人工智能社会学的行动论视角

——人工智能体(AI Agent)技术对社会学理论的启示

陈　氚

摘要：人工智能体技术将极大提升人工智能独立行动的能力。人工智能体的方法论和技术视角给社会学研究带来新的启示，应当把看待人工智能的方式从工具转变为社会行动者。在经典社会行动理论看来，社会行动主体的变化将带来社会结构层面的系统变化。人工智能成为新社会行动者带来的影响包括：在微观社会行动层面，将提出关于人工智能行动类型，尤其是人工智能理性行动与人类行动者行动类型差异的议题；在社会互动层面，将引发人与人工智能的社会互动和主体间性研究；在宏观社会结构层面，将带来关于人工智能适应、重塑和校准社会规则与社会制度的一系列研究。人工智能不仅能够通过对人的影响间接影响社会，而且能够作为行动主体直接影响社会变革。

关键词：人工智能；人工智能社会学；人工智能体；新社会行动者；信息行动理论

DOI：10.19836/j.cnki.37-1100/c.2026.03.009

在人工智能研究领域，人工智能体(AI Agent)技术的应用与研究正成为未来发展的重要趋势。目前大部分关于人工智能的社会学研究都是以算法社会学为主要进路，重点关注人工智能的算法部分、算法的社会影响以及社会因素对算法的作用，而较少关注作为算法、数据、算力硬件整体的人工智能系统。算法研究抓住了人工智能的关键要素，以及人工智能作为计算机程序的逻辑实在，强调人对算法的设计与操纵，以及算法作为强化资源和能力不平等的一种权力工具，对社会产生的各种影响[①]。这种算法社会学视角并不会过时，因为在计算机算法应用于社会场景的背后，体现出人与人之间的社会关系。同时，在人工智能的独立行动能力尚未完备之前，人工智能是人的工具性产物。对算法和技术而言，研究者应当将其视为工具，不应过高估计其独立的自主性和能力。但是，随着人工智能技术的高速发展，有必要进行一种前瞻性的研究，在不否定传统算法社会学的基础上，可以容忍一种不同于算法社会学的视角：将人工智能视为社会行动者，从经典的社会行动理论这一自社会学诞生之初就存在的理论视角，重新审视人工智能问题，重新考虑数字社会中的根本性问题，即不仅是人，人工智能同样具有成为社会行动者的可能性。新的社会行动者的加入可能带来从微观社会行动到宏观社会结构层面的一系列变化，对人类社会产生深远的影响。

一、人工智能体——新社会行动者的可能性

(一)智能体的社会学内涵

智能体并非一个新概念，而是自人工智能诞生之初就存在的理念。在经典的人工智能教程《人工智能：现代方法》中，罗素和诺维格明确地提出了“理性智能体”的概念，并将其确定为“研究人工智能的方法的核心”[②]。所谓智能体，英文单词为agent，在拉丁语中为agere，意为“做”，就是指某种能够采取行动的事物(能动者)。在人工智能学科中，“任何通过传感器感知环境并通过执行器作用于该环境

作者简介：陈氚，中共中央党校(国家行政学院)社会与生态文明教研部社会学理论教研室副教授(北京 100091；cxh100775@163.com)。

① Burrell J., Fourcade M., “The Society of Algorithms”, *Annual Review of Sociology*, 2021, 47, pp. 213-237.

② 斯图尔特·罗素、彼得·诺维格：《人工智能：现代方法》(第4版)，张博雅等译，北京：人民邮电出版社，2022年，第32页。

的事物都可以被视为智能体”[①]。智能体具有机器感知能力，同时具有作用于外界环境的能动性。这一定义在某种程度上接近于社会学的经典方法论个体主义视角下的行动者（能动者）[②]定义。例如，在吉登斯的结构化理论中，主体的能动行为这一概念即为agency[③]。只不过人工智能学科的智能体概念，将传统的人类行动扩大到非人的领域：在这一概念下，人可以是agent，机器也可以是agent，人与非人均可以被视为行动者。这也表明，从人工智能技术诞生之初，人工智能科学家所追求的就是创造出一种可以具有行动能力的、独立于人的、具有智能的行动者。

为避免语义混乱，可以对人工智能和人工智能体作出进一步的界定。除了作为学科的人工智能外，人工智能还是一个宽泛的概念，对应于人类智能。而在日常生活用语中，人工智能指代一类可以让人造物产生智能的事物集合。智能体则是在具体应用场景中指代人工智能行动者，与社会科学术语中的人类行动者相对应，强调二者的能动性。

从人工智能学科来看，智能体具有双重含义。一方面，智能体，尤其是理性智能体是人工智能学科研究的基本方法论。根据罗素和诺维格的总结，人工智能研究在历史上有四种方法论，分别是图灵测试方法（类人行为）、认知建模方法（类人思考）、思维法则方法（理性思考）、理性智能体方法（理性行为）。近年来，基于概率论和机器学习的方法可以使智能体在不确定条件下作出决策，以获得最佳期望结果。理性智能体的方法逐渐成为现代人工智能的主流，而人工智能研究追求的标准模型也体现为“研究和构建做正确的事情的智能体”[④]。在这一意义上，人工智能学科研究出来的具体“机器”或者“产物”，就是“智能体”。这一智能体在不同研究领域和使用场景中表现不同，既可以在数字空间中体现为一段程序或代码，又可以体现为具身智能的智能机器人。

