

◇ 法律研究 ◇

## 在“功利”与“道义”之间： 中国人工智能立法的科学性探析

王荣余

**摘要：**人工智能的立法挑战兼具“功利”与“道义”双重属性，这要求立法超越近代以来的“科技—经济”立法范式，而转型至“在功利与道义之间”，并以此引领和保障人工智能立法的科学性。其中，道义论约束设定了基本门槛，人工智能活动只有在经过道义论检验之后才能转向功利性考量，否则就将是否定性评价。具体而言，立法必须注重人工智能在“安全问题”、“权利问题”以及“人性问题”三领域构成的道义论挑战，并以此为基础在人工智能的研究和开发、运用和转化阶段进行鼓励、支持和引导的立法选择，同时辅以事前、事中与事后评估的管理和监督。在具体问题的抉择上，立法科学性要求立法者必须充分发挥好“在功利与道义之间”的立法技艺，最终实现功利与道义的双赢。

**关键词：**人工智能；科学立法；道义论；功利主义

**收稿日期：**2020-03-15

**基金项目：**国家社科基金后期资助项目“人权话语的中国表达与体系建构”(20FFXB042)；重庆市哲学社会科学重大规划项目“权利的制度之维”(2021ZDZJ04)

**作者简介：**王荣余，西南政法大学哲学系，博士，主要从事法哲学、制度法理论和科技伦理研究，E-mail: vincenty0322@yeah.net。

### 一、问题的提出：人工智能立法的范式转型

随着信息技术和生物科技的迭代发展，“第四次工业革命”正逐步成为一个被广泛接受的事实。其中，信息技术经由“世界互联网化”和“大数据”两个阶段的发展而开始转向以机器学习和算法等技术为基础的“人工智能时代”。2016年3月，智能机器人AlphaGo以4:1的战绩击败世界围棋冠军李世石，这预示着“人工智能进入突破阶段”，2016年成为人工智能的元年<sup>[1]</sup>。自此以后，人工智能(Artificial Intelligence, AI)的讨论总是笼罩着一种潜在的忧虑，霍金、比尔·盖茨等人便表达了这样的不安。中国法学界关于人工智能法律责任问题的讨论更是直接性地表达出“规制”意味，而对于其中较为棘手的人工智能体的法律主体地位问题也是要么忙于肯定要么忙于否定。何以如此？其原因就在于与一般的科技实践引发的风险不同，人工智能这一事物本身便具有特殊性，主要表现在自身潜在拥有的“认识能力”和“控制能力”使其有可能会超越科技与人之间传统的“手段—目的”或“客体—主体”关系框架。这即是说，在人工智能时代，作为地球主宰的“人”可能会沦为人工智能的“对象”，学

界当前所讨论的“大数据杀熟”以及“算法权力”等议题便是一个重要的表征。无论是对人们的外在行为还是内在尊严,这都是我们无法忍受的。更根本的是,如果这一挑战成真的话,那问题就不再是“智能机器人是否具有法律主体地位”等问题,而是“我们是不是人”以及“我们凭什么是人”<sup>[2]</sup>。当代哲学前沿关于神经生物学、意向性以及自由意志等问题的讨论便涉及对诸如此类问题的尝试性解答<sup>[3]</sup>。此类评论并非危言耸听,雷·库兹韦尔就认为,在 2045 年前后会出现人工智能的一个“奇点”,而跨过这个点人工智能就将会超越人类的智慧<sup>[4]</sup>。即便“奇点”何时来临尚无共识,但人工智能时代的步步逼近却是一个不争的事实,而随着“强人工智能”或“智慧人工智能”的到来和挑战,如何通过法律手段更好地应对这一潜在的“实质性危机”,这是我们当前应予进行讨论的课题。

法律的功能在于“以各种各样的方式,导引、计划和控制人们的生活”<sup>[5]</sup>,而“立法”则是处理人工智能及其与人之间关系的首要环节。由于人工智能可能引发的实质性挑战与固有科技立法思维之间存在较大的出入,既有范式便面临着转型的现实需要。现行“科技法”,包括《中华人民共和国科学技术进步法》以及《中华人民共和国促进科技成果转化法》在内的诸法律,都受到了近代以来科技立法中“技术—经济”范式的影响。该立法范式的一个主要表现在于,以鼓励、刺激科技的经济实用价值为目标,而对科技引发的社会问题则只是附带性地提及<sup>[6]</sup>。更明确地讲,它比较侧重科技经济主义这一维度,旗帜鲜明地以效率为导向。这不仅轻视了人工智能科技可能引发的前所未有的“价值关切”,更导致了技术的自主化滋生,即技术在这种自主性结构中为所欲为,进而发展为失去人的控制的主体化僭越<sup>[6]</sup>。这种后果恰恰是法律试图极力避免的情形。对于立法者而言,这都内在地要求他们进行立法范式的转型,即转向“在功利与道义之间”的新范式,这是为保障人工智能立法科学性而对立法者提出的新要求。

2017 年 7 月,国务院颁布的《新一代人工智能发展规划》(以下简称《规划》)提出了“三步走”的战略布局,明确在部分领域初步建成人工智能伦理规范和政策法规的第一步战略要求。在“从规划到立法”的这一时期内,寻求立法的科学性继而进行科学立法将是立法者的关键任务,而如何保障立法的科学性更是关键中的关键。然而,当前关于人工智能立法的既有讨论主要集中于部门法领域,但由于人工智能所引发之挑战的根本性和实质性,部门法学者基于法律教义学所提供的说明性理论实际上就将变得有限<sup>[7]</sup>。这意味着,关于人工智能立法科学性的一般性讨论必须转入一般法理学层面来进行,并以此为基础提供一种规范性路径,也正是在这个意义上,我们需要借助规范伦理学的资源,尤其是功利论和道义论。具体而言,就中国的具体语境来看,人工智能领域的立法遵从什么样的原则来体现科学性?立法活动如何合理地处理人工智能所涉及的伦理风险与经济发展问题,以及应如何展现道义论约束和功利论刺激?这是本文试图通过新立法范式进行回答并试图从哲学和法学层面展开论证的问题,但在此之前,本文将首先考察一个在当下学界讨论不充分的论题,并以立法思维进行重新审视。

