

人工智能的主体地位探析

张友恒

武汉理工大学马克思主义学院, 湖北 武汉

收稿日期: 2022年9月28日; 录用日期: 2022年11月1日; 发布日期: 2022年11月8日

摘要

对于主体的资格判定通常都归结于人类自身, 哲学史上对于人的主体地位的确立是在马克思那里才得以实现。但随着科技的发展, 人的主体地位的至上性受到了来自智能机器的挑战。目前学界对于智能机器能否获得主体地位多持否定态度, 但仍有诸多学者对智能机器可以成为“新主体”给出见解。本文从科学哲学的角度对人工智能的主体性研究进行分析, 并说明人工智能机器可以成为主体的关键性因素与可能的基本设定。

关键词

人工智能哲学, 主体, 自我意识, 人工道德体

Artificial Intelligence Subjectivity Probe

Youheng Zhang

School of Marxism Studies, Wuhan University of Technology, Wuhan Hubei

Received: Sep. 28th, 2022; accepted: Nov. 1st, 2022; published: Nov. 8th, 2022

Abstract

The determination of the qualification of the subject is usually attributed to human beings themselves, and the establishment of the subject status of human beings in the history of philosophy was found only in K. Marx. However, as technology advances, intelligent computers have begun to threaten the supremacy of human subjectivity. Although the majority of academics now hold a negative view of intelligent machines gaining “subject status”, many continue to express their views on whether intelligent machines can become “new subject”. From a standpoint of the philosophy of science, this paper analyzes the research on artificial intelligence’s subjectivity and describes the fundamental factors and basic settings in which an artificially intelligent computer can become a subject.

Keywords

Philosophy of AI, Subject, Self-Conscious, Artificial Morality

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).

<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

1. 引言

二十世纪四十年代计算机的问世引起了巨大的轰动。在达特茅斯会议中麦卡锡提出了人工智能概念，其发展为一门独立学科至今不过百年，却已取得了举世瞩目的成就。在人工智能的历史上，早在二十世纪五十年代，首个人工智能语言 Lisp 就被发明并可以用于实现自修改应用程序；人工智能第二次被称为前沿技术是在二十世纪七十年代，当时专家系统(Expert Systems, ES)被开发出来；随后日本政府举国之力以 Prolog 语言开发第五代计算机系统(FGCS)，又一次点燃了人工智能的浪潮；在当今时代，人工智能以机器人应用、智能数据分析、自我学习等广阔的应用处在时代的浪尖。愈加自主化的人工智能机器能否具备主体地位一直都是人工智能哲学的热门话题，那么人工智能机器有可能获得主体地位吗？

2. 主体概念的内涵及其历史演变

“主体”(“subject(英)、subjekt(德)”)这个概念是近代哲学的产物，“subject”在中文里可以被译为主体、主词、主项。其实在古典哲学时期，亚里士多德就已在《范畴篇》中有提到过类似于主体的概念，即ὑποκείμενον，此词更多指的是一种实体，但也包含着“主词”和“基体”的含义[1]。明确的主体概念源于笛卡尔《第一哲学沉思录》中的“我思故我在”。笛卡尔解释道：“我是一个在怀疑，在领会，在肯定，在否定，在愿意，在不愿意，也在想象，在感觉的东西。”[2]即“我”是一个具有主观能动性的东西，也是一个有意识、有情感、能思考、在感觉的东西。笛卡尔的这个观点开启了近百年经久不衰的人本主义思潮，从“我”出发来对一切事物进行反思，由此确定了哲学中“我”作为主体的必要条件和性质；而在黑格尔那里，实体即是主体，也即绝对精神，其作为实体既是可以承担的基质、也是作为主体具有的能动意识；马克思在其著作中把这个词分别与“objekt、Pradicat、Accidenz(德)”相搭配，其含义可以分别为主体、主语与基体[1]。在《黑格尔法哲学批判》一书中，马克思对黑格尔主体观念进行了彻底的批判，指出其主谓颠倒的实质，并创造性的提出了“主体-客体”的说法[3]。在创立唯物史观前期，马克思对费尔巴哈的人本主义和黑格尔的“实体即主体”的形而上学体系进行批判反思，从抽象理性过度到现实感性，指出现实的主体应为“人民理性、市民社会、类本质”。在唯物史观时期，马克思则打碎了形而上学思辨产物的“美梦”，尖锐地呼唤出“历史主体”的出场，即具有能动性和阶级性的“无产阶级”和“劳动者”主体[4]。