另一方面，在具体的技术层面，智能体技术是一种使人工智能具有更强独立性、自主性和具备行动能力的技术[⑤]。在2025年，这一智能体技术主要是将大语言模型等多个模型系统化，让人工智能模型具有环境感知、学习、记忆和行动能力。现实的行动能力是当前人工智能体的主要发展目标，而为大模型添加环境感知、记忆、强化学习等功能模块或者一体化解决方案则是实现智能体行动的手段。因此，人工智能体技术在当前被特别强调，并不意味着之前的人工智能，例如阿尔法围棋（AlphaGo）或者一辆自动驾驶汽车，本质上不能被称为智能体，而是新的智能体技术可以让这些人工智能体的自主行动能力更强，以及向更加通用的行动能力方向演进。李飞飞等学者关于人工智能体技术进展的文献综述，总结了智能体的范式框架图。在智能体互动的封闭式循环（agent interactive closed-loop）中，环境、感知、学习、记忆、认知、行动再反馈到环境等多个环节要素，构成一个封闭的循环[⑥]。基于多模态，尤其是大语言模型和视觉学习模型的人工智能系统，可以具有相对独立的感知能力、学习能力、记忆能力、决策能力和行动能力。而国内AI团队推出的Manus人工智能体，已经在应用层面实现了从收到简单指令到独立完成任务的行动流程[⑦]。

（二）从算法视角到行动者视角

人工智能学科中智能体技术的广泛展开，将为新社会行动者的诞生奠定现实可能性。随着人工

① 斯图尔特·罗素、彼得·诺维格：《人工智能：现代方法》（第4版），张博雅等译，第4、32页。

② 在社会科学中，行动和能动的概念差别较为细微，行动多指主体从目的到手段再到结果的全流程，能动则更强调主体对结构的能动作用。但在人工智能研究领域，行动一般是与感知、判断、决策、学习、反馈相区分的具体操作环节。为与传统社会学理论对话，这里沿用社会科学对行动的用法。

③ 安东尼·吉登斯：《社会学方法的新规则——一种对解释社会学的建设性批判》，田佑中、刘江涛译，北京：社会科学文献出版社，2003年，第155页。

④ 斯图尔特·罗素、彼得·诺维格：《人工智能：现代方法》（第4版），张博雅等译，第5页。

⑤ 当前，人工智能体技术并不是人工智能技术的全部，而是更偏重基于多模态的系统集成。

⑥ Durante Z., Huang Q., Wake N., et al., “Agent AI: Surveying the Horizons of Multimodal Interaction”, https://arxiv.org/pdf/2401.03568，访问日期：2026年3月10日。

⑦ 祝凤岚：《AI新秀Manus火爆市场》，《中华工商时报》2025年3月9日，第4版。

智能体技术的进一步发展,人工智能的智能化程度越高,独立完成任务的能力越强,就越具有成为相对独立的非人行动者的可能,进而越有可能与人类行动者之间产生一系列社会行动层面的互动,卷入人类的社会生产生活实践之中。例如,在人工智能驾驶领域,随着人工智能汽车感知外界能力、判断决策能力的提升,人工智能驾驶智能体在驾驶时,构成了和人类驾驶行动者、其他人工智能驾驶智能体、行人等交通参与者之间的多重社会互动,包括对路权的争夺、空间的博弈、对既定规则的遵守与突破等多种社会行动。此外,类似的社会变革同样会发生在金融市场交易、军事与外交战略博弈、企业间商业策略竞争、公司治理、社会公共治理等领域,涉及人工智能体和人类行动者之间的合作、竞争、博弈等诸多维度,最终将在更加宏观的社会组织、社会制度等结构层面上产生影响。

技术和社会现实层面的演进,也理应带来人工智能社会学理论视角的相应转变。从以算法为核心研究对象的算法社会学,到以智能体行动为研究对象的社会学,必然会存在不同的切入视角和理论基础。

首先,从研究对象即算法和人工智能的差异性来看,算法固然是人工智能的核心要素,但是人工智能还包括数据、算力和硬件设施。仅对算法进行社会学研究,并不能充分探讨人工智能发挥现实作用的全部机制。在人工智能体技术初步具有实用价值后,人工智能体区别于传统计算机程序的重要特征就在于其具有传感器、记忆存储器、执行器等。因此,有必要在考察人工智能对人类和社会结构的影响时,将其视为一个完整体。

值得注意的是,在基于大数据和概率论的实现路径下,当前即使具有相同算法的人工智能体,由于其算力不同、投喂的数据不同,也会成为具有不同现实行动能力的行动者,发挥截然不同的社会建构作用。例如,在同样使用DeepSeek开源算法的情况下,由于训练数据和算力不同,会造成人工智能体在理解能力、推理能力和输出结果上的显著差异,进而在社会影响上产生差异。

其次,在社会学方法论意义上,如果仅将人工智能视为一段固定化的计算机程序算法,或视为人工智能创造者或者使用者利益的代表,那么这种视角可以被称为一种工具论的方法视角,即人工智能归根结底是一种技术工具,其本身不具有特定的利益诉求,其行动也必然反映设计者或者使用者的动机。在工具论视角下,人工智能与人类以往的其他工具并无本质上的差别,仅是在自动化程度和能力层面上更加高级。对人工智能的社会学研究,归根结底是对人与人之间社会关系的研究,以及对人和工具技术之间关系的研究。