## 二、一个不充分的讨论:人工智能体的法律主体地位

人工智能体法律主体地位的相关讨论旨在回答两个基本问题,一是能否赋予其法律主体地位,二是相应的理由说明。就后者而言,既有讨论主要围绕“法人拟制”与“人畜关系”展开,但这对立法又意味着什么呢?

对于物理层面上的“人”,我们可以从多个方面予以审视,如生物学、人类学、哲学、社会学、法学只是其中的视角之一。一般而言,一个对象要成为法律意义上的“人”,必须满足“事实”与“规范”两方面的条件:前者要求它必须是具备意志能力的客体,这是法律赋予其法律主体资格的不充分的逻辑前提;后者则要求该客体拥有享受权利和履行义务的法律资格,这最终将取决于立法层面的考量。因而,非生物意义上的“人”完全有机会成为“法律意义上的人”,而生物意义上的“人”也完全可能无法成为法律意义上的人,如古代的奴隶<sup>[8]</sup>。这在很大程度上可以解释为什么“法人”能够具备法律主体资格——尽管它肯定不是生物意义上的人,就规范层面而言,这一认定是完全可能的,只不过需要注意的是,规范层面的认定必须有事实层面的支持。由于法人的意志能力实际上能够通过“法律拟制”的立法技术与自然人相联结来予以克服,从而基于“实体屏蔽”“清算保护”等便利考量,“法人”就完全可以成为“法律上的人”。这实际上是在提醒我们,最终承担法律责任的仍是生物意义上的人,在“揭开公司的面纱”时,提供实质性责任保证的不是法人而是生物意义上有血有肉的自然人。当然,基于同样的理由,以“人畜关系”展开的说明同样否定诸如大猩猩等灵长类动物的法律主体资格。因为动物不具有对于“应当如何”的规范认知能力——即便一些动物可能具备规律性地进行学习的能力,但正如有学者所说的那样,或许正是由于缺乏那1%的规范认知能力,使得它们永远无法成为真正的人<sup>[9]</sup>。

此类关于人工智能体法律主体地位的否定性论证是有说服力的,至少它强调了“事实”维度的不可能性,遵从了“从事实到规范”的路径。但或许正是如此,它轻视乃至忽略了法律自身的性质,即法律是一种制度性事实<sup>[10]</sup>。法律的这种性质使得“法律主体”具有了他者不曾拥有的规范意味。就此而言,即便能够证明人工智能体在事实层面最终无法具有“规范性认知”,但这仍然无法排除人工智能体可以通过诉诸“法人”那样的立法技术而被赋予法律主体资格的可能性。法律语言具有一种“构成性规则”意义上的宣告式言语行为的能力,当它进行“X在C中算作Y”的表达时,它就赋予了C以Y的地位功能。比如说,当法律以宣告式语言表达“这张纸算作一张一百元的人民币”时,它就赋予了这张纸(C)以一百元人民币(Y)的功能地位,这张纸也就不再是纯粹的一张纸,而成为法定货币。法律作为一种制度性事实,也即是一套构成性规则体系,这种体系有创立其他制度性事实的可能,这即是“语言向世界”的适应指向<sup>[11]</sup>。这要求我们法律人更加地关注法律自身的性质与逻辑,而不要过多地纠缠在纯粹事实的繁杂表象之中。更何况,人工智能是否具有规范性能力在目前看来仍是一个未知的事实,我们没有必要先入为主地进行否定性预设。因此,人工智能体能否被赋予主体资格,这首先不是一个事实问题,而是一个“是否应当赋予其法律主体资格”的规范性问题。这与赋予法人的法律主体资格之历史是完全一致的,只有在考虑了是否应当赋予法人以主体地位之后,我们才会进一步技术化地处理其中的事实承担问题。“由规范到事实”而不是“由事实到规范”更契合法律思维,立法思维尤其如此,其中,尤为重要的是法律如何为人们的人工智能问题上设定行动理由,因而是一个关乎实践理性的问题。

立法者不是人工智能的理论专家,但这根本不妨碍法律为该领域设定行动理由的实践权威<sup>[12]</sup>。从规范上讲,人工智能诸问题肇始于立法,而立法的背后则涉及立法者“应当如何进行选择”的规范伦理学问题。面对人工智能这一情势,借鉴规范伦理学的研究成就,“在功利与道义之间”则为如何行动和选择提供了指引,下文将对此分别做出哲学和法学层面的论证,这种外部论证对包括法律主体地位在内的人工智能立法的讨论具有重要的方法论意义。

### 三、“功利”与“道义”:人工智能立法科学性的哲学论证

人工智能立法的科学性至少蕴含着法律条文具备明确性、公开、不溯及既往且能为人所遵守等程序法治的基本原则<sup>[13]</sup>,但由于人工智能问题具有前文所说的那种特殊性,这使得相关立法在遵循基本的法治考量因素之下还必须更明确地认识到其中所涉及的两个关键性问题:其一是人工智能所可能引发的道德伦理问题,即道义维度;其二是人工智能及其立法与科技发展或经济进步之间的共变关系,即功利维度。如果近代科技立法中的“技术—经济”思维本质上是一种经济目的导向从而侧重功利进路的话,那么未来人工智能的科学立法则必须明确地参考道义维度,二者共同构成人工智能立法转型的新方向,继而成为其新的思维框架。

