



关于人工智能发展与治理的若干哲学思考

清华大学科学技术与社会研究所教授 吴 彤

【摘要】针对人工智能发展的现状，着重分析如下问题：第一，智能与人工智能的概念；第二，人工智能发展背后的哲学意义，其中包括人工智能发展的本体论与认识论意义、人工智能发展面临的伦理学和政治哲学问题；第三，人工智能可能带来的社会治理问题，以及若干可以参考的治理建议。

【关键词】人工智能 科学技术哲学 意义 治理建议

【中图分类号】TP311

【文献标识码】A

【DOI】10.16619/j.cnki.rmltxsqy.2018.10.002

2016~2017年可以说是“人工智能”年。在这近两年的时间里，以阿尔法狗（AlphaGo）及其系列为代表的“人机大战”开启了人工智能热，从学术界到企业界，从精英到平民，大街小巷所谈的大多是“人工智能”。其中，最主要的话题就是人工智能是否会超过人类，甚至会形成“人机杂合体”，从而变成新新人类，奴役或取代现在的人类。对于人工智能及其发展众说纷纭，其中既有深刻之见，也有浅陋之言，不管如何，讨论背后都蕴含着极为深刻的意义，如不深入揭示，则可能有意无意地产生某种遮蔽，因此有必要进一步讨论这个重大问题。

关于人工智能的概念分析

人工智能，这个概念的内涵与外延其实都非常丰富。

首先是智能（intelligence）概念。何为智能，众说纷纭：就内涵而言，智能是一个历史和心理测评概念，主要是比对整体人类智力发展水平，以及依据一定的问题求解比个体智力发展水平；智能概念是一种测评、手段性的概念，运用最多的是以人类智力发展水平比对其他物种特别是灵长类动物的问

吴彤，清华大学科学技术与社会研究所教授、博导。研究方向为科学实践哲学、系统科学哲学、科技与社会。主要著作有《复归科学实践——一种科学哲学的新反思》《自组织方法论研究》《现代科学技术与当代社会》等。

题求解能力如何；肖恩·莱格与胡特一起提出了对智能的理解——智能就是在各种各样的纷繁复杂的环境中实现目标的能力（集智俱乐部，2015：122）。实际上，以上对于“智能”的定义和理解，是一种功能意义或行为主义的理解。这让智能的操作变得可能，但并没有在根本上、实质上理解何为“智能”。人类智能实际上包含了“逻辑思维”能力、“直觉”能力、自主学习能力，以及许多应变策略甚至犯错误纠错的能力，就这个意义上讲，人类智能是一种通用的、历史演化的智能。人类的能力，就单个方面而言，不如许多其他物种，例如大脑协调小脑进行奔跑及其改变路线的能力不如捕杀猎物的豹子，视觉辨别的能力（它也需要大脑的思考与理解）不如空中盘旋搜寻的鹰隼。就历史演化而言，早期生活在野外的人类感知能力可能强过现代都市的人们的感知能力，例如，卡西尔在其著名的《人论》中就讲述过原始先民对于空间的感受是一种行动空间的感受，这种感受明显强于现代人类，然而，他们却没有几何空间或逻辑空间的理解能力；而现代人建立的空间概念，明显是欧几何的空间，是逻辑的空间（如百度地图，或高德地图）。有鉴于此，下述对于智能概念的理解似乎不错：

1) 一个人的智能就是他理解和学习事物的能力（Someone's intelligence is their ability to understand and learn things）；

2) 智能是一种思考和理解的能力，而不是依赖本能或无意识地做事（Intelligence is the ability to think and understand instead of doing things by instinct or automatically, Essential English Dictionary, Collins, 1990, 转引自M. Negnevitsky, 2005: 1）。

其次是人工智能（Artificial Intelligence, AI）概念。自第一台电脑诞生，人们就开始梦想制造一台像人脑一样会思考的机器。由于这个概念是将机器类比为人类，单指被人类建造的机器具

有智力水平，具备智能，即人类建造的类似具有人之智能的机器，所以在概念上被称为“人工智能”。一般的人工智能概念是一种测试和操作意义的概念，最典型的测试是图灵测试。一般认为通过了图灵测试的人工实体是具有智能的（R. Neapolitan & X. Jiang, 2012: 2）。从功能上看，人工智能就是让计算机完成人类心智能做的各种事情（玛格丽特·博登，2017: 3）。