但是,这一视角也将面临一些新挑战。仅将人工智能视为程序和算法,并不能充分体现人工智能与人类历史上的其他工具在自动化能力上的显著差异。一旦人工智能在技术突破后呈现相对的独立自主性和能动性,无论是在微观的社会互动层面,还是在相对宏观的社会结构层面,都有可能带来其他工具所不能引发的重大变革。此外,仅将人工智能简化为使用者或者设计者意志的表达,很容易忽视人工智能在微观行动层面所具有的独特性,例如人工智能在微观行动逻辑上的幻觉问题、不可解释性问题、理性化冲突问题、伦理问题等。这些微观行动层面的独特性,使得人与人工智能无论在协同合作关系还是在竞争关系中,均面临新的社会挑战。

由此,一种将人工智能体视为非人行动者的社会学视角将会是对既有算法社会学的有益补充。将人工智能体视为微观的社会行动者,而非简单的代码程序,是人工智能体的方法和技术对社会学的重要启示。从社会行动理论传统来看,非人行动者的概念最早由拉图尔提出。在科学技术社会学的“行动者-网络”理论(ANT)中,拉图尔最早将技术、工具、观念等非人主体视为具有建构作用的非人行动者①。将视野扩大到更早的自然科学领域,在控制论和信息论的相关研究中,维纳和罗森布鲁等学者在《行为、目的和目的论》中提出,从广义上讲,任何执行把输入转换为输出的事物都是机器,如果机器由反馈控制,那么完全可以把这些机器描述为有目的的行为主体②。这些思想也深刻影响到塔尔科特·帕森斯的社会系统理论和赫伯特·西蒙的组织理论与人工智能思想。近年来,国内也有学者

① Latour B., *Science in Action*, Cambridge: Harvard University Press, 1987.

② Rosenblueth A., Wiener N., Bigelow J., “Behavior, Purpose, and Teleology”, *Philosophy of Science*, 1943, 10, pp.18-24.

陆续以人工智能社会学为主题，开展区别于算法社会学研究的学科构建工作。例如：王宁提出，新的人工智能社会学分析框架应当对以往隐含的“技术与人的主体性对立、机器不具有人的主体性”的假设进行超越，重新考虑结构和能动的关系①；景军在论述什么是人工智能社会学时也强调，应将生成式人工智能视为社会学的拟人研究对象②。

在数字空间和社会空间融合发展的趋势下，越来越多的社会行动以信息流的形式展开。在这样一种信息行动理论的视角下，发出社会行动即生产信息流的主体在逻辑上不仅可以是人类行动者，也可以是人工智能等非人行动者。例如，在2025年智能体技术的典型应用中，在数字空间中进行点外卖、比价购物全流程的行动工作流，不仅可以由真实的人完成，也可以由人工智能体独立完成。人工智能在现实应用中的逐渐普及，印证了数字社会中信息行动的重要性。在信息行动的场景下，人工智能可以被视为数字社会中信息行动的新主体。在虚实融合的数字社会中，人工智能以产生信息流的方式完成社会行动，不仅在理论逻辑中，更重要的是在现实中正逐步实现。恰恰是技术变迁后，人类社会行动方式和时空场景的转变，为人工智能成为行动主体提供了另一种新的可能性③。而未来的具身智能，则有可能进一步弥补人工智能传统物理行动的短板。

值得注意的是，在方法论意义上将人工智能视为社会行动者，并不是在哲学本体论意义上将人工智能直接等同于社会行动者，也不等于将人工智能在本体论上等同于人。在本体论层面上，人工智能是否等同于社会行动者和人类意识，取决于社会行动者的定义本身，尤其取决于意识、动机、意向性、决策等概念的定义，是一系列值得进行哲学层面探讨的元问题。如果在哲学本体论层面尚无法对此达成共识，那么在社会学研究中可以暂时采用一种悬置的态度。正如迪尔凯姆在设定实证主义社会学研究的对象时提出将社会事实当作物，而在本体论意义上将社会事实视为一种集体表象，仍然是一种集体意识而非物④。在对人工智能的社会学研究中，在特定的社会场景下，如果人工智能体A在社会关系中产生的能动作用等同于人类行动者B，且A的社会行动逻辑及其与其他社会行动者互动产生的社会结果值得进行考察，那么A就应当被视为一种社会行动者，其在逻辑上的地位等同于B。

二、社会学中从行动到结构的理论——人工智能行动者的结构化作用

将人工智能视为社会行动者，并不是说人工智能等同于人类行动者。需要研究的恰恰是人工智能这一非人行动者和人类行动者之间的差异性，考察当非人行动者介入到社会的生产、交往等实践活动中以后，人类社会从最基本的社会行动单元到宏观的社会结构可能发生的重要变化。从微观的社会行动上升到相对宏观的社会结构这一研究路径，正是自现代社会学诞生以来的基本分析思路。无论是帕森斯、吉登斯、布迪厄、科尔曼，还是我国的社会学家费孝通和郑杭生先生等都或多或少试图采用不同的理论建构来弥合微观与宏观、能动与结构、主观与客观之间的鸿沟。这些统一性的社会学理论体系的创立和发展，最终目的都是要解释人类社会从日常生活的基本社会行动单元，到我们所面对的可以感知的日常生活现象，以及那些遥远的宏大的抽象的社会结构和社会议题之间的关联性和社会历史过程。这其实也正是米尔斯在《社会学的想象力》一书中所倡导的：将日常生活的微观层面与人类社会的重大历史变迁勾连起来⑤。