#### (一)功利论的内在要求

功利论是规范伦理学中目的论的典型代表,在当代具有各种不同的形式和分支,后果主义便是其中最为重要的一种。后果主义理论由两方面内容构成:(1)根本的价值;(2)达致这一价值的适当进路<sup>[14]</sup>。前者是特定的价值理论,它认为价值内在地包涵于事物/事态之中,实现这种价值就必定能够促使这一事物/事态的实现;后者是特定的正确选择理论,它认为,如果一个行动促进了更好的可能事物/事态的实现(比较意义上的),那这一行动的正确性便得到了辩护<sup>[15]</sup>。标准后果主义(standard consequentialism)的理论形态据此向我们承诺了这样一件事情:通过评价事物/事态对其价值的促进,我们便获得了最终的辩护根据,不管这是道德辩护、法律辩护还是其他类型的辩护事项,它都能够得到适用。根据“一般形式的客观的整全后果主义”的观点,对象的意向、品格特征、动机等或行为的道德属性都可以依据客观后果主义的标准来进行评价<sup>[14]</sup>,这使得后果主义的理论承诺会更加地具有吸引力。

在此理论路径之下,三方面的特质型塑着后果主义。首先,它所指的价值内在地具有整体利益的倾向,即功利主义所说的“最大多数人的最大幸福”<sup>[16]</sup>。比如,就赋予人工智能体主体资格的立法动议而言,它必须首先证明赋予其主体资格比不赋予其资格能够带来整体上的更大利益。德国工程与机器人协会(VDMA)也正是以此为指引论证说,立法者不应通过颁布规制措施来阻碍工业 4.0 的可能发展,因为人工智能会给人更多的益处而非不利<sup>[17]</sup>。在此判断下,便隐含着人工智能法律主体的证成语境,美国的《人工智能未来法案 2017》(*Future of Artificial Intelligence Act of 2017*)亦是作了如此判断。其次,它将行为所获得的效果作为评价的标准,因而结果本身只是手段而不是目的,它是朝向未来的可欲手段。在这个意义上,“判断道德意义上的正当、不正当或尽义务等基本或最终的标准是非道德价值,这种非道德价值是作为行为的结果而存在的。最终的直接或间接的要求必须是产生大量的善,更确切地说,是产生的善超过恶。因此,在任何有效的选择中,一个行为是正当的,当且仅当它或它的指导准则能够促成或趋向于促成的善至少超过恶;反之则是不正当的”<sup>[18]</sup>。最后,它认为“善”优先于“正当”。在道德评价过程中,“正当”表现为康德式的“绝对命令”,即“要只按照你同时认为也能成为普遍规律的准则去行动”<sup>[19]</sup>。与之不同,“善”则凸显为它的诱人的可欲性,在此之下并无绝对的命令或义务,而只有相对义务。善优先于正当便是要求着眼于最大利益。

功利或后果是一个带有评价性质的规范性概念,备受指责之处主要在于,它所承诺的“功利价值”



如何辨识以及客观性评价基础如何还原。极端的功利主义不利于社会基本善的形成,更不适合人工智能这一情势的特殊性,因而要完整地进行评价,我们还必须引入“道义论”。

## (二)道义论的内在要求

一般而言,“道义论约束”往往是通过一些否定性的方式来予以陈述和表达的道德理由,是对实践活动所提出的禁为性理由。因而,道义论从根本上关注的是开展这类行动的不可允许性问题<sup>[20]</sup>。道义论与功利论极为不同。首先,道义论以观念形态的义务和应当为出发点,“要求个人应当按照某种主观上既定的原则或主观上认定属于某种现象本身固有的正当性去行动,从而将人们的义务和责任主观化、普遍化和绝对化,并以此来限制和规范个体”<sup>[21]</sup>。其次,道德义务本身便是目的,是“为义务而义务”,从而对道德自身所蕴含的尊严以及人格中的人性等基本价值予以肯认,因此,道义论所提出的义务是“绝对的和无条件的”。这即是说,在进行选择的时候,我们一开始便受到限制而不是随心所欲的。如果赋予人工智能体以法律主体资格会对作为目的本身的人/人类之尊严造成贬损,那道德律令就将禁止我们这样做。最后,道义论预设了“正当优先于善”。按照这一形式标准,行动的理由和根据不在于任何功利性的价值,相反,事物本身便能确定何种欲求是正当的。尽管学界对于“正当”存在不同的解释,但却能够就以下观点取得基本共识,即“一个行为是否正确,是由它所固有的某些特点来确定的,或者是由与它相关的某个规则来确定的”<sup>[22]</sup>。

在道义论理论中,当前最具说服力的一种进路表现为“牺牲者中心”<sup>[23]</sup>。它认为,作为共同体中的每一个人,都有可能被他者或共同体伤害,即都可能成为潜在的牺牲者(victim),与此同时,我们每一个人都拥有优先于共同体本身的资格(status),这种资格使得我们有很强的理由避免成为他者或共同体本身的可能牺牲者。就人工智能法律主体这一事情而言,不管人工智能能否被授予法律主体资格,只要每一个个体本身的资格得到了保障,那这种设置就通过了合法性的检验,如果人工智能法律主体的制度设计危及或者是必定使每一个个体成为潜在的牺牲者,那它就不具备道德合法性。与之相关的另一种道义论理论即契约主义的道义理论认为,“资格”概念的设置并不存在实质意义上的必要性,因为在契约签订过程中,共同体中的每一个人只要具有“自由且平等”这个事实便足以对每一个主体施加保护。相应地,能否被授予主体资格从而拥有一项权利只需取决于它能不能满足共同体成员所设置的假设性契约条件。因此,人工智能体能否被授予法律主体资格,这必须取决于由理性的共同体成员所设置的条件要求:满足其要求则被授予法律主体资格,不满足其要求则不能被授予法律主体资格。可以发现,“牺牲者中心”和“契约主义”的道义理论在人工智能体能否成为法律主体的道路上设置了“防御性”的条件;与通过“预防原则”应对科技风险不同,它所提供的方案具有根本性;与后果主义所诉诸的实质利益不同,它所捍卫的是形式性基本原则。这一方案不仅适用于人工智能法律主体资格问题,而且适用于关于人工智能的整个议题。