在学科技术种类上，人工智能是计算机科学的一个分支，是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究主要包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能从诞生以来，理论和技术日益成熟，应用领域也不断扩大，可以设想，未来的人工智能，将会是人类智慧的“容器”，也可能超过人的智能。例如，有“现实版钢铁侠”之称的SpaceX公司，其创始人埃隆·马斯克（Elon Musk）提出“人工智能威胁论”，声称AI将很可能引发第三次世界大战。长久以来，对人工智能发动战争的恐惧一直是马斯克在公共场合发表言论的主论调。2017年8月，他与其他一百多人联名签署了一封建议信，呼吁联合国禁止研发“机器人杀手”（能致人死亡的自动武器）（埃隆·马斯克，2017）。

在研究发展分类上，人工智能研究有两类：通用人工智能（Artificial General Intelligence, AGI）研究与类脑智能（Brain-Like Intelligence, BI）研究。这是完全不同的两类人工智能研究。

通用人工智能特别强调两个特征，即强调与专用领域无关的通用性智能，强调具有多功能协同的整体性智能；但通用人工智能并不否认专用人工智能的价值。目前，通用人工智能的进展很慢，水平很低，离具有人类智能的目标还有很长的路要走。



而类脑智能主要指，“受脑后发，以计算建模为手段，通过软硬件协同实现信息处理机制上类脑、认知行为和智能水平上类人的机器智能。目标是使机器以类脑的方式实现各种人类具有的认知能力及其协同机制，最终达到通用智能”（中科院曾毅研究员语）。类脑智能的进展目前也并不是很大，用目前居于前沿的研究人员的话说，只是万里长征迈出了小小的半步。

就人工智能发展的阶段或目标而言，有三个。

第一，狭义人工智能（Artificial Narrow Intelligence, ANI）阶段，ANI的特征：一是仅具备某项认知能力，实现某项技能；二是无法推广至实现其他功能。所以，狭义人工智能又称“弱人工智能”。例如，AlphaGo围棋程序这种人工智能，就是典型的狭义人工智能。即便是后来全新的强化学习算法阿尔法元——可以从零开始，通过自我对弈精通棋艺，堪称“通用棋类AI”——也是狭义的人工智能。凡是具有特定、专门用途的机器人，都属于狭义人工智能，如语音/面孔识别、人工智能翻译系统、无人机、自动驾驶汽车、工业机器人、家用扫地机器人，等等。目前这个领域的人工智能发展最为迅猛，而且有望在各个领域获得广泛应用。2016~2017年引起社会轰动的也主要是这个方面，这也是人工智能界发展的主流目标。

第二，通用人工智能阶段，又称为强人工智能阶段，即指具有人类水平的人工智能。它的主要特征：一是具备所有的人类认知能力；二是可以由同一智能系统执行不同的认知功能。这个阶段的研究正在展开，但是目前远未达到其愿景目标。

第三，超级人工智能（Artificial Super Intelligence, ASI）阶段，其预期的主要特征：一是在所有领域超过人类；二是所有的认知效率超过人类认知。这个阶段还远未到来，是一种遥远的未来预期。而且，究竟是否需要把人工智能发

展到此阶段，目前存在重大争议。

人们对人工智能的争论恰好就处于这样的境地：身处人工智能发展的第一阶段、初级阶段，即狭义人工智能阶段，而思想已经飞跃到第三阶段，即超级人工智能阶段，想象着这种强大无比的人工智能是否会取代人类，或奴役人类。人们既想发展人工智能代替人力劳作与思考，又担忧有一天强大的人工智能把人类作为自己的奴隶而奴役。大多数人工智能专家不主张发展强人工智能，当然也会有人工智能专家认为应该发展，不用担忧。

