

从“关系情境主义”到马克思主义道德学说的“语境主义”

——人工智能伦理范式基础的转换

王 亮

摘要：人工智能伦理研究最根本的目的是增进人类的繁荣福祉，但是，学界存在着不同的理论主张：“人工智能体中心论”主张应当使人工智能体符合人类伦理规范，有必要对它们进行伦理设计、监管；“关系情境主义”主张人是人工智能伦理的中心而非智能体，人工智能伦理应当是关于“人类善的伦理”；马克思主义道德学说的“语境主义”则从唯物史观的高度将生产力、道德、人的自由联系起来，从而从新的高度奠定了人工智能伦理范式基础。

关键词：关系情境主义；语境主义；人工智能伦理

中图分类号：N031 **文献标识码：**A **文章编号：**1000—8691（2022）03—0020—09

不同的人工智能伦理研究范式有着不同的理论前提预设，即人工智能伦理主要是立足于人工智能体，还是立足于人本身的美好生活、人的自由等，这些理论前提构成了人工智能伦理研究范式的基础，它不仅决定了人工智能伦理研究的方向，也体现着人工智能伦理的价值底蕴。因此，对这些理论前提的探讨有助于推进人工智能伦理的深入研究。

一、“以人工智能体为中心”的人工智能伦理

目前的人工智能伦理研究主要是围绕人工智能体进行，它的研究分为两个层面：第一个层面是对人工智能体是否可以承担伦理责任进行追问；第二个层面是对人工智能体的伦理设计、监管等问题进行探讨。相比较而言，第一个层面更为基本，它是第二个层面的研究得以开展的前提。

一些学者认为人工智能体可以承担伦理责任。苏尔林斯（John P. Sullins）认为，只要机器人满足相对独立性、行为的可解释性、理解力三个条件，它的道德权利和责任就与其身份无关。^①费尔贝克（Peter-Paul Verbeek）认为，人工物和设计者、使用者之间具有交互式伦理责任。^②柯恩科威克和舒科吕（Gordana

基金项目：本文是国家社会科学基金青年项目“跨文化视角下人工智能的伦理嵌入机制研究”（项目号：19CZX018）的阶段性研究成果。

作者简介：王 亮，男，西安交通大学副教授，硕士生导师，主要从事马克思主义哲学、人工智能伦理研究。

① Sullins J P, When is a robot a moral agent, *Machine ethics*, 2006, pp.151-160.

② Verbeek P, Design Ethics and the Morality of Technological Artifacts, Vermaas P E., Kroes P., Light A., et al. *Philosophy and Design: From Engineering to Architecture*, Dordrecht: Springer Netherlands, 2008, pp.91-103.

Dodig Crnkovic & Baran Çürüklü)认为:“所有的人类利益相关者(机器人设计师、制造商、维护人员和用户)都有特定的责任来保证安全、适当的性能,而机器人则需要不断地发展人工伦理能力,包括为自己的行为承担功能性责任的能力。”^①波尔森等人(Adam Poulsen et al.)认为:“机器伦理领域关注的是如何将伦理行为或确定伦理行为的方法嵌入到人工智能(AI)系统中。其目标是生产出具有隐性伦理(旨在避免不道德后果)或显性伦理(旨在行为合乎伦理)的人工道德代理(AMAs)。”^②正如波尔森所言,在肯定人工智能体可以承担伦理责任的基础上,一些人工智能伦理研究者分别从人工智能体的伦理设计路径、伦理数据库构建、价值观嵌入等方面研究了人工智能体的伦理设计问题。就人工智能体的伦理设计路径而言,瓦拉赫和艾伦(Wendell Wallach & Colin Allen)可以说是该领域的奠基者,他们在《道德机器:如何让机器人明辨是非》一书中详细对比了“自上而下式(top-down)”进路和“自下而上式(bottom-up)”进路,认为“自上而下式”进路由于涉及抽象的伦理“原则”或“规范”,因此具有“全面的解决方案”的优势,但同时面临着“繁重的计算需求”的劣势;“自下而上式的策略有望产生对人工道德智能体整体设计而言不可或缺的技术和标准,但这些策略相当难以进化或发展”^③。在综合分析“自上而下式”和“自下而上式”进路的基础上,瓦拉赫和艾伦提出了一种“混合式美德伦理”进路,并认为它综合了“美德的自上而下进路”和“联结主义式的美德”进路,适合作为人工道德智能体的有效设计进路。^④就人工智能体的伦理数据库构建而言,德加尼等人(Morteza Dehghani et al.)倡导通过创建数据库来为人工智能的道德决策提供参考依据。^⑤穆恩等人(AJung Moon et al.)提出采用开源模式共建人工智能的伦理知识库。^⑥徐英瑾深刻分析了人工智能中的“人工神经网络”计算模型与儒家传统的“德性熏养”以及“儒家德性样板库”的隐喻性投射之间的关系。^⑦价值观嵌入也是人工智能体伦理设计的重要内容之一,施皮克曼(Sarah Spiekermann)将设定和理解价值目标作为人工智能系统伦理设计的第一步。^⑧拉根伯格(Michael Nagenborg)认为,在人工道德代理人的设计中价值观总是被有意识地嵌入到“伦理子程序”中。^⑨李伦和孙保学认为,人工智能系统应当内置价值纠偏机制。^⑩