## (三)一个综合评价

由此,在人工智能立法科学性问题上,人们就拥有了功利论和道义论这两套可供选择的辩护方案,虽然它们各自在一些关键问题上针锋相对,但这并不意味着只能非此即彼地进行选择,相反,我们可以而且应当对其予以综合考察,并将其灵活地运用于人工智能立法科学性的辩护议题上。这就好比说,人工智能的立法者必须左手握着道义论的“利剑”,右手握着后果主义的“天平”,在此二者之间来回地细致权衡和拿捏,并以此来保障人工智能立法问题的科学性。仍然以人工智能体法律主体地

位立法问题为例,它内在地要求:

首先,人工智能立法必须全面权衡基于功利主义的整体利益的进取型要求与基于道义论的个人资格的防御性要求之间的关系。如果人工智能法律主体地位与个人资格的道义论要求根本冲突,那就必须捍卫“人类中心主义”;如果不相冲突,那就应当基于功利成本进一步考虑是否赋予其主体资格。其次,在手段与目的关系问题上必须保证“人是目的”这一基础性的道义论原则。虽然“从一个角度看,我们创造的机器人貌似只是机器而已,而从另一个角度看,它们似乎是复杂而具有适应性的实体,甚至对我们在道德关心的问题有着或多或少的敏感性”<sup>[24]</sup>。人工智能体虽然不一定就只能作为一种工具存在,但由于“人是目的”这一绝对命令的存在,不管赋予还是不赋予其法律主体地位都不得逾越这一底线,一旦触碰便是无效的,与之相关的是,在人机伦理关系中,立法者不仅要考量智能机器人的“外部进攻”,同时还应考量作为主体的人本身在主体问题上可能出现的“消极抵抗”。人工智能时代会引发一种危险,作为人类的我们不再将自我视为目的本身,而是视作为人所利用的装置和被人利用的工具,从而仅仅只是像计算机那样思考问题<sup>[25]</sup>。最后,在正当与善的优先性权衡上,正当是根本,善是长远发展的指引。人工智能时代是大势所趋,如果人工智能法律主体地位更符合长远发展的要求,至少就可以考虑赋予其主体资格,但这并不意味着丢弃其内在的正当性要求。

#### 四、人工智能立法中的“道义”与“功利”

通过辩证看待功利论和道义论这两种规范伦理学进路并予以综合审视,我们获得了达致人工智能科学立法的一套辩护工具。在这里,二者在人工智能立法问题上绝非相互对立,而是可以相互补充,而且综合出的这套辩护工具也并不囿于某一思想家的学说或某一学派的主张,而是以实事求是的姿态对其予以客观评价,具有切实的可行性。此外,我们还以“人工智能法律主体地位”这一当下为学界广泛关注的论题为例进行了针对性的解读和说明,但法律主体地位问题毕竟只是人工智能立法的面向之一,因此我们还需要进一步考察如何将这一套辩护工具适用到人工智能立法中去。

##### (一)人工智能立法科学性的“道义”考量

有学者将新的立法思维归结为“技术—社会+经济”,相较于既往的“技术—经济”范式,它更加强调人工智能引发的“强社会面向”,从而“展现了前所未有的社会规范需求性”<sup>[6]</sup>。实际上,“社会维度”在实际操作层面难以和“经济维度”进行有效的区分,这不仅是因为经济系统本身就是社会中一个重要的子系统,更在于不管经济维度还是社会维度都可以被纳入功利论范畴中进行讨论,新旧范式之间就无法进行有效区分,从而无法揭示人工智能带来的根本性挑战。相反,在尊重道德法则的行为必然性意义上,只有“道义”维度的考量或许才能直接揭示和呈现出由人工智能引发的法律挑战的特殊性。通过对这种挑战进行由外到内的处理,相关道义论挑战可以类型化为以下三个方面:

##### 1. 安全问题

安全问题包括人工智能技术本身所带来的挑战,如由于“黑箱(black box)”系统的存在,输入环节是透明的,输出也得到了理解,但两者之间发生的事情,连程序设计者在内的人都无法确切地理解自己设计出来的人工智能的具体决策过程和决策逻辑<sup>[26]</sup>,从而致使技术本身可能就是不安全的。再者,当自动驾驶汽车脱离制造商控制而进入流通领域后,由于新的数据输入导致其行为和决策超出预

先设置的规则,从而可能会打破原先设定的规则<sup>[27]</sup>。此类安全来自技术本身的“不确定性”维度,这使得我们“无法提前预估可能遇到的全部问题,也不可能绝对排除随机生成的稳定性故障,因而带来不同于传统技术运用所制造的新型风险”<sup>[28]</sup>。对于此类不确定和不可知的现实问题,欧盟提出的立法动议首先便强调要确定人工智能领域的伦理准则,这既包括具有全局性的伦理框架设计,也包括诸如“一键关闭”这样的具体考量,可以说,人工智能技术本身的安全性问题将是立法者在进行立法设计时绕不开的首要问题之一。