关于人工智能发展背后的深刻哲学意义

人工智能发展的本体论与认识论意义。人工智能，是人类创造的机器智能，首先仍涉及这个智能为何种智能的认识论问题——是类似人的智能还是非人的机器或其他类型的智能。这涉及发展人工智能是否必须依赖于人类智能，或者必须仿造人类智能而去“克隆”机器。其次，人类能否发展非人的“其他人工智能”？假如我们不是按照本体论的意义定义人工智能，而是从功能角度定义人工智能，那么，比人解决问题更为快速、更为清晰和更为强大的智能就是我们需要的智能，狭义的人工智能是可以实现这一目标的。智能的核心是意识，是心智，让人工智能具有类似人的意识，或者非人的意识，或者超人的意识，实在是一个极为深刻但又极难的问题。类脑人工智能走的道路是发展类似人类的智能，但是其他人工智能研究走的路线不一定是类人的路线，就功能主义而言，其他人工智能解决某种类型问题的能力，可以肯定的说，必然超越人类；因为，人工智能在可形式化的、可表征化的数据处理方面，的确大大地强于人类，而且这种优势仍然在不断地扩大。这在本体论意义上并不可怕，可怕的是其社会学意义，这意味着它替代了人类许多行业，可能引起大规模失业。而事实上，人工智

能之所以能够让大量的人失业，是因为它确实各个产业内巨幅降低成本，从而增加更大利润空间。在这个意义上说，引起大规模失业，问题不在人工智能本身，而在资本的利润驱使。这类问题是社会如何预先应对失业和风险、管控资本的野心的问题，而不是产生新的“物种”可以奴役人类的问题。

人工智能专家试图在本体论上解决此类问题的研究进路，即类脑人工智能研究进路。所谓类脑，就是让人工智能在结构机理上与人脑相似；在实现机理上全部采用类脑神经网络，使之在网络结构、脑区、神经环路、神经元功能和计算模式上接近人脑，而且从单任务态向多任务、多通道、多脑区协同处理发展。此外，类脑还表现在认知行为上类人，即能感知现实世界，并对环境做出实时反应，自适应能力不断增强；能对不同类型的问题进行识别，设计问题求解步骤，自行解决问题；从自主学习过程上模仿人，包括学习人类动作。通过类人的智能行为实现与现实世界人一机一物的持续交互、不断地自主学习、自主地智能生长与演化。所以，从哲学上看，类脑人工智能的科学发展方向就是通过类脑机理研究达到类人行为上的强人工智能；其他人工智能研究，包括通用人工智能，基本上是一种功能主义的研究，即人工智能在解决问题的领域足以和人类一样聪明，或比人类更强。问题在于，我们人类在面对未来时，是否会满足于有一个只在功能主义意义上解决问题的“伙伴”，这个答案分析起来其实也是极其复杂的。我们手里握有一个智能手机时，通常会满足于它是一个功能主义的“工具”，是为我所用的工具。但是我们对面有一个“人”，它是你的智能助手，或是在未来社会里（假定实现了通用人工智能）你偶遇的一个聊得开心的“人”，你想要判断它是人还是非人吗？这是否是一个问题呢？是否是一个极其重要的先决条件呢？人工智能所涉及的认识论或本体论问题，就是它是否具有人的意识与身体。近年来，

反对人工智能一定要具有人之意识的人，从认知科学哲学角度提出涉身认知观点，认为机器不可能具有人之肉身，因此机器永远不可能有人的意识与智慧。但是如果意识如同软件可以上载到不同硬件上的观点可以成立的话，那么，机器意识依然是可能的。

由于有上述担心，所以有人提出一种解决方案，即以实施人机结合的方式，如脑机接口，生成某种半人一半机器的“人”（赛博格），这被称为“人机协同的混合增强智能”。在某个微信公众号上关于该不该研究强人工智能的争论中，有学者认为，人机融合可以让人的智能和机器的智能协同发挥作用。人有知识，机器长于采集数据；人有经验和常识，机器则长于进行公理推理；人有直觉，而机器长于逻辑。当人和机器有了足够的默契，人能理解机器如何看待世界，而机器也能熟悉人的所思所想，未来的机器也可以有一些特殊特定的意向性（弥补人类认知的不足），而当两者成为搭档甚至知己时，强人工智能也就到来了。“人机协同的混合增强智能”作为人工智能2.0的一个重要方向，旨在通过人机交互和协同，提升人工智能系统的性能，使人工智能成为人类智能的自然延伸和拓展；通过人机协同更加高效地解决复杂问题，具有深刻的科学意义和巨大的产业化前景。人机融合还受到了始于2000年的四大融合技术（纳米技术[Nanotechnology]、生物科技[Biotechnology]、认知科学[Cognitive Science]和信息技术[Information Technology]）观念发展的支持。但是，在本体论和认识论上，这种方案仍然有一些需要讨论的问题。首先，如何融合？是一种松散地联合还是在硬件上紧密地融合？其次，融合中谁起主导作用，是“人的部分”指挥“机器的部分”，还是相反？（Weld D.S.; Mausam C.H.L. and Bragg J., 2015: 89）