在不同的领域中，主体都具有其特定含义，因此我们可以区分出不同类型的主体，如认知主体、行动主体、伦理主体、法律主体等[5]。在关于主体的讨论中，哲学重点关注的是认知主体和伦理主体两类，因为其所对应是认识论和伦理学两大重要板块。无论是在认识论还是伦理学中，所讨论的主体都是人本身，“这个概念要解决的问题，是一个‘谁’的问题：是谁在认识、思维，是谁在行为，依据这个‘谁’，就可以界定主体和客体”[6]。笛卡尔的人本观念否定了上帝对人的一切主导性，从繁琐的经院哲学中脱离出来，对一切所被灌输的“不证自明”的知识进行怀疑，直到怀疑到不可怀疑的“我思”才是“我”

所得到一切认识的开端。经过康德、黑格尔、费尔巴哈、马克思对主体观念的充实和改造，才以唯物论的观点确立了客体来源于主体的经验，而唯一能经验到客体并获得知识的只有人。“主体的问题，存在于人对自己存在的各个层次的矛盾的认识和规定中，在各个层次对主体的认识又是内在统一的，因为人的存在就是一个总体，各个层次是总体的人为了认识自己存在而按照抽象程度区分的。”[6]。作为认识主体的人首先获得的是对自身的认识，这种认识不仅仅限于物质层面还表现在意识层面，例如对于自身存在的认识和自我价值的认识等。这些非物质性的认识，即个人的世界观、价值观、人生观等意识内容，可以指导人的行为，从而引出了主体概念的另一个重要含义，即伦理学中的道德主体问题。“道德主体”是指具备道德意识、可以进行道德判断和推理、能进行道德行为并且为自身行为负责的社会主体。对于“道德主体”而言，其首先必须拥有自由意志，一切行为的动因来源于自身而不是外部；其次，它必须可以为来自于自身的行为负责；最后，它还必须具有道德意识。对此，肯尼斯艾娜·希玛(K.E. Himma)指出：“对于缺乏意识心理状态的事物，无论它的计算能力在其他方面有多么复杂，赞扬或谴责都是没有意义的。赞扬、奖励、谴责和惩罚是理性的反应，只有对那些能够经历骄傲和羞愧等意识状态的人。”[7]。道德意识的作用在于评价某行为是否符合道德标准，从而可以规范主体自身行为以及决定主体是否进行某些行为，从而引出主体也是必需在有意识的精神状态下才能进行评价活动以及实践活动。

由此可见，成为主体的首要条件就是必须能够拥有意识，其次必须是一个自由行动者，最后还必须能为其行为负责，而满足这诸多条件的只有人本身。然而，随着近年来人工智能技术的快速发展，智能机器表现出了越来越强大的信息处理能力和自主化的智能行为，人的主体地位正在遭到挑战，这引起诸多学者对人工智能未来发展的可能性进行深度思考：是否会出现人工智能代替人的现象，是否会出现无法掌控的人工智能，是否会出现成为独立主体的人工智能等诸多问题。诚然，这些问题背后不仅有其哲学依据，更有现实的意义。我们可以看到目前的人工智能汽车对于目标的识别错误而造成的交通事故、军用无人机对于定点目标摧毁而忽略生命识别、商业公司利用算法对客户个人权益侵犯等事件。但是任何智能机器一旦设定和制造完成，它的存在不仅是由于设计者、工程师、数据提供者、技术人员、制造者，甚至还由于监管机构、政府部门、经销商、使用者等人的共同目的而导致的，并且随着计算机技术日趋复杂，其中任何一方都无法直接决定何种智能机器应该产生，人类直接控制或干预这些技术行为的能力也就越来越少。具有自主性的机器的产生是不可避免的，但是我们要注意的是这种“自主性”与人的自主性并不相同：在计算机科学中，它所指的是智能机器或计算机系统不可预测的环境中长时间独立执行复杂任务的能力。因此，这使得人工智能的主体性问题成为了人工智能哲学中最具有争议的问题。一方面，对于人工智能所评估的伦理、政治、法律问题皆是基于人工智能可以成为某种形式的主体的基础之上，如果无法出现具有自主能动性的人工智能机器，那么对于人工智能的思考就变成了对“非智能的纯机械”或是“带有智能的算法”的思考，这样对于人工智能的发展研究会出现很大的局限性。另一方面，如果一台智能机器可以成为某种形式的主体而不把其作为主体，那么由其引发的伦理、法律、政治等问题将被追究于机器的创造者或是数据的提供者，这样对于技术创新和科技的发展将有严重的打击。因此，拥有主体性是机器走向真正智能的标志，是人工智能哲学的基本问题[8]。