与此同时,“安全问题”也包括由于技术滥用所带来的极大风险,其中广为人所忧虑的一个便是“人工智能武器”。伦理学界早已对此问题及其未来情势表达过忧虑,这种武器及其滥用“很可能成为人类自作自受的掘墓人”<sup>[29]</sup>。技术滥用导致的危险与技术不确定性导致的危险具有本质区别,立法无法规制技术本身的不确定,但却完全有能力规制使用或滥用人工智能的人的行为,从后者寻求突破将最为可行。在国际范围内,以“公约”的形式限制使用或禁止人工智能技术的滥用将是一个发展趋势。就国内来看,对于智能科技自身所存在的安全问题,“合理的对策就是为了保障技术进步而适度容忍风险,并立足于现有法律的基本体系,将风险进行成本最低的分配”<sup>[28]</sup>。

## 2. 权利问题

“权利问题”关涉人的地位和资格,因为权利是现代人生存和生活的基本方式<sup>[30]</sup>。人工智能引发的相关新兴(型)权利问题正在对现有权利义务体系提出挑战,如人工智能生成内容的权利定性、自动驾驶汽车引发的相关责任认定等。在既有的权利意识形态中,“X有做 $\varphi$ 的权利”或“X有关于 $\varphi$ 的权利”表述中的“X”被默认为是生物学意义上的人或以它为基础,既有权利体系的哲学基础在于“人的主体性”,甚至可以说“当代权利理论的发展过程正是主体性通过权利概念和法学话语的表征过程”<sup>[31]</sup>。人工智能所可能拥有的“认识能力”与“控制能力”,即人工智能所可能涌现的自主性与主体性问题会通过挑战现有权利体系对主体性哲学基础构成威胁,从而最终威胁到“人”本身这一权利的道义论根基。随着人工智能的深入研究和推进,法学界终将面临“人工智能能否拥有权利”以及“凭什么能拥有权利”这类法理学问题(实际上法理学界已经在开始处理这一问题了)。

如果这种挑战因其具有基础性以致目前还无法进入立法视野的话,那伴随着“算法权力”“算法规训”“算法歧视”等带来的权利挑战则无疑具有十足的现实性。人工智能的本质是算法,由于算法的控制者与作为数据主体的个人之间存在信息不对称,因此算法不断增强的权力态势会对个人权利形成霸权压制,如对数据所有权以及数据隐私权的侵犯和歧视,它使得“个人越来越透明,而权力行使者却变得越来越隐秘”<sup>[32]</sup>。如无算法治理的介入,人的权利将遭受全面打击。面对此情势,欧洲议会发布的《关于欧洲合作智能交通系统战略协作、连接和自动移动性的决议》便指出:“必须使用户确信他们的个人数据并非商品,用户能够有效控制数据的使用方式和目的”<sup>[1]</sup>。的确,算法的法律规则只不过是人工智能立法的方面之一,但算法对权利的挑战却已揭示出这一情势的根本性,因而有必要以算法正义的立法理念为导引捍卫表征人之自主性的权利的道义之维。

## 3. 人性问题

阿西莫夫提出了“机器人三定律”,即“不得伤害人类,必须服从于人类,必须保护自己的生存”<sup>[33]</sup>。这三条规制机器人的定律预设了智能机器人已经具备自我意识和自由意志的能力,这或许是过往立法实践较少明确予以考量的问题,但却是法哲学一直在面对和捍卫的一个命题,即“人是目的”。然而,人工智能的深入研究却揭示了另一番景象。在技术研究层面,这一问题表现在“深度学



习”和“神经网络”所隐含的一个基本假设,即智能及其思想意识本身均源于大脑的复杂性<sup>[34]</sup>,也即是“神经元决定了思想”<sup>[35]</sup>。在具体研究中,它通过“虚拟神经元”进行人脑模拟并在此基础上构建“人工大脑”。基于对人之大脑的心智模拟,AI 程序就能够实现“像”人类一样地进行认知和思考,达到甚至超过人的智能水平。在当代哲学前沿理论中,强人工智能与人的心灵皆被理解为一种“计算机功能主义”。该理论认为,“整个系统工作的方式就是大脑是一个数字化的计算机,而被我们称为‘心灵’的东西则是一个数字化的计算机程序,或是一个程序序列”,即“心灵状态乃是脑的计算状态。脑是一个计算机,而心灵是一个程序或程序序列”<sup>[3]</sup>。前沿量子力学对人之生命的解释以及存在非碳基生命的理论可能性,诸如此类的研究尝试和推进与哲学层面的理论辩护间接性地强化了人工智能的主体性地位,同时也在很大程度上弱化了人之地位的道义论基础,于是,问题就逐渐地变为“人工智能是不是人?如果它不具备人的地位,那么我们为什么因为拥有同样的或类似的智能能力而被叫作人?”<sup>[36]</sup>正如赵万一教授所言,“法学应当重新思考人”<sup>[37]</sup>,尤其是法律的人性基础问题。

在极端目的论支持者看来,上述三类议题都可以在功利范畴中进行讨论和解决。然而,不管是安全问题、权利问题还是人性问题实际上都无法简单地适用最大多数人的最大幸福这一判准,因为它在很大程度上挑战甚至会超越既有法律体系赖以存在的基础,比如说,既有科技立法的对象本身是可控的,但人工智能技术愈加显示其不可控的黑箱效应。因此,有必要首先将此类问题置于道义论约束中予以审视,而不能仍停留于功利思维层面。就立法设计而言,“安全问题”中的技术滥用与“权利问题”中的算法利维坦现象是立法必须予以明确回应的,而理论界则需要在不确定的技术、权利的主体性挑战以及人性问题上展开进一步的研究,并在此基础上为未来的立法回应提供理论层面的支撑,结合功利取向进行相应立法。