也有一些学者认为人工智能体无法承担伦理责任。西普内恩(Mikko Siponen)认为,目前的人工智能实体不能视为具有伦理责任的代理,将伦理责任转移至人工代理也不利于对开发人员或用户进行追责。^⑪米勒(Keith Willam Miller)认为,人工智能体的伦理责任应当由设计者、开发者、评估者和使用者来承担。^⑫温斯伯格和罗宾斯(Aimee van Wynsberghe & Scott Robbins)认为,构建人工道德代理人的理由不够充分,没有嵌入伦理的机器人也可以发挥作用。^⑬弗洛里迪和桑德斯(Luciano Floridi & J. W. Sanders)通过抽象

① Crnkovic G D, Çürüklü B., Robots: Ethical by Design, *Ethics and Information Technology*, 2012, Vol.14, No.1, pp. 61-71.

② Poulsen A., Anderson M., Anderson S L., et al., Responses to a Critique of Artificial Moral Agents, <https://arxiv.org/ftp/arxiv/papers/1903/1903.07021.pdf>, 2019/2019-05-15.

③ [美]温德尔·瓦拉赫、[美]科林·艾伦:《道德机器:如何让机器人明辨是非》,王小红等译,北京:北京大学出版社,2017年,第71—99页。

④ [美]温德尔·瓦拉赫、[美]科林·艾伦:《道德机器:如何让机器人明辨是非》,第102—109页。

⑤ Dehghani M, Tomai E, Forbus K, et al., *An Integrated Reasoning Approach to Moral Decision Making*, New York: Cambridge University Press, 2011, pp.422-441.

⑥ Moon A, Caliskan C, Operto F, et al., Open roboethics: Establishing an online community for accelerated policy and design change, http://robots.law.miami.edu/wp-content/uploads/2012/01/Moon_et_al_Open-Roboethics-2012.pdf, 2012/2019-04-05.

⑦ 徐英瑾:《儒家德性伦理学、神经计算与认知隐喻》,《武汉大学学报(哲学社会科学版)》2017年第6期。

⑧ Spiekermann S., *Ethical IT innovation: A value-based system design approach*, New York: Auerbach Publications, 2015, pp.239-241.

⑨ Nagenborg M., Artificial moral agents: an intercultural perspective, *International Review of Information Ethics*, 2007, Vol.7, No.9, pp.129-133.

⑩ 李伦、孙保学:《给人工智能一颗“良心(良心)”——人工智能伦理研究的四个维度》,《教学与研究》2018年第8期。

⑪ Siponen M., A pragmatic evaluation of the theory of information ethics, *Ethics and Information Technology*, 2004, Vol.6, No.4, pp.279-290.

⑫ Miller K W., Moral Responsibility for Computing Artifacts: "The Rules", *IT Professional*, 2011, Vol.13, No.3, pp.57-59.

⑬ Van Wynsberghe A, Robbins S., Critiquing the reasons for making artificial moral agents, *Science and engineering ethics*, 2018, pp.1-17.

层次方法 (LoA) 将责任和道德代理相分离, 认为人工道德代理尽管不能承担责任, 但作为道德行为能力者仍有必要接受相关的道德行为规范。^① 因此, 这也为人工智能伦理研究开辟了另一条路径, 即从外部对人工智能体进行伦理规范, 而非其内部的伦理设计, 相关研究主要集中于工程师伦理责任、伦理准则、伦理监管等方面。赫图宁等人 (Anniina Huttunen et al.) 主张通过工程人员学习“新阿基米德誓言 (New Archimedes' Oath)”来解决人工智能体的伦理问题。^② 木村刚 (Takeshi Kimura) 认为机器人工程师应当熟悉不同的社会、文化和伦理内涵, 要具备良好的伦理责任意识。^③ 温菲尔德 (Alan Winfield) 对世界上部分人工智能研究机构或政策机构所制定的伦理准则进行了总结。^④ 伍敏敏总结了国外人工智能伦理标准的制定情况, 并分别从“人类社会伦理标准优化、人类对人工智能技术探索的伦理标准、人工智能体的伦理规约标准”三个层面对人工智能伦理标准化体系构建进行了思考。^⑤ 孙保学认为, “自上而下”和“自下而上”的人工智能道德决策研究进路并不能完全解决问题, 还需要结合“外部制约”和“内部监控”。^⑥ 苏令银认为, “自上而下”和“自下而上”的伦理嵌入路径存在困难, 应当采取“法律规制、技术监管、伦理建模、程序正义等综合策略”。^⑦