纵览社会学史中的社会学理论创建历程，可以简要分析三个比较知名的联结微观与宏观的理论逻辑，进而探讨人工智能作为微观行动者加入后，从社会微观层面上升到宏观层面的潜在理论逻辑变化。

第一种占据社会学历史重要地位的理论逻辑，可以称为社会建构论的传统。在20世纪70年代后

① 王宁：《AI时代的智力物替、主体重塑与结构转型——一个人工智能社会学的分析框架》，《探索与争鸣》2025年第3期。

② 景军：《什么是人工智能社会学?》，《智能社会研究》2025年第1期。

③ 人工智能成为新的行动主体反过来将进一步促进社会数字化、自动化和智能化发展，二者构成相互促进的加速进步的循环。

④ 迪尔凯姆：《社会学方法的准则》，狄玉明译，北京：商务印书馆，1995年，第7—10页。

⑤ 赖特·米尔斯：《社会学的想象力》，陈强、张永强译，北京：生活·读书·新知三联书店，2005年，第6—9页。

期,建构主义流派在美国社会学理论界产生重要影响的著作是《现实的社会构建》。在这一著作中,伯格和卢克曼相对体系化地阐明了社会学中社会建构论的历史逻辑,提出社会现实是由社会个体在历史的进程中建构而成的,这种社会现实既包含主观性和能动性,又包含客观性和制约性,但归根结底强调"制度世界的客观性,不管它看上去对个体来说有多巨大,都是由人创造、构建的客体"①。这一社会建构论的总结和宣称,实际正是沿袭了古典社会学时期韦伯等人的诠释学社会学的传统。尽管在理论逻辑上并非原创,但是该书关于社会建构主义的宣言和概括,对社会学的经验研究视角影响颇深。例如,经济社会学领域中市场、理性观念的社会建构,性别社会学中性别相关制度的社会建构等研究思路②,都是在社会建构主义视角的启发之下得以展开的。

实际上,从法国的早期社会心理学家塔尔德开始,从个体心理上升到集体心理,进而解释社会宏观现象的方法论个体主义思路,就在社会学中具有重要的影响力。只是微观个体和宏观社会结构之间,如同一个难以观察且超越个体时空经验的巨大黑箱,究竟是简单的个体几何聚合,还是经历何种复杂的有或无规律性的历史过程,很难在理论的描述中清晰界定。社会建构论最终的理论价值也并非精确地量化地概括出一种人类社会确定性的普遍性的建构过程,而是可以让研究者揭示出看似理所当然的社会宏观现象背后的隐蔽之处,即如何让某一社会现象成为今日之现状的特定过程。这一历史的社会的建构过程本身被揭示出来,就具有重要的现实批判意义和理论意义。

当人工智能行动者加入到这一建构主义的理论视角之中,可以发现,从逻辑上讲,影响人类制度的历史建构进程中,出现了新的能动性因素。这种宏观社会制度和结构层面的变迁,不仅是人的行动所建构的,也有可能是人与非人行动者的行动所共同建构的,将来抑或是非人行动者独立建构的。社会建构的具体历史过程将更加复杂化,非人行动者的社会建构作用很可能会更加凸显。不仅如此,面向未来社会的社会建构论,不仅能揭示出社会结构被建构出来的历史逻辑,更有可能发现和探索具有不同行动类型的非人行动者对人类社会制度变迁的潜在可能性。

第二种从微观到宏观的理论逻辑来自社会学中的理性选择理论,借鉴自经济学中的理性人假设。这一理论路径同样从方法论个体主义出发,只不过假定每一个微观层面的行动者都是具有特定偏好的理性个体,遵照目的理性或者某种效用最大化的行动逻辑行动③。理性选择理论不仅在社会学中占据重要地位,同样也对政治学和管理学等学科影响深远。与一般意义上的社会建构论相比,同样从微观行动者出发的理性选择理论的最大优势,是让每一个微观行动个体的行动都遵循固定的可量化的数学逻辑。这样一来,无论是微观层面的行动者行为预测,还是由多个行动者相互进行社会互动,以及社会互动中的合作、博弈等行动,都可以用数学的方法来进行仿真和模拟。宏观层面的社会结构,也由此可以还原为这种理性人行动的产物,而新的宏观层面的社会结构,也可以通过对个体以及群体行动的预测和计算进行建构。

值得注意的是,社会学中的理性选择理论,将复杂的人性和社会简化为可以用数学表达的标准化理性过程,而这一过程恰好与人工智能体技术的发展方法路径相契合。人工智能体技术试图创造的行动主体,在当下乃至很长的时间内很难做到对人类的复杂人性进行模仿,但却可以对理性化的、目标明确的如机器一般的人进行替代。换言之,如果人类的行为抛弃了人性的复杂性,都能像理性选择理论预测的一般按照理性人的行动逻辑运行,那么人工智能行动者替代人类行动者的逻辑可能性将大大提升。个体层面的理性化过程,也将带来宏观的社会层面的结构变迁,并且这些结构的变迁更容易被计算和推导出来。