## (二) 人工智能立法科学性的功利考量

人工智能之所以具有广泛的热度和关注度,不仅因为它会引发上述道义层面的挑战,更在于它具有丰厚的功利价值,即创新驱动和经济发展这一目的指向,可以说,谁率先占领了人工智能这一广阔领域,谁就赢得了第四次工业革命的主导权。人工智能技术作为重要的生产力要素,对于推动经济和社会发展发挥着无可替代的关键作用。《规划》将人工智能定位在“国际竞争的新焦点”“经济发展的新引擎”“社会建设的新机遇”的国家战略高度,党的十九大报告更是提出“推动互联网、大数据、人工智能和实体经济深度融合”的目标要求。可以预见,随着人工智能尤其是通用人工智能的发展,自动驾驶、远程医疗等新领域正在逐步变为现实,这都使得人工智能成为未来国家实力和国际竞争的一个关键所在。因此,“如何在确保人工智能技术不至于触及法律底线和引发伦理道德问题的前提下,最大程度地发挥其促进生产力发展、促进社会进步的功能,就必然成为亟待解答的时代课题”<sup>[38]</sup>。这一时代课题要求立法科学性进行与之相适应的制度跟进,要求人工智能的发展能够有益于最广大人民的根本利益,这是立法科学性的功利考量。具体而言,在经由道义论约束的检验之后,人工智能立法应当在以下三个阶段进行权衡设计,以此来保证其科学性:

### 1. 研究和开发阶段

在人工智能的研究和开发阶段,法律 and 政策的积极或消极态度将对其产生关键性的影响。现实中,不同的国家和地区对人工智能也是采取了不同的应对思路。德国工程和机器人行业协会(VDMA)指出,人工智能目前给人类带来的好处要远远胜于其弊端,立法者不应该仓促出台规制措施来阻碍其发展,更没有必要制定详细的法律规则。美国的《人工智能未来法案 2017》更是旗帜鲜明地



主张要大力发展人工智能,如建立有利环境、优先发展人工智能、支持和推动其发展应用等。与之相反,欧盟议会则主张通过制定民事规范来限制机器人的生产。中国在《规划》中进行了“发展为主、控制潜在风险为辅”的战略布局,就此而言,相关人工智能立法的基本取向应当是支持、赋权和引导并营造有利的法律环境,以便响应“以提升新一代人工智能科技创新能力为主攻方向”的政策号召。

科技发展有其自身的规律,立法者只有在此客观基础上进行审视而不是基于主观臆想,这样才能符合科学立法中的功利考量。在鼓励人工智能发展的前提下,立法者应当突出人工智能发展的重点方向和关键领域并进行具体而有针对性的立法,这意味着,目前尚不具备足够成熟的条件针对人工智能进行体系化的立法,因而有必要为技术发展留下足够广阔的想象力空间<sup>[39]</sup>,这是体现当下立法科学性的内在要求。与之相关,在弱人工智能时期或通用人工智能阶段,立法者应当就具体领域进行相应的立法设计以便回应系列现实问题。比如说,在自动驾驶汽车立法问题上,既有法律体系和立法思路具有重要参考价值,而不需要完全重新开始,而是可以选择诸如通过修改交通肇事罪、危险驾驶罪或增设人工智能重大安全事故罪等方式来进行局部性、针对性立法<sup>[40]</sup>。与此同时,还可以通过立法解释、法律修改、条文增设等法律技术手段将诸多涉及人工智能问题的分散性内容有序地纳入既有法律体系之中,以便于维持法律体系的稳定性、保障立法的科学性,同时也能为将来人工智能立法提供一个缓冲。在更为细致的问题上,由政府制定标准,省市进行试点,并在此基础上通过总结经验来凝练法律条文,这也不失为一条可行的进路,且也符合中国当下的相关实践,如北京、深圳等地对自动驾驶汽车就进行了这种意义上的立法尝试。

## 2. 运用和转化阶段

如果说人工智能的研究和开发阶段,立法者可以通过支持或抑制来进行“门槛把关”的话,那法律在人工智能运用和转化阶段所要处理和回应的则会相对比较复杂。在宏观层面,虽然人工智能的研究和开发能力是国家软实力的重要组成部分,但作为一门技术的人工智能只有通过转化为现实的生产能力时才能真正发挥实效,因此,如何加快人工智能与经济、社会、国防等领域的深度融合,这是立法者在进行立法设计时应予考虑的问题。人工智能与经济的深度融合有助于经济的转型升级,发展智能经济,从而提升国家的综合经济实力和生产力;人工智能与社会的深度融合有助于优化社会资源,建设智能社会,进一步加快国家治理体系和治理能力的现代化;人工智能与国防领域的深度融合有助于维护国家安全,提升应对国际风险挑战的国家竞争力。《规划》指出的“发展智能经济,建设智能社会,维护国家安全”是人工智能立法在运用和转化阶段的法理依据和基本的功利考量。在促进人工智能良性转化的同时,立法者也应考虑到人工智能在向现实转化时所带来的危机和挑战,这构成了功利的背面。由于人工智能在向现实转化时所带来的危机和挑战相对比较复杂和多样,这会使得立法回应同样变得更为复杂,这也是立法者实践智慧得以展现的契机。