3. 从人工智能研究范式看其主体资格

现代人工智能的发展使意识的属人性或是说“人作为自然中唯一具有主体地位”这个观念，遭到了巨大的冲击和挑战[9]。这一现象的出现源于人工智能在算法和应用上的大力发展，还有科幻小说家、影视作品对于人们传统观念的变革。在人工智能领域研究的主有符号主义(也称逻辑主义)和联结主义(人工神经网络)。符号主义理论主要应用于人工智能的决策和推理，联结主义理论主要应用于人工智能的识别功能。在不同的理论体系下也发展出不同的算法模式，用于被设计的人工智能机器从事不同领域的活动

[9]。人工智能在其活动中行为问题的探讨来自于机器人伦理学与机器伦理学。机器人伦理学研究的是作为特殊设备的机器在应用方面所产生的伦理问题，而机器伦理学研究的是如何使赋予自我调节的机器以道德行为[10]。对于这种人工道德主体的设计主要有由瓦拉赫(W. Wallach)和艾伦(C. Allen)等人提出的“自上而下”和“自下而上”以及两者综合的三种进路：“自上而下”伦理理论将道德法则或“善”定义为机器人所能接受的某种客观的东西；“自下而上”伦理理论认为是人工主体选择了价值观从而指导其行为，并在经验的学习过程中逐步完善这些价值观；“在建立特定的机器伦理解决方案时，这两种二分法也可相互综合”[10]。在主要的两种进路中，前者是将道德法则作为“设定”嵌入在人工智能的运行框架中；而后者则是通过人工智能的自主学习，从而拥有道德的意识。“如果机器越来越智能、具有自主性，那么让其具有一定的道德判断和道德行动能力，是未来人工智能伦理无法回避的议题”[11]，因此对人工智能主体的道德问题的担忧并非是空穴来风。人工智能的研究范式共同关注的问题仍归属于意识与大脑、意识与语言、意识与思维以及意识与实践之间的交互关系问题[12]。

人工智能的符号主义范式主要来源于纽厄尔(A. Newell)和司马贺(H.A. Simon)的工作，《作为经验探究的计算机科学：符号和搜索》这篇文章可以作为其正式确立的标志。作为人工智能研究的首个基本纲领，它在上世纪50年代到90年代一直占据学界优势地位。符号主义主张：人类所获得的一切认知都可以看作某种信息，这些信息可以用语言或者非语言的逻辑符号来进行表示，那么人的认知过程就是符号处理的过程[13]。既然人的认知过程被认为是一个物理符号系统，那么以物理符号系统为基础的计算机就可以对其进行模拟，并通过这种模拟方法赋予智能于程序。这种智能实现的方法是通过启发式搜索来完成的，但启发式算法的问题域是圈定范围的。因此，要使符号主义人工智能能够成为主体，首要条件就是这个问题域的范围要足够大，其次则是算法的逻辑要能应对这样范围的问题域。在这种模拟方式中，符号主义流派所接受的人工道德体的设计就为“自上而下”进路，但其问题在于“道德可以模拟为经验性的符号语言被接收从而使主体拥有道德吗”？首先，道德标准不是普适的，不同标准中具有差异性，“即使选取最大公约数，让人工智能遵守某种约定俗成的基本道德法则，但若是让人工智能遵守更为精确的道德法则，仍然会引来各方的争议”[11]；其次，对于伦理和道德的客观性我们并未获得统一性的理解，基于“0”和“1”的算法设定未必能完全描述出尽可能完备道德法则。符号主义在研究发展过程中被联结主义所“击败”自然有其原因：“规则永远无法被定义完全或囊括穷尽，不管可以为人工智能输入多少符号信息、组成多大的逻辑系统，但是也总有遗留在此框架之外的东西”[14]。符号主义把人的认知和行为简单的还原为数理的逻辑符号，虽然其具备一定可行性和实现性，但是“人工智能的目标是为了其能实现并拥有跟人一样的情感、意识和能力”[13]，因此经验性的符号主义能否达成这一目标，仍取决于符号主体能否实现对客体进行更完备的知识表示。

