产法人格底蕴的“权利能力”之下,致使两种人格在自然人上呈现出一种融合或纠缠的状态,并体现为一种位于伦理性 and 财产性之间的序列人格体,其中法律对于自然人的“抽象度”成为度量人格具体形象的重要指标。基于此,未来人工智能是否也会呈现同样的特征,以及呈现何种人格融合状态,值得我们思考。以下就传统民法人格和人工智能人格的抽象度问题做一简述。

##### (一)传统民法基于“抽象度”形成的人格序列递进现象

此处所称人格的“抽象度”,指的是民法上在对于人格进行抽象形成具体人格的过程中主体的被抽象程度。从传统民法体系上看,人法部分看似单一,实则呈现出从家庭身份到法人直至商法上的上市公司等一系列人格系列,从亲权到上市公司,其主体形态已判若云泥,其中决定这种人格序列顺位的是私法对主体的抽象程度。它意味着从亲属关系到上市公司经历了法律一次次的抽象,将自然人从现实中的家庭成员逐步抽象至普遍性的人格权主体、自然人财产

主体、法人财产主体直至商法上的商主体等。通过法律的逐步抽象,自然人的生物、伦理和社会属性逐步褪去,逐渐嬗变成无色无味的“抽象人”,最后在财产法上“自然人”亦被抽象掉,成为与法人同等的“符号”化财产人格。从人到“非人”、从伦理人至“符号”,其间经历的一个实质性断裂就在于伦理人格完全变成工具性、形式性的人格。但在民法上这种立体性的人格体系被平面化的统一人格遮蔽了,致使我们对于亲属法、人格权法与财产法的结构区分失去参考支点,亦导致在论证民法和商法关系上失去了一个重要参考因素。尽管从抽象度上我们可以列举私法人格的递进序列,但财产法上的人格抽象还是根本性的、彻底性的,它与伦理人格最后仍然保留社会化场景中自然人的生物性和伦理性品格不同,而是成为一种完全不同的格式化产物。这种工具人格并不能通用于其他法律领域,而伦理性人格却可以与整个法律体系分享同样的价值体系。基于此,两种人格的区分在对私法上的人格体系的解释上,既有“渐变”的一面,亦有“突变”的一面。实际上我们可以观察到,不同的人格对应不同的私法规则体系,这有利于我们从人格谱系角度区分人身法和财产法、民法和商法的大致界限,亦有利于我们正确理解由两种人格因素相互渗透带来的“民法商法化”和“抽象人格具体化”的私法发展现象。

对民法上的两种人格进行理论区分,并非要刻意将这种区分绝对化,以解构传统私法的结构。这种区分类似于几何学上的辅助线具有增进理解和解决问题的意义,并不能用来“切割”现实生活。实际上,自然人在私法人格上担负着基础的功能和角色,它同时具备两种人格生活,并在人格类型间自由切换且应付裕如,并构成近代以来市民生活的底色。就私法价值体系而言,技术性人格只是法律的创造,它解决的只是财产领域的归属和效率问题,尽管具有刻板化和程式化的特点,但最终还是应服务于或回归于社会伦理生活中的基本价值。如拉伦茨认为,民法最基本的价值还是在于对他人的尊重。<sup>[44]</sup> 伦理性人格是技术性人格的基础和归宿,因为工具人格并不要求自然人在超市购物时设想对方毫无人情可言,也不会要求公司在进行交易时完全漠视社会的礼节和社会责任。一旦技术人格发展过快,脱离社会基本面并影响社会正常运转时,技术人格又将披上社会化的具象外衣,来平抑这种失衡。如当代民法频繁关注诸如消费者、承租人和劳动者等交易角色并予以特殊保护,即当产生“强而智”之人对“弱而愚”之人的压迫现象时,民法对之干预便成为必要,它是民法对市民社会自我修复功能的体现。但另一方面,当代商业社会的发展又进一步重塑了传统市民社会的价值观念和生活方式,尤其是互联网商业模式的普及和演化,使传统社会生活和交易的场景化特征完全消失,以“卡通式”形象出现的个体电商游走于民事人格和商法人格之间,呈现出“民法商法化”加剧的趋势。总体而言,伦理人格决定技术人格的底色,而技术人格又将重塑伦理人格的内涵,是未来私法人格发展的主旋律。