由此可见，人工智能的发展在理论上和实践上都面临一大堆问题。那种对于人工智能发



展的极端担心、狂热追捧，以及不加分析的一概而论的意见，都如同一个硬币两面中的一面，都是不切实际的，都无法透彻理解人工智能及其发展带来的社会问题。我们需要细致、深入的思考。

人工智能发展面临的伦理学和政治哲学问题。人们对人工智能发展最主要的担心，是人工智能强大到奴役人类。而关于奴役的思想，源自人类本身。这种担心其实不仅有深刻的人类中心主义观念，而且带有强烈的、根深蒂固的“等级—区隔”的政治哲学观念。在进化论的视野里，人类这种高等灵长类的智能似乎居于所有智能最高等级的位置。人类对其他物种的支配、使用或奴役，似乎有一种自然正确性（natural right）。人类还把这种自然权利置于生命安排中，似乎有一种自然的等级制度或秩序、一种天然的高等生命支配低等生命的权力与秩序。这是人类中心主义的理念。在人类社会，平等、自由本来应该是一个民主社会的共有理念，社会主义社会作为民主社会的一类也不例外。人生而平等，是人权的基本原则。但人类中心主义的“等级—区隔”理念何以渗透到人类社会呢？福柯、阿伦特以及近年来的阿甘本都分析过这种等级镶嵌的权力如何以“政治正确”的名义渗透到生命层面、物种层面和种族层面。不仅如此，人类在自己的生活和社会内部也进一步制造并隐秘地实施这种“等级—区隔”，把社会达尔文主义视为当然。例如，第二次世界大战期间的纳粹，就把犹太人视为低等的人、“人类中的渣子”，把犹太人从人类中“区隔”出去，关入集中营，甚至实行集体屠杀。“人一动物”关系中的等级制人类中心主义在这里变成了“人—人”关系中区隔性的人类等级主义，在这种观念看来，低端的人是可以驱逐的，没有什么不妥，这种思想也可以合适地移植到“人一机”关系中。

人工智能中“人工”（artificial）一词的含义，

已经清晰地表明，人工智能是人创造的，因此，创造出来的人工智能是为人服务的。如吴冠军认为robot应该译为“机奴”，事实上人类一开始就把机器人看成为人服务的“机奴”（吴冠军，2017a）。我们害怕机器人超越人类，对人类实施奴役，而提倡对人工智能的智能进行伦理规范以形成机器人的“善智”，吴冠军甚至认为“善智”实则是让机器人更好地成为服服帖帖的“机奴”。换言之，人工智能作为智能即使再发展，也应该是一种低于人类、为人类服务的智能。所以，这仍然是一种政治学意义上的人类中心主义统治下的人工智能发展方案，或是阿甘本意义上的“人类学机器”的体现（吴冠军，2017b）。

在人工智能的发展中，人们还担心人工智能会取代人类社会的许多职业。人们从许多角度解读这种取代的可能性，预测被取代工作、职业的种类，然而很少有人从政治哲学的视角看待这种可能的取代。

“取代”有两个政治哲学后果。第一，被取代的工作，据说都是低端的、可以成规模的、无法个性化的工作，如出租汽车司机，被取代工作的这类人群，从此成为“闲人”；而且当代所谓高端的工作、职业，如医生、律师，也可能被人工智能取代。这必定使大量的人口落入被取代的人群，从而加速了政治哲学意义上的“等级—区隔”，这意味着，人类的等级从符号性社会学层面转变成成为真实的生物学/人类学层面。第二，取代还意味着种类的变化，这种取代更为可怕，因为它可能是人机协同的混合增强智能取代纯粹生物学的人。加上器官移植、再生医学、基因工程等技术，人之中的一小部分“权—贵”则可能在生物学+人工智能融合的新生命体系中完全不同于现在的人类。所以，如何使全人类具有平等机会享受最新科技带来的利益，才是最新科技，特别是人工智能发展的最好“善治”。