人工智能的联结主义纲领源自认知科学中的一种运动，其试图通过模拟人类大脑的神经元从而使机器模拟出人的智能与认知能力。随着脑科学的研究发展，联结主义也实现新的突破，2006年深度学习算法的破壁使其拓展到更广泛的应用平台中。“alphago Lee”击败人类顶尖围棋选手已经让我们瞠目结舌，而抛弃人类经验和“从零开始”自我学习和训练的“alphago Zero”则是让我们惊艳于深度学习算法与强化学习算法的强大能力。但联结主义依赖于脑科学对脑部的研究深度，而大脑对于人类仍是一个黑箱，因此它的发展仍有一定限制[15]。与符号主义对心灵的模拟不同，这种计算神经节点对于神经元的数字模拟实际上是限制了智能的范围，由此引发出“无意识的智能”以及“行为智能”等问题。目前人工神经网络研究中偏重卷积神经网络层数，因此联结主义的主体智能需要强大的硬件能力保持算力，由此可见，联结主义的主体实现的“条条框框”并不会比符号派少。在联结主义的人工智能对大脑的模拟中，道德意识实际上被作为是主体本身所可以具有的、通过学习所强化的而构成的一个框架，因此所接受的是“自下而上”的人工道德体的设计。但这种设计仍有诸多问题：首先，“人与人工智能的学习方式有根本性

差别,人的学习过程以语言和符号为中心,就道德学习而言,概念在人的道德判断中起到决定性作用”[11],人是在与社会的相互作用中获得道德意识和道德情感,而基于数据学习的人工智能是否能获得“仁、义、礼、智、信”、“羞耻、惭愧、不安”等这样的道德意识从而激起道德情感还很难说;其次,即使可以将这些道德意识和情感数据化的表述便于人工智能学习,但是在无相互作用的对象的情况下,这些道德意识和情感并不能获得任何指向性。意识的产生从其本质上来说是人脑这个高密度、极复杂的融合体的“飞跃”,人脑万亿个神经元的模拟难度不言而喻,况且机械性的神经元与生物性的神经元是否具有同一性还很难说。即使脑科学的研究出现重大突破,但是由于深度学习算法需要广泛的数据支撑以及其作为“无监督特征学习”特征的黑箱性质,从机械化大脑到生物化的大脑并不能一蹴而就[12]。

4. 人工智能何以获取主体资格

“人之所以获得主体地位,是因为人类将自己当作目的,而不是像一般动物一样仅仅依据本能生活;如果人工智能的发展使得自己具有目的意识,不再将自己视为人类的工具或者创造品,那么,我们可能就不不得不承认人工智能具有主体地位。”[5]人工智能成为主体的构想可以归结为通用人工智能的“圣杯”,这并不是一个幻想,而是初期人工智能专家们的研究目标,但是这个目标是如此的遥远,以至于目前大多人工智能专家闭口不谈,从而认真耕耘人工智能学科更细致的领域。如果我们假设这个“圣杯”有一天真的能够实现,那么何以判定某个人工智能可以被称作主体呢?这显然不是一个科学问题,而是哲学问题[16]。回忆对比一下两个古老的案例:作为人工智能以及人工智能哲学的开创者,图灵(A.M. Turing)在《计算机与智能》此文中以“模仿游戏”为例,从功能的方面阐述了电子计算机可以获得思维;而没有计算机学科背景的哲学家塞尔(J.R. Searle)的“中文屋”,则是从语法和语义做出区分,也就是从功能的角度反驳人工智能的“思维”。思维就是意识,不存在无意识的思维或无思维的意识,人工智能哲学问题的争议不仅存在于两种不同的学科背景之中,也存在于哲学本身的流派之中,如“分析哲学与现象学对人工智能的争论,一方以科学分析的态度来对待这个问题,另一方则是反对用有意识的科学意识去解释意识其本身”[16]。本文无意详细探讨哲学本身的争论,而更赞同以科学的“分析”来对待人工智能主体问题,即从功能角度分析人工智能的意识、自主性以及责任问题。