但是,这并不意味着理性选择理论以及基于该理论的博弈论,乃至计算社会科学会成为人工智能行动者介入社会后的"大一统"理论。因为真实的历史实践中人类行动者无法被简单还原为理性行动

① 彼得·伯格、托马斯·卢克曼:《现实的社会构建》,汪涌译,北京:北京大学出版社,2009年,第52页。

② MacKenzie D., Muniesa F., Siu L., *Do Economists Make Markets? On the Performativity of Economics*, New Jersey: Princeton University Press, 2007, pp. 2-4.

③ 詹姆斯·科尔曼:《社会理论的基础》,邓方译,北京:社会科学文献出版社,2008年,第15—18页。

者，诸多复杂多变和不确定的人性因素和能力因素，导致人类社会的真实运行逻辑和理性化逻辑之间存在着相当大的现实差异性[①]。一旦未来的人工智能行动者所遵循的行动逻辑仅是假设中的人类行动者逻辑——理性逻辑，那么人类行动者所建构的宏观社会结构，诸如规则和制度等，与人工智能行动者介入后的规则和制度之间，将产生更多不确定性的矛盾和冲突，这些正是人工智能社会学应当分析的现象。

第三种在社会学历史中产生广泛影响的理论逻辑来自吉登斯的结构化理论。结构化理论同样被认为是解决或者至少回应了社会学理论阵营中长久存在的从微观到宏观、能动到结构之间的对立问题。与同时代号称结构主义的建构主义者与建构主义的结构主义者的布迪厄类似，吉登斯本人也在试图调和个体行动的自主性、能动性与宏观社会结构的制约性之间的矛盾。与传统的从方法论个体主义出发的社会学行动理论有所不同，吉登斯结构化理论中至少出现了两个重要的理论创新，分别是对个体行动的意图之外的社会结果的强调以及将时间空间场景的概念引入到抽象的理论构建之中。微观的社会行动所产生的动机、意图、意识之外的社会结果，很可能是一种未被行动主体预期的结果，这种行动所导致的宏观系统性变化，刚好解释了受到结构制约的个体如何建构出日常生活场景和社会系统制约之外的社会结构，使得社会产生结构性的变革[②]。而在时空维度上，在日常生活的时空场景被推广到更广阔的现代全球化场景的历史现实中，吉登斯将日常生活的具体的可经验的时空体验与脱域的、抽离化的社会系统时空进行了分离和区分，但是又阐释出其内在的一以贯之的关联性[③]。

从人工智能在结构化理论中的可能位置思考，一方面，可以将人工智能视为人类行动主体行动的产物，但是一旦人们不能保证人工智能这一人造产物的行动符合预期，人工智能同样会产生一种人类行动主体非预期的社会后果，从而成为社会结构变迁的革命性因素。另一方面，如果直接将人工智能提升一个理论逻辑层级，将其视为新的社会行动主体——相对独立于人的行动主体，那么人工智能本身在行动上的不确定性，例如人工智能失控的现象、出现大模型幻觉的现象，以及不同于人的行动逻辑——目前技术条件下的部分行动不可解释性，都会带来结构化过程和结果的重大变化。而从时空的场景变化来看，数字社会带来的时空场景变化模糊了行动在场和缺场的传统时空边界。更重要的是，数字社会中的信息流的全球范围流动使得数字社会时代的社会行动，得以在新的时空情境下以信息流的方式完成。发生信息流的主体无论是人还是人工智能，都处于相对平等的逻辑位置上。因此，考虑到社会时空场景扩展与变迁的因素后，人工智能在新的时空场景下直接介入社会行动、改变宏观社会行动的能力会得到更充分的理论支撑。超越空间的信息行动能力，以及超越时间的“无时间”（timeless time）性[④]，使得人工智能行动者在数字社会的场景中，很可能发展出更加强大的改变和重塑社会现实的能力。

以上三种相对基础的社会学行动理论流派，都在试图解释人类社会历史中从微观行动到宏观社会结构之间的基本演变逻辑。而这些理论都有可能因为人工智能行动者的出现而产生重大的变化，但是又并未排除人工智能作为新的行动者加入其中的理论可能性。在这种人工智能行动理论的背后，是人类社会进入数字时代，在新的时空场景下和新的行动方式变迁的历史进程中，当人造之物的自主化能力不断提升，直到突破以往的“智力护城河”后，必然出现的一种社会变迁的力量。

三、人工智能社会学的新议题——从新行动到新结构的社会现象

与算法社会学的视角不同，把人工智能体在方法论意义上视为社会行动者，将更容易发现人工智

① 刘少杰：《理性选择理论的形式缺失与感性追问》，《学术论坛》2005年第3期。

② 安东尼·吉登斯：《社会的构成——结构化理论纲要》，李康、李猛译，北京：中国人民大学出版社，2016年，第25—26页。

③ 安东尼·吉登斯：《社会的构成——结构化理论纲要》，李康、李猛译，第125页。

④ 曼纽尔·卡斯特：《网络社会的崛起》，夏铸九等译，北京：社会科学文献出版社，2006年，第404页。

能融入数字社会后产生的一系列新的现实和理论问题。从上一部分所述的行动-结构的经典社会学理论框架出发,不难发现,无论是在微观的社会行动层面,还是在相对宏观的社会规则、法律和制度层面,人工智能体介入人类社会后,会带来诸多需要解释和研究的新议题。