## (二)未来人工智能私法人格发展的展望

随着人工智能的快速发展,人工智能在法律上的主体地位问题会一直存在,并日趋凸显。

[44] 参见(德)卡尔·拉伦茨:《德国民法通论(上册)》,王晓晔、邵建东、程建英、徐国建、谢怀栻译,法律出版社2003年版,第45—48页。

依上文所述,在缺乏自主意识的通用人工智能真正出现以前,人工智能主要体现为人类的智能增强工具,没有办法在法律上赋予其独立的伦理性人格,并把它当作完全独立的伦理主体来认定。同时在技术性人格的塑造上,人工智能又面临如何用自身独立财产来承担自主行为的责任的问题,因为传统民法上的拟制人格最终需要自然人的自由意志和所投入财产的支撑,即使将人工智能本身当作独立“财产”处理,亦需要自然人意志的支持。对于未来人工智能而言,它的主体性发展在法律上呈现两种人格并行发展的趋势,且相互促进。就伦理性人格而言,在通用人工智能出现以前,各类专门人工智能将日益普遍地担当服务人类的各种角色,如上文所述的情侣机器人、助老机器人或电子宠物等将更加深入地融入到人类生活,并产生人机交互伦理的诉求,从而客观上产生是否应将机器人当作类似“动物”的第三类主体问题。就财产性人格而言,尽管目前为促进技术发展,尚不能将相关风险完全抛给社会,财产人格仍有发展空间,只是它的理论基础与传统民法拟制人格有所不同。传统法人制度依然是自然人和财产的结合,通过有限责任来形成资本融合和风险限定,而人工智能本身兼具“智能体”和“财产”的属性,只是体现为“个体”,与“组织”无关,在人工智能体本身尚不能承担责任且无法由自然人有限分担的情形下,法律可能会创造出超越“自己”责任的另一种社会化风险分担方式,这有点类似于严格责任对过错责任的跨越,只是这次的跨度更大。在人工智能发展到一定阶段,通过人类的努力已经无法由特定主体承担由“算法黑箱”产生的自动化风险,且这种风险通过大型算法的连接已经成为系统性的社会风险时,基于公共利益的考虑,一种全新的社会风险集体承担机制将会出现。

在两种人格以及抽象度问题上,人工智能在法律上与自然人遵循相反的发展路径,即与私法首先将自然人当作生物人和社会人不同,法律对于人工智能是以技术性人格的探讨作为起点。人工智能一开始即以极抽象的客体形象出现,目前它仍旧处于人类的“财产”领域,并逐步往伦理化方向发展,在此过程中人工智能会逐渐被具象化并赋予伦理意义。在人工智能具象化的过程中,人机交互充当了重要媒介。人机关系是人类从未遇过的人工智能时代的基础关系,它使人类同时在“智能体”和“财产”之间进行动态权衡,既渴望和担忧强人工智能的出现,又追求人工智能的利用效益和排斥自动化带来的风险。<sup>[45]</sup> 无论如何,人工智能一旦被人类开启,即如凯文·凯利(Kevin Kelly)所言,技术本身具有生命,它将坚定地朝着自身固有的方向自动演化。<sup>[46]</sup> 同时人工智能兼具“智能体”和“财产”的双重属性,对于法律上的主客体理论以及线性思维造成了较大冲击,它将使未来社会生活呈现出原子化和系统化相互交织的图像,人格理论在此种社会场景中亦将呈现出多元化和功能化的趋势,抽象化和具体化同时成为人格形象的两种工具。可以想象,一旦强人工智能时代来临,公法首先将确定人工智能的伦理主体意义,同时人工智能的自主行为和独立财产会随之产生,届时人工智能的财产属性将逐渐

[45] See Shakuntla Sangam, “Legal Personality for Artificial Intelligence with Special Reference to Robot: A Critical Appraisal,” *Indian Journal of Law and Human Behavior*, Vol. 6, No. 1, 2020, pp. 15-22.

[46] 参见(美)凯文·凯利:《科技想要什么》,熊祥译,中信出版社2011年版,第12—18页。

减弱。至于人工智能是否具有类似人类的伦理生活(如亲属和人格),以及是否也会分享人类的财产人格和交易法则,只能留待未来去验证了。

---

**Abstract:** At present, the theoretical discussion of the legal subject status of AI follows two paths: one is to treat AI as the “population” similar to human beings with advanced intelligence; the other is to grant the legal personality, to a certain extent, to the weak artificial intelligence, so as to assume responsibility and rights. It’s necessary to puts forward a theory to distinguish the “ethical personality” and “technical personality”, to define the different theoretical orientation of AI. The differentiation of personality already exists in the traditional civil law, which is demonstrated as a progressive sequence marked by the abstract degree of personality, and there exist the phenomena of “gradual change” and “fracture” of these two kinds of personalities. As ethical subject, AI encountered the problem of “self-consciousness”, and thus the current weak AI cannot obtain the independent ethical personality, so it is the technological ethics regulating technology developers to dominate this issue, meanwhile, the ethics on human-computer interaction may grant the robot a certain kind of moral “rights”. With specific reference to the technical personality of AI, if it is studied as an “individual”, it can not make an expression of its intent and own property as natural person, and consequently it is not fully qualified to grant technology personality to a robot. In the future, the subjectivity of AI will develop in the path of the two personalities above mentioned, namely it’s development will be led by technical personality, and be gradually shaped by the ethical personality. The ethical personality of human beings or robots will eventually become the final destination of technical personality.

**Key Words:** AI; Ethical Personality; Technical Personality; Legal Fiction; Human-Computer Interaction

---

(责任编辑:贺 剑)