对于意识的探究一直存在于哲学领域,没有哲学家对意识的理解能让所有人的信服,为避免这个无休止的争论,对人工智能的意识探讨应该限于“机器意识”的范围内。对于生物能有意识是毋庸置疑的,但是要讨论某个软件或硬件可以有意识则必须从其显现出的行为过程(运行过程)来推论。因此“机器意识”则是从功能方面对意识的模拟,但“机器意识”并非超出了人的意识,“如果要创造具有意识的人工智能,就只能以人的意识作为范本”[17]。我们无法从计算术语中分析的争论中得出意识到底是什么,但意识的功能则可以从人自身得出,“如意向、思考,反思、创造等能力,都可以作为意识的功能表现”[16]。巴尔斯(B.J. Baars)曾提出“全局工作空间理(GWT)”作为功能整合的认识架构,把意识的功能整合在一个框架系统内,从而避免了一些哲学争论议题,“GWT将大脑视为一个分布式系统,其中并行运行的大量专用子系统竞争访问工作记忆”[16]。鉴于人工智能的目标不是创造人造人,而是拥有智能行为的软件或硬件,因此通过将意识的功能分布为各个模块加以整个是个理论上可行的方案。人工智能可以有意识,因为我们在制造它的时候有意向,预想它完成某种任务,那么就是人的意识参与进去了[14]。

人工智能自主性问题可以被归结为自由意志问题。目前的弱人工智能程序一旦设定好便决定了它是如何运行的,在任何情况下都不会产生本质性的变化[5]。通常认为只有人类才具备自由意志,但是人类到底有什么能力可以使其自由行动且是其他实体所不具有的?这种能力到底是源于理性、情感、意向性还是认知?哲学界似乎并没有达成共识。在伦理学中的一个重要议题就是人类如果没有自由意志,那道德行为又该归因于何物呢?人工智能的无自由的想法,实际上是一种被设定的决定论。但是在决定论的

世界里人类也不具有自由意志，一切皆是由因果关联所引起。若是持此观点，人类的“程序”似乎也是被设定好的，人类的行为也是被决定的。即使可以不假思索的把自由意志直接归因于人自身，但问题在于并非每个个体在社会中都具有同样充分的“自由意志”。成年人通常能比幼儿拥有更多程度的“自由”，因为成年人可以有更多的自主权、可以更少的被外界所影响；有心理疾病或是精神问题的人也比正常人拥有更少的“自由”，因为其行为和判断会受到精神状态的影响。因此，人类在不充分的自由意志下似乎也可以获得的自由，那么以功能为导向的人工智能同样可以获得。在此笔者赞同蔡恒进教授所提出的内核(seed)构想，“这种机器主体性的内核就像是一颗具有自我意识的种子，在人工智能与外部交互环境中就能够自然成长[14]。

人工智能的责任问题可以归结为人工道德体(AMAS)问题。在构建 AMAS 的问题上，哲学界的反对声音是大于支持的声音，这些反对虽然具有破坏性，但反对者们没有给出任何新的建设性理论[18]。对 AMAS 的建模并非易事，责任问题会涉及到主体的内在道德和外在道德，且还必须考虑人工智能技术问题。菲利克(A. Kuflik)曾认为：“智能机器无法接受惩罚，因此就无法承担责任” [18]，这个想法的指向并非是今日的弱人工智能而是未来的强人工智能。道德责任会涉及到自我、身体、心灵、意识、自由等诸多哲学问题，因此，我们同样可以反问，到底什么才是主体承担责任所必要的？人类可以“惩罚”其他生物，是因为它们可以承担责任吗？从功能的角度来看，“惩罚的关键并不在于被惩罚者能否体验到威慑，而在于对其后续行为产生的实际影响” [19]。在惩罚关系中，惩罚者和被惩罚者是对立的两端，而惩罚手段则是惩罚关系的具体体现。对人工智能来说，重要的是如何对惩罚手段做出正强化或负强化的设定。今天的强化学习算法发展使得人工智能通过训练能做出相对正确的决定，从其行为看来已经能起到正强化的作用。图灵在人工智能研究初期就已提出：“机器的构造必须使惩罚信号发生前不久发生的事件不太可能被重复，而奖励信号则增加了导致惩罚信号发生的事件的重复概率” [20]。