而且，在有平等机会享受人工智能带来的利益的同时，人类需有勇气考虑，人工智能的非

人工智能和非人“意识”的研究是否可能呈现或凸显出来。这是一种非人类中心主义视角的思考。当然，其可能的风险评估也需要在研究进行中提早纳入议事议程。

关于人工智能可能带来的社会问题与社会治理

人工智能可能带来的社会变化与社会问题。

人工智能超越人类、可能奴役人类的问题还可能比较久远，目前对人类社会构不成威胁；而且人工智能的发展并非都是不好的，也有利好人类社会的可能性与现实，所以我们需要提前了解和密切关注人工智能研究的理论进展与实践研究进展。目前大多数狭义人工智能的研究都是基于利好人类社会的工作的，比如智能机器人可以用于救灾，如核泄漏；可以用于家庭服务和养老，如语音/面孔识别、人工智能翻译，还有无人机、自动驾驶汽车、工业机器人、家用扫地机器人，等等。大略来说，人工智能可能给社会带来这样一些变化。

第一，一些重复性强的工业化研究与生产，可能越来越被智能机器取代。其实这个过程早就开始了，比如越来越多的公司，其真正在一线工作的工人越来越少，而且工人也越来越需要知识化。有些行业整体失业发生的概率由于人工智能的介入而加大了。这是第一级效应。这一级效应还可能引发更大的、更深的效应，如“蝴蝶效应”那样波及其他相关行业和社会生活的方方面面。特别是人工智能可能带来工作的本质改变（杰瑞·卡普兰，2016：xviii）。

第二，用于利好人类社会的人工智能也可能被滥用。如智能识别（语音和图像识别），也可能被用于破解识别码、安全锁和网上支付系统。再比如，智能无人机，可以民用，也可以军用，最近一些人士呼吁禁止发展自主性智能无人机，表明这种滥用并非没有可能。这些人工智能在社

会生活方方面面可能滥用的现象，从根本上来说，取决于高端的人工智能研究人员的职业伦理和职业行规。

第三，人工智能的使用可能带来人类社会等级区隔的加大。如果马斯克等人设想的人机融合的四大技术能实现，那么就可能在资本一权力的人为社会等级之上，再进一步增加生物—机器智能的生物等级区隔，由此造成不同的“新人类”。这些“新人类”的出现虽然就其整体而言似乎还是个神话，但是局部的生物—机器融合其实已经开始，如给大脑植入芯片、给肌肉加装增强技术产物。这种局部的人机融合的“赛博格”的出现，有极强的两面性：一方面它可以补全患有盲视、失聪、肢体缺失以及瘫痪的人的机能；另一方面，当它用于所谓的健康人身上，则极大地增强了人的能力，使人类物种中出现了一种“增强人”——赛博格，如同体育比赛中吃了增强药物的运动员，可以跑得更快，跳得更高，有更为持久耐用的身体。这对没有吃增强药物的运动员的公平性可想而知。表面上带来的新的社会不公平问题，实际上是造成了人类物种的新生物—机器的不平等区隔。

人工智能发展的社会治理。如上所说的人工智能发展的社会问题只是冰山一角，人工智能的发展势不可挡，因此有必要对人工智能的发展未雨绸缪，及早做出可能的社会治理预案。

第一，谨慎研究和发​​展人工智能，建立有区别的人工智能研究与发展行业政策。

当代人工智能的研究势在必行。各国在权力与资本逐利过程中，不可能不发展人工智能，就科学研究而言，也不可能研究上过早设置禁区。人工智能不同类型的研究（如狭义人工智能、通用人工智能、类脑人工智能、人机结合等），不同发展阶段的研究（弱人工智能、强人工智能、超级人工智能），不同行业的研究（一般工业、特殊工业、民用和军用），都应该建立相应的人工智能研究政策、行业规则以及相应的



伦理规约。我们不仅要关注国际人工智能界的研究进展，同步建立相应的研究与发展政策，而且应该提早研究由研究和发​​展人工智能带来的各类社会风险问题，特别是人工智能在应用到不同行业、不同领域和不同部门时可能带来的社会风险（物质安全、信息安全和网络安全）。在可以大力发展人工智能的领域，也需要研究相应的可能的社会风险。

第二，认真研究和注意人工智能发展的环境代价与效应。任何工业制造业都有环境效应，发展人工智能产业可能会有更大的环境效应，这要根据国情综合考虑。如果发展代价太大，就需要对其发展有所限制，特别需要提早考虑可能的环境代价。