(一)微观社会行动层面

从社会学行动论的理论传统出发,人工智能社会学的首要研究议题是新的社会行动者的行动类型问题。除了理性选择理论之外,大部分社会学流派认为人类社会行动者的理想类型并非一种。其中,韦伯关于人类社会行动的分类最为经典,他将微观层面的人类社会行动者的行动类型划分为四种,分别是目的理性行动、价值理性行动、情感行动、传统行动①。而人工智能行动者的行动类型,按照当前人工智能体的技术能力和路线,更多是基于"理性智能体"的方法研发。这意味着,人工智能行动者的行动类型与人类行动者存在显著的差异。设计为理性的人工智能行动者,正在试图替代在现实社会生活中有限理性的、感性的或者多重复杂人性因素影响下的人类行动者。这一过程如何发生,是否应该发生或者有所限制,会产生何种深远的社会影响,这些正是值得研究者去思考和解释的社会现象。

此外,回到人工智能的理性行动本身,究竟是否能用简单的理性行动概念去概括人工智能体的行动方式,在不同的人工智能技术发展路线下似乎也成为一个需要追问的问题。例如,在基于机器学习的大模型路线下,在没有人工干预的情况下,人工智能会基于既有的海量数据学习人类已有的经验和知识,模仿人类的行动方式。具体到人工智能驾驶领域,基于人类驾驶经验数据的端到端大模型,很可能出现人工智能学习到人类驾驶员普遍存在的违法驾驶行为以及一些非理性驾驶行为。更重要的是,由于大模型中间黑箱环节的不可解释性等问题,人工智能可能会出现幻觉,导致在具体的社会行动中出现非预期行动,甚至出现一些低级的难以解释的错误,这或许可以类比为人类行动者的非理性行动或者行为谬误。而其他的人工智能技术路线,例如AlphaGo的后期版本AlphaGoZero,在围棋的机器学习中并没有像早期版本那样大量学习人类棋手的对弈经验,而是从零开始自我构建博弈训练,这有可能会产生不同的智能体行动类型。

解决这些人工智能行动者的复杂行动问题,使其回归理性的状态,又涉及社会的因素,如社会规范和价值观、法律、其他社会规则等,如何决定并制约人工智能的行动类型。例如,在人工智能设计中非常重要的"对齐"环节,就是让人工智能行动者回归到人类社会正常的价值观。但是这一过程同样会充满社会学和伦理学的争议,如何平衡法律底线与道德高线、选择模仿人类行动还是摒弃人类的非理性行动等,这些不同技术理念产生的社会后果都需要社会学研究者进一步去探讨。

(二)社会互动层面

以上是从人工智能行动的单向视角出发,如果考虑到人工智能行动者介入社会交互性的实践,那么就需要考虑一种新的复杂社会互动过程。这一议题是新行动主体出现后的主体间性问题,既涉及人对人工智能的理解和人工智能对人的理解,又涉及双方考虑到彼此后的交往、对话和行动问题。

在社会互动论视角下,人与人工智能之间、人工智能与人工智能之间、人工智能协同人之后的行动体与其他行动体之间的互动过程都会呈现出新的变化。完全理性行动、有限理性行动和多重行动逻辑可能会混合交织,也将会带来新的社会博弈。不同的人与非人行动主体组合之间的合作、竞争和博弈,以及不对等的理性行动能力将会带来新的权力问题。人类运用人工智能能力的强弱、人工智能本身能力的差异性、数据资源和算力资源的差异性,都会为传统的社会竞争和合作增加新的不确定性。

而从社会交往的层面上看,人与人工智能之间的相互交流,很可能会随着人工智能拟人化思维和表达的进步,导致人对人工智能理解的偏差。将非人行动者误识为人类行动者,会导致一系列情感投射问题。人对人工智能的心理和情感依赖,人工智能应对人类行动的合理边界,人工智能的理性行动对情感行动和传统行动的模拟和替代,甚至技术理性行动与情感行动之间的冲突,都是在未来社会中值得观察和思考的问题。

① 马克斯·韦伯:《社会学的基本概念》,顾忠华译,桂林:广西师范大学出版社,2005年,第32页。

谷歌公司在关于人工智能的“哈贝马斯机器”研究中提到了交往理性和人工智能之间的关系。从乐观主义视角出发，人工智能与人之间持续谈话交流，有望形成一种理想的数字空间和数字公共领域[①]。这一预期建立在哈贝马斯著名的交往理性基础上，即人与人工智能的交流遵循交往理性的三个有效性宣称——真实有效性、规范有效性和真诚有效性[②]。当然，在现实中发生的社会现象却与交往理性的理性构想不尽一致，基于人工智能的虚假信息编辑，尤其是生成式人工智能被应用于网络诈骗、虚假新闻生产、网络舆论引导和认知作战的现象并不罕见[③]。这些都是新的行动主体加入社会互动后，其独特的行动能力可能会带来的新现象。