5. 结论

“主体性意味着主体自身是自律的、自为的，这种主体性应该是从自身中生长出来、自我意识到的，显然这种自生的主体性在机器没有诞生意识的情况之下是不可能的。” [8]但是对人工智能的未来发展持悲观态度却还为时过早。科学研究要实事求是的以唯物主义的角度来看待研究的进展，过去的研究不仅是指向当下的成功，还应该指向更远的未来。对于通用人工智能的设计方案，一部分人工智能专家早已提出并在进行实验。这些学者所持有的基本态度可以归结为：假若人工智能无法获取主体资格，就如我们日常使用的风扇、空调等电器一般，其仍是被当作工具和手段，而研究的目标应指向的是学科的“圣杯”。不可否认作为工具的人工智能为我们的生活带来了诸多便利，但是其显现出的危险也日益揭开。过去许多科技，例如核电、克隆、转基因等，虽然其危险性已经引起了我们的讨论，但更多的是对这些技术进行监管和限制，若是技术实现了具备主体地位的人工智能，其所引发的认知、法律以及伦理问题并不是通过监管和限制就能简单解决。因此，无论是哲学还是人工智能学科对于人工智能获得主体设计方案，并非是无意义的猜想。凡是过往、皆为序章，人工智能的未来永远是危险和希望并存。

基金项目

中央高校基本科研业务费专项资金资助(项目批准号：2021VI011)。

参考文献

- [1] 徐长福. 论马克思早期哲学中的主谓词关系问题——以《黑格尔法哲学批判》为解读重点[J]. 哲学研究, 2016(10): 22-30.
- [2] 笛卡尔. 第一哲学沉思集[M]. 北京: 商务印书馆, 1998.

- [3] 马克思, 恩格斯. 马克思恩格斯全集 第3卷 1842年11月-1844年8月[M]. 第2版. 中共中央马克思恩格斯列宁斯大林著作编译局, 编译. 北京: 人民出版社, 2002.
- [4] 孙琳. 唯物史观视域中的“实体”与“主体”之辨[J]. 学习论坛, 2021(6): 108-115.
- [5] 孙伟平, 戴益斌. 关于人工智能主体地位的哲学思考[J]. 社会科学战线, 2018(7): 16-22.
- [6] 孙晓飞. 主体、主体性概念析辨[J]. 社会科学论坛, 2019(1): 125-136.
- [7] Himma, K.E. (2009) Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to Be a Moral Agent? *Ethics and Information Technology*, **11**, 19-29.
<https://doi.org/10.1007/s10676-008-9167-5>
- [8] 陶锋. 当代人工智能哲学的问题、启发与共识——“全国人工智能哲学与跨学科思维论坛”评论[J]. 四川师范大学学报(社会科学版), 2018, 45(4): 29-33.
- [9] 肖峰. 人工智能与认识主体新问题[J]. 马克思主义与现实, 2020(4): 188-195.
- [10] Lumbreras, S. (2017) The Limits of Machine Ethics. *Religions*, **8**, 1-10. <https://doi.org/10.3390/rel8050100>
- [11] 阮凯. 机器伦理学的当代争议及其解决方案[J]. 自然辩证法研究, 2021, 37(11): 42-48.
- [12] 刘宇. 关于人工智能认知限定与限度的多重思考——基于马克思主义认识论的分析视角[J]. 广西社会科学, 2021(2): 76-83.
- [13] 林艳. 人工智能的符号主义纲领及其困境[J]. 求索, 2019(6): 186-193.
- [14] 蔡恒进. 行为主义、联结主义和符号主义的贯通[J]. 上海师范大学学报: 哲学社会科学版, 2020, 49(4): 87-96.
- [15] 王广赞, 易显飞. 人工智能研究的三大流派: 比较与启示[J]. 长沙理工大学学报(社会科学版), 2018, 33(4): 1-6.
- [16] 玛格丽特·博登著. AI人工智能的本质与未来[M]. 北京: 中国人民大学出版社, 2017.
- [17] 赵汀阳. 人工智能的自我意识何以可能? [J]. 自然辩证法通讯, 2019, 41(1): 1-8.
- [18] 王淑庆. 人工智能能动性的三种反驳进路及其价值[J]. 哲学研究, 2021(4): 119-126.
- [19] 吴童立. 人工智能有资格成为道德主体吗[J]. 哲学动态, 2021(6): 104-116.
- [20] Turing, A.M. (1950) Computing Machinery and Intelligence. *Mind*, **59**, 433-460.
<https://doi.org/10.1093/mind/LIX.236.433>