第三，特别需要注意研究人工智能发展的政治哲学后效。前文已经论证了人工智能作为“机奴”可能的政治哲学后效，有两个需要提早进行规约的方面。

其一，时时刻刻注意人工智能应用的社会平等性。无论是经济方面还是政治方面，都需要给予社会公民平等享用人工智能利好的机会。当然这主要是政府需要注意的事情，通过汇率或利率调节的方式，向不能直接享用人工智能利好的阶层转移支付利好比例。就人工智能的研究者而言，也需要注意，不要让人工智能研究带来“等级一阻隔”的扩大与延伸。

其二，当人工智能发展存在取代某些行业/职业或人群的可能性已经出现时，智库类研究机构和政府就需要做出评估，并且对这些行业/职业涉及人群的可能转型给出合理建议与安置。

涉及人工智能发展的社会治理是一个极为复杂的巨系统工程这种说法，实际上缺失了人文和社会思考的维度，而且常常容易使人把这种问题理解为可以按照工程和技术的方式加以解决。事实上，这种工程不是技术本身可以解决的，需要人文、社会科学多学科的交叉和合作研究，需要

科学、技术、伦理、政治学（包括政治哲学）的多学科视野，我们特别建议，建立相应的全国委员会，除人工智能专家外，还需要包含哲学家（伦理学家、科技哲学家），特别是政治哲学家。

如若在起始阶段，由人类研究的人工智能可以自主地发展出非人类的智能与意识，那人类能够容忍其自主发展吗？如若人类不容忍，那此时不容忍是对的么？如若这时的人工智能已经比人类更为聪明，那人类还要消灭它们吗？人类还能消灭它们吗？比人还聪明的人工智能此时还是人工智能吗？它要奴役、消灭人类还是与人类合作？它把人类看成为伙伴吗？它会建立起一套相应的“社会政策”与“伦理原则”吗？这些问题虽然还很遥远，但可以在将心比心的思考中帮助我们提前理解当代人工智能发展的哲学问题。有人就认为，一个独立的机器世界由于是人类建立和发展起来的，所以必然继承了人类文明，那么由机器掌管整个世界又有什么不好呢？（集智俱乐部，2016: 208）继续在理性与情感、自我与他者的交集中争论人工智能的哲学问题，是极其有益的。

参考文献

M. Negnevitsky, 2005, *Artificial Intelligence: A Guide to Intelligent Systems*, Second Edition, Boston: Addison-Wesley.

R. Neapolitan and X. Jiang, 2012, *Contemporary Artificial Intelligence*, Florida: CRC Press.

Weld, D.S.; Mausam, C.H.L. and Bragg, J., 2015, "Artificial Intelligence and Collective Intelligence", in Thomas M. and Michael B., (eds.), *Handbook of Collective Intelligence*, Cambridge (Massachusetts): the MIT Press.

[美] 埃隆·马斯克，2017，《人工智能将引发第三次世界大战》，<http://www.24en.com/e/DoPrint/?classid=66&id=196183>。

[英] 玛格丽特·博登, 2017, 《AI: 人工智能的本质与未来》, 孙诗惠译, 北京: 中国人民大学出版社。

[美] 杰瑞·卡普兰, 2016, 《人工智能时代》, 李盼译, 杭州: 浙江人民出版社。

集智俱乐部编著, 2015, 《科学的极致——漫谈人工智能》, 北京: 人民邮电出版社。

集智俱乐部编著, 2016, 《走近2050: 注意力、互联网与人工智能》, 北京: 人民邮电出版社。

出版社。

吴冠军, 2017a, 《人工智能与未来社会: 三个反思》, 《探索与争鸣》, 第10~13页。

吴冠军, 2017b, 《神圣人、机器人与“人类学机器”——二十世纪大屠杀与当代人工智能讨论的政治哲学反思》, 《中国青年哲学论坛(2017)暨首届贺麟青年哲学奖评审会议论文集》, 第155~176页。

责编/刁娜

Several Philosophical Reflections on the Development and Governance of Artificial Intelligence

Wu Tong

Abstract: According to the current situation of the development of artificial intelligence (AI), this paper mainly analyzes the following issues: first, the concept of intelligence and AI; second, the philosophical significance of the development of AI, including the ontological and epistemological significance, and the ethical and political philosophical issues affecting the development of AI; third, the social governance problems that AI may cause, and several governance suggestions that can be used for reference.

Keywords: Artificial intelligence, philosophy of science and technology, significance, governance suggestions