（三）社会制度层面

按照社会学中行动理论传统的基本逻辑，微观层面的社会行动是宏观层面社会结构变迁的重要推动力量，社会行动层面的变化必然会导致诸如组织、规则、法律、制度、社会秩序等更加宏观层面上的变化。当人工智能作为新社会行动者加入复杂社会系统、嵌入社会关系网络、融入人们的日常生活之后，随之而来的必然是一系列更加复杂和不确定性的相对宏观层面的社会变革。

在最容易观察的层面上，人工智能作为新行动主体，会直接引起与行动规则有关的一系列变化，这些正是人工智能社会学必须面对的社会规则议题。原有的有关人类行动者的社会规则，能否适用于非人行动者以及人与非人行动者之间的协作行动？除了行动主体风险责任划分的要求外，新的非人行动者的理性行动逻辑，往往会将人类社会中最隐蔽的潜规则和合法性规则之间的冲突暴露出来。最典型的现实案例莫过于人工智能介入人类驾驶的社会博弈。如果人工智能完全遵守交通规则，则很可能在一个复杂的不完美世界中难以获得具有实用性的通行效率。而人类行动者则会权衡规则的应用，选择突破规则的限制来达到目的。这实际上正是人类社会中名义制度的合法性机制与真实制度的效率性机制之间的冲突。

人工智能的自动化逻辑和非人特性，将很有可能倒逼人类社会的潜规则和模糊性制度规定进一步清晰化。换言之，在人工智能设计者对人工智能行动者进行人类社会法律和价值观“对齐”的同时，社会规则本身可能被人工智能行动者所“校准”，从而呈现出一种“逆对齐”的现象。

而在更广泛的社会治理领域，人工智能作为行动者加入新的治理体系之中，其严格遵守规则、非人格化以及利益倾向的隐蔽性，都可能会对人类社会的治理系统产生深远的影响，这些都是未来社会学无法回避的重要议题。例如，在2025年马斯克主导的美国政府最新一轮机构效率改革中，大幅裁员背后已经隐约可见人工智能行动者参与国家治理和政治效率规划的身影。此外，在法律领域、税收领域、公司治理领域，假如人工智能将自主化、去感情化、去人格化的逻辑演化到极致，一定还会带来不同于前人工智能社会的新制度冲突和变革。

以上仅是在社会制度层面人工智能作为社会行动者带来社会结构变革的冰山一角，更多相关的现实议题仍有待更多学者进一步发掘和探讨。但是至少可以看到，从社会行动理论的视角出发，已经可以触及到很多在数字社会的变革阶段，人工智能社会学应当发现和关注的全新议题。

四、余论：人工智能行动者的潜在变革意义

为什么要将人工智能体视为一种非人的社会行动者？是一种画蛇添足的文字游戏，抑或是如同物理学历史上的“以太”概念一样，需要被“奥卡姆剃刀”剔除？用简单的人类行动者加上工具的概念，能不能替代人工智能行动者的概念呢，或者人工智能行动能不能被还原为人的行动呢？

① Tessler M., Bakker M., Botvinick M., et al., “AI Can Help Humans Find Common Ground in Democratic Deliberation”, *Science* 386, eadq2852（2024）. https://doi.org/10.1126/science.adq2852，访问日期：2026年4月24日。

② 尤尔根·哈贝马斯：《交往行为理论》第1卷，曹卫东译，上海：上海人民出版社，2018年。

③ 张文祥：《生成式人工智能虚假信息的舆论生态挑战与治理进路》，《山东大学学报（哲学社会科学版）》2025年第1期。

作为一种前瞻性的社会学理论,将人工智能视为社会行动者,为未来社会的新社会行动者留下了社会结构中的逻辑空间。换言之,如果在未来社会中,关于何为智能、意识、主体性的定义不以人类既有意识为标准的话,那么人工智能所具有的思维方式、学习方式和行动方式都可以被视为一种新的独立意识的体现。人工智能具有意识的时刻乃至硅基生命降临的时刻,或许会随之到来。当然,或许也有可能因为技术和社会制约等因素,这一时刻永远都不会到来。除了人工智能的技术边界之外,人类终将会为人工智能划下社会边界,未来的智能社会也因此充满着不确定性。但是无论如何,其非人行动的独特属性影响和变革社会结构的能力应当被充分重视。

在现实层面上,随着人工智能体技术的广泛应用,人工智能作为行动者发挥相对独立的能动作用的社会场景越来越多,更多和人工智能行动者相关的社会现象也应当进入社会学研究的视野。在现实社会中,人们即将面对和人工智能行动者进行协同、竞争和博弈的新社会现象。而这些人工智能行动者并不是被虚构出来的风车,与之博弈的人类也并非堂吉诃德,而是会真实地感受到人工智能行动者带来的社会影响。无论是人工智能对某些人类工作岗位的替代,还是在社会交往中与人类产生"虚假"的情感纠葛,抑或是在不对等的社会博弈中出现对既定法律规则甚至社会秩序的重塑,都是有可能发生或者是正在发生的新社会现象。