在经济层面,通用人工智能的大量应用可能会导致大量失业问题,而现有的劳动方面的法律并无相应规定。此外,掌握人工智能先进技术的大企业会使传统制造业以及其他与互联网不具直接相关性的行业明显处于劣势,中小企业同样会面临边际效应不断递减的情况,强者会越强,弱者会越弱。在社会和国家安全层面,人工智能对社会治理所构成的危害源于“算法”的非中立性和偏私性,“它毫不遮掩地服务于设计者植入其中的目的”<sup>[41]</sup>。通过人工智能进行的社会治理很容易使得法律的行使和社会的监管被牵着鼻子走,从而使得人被算法所统治,并导致社会治理的异化和国家安全存在本身之“目的因”的消退。如果说发展智能经济、智能社会和智能国家安全的最大功利在于最广大人民的

根本利益的话,那么人工智能在转化阶段的立法就必须回到最大幸福的承载者,即人本身,并以“以人为本”为底线防止人工智能对人造成的异化,进而推动人工智能的可持续发展和智能化成果的共享。

### 3. 评估阶段

由于“人工智能”这一事物本身的特殊性,立法者有必要在“评估”问题上对人工智能进行设计,以管理和监督人工智能的可行性和安全性。评估是当前法律认知的重要途径,人工智能评估更应是法治评估的组成部分。法律层面的人工智能评估不同于技术评估,更不是对人工智能相关法律法规的评估,它主要是对“人工智能”这一对象及其活动的风险评估。具体而言,评估包括事前评估、事中评估以及事后评估,所评估的内容可分为安全性、经济性和伦理性三个方面。首先,在事前评估阶段,根本性地侵犯伦理性的人工智能应当在事前评估阶段就给予否定性评价,就像对基因编辑技术的处理一样,对于安全性不确定而经济效益巨大的人工智能应基于功利的考虑进行相对宽松的评估环境设置。其次,对人工智能在现实社会中运用情况的事中评价,立法设计应主要集中在安全性和伦理性方面,如不安全使用的频率、已造成的伦理风险等内容,并以此使人工智能技术在安全的前提下创造更大的功利价值。最后,事后评估应是综合评估,立法设计应在功利导向、安全为基、道义为辅的框架下进行全面审视,并将信息反馈到新一轮的人工智能规划中。应该说,法律层面的人工智能评估从纵向上涵括了人工智能的整个阶段,同时也有能力从横向上对人工智能的整个内容进行评估化处理,因而同时可以兼顾伦理风险和经济发展问题。

由于当前的人工智能立法要响应“发展为主、控制潜在风险为辅”的战略布局,因此类似“人工智能法”之类的综合性法律之制定肯定是不现实的(此前已作考察,故不赘述),但却可以以“评估”为基本思路颁布或拟定“人工智能评估法”性质的法律或部门规章,以便于实现对人工智能全过程的法律管理和监督,这一路径既能够实现发展的功利需求,也能够容纳道义论限制的内容。这意味着,在通过道义论约束针对性地应对人工智能引发的法律挑战的特殊性之后,既有立法范式和思维在可见的一段时期内仍大有可为。

## 五、在“功利”与“道义”之间寻求立法的科学性

人工智能兼具功利与道义的双重属性。作为一种日新月异的科技,它“具有为行业和社会创造巨大效益的潜能,因此有必要积极推动其发展”<sup>[1]</sup>。对此,最有效的法律方式就是进行支持、鼓励和赋权式立法。但其潜在的道义论挑战又使得立法者不得不保持一种“远虑”,因为它可能会使得“人类在世界存在系统中失去了地位,人类不再重要,历史将失去意义,人类文明将成为遗迹,未来也不再属于人类,人类文明数千年的创世纪将被终结而开始人工智能的‘创世纪’”<sup>[29]</sup>。此类“远虑”初看起来似乎有些危言耸听,但也并非空穴来风,实际上这种远虑正是立法者凭借其实践智慧应予以重点关注的主要内容所在。道义论论辩所提供的“绝对义务”或“绝对命令”会因为人工智能实质性挑战的变现而消解功利价值的正当性,从而转入捍卫人类中心主义的道路,但是,在弱人工智能阶段,法律无疑会因其功利价值的刺激而予以鼓励,而实质性挑战的“潜在可能性”又不得不使立法者必须对其予以特殊对待。一方面,法律应当设置一些基本性的门槛检验以捍卫法律的安全、权利和人性基础,在未通过其检验时,无论功利价值如何巨大的人工智能都不能获得成立;另一方面,如果通过了道义论的门槛检验,那么人工智能发展就可以基于功利要求寻求获得肯定性支持。在具体问题的抉择以及具体条