人工智能能否还原为人类的工具,或者被还原为设计者、使用者利益的代表,并不影响我们观察和研究其在社会结构中的建构和塑造作用。正如这个社会中越来越多的人也被称为"工具人",越来越多的人类行动者也正在按照机器一般的理性行动者模型去行动,人类也在现代社会的工具理性逻辑中自我异化,失去了所谓的独特的人性,成为他人利益的代表,抑或是整个系统链条上的一个去人性化的环节。所谓工具和行动者的区别,在现代高度技术理性化的社会中成为既对立又统一的概念。在这一意义上,与其哀叹和拒绝承认人工智能行动者对人的替代,不如感慨和反思人为何变为与人工智能一样的非人行动者。

从社会行动论的视角看,将人工智能视为社会行动者,重新理解和判断人工智能时代社会在宏观社会结构层面的潜在变化,是一种不同于算法社会学的人工智能社会学理论视野,至少可以成为前者的有益补充。从微观与宏观相结合的社会行动层面理解数字社会的变迁,才能更加直接地认识到为何人工智能技术具有产生潜在重大社会变迁的可能性。

在人工智能技术何以引起社会变迁的历史判断中,我们往往容易通过日常生活中的经验,从人工智能替代人类工作以及人工智能违背伦理道德的角度,判定人工智能对社会发展的影响。然而社会学传统中的社会行动理论,则可以更加清晰地揭示出为何人工智能不同于人类历史上的其他任何工具和技术。这是因为,一旦人工智能出现突破性的技术进步,达到技术的临界点,人类历史上将第一次出现替代人类主体行动的独立行动者。而微观层面新的社会行动者的出现,最终会导致社会宏观层面结构和秩序的重构。与马克思主义理论中生产力的发展最终导致生产关系发生变化的判断相比,人工智能本身既作为新质生产力又作为生产关系中的一部分,直接对社会结构产生作用。如果以往技术对社会的改变,是技术通过对人产生作用,进而引发社会结构变化的话,那么在新的逻辑链条中,人工智能作为行动者,直接对社会结构的变化产生作用,这种影响会更加直接和激进。

当然,我们或许也不应在当下过高估计人工智能的发展速度,未来依然充满技术和社会的多重不确定性。不排除这样一种可能性:也许在相当一段历史时期内,存在一种技术抑或社会层面的玻璃天花板效应,导致人工智能在现实应用中不会成为独立于人的社会行动者。也有可能存在一种社会缓冲效应,会有效对冲人工智能带来的冲击力和社会结构破坏力。如果是这样,这也同样构成了人工智能社会学有待研究的重要议题。但是无论如何,都应在社会学理论层面作好准备,以一种前瞻性的视角去发现和研究极有可能发生的社会变迁。

The Action Theory Perspective of the Sociology of Artificial Intelligence: The Implications of Artificial Intelligence Agent Technology for Sociological Theory

Chen Chuan
(Party School of the CPC Central Committee (National Academy of Governance),
Beijing 100091, P.R.China)

Abstract: Artificial intelligence (AI) agent technology will greatly enhance the capacity of AI to act independently. The methodological and technical perspectives emerging from AI agent study bring new insights for sociology, calling for a paradigm shift from viewing AI as tools to recognizing it as a social actor. Existing sociological research on algorithms has focused on the social impact of algorithms. It is necessary to treat AI agent as a new social actor to supplement the study of algorithmic sociology. From the perspective of classical social action theory, schools of thought such as social constructivism, rational choice theory, and structuration theory all indicate that changes in social action subjects will lead to systematic changes in social structures. As new actors, AI agents have the potential to fundamentally reshape these theoretical frameworks. The research agenda concerning the implications of AI agents as new social actors encompasses three dimensions. First, at the micro level of social action, new questions arise regarding the typology of AI actions—particularly the differences between AI's rational actions and the action types characteristic of human actors. Under the prevailing paradigms of big data and machine learning, AI does not necessarily achieve rational action in practice. The forms of action that emerge when AI mimics human behavior are considerably more complex and must be aligned with social rules and values. Yet whether alignment should adhere to a rational baseline or aspire to a moral high standard remains an open and pressing question for sociological inquiry. Second, at the meso level of social interaction, AI agents prompt research on human-AI interaction and intersubjectivity. This includes the dynamics of cooperation, competition, and strategic engagement among various configurations of human and non-human actors, as well as the new power asymmetries generated by unequal capacities for rational action. Third, at the macro level of social structure, AI agents give rise to a series of investigations into how they adapt to, reshape, and recalibrate social rules and institutions. As new actors, AI agents may come into conflict with established rules. Moreover, the automated logic and non-human characteristics of AI may also force the implicit rules and ambiguous institutional provisions of human society to become more explicit and codified. When AI agents enter governance systems—with their strict rule adherence, impersonal operation, and concealed interest orientations—they may exert profound effects on human governance arrangements. These constitute critical issues for future sociological research. Whereas previous technologies transformed society indirectly, by acting upon human beings who in turn altered social structures, in the emerging logical chain AI agents act directly upon social structures themselves. Conceptualizing AI as actors thus opens logical space for accommodating various new types of actors in future societies. Nevertheless, factors constraining the agency of AI in practice persist, and social conditions may significantly shape and limit the capacity of AI to act independently.

Keywords: AI; Sociology of artificial intelligence; AI agent; New social actor; Information action theory

[责任编辑:孔令奇]