文的设计上,立法科学性要求立法者必须充分发挥好“在功利与道义之间”的立法技艺,从而实现功利与道义的双赢。

#### 参考文献:

- [1] 龙卫球. 科技法迭代视角下的人工智能立法[J]. 法商研究, 2020, (1): 59, 65, 65.
- [2] John Cheney-Lippold. *We Are Data: Algorithms and the Making of Our Digital Selves* [M]. New York: New York University Press, 2017: 244-268.
- [3] 约翰·塞尔. 心灵导论[M]. 徐英瑾, 译. 上海: 上海人民出版社, 2019: 108-193, 64.
- [4] 夏妍娜, 赵 胜. 中国制造 2025: 产业互联网开启新工业革命[M]. 北京: 机械工业出版社, 2016: 146.
- [5] H. L. A. Hart. *The Concept of Law* [M]. Oxford: Clarendon Press, 1994: 40.
- [6] 龙卫球. 人工智能立法的“技术-社会+经济”范式[J]. 武汉大学学报(哲学社会科学版), 2020, (1): 68, 69, 72.
- [7] 王荣余. “机器人也是人”的法理拷问[J]. 社会科学动态, 2019, (11): 18-23.
- [8] 凯尔森. 法与国家的一般理论[M]. 沈宗灵, 译. 北京: 中国大百科全书出版社, 1996: 107.
- [9] 冯 洁. 人工智能体法律主体地位的法理反思[J]. 东方法学, 2019, (4): 43-54.
- [10] 王荣余. 重访法律制度主义: 新旧之别与异同之辨[J]. 学术月刊, 2020, (9): 114.
- [11] 约翰·塞尔. 人类文明的结构: 社会世界的构造[M]. 文学评, 盈 俐, 译. 北京: 中国人民大学出版社, 2014: 44-64.
- [12] Joseph Raz. *The Authority of Law: Essays on Law and Morality* [M]. Oxford: Oxford University Press, 1979: 10.
- [13] Lon L. Fuller. *The Morality of Law* [M]. New Haven: Yale University Press, 1969: 46-91.
- [14] 德莱夫. 后果主义[M]. 余 露, 译. 北京: 华夏出版社, 2016: 2, 179-180.
- [15] 张 曦. 后果主义与道德约束[J]. 伦理学研究, 2014, (4): 15.
- [16] 杰里米·边沁. 政府片论[M]. 沈叔平, 译. 北京: 商务印书馆, 1995: 92.
- [17] 潘铭方. 人工智能: 从规划迈向立法[J]. 电子知识产权, 2018, (4): 85.
- [18] 弗兰克纳. 伦理学[M]. 关 键, 译. 北京: 三联书店, 1987: 28.
- [19] 康 德. 道德形而上学原理[M]. 苗力田, 译. 上海: 上海人民出版社, 2018: 72.
- [20] 张 曦. “道义论约束”与“行动者中心”进路[J]. 哲学动态, 2014, (3): 64.
- [21] 方 毅. 功利论与道义论的对立及其超越[J]. 学术交流, 2008, (8): 21.
- [22] 徐向东. 自我、他人与道德: 道德哲学导论[M]. 北京: 商务印书馆, 2016: 357.
- [23] 张 曦. 理解道义论: “牺牲者中心”进路与人的道德地位[J]. 道德与文明, 2014, (1): 41-48.
- [24] 温德尔·艾伦. 道德机器[M]. 王小红, 等, 译. 北京: 北京大学出版社, 2017: 188.
- [25] 马长山. 人工智能的社会风险及其法律规制[J]. 法律科学, 2018, (6): 49.
- [26] 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[M]. 阚 佳, 译. 北京: 电子工业出版社, 2016: 239.
- [27] 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[J]. 法律科学, 2017, (5): 169-170.
- [28] 李 晟. 人工智能的立法回应: 挑战与对策[J]. 地方立法研究, 2019, (5): 65, 68.
- [29] 赵汀阳. 人工智能“革命”的“近忧”与“远虑”: 一种伦理学和存在论的分析[J]. 哲学动态, 2018, (4): 8, 10.
- [30] 汪太贤. 权利泛化与现代人的权利生存[J]. 法学研究, 2014, (1): 7-9.
- [31] 黄 涛. 主体性时代的权利理论[J]. 法制与社会发展, 2019, (1): 52.
- [32] 郑 戈. 在鼓励创新与保护人权之间: 法律如何回应大数据技术革新的挑战[J]. 探索与争鸣, 2016, (7): 79.
- [33] Isaac Asimov. *Runaround* [M]. New York: Doubleday, 1950: 40.
- [34] 皮埃罗·斯加鲁菲. 人工智能的本质: 人工智能与机器人领域的 64 个大问题[M]. 任 莉, 译. 北京: 人民邮电大学



出版社,2017:23.

[35] 乔治·扎卡达基斯. 人类的终极命运:从旧石器到人工智能的未来[M]. 陈朝,译. 北京:中信出版社,2017:276.

[36] 陈景辉. 人工智能的法律挑战:应该从哪里开始? [J]. 比较法研究,2018,(5):144.

[37] 约翰·韦弗. 机器人也是人:人工智能时代的法律[M]. 郑志峰,译. 台北:元照出版公司,2018:中文序言.

[38] 武雪健. 人工智能立法的海外状况及难点分析[J]. 互联网经济,2019,(4):49.

[39] 郑戈. 如何为人工智能立法[J]. 检察风云,2018,(7):17.

[40] 李振林. 人工智能刑事立法图景[J]. 华南师范大学学报(社会科学版),2018,(6):130-133.

[41] 郑戈. 人工智能与法律的未来[J]. 探索与争鸣,2017,(10):81.

## Between Deontology and Utilitarianism: The Scientificity of China's Artificial Intelligence Legislation

WANG Rongyu

**Abstract:** The challenges coming from Artificial Intelligence technology have the features of both Deontology and Utilitarianism, which requires that the lawmakers on Artificial Intelligence should go beyond the modern Techno-Economism pattern and turn to a new pattern which stands “between Deontology and Utilitarianism”, so as to guide and safeguard a scientific legislation of AI. The constrains of Deontology set the basic requirement that AI should be tested by Deontology firstly and then turned into utilitarian considerations, otherwise it will not work. Specifically, Artificial Intelligence legislation should focus on the Deontology challenges coming from the issues of Security, Right and Humanity, and then make a legislation choice to encourage, to support and to guide during all the stages of researches, exploitations and applications. In addition, it should be managed and supervised supported by the AI evaluation before, in and after. At the matter of special issues, scientific Artificial Intelligence legislation requires legislators to make a balance legislation between Deontology and Utilitarianism so as to acquire win-win results.

**Key words:** Artificial Intelligence; scientific legislation; Deontology; Utilitarianism

(责任编辑:陈艳艳)