可以看出, 对人工智能体的伦理责任追问是“以人工智能体为中心”的人工智能伦理研究的前提问题, 它直接影响着对人工智能体的伦理设计、监管等问题的探讨, 并由此形成了人工智能伦理研究的两条不同的路径: 内部伦理设计和外部伦理监管。而人工智能体的伦理责任追问所对应的深层次理论问题是“人工智能体是什么”? 是纯粹的人工智能物 (系统) 还是具备伦理责任的人工道德智能体? 这些问题的答案构成了“以人工智能体为中心”的人工智能伦理范式的理论基石。然而, 科克尔伯格 (Mark Coeckelbergh) 并不满足于对“人工智能体是什么”的探讨, 而是深入分析人工智能体与人之间的关系, 从美德伦理的视角分析“人工智能体应该是什么”, 因此, 他将人和人工智能体同时融入人工智能伦理范式之中, 并且重点突出了伦理视域下人工智能体对人的影响。

二、科克尔伯格“关系情境主义”视域下的“外观伦理”和“人类善的伦理”

与主张伦理设计和伦理监管的人工智能伦理范式不同, 科克尔伯格明确提出, “将个人机器人伦理学作为一种外观伦理学、美好生活伦理学、体验伦理学和想象力伦理学”^⑧。他这样做的依据何在? 其伦理范式与“以人工智能体为中心”的人工智能伦理范式的区别是什么? 为了更好地回答这些问题, 科克尔伯格首先对传统的人工智能伦理范式进行了分析, 他认为: “伦理学通常被理解为应用伦理学。从这个角度来看, 如果要评估人机交互 (可能包括人机关系), 应该将伦理标准——伦理理论提供的道德原则——应用于手头的问题或案例。这就允许人们判断所发生的事情在伦理上是否可以被接受。在机器人伦理学中, 这种方法意味着外在于人机交互的伦理标准应用于其中。”^⑨ 显而易见, 科克尔伯格认为传统的人工智能伦理范式的核心是对“伦理标准”的运用, 而这一运用的方式正如之前所探讨的, 可以分为

① Floridi L, Sanders J W., On the morality of artificial agents, *Minds and machines*, 2004, Vol.14, No.3, pp.349-379.

② Huttunen A, Brace W, Kantola V, Lechner L, Kulovesi J, Silvennoinen K., Cross-cultural Application of Ethical Principles in the Design Process of Autonomous Machines, http://www.iiis.org/CDs2010/CD2010IMC/ICEME_2010/PapersPdf/FB476NT.pdf, 2010/2019-4-2.

③ Kimura T., *Roboethical arguments and applied ethics: Being a good citizen*, Cybernics. Springer, Tokyo, 2014, pp.289-298.

④ Winfield A., An Updated Round Up of Ethical Principles of Robotics and AI, <http://alanwinfield.blogspot.com>, 2019-04-18/2019-05-02.

⑤ 伍敏敏:《人工智能伦理标准化体系初探》,《中国标准化》2018年第21期。

⑥ 孙保学:《人工智能如何进行道德决策》, http://www.xinhuanet.com/tech/2017-09/11/c_1121640568.htm, 2017-09-11/2019-05-02。

⑦ 苏令银:《能将伦理准则嵌入人工智能机器吗?——以无人驾驶汽车为例》,《理论探索》2018年第3期。

⑧ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, *International Journal of Social Robotics*, 2009, Vol.1, No.3, pp.217-221.

⑨ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

外部伦理监督和内部伦理设计两条路径,前者主要体现为人类运用“伦理标准”来评判人工智能体的一些行为特性“在伦理上是否可以被接受”,后者主要体现于“伦理标准”的内化,“对于设计师来说,这种‘外在的’方法意味着他们的目标是尝试在机器人中‘内置’规则”。^①同时也要注意,科克尔伯格认为传统人工智能伦理范式中的“伦理标准”是外在于人机交互的,因此,“这种伦理方法,无论是应用于人类还是应用于机器人,不仅在具体的情境中遇到麻烦(正如阿西莫夫自己的故事所显示的),而且还忽略了其他类型的伦理问题”^②,“这种方法忽略了更广泛的伦理问题,比如人们想要什么样的生活”^③。至此,科克尔伯格指出了传统人工智能伦理范式的两大缺点:第一,缺乏人机交互情境的伦理解释力;第二,忽略了与人类美好生活相关的美德伦理问题。而这也正是科克尔伯格想要通过新的人工智能伦理范式所要解决的问题。他认为:“除了上述方法论取向之外,另一种可能的选择是转向外观伦理(ethics of appearance)和人类善的伦理(ethics of human good),即在体验和实践中出现的伦理。”^④可以看出,科克尔伯格新伦理范式的内核就是“外观伦理”和“人类善的伦理”。

何为“外观伦理”?科克尔伯格提出,“我建议我们从人类与基于表象而非真实的人形特征(智力,意识,情感等)的机器人相互作用的研究开始,而不是探究类人机器人如何能成为具有代理、自主性和责任感的机器人。例如,一个特定机器人的脸看起来像人吗?如果是,我们如何在与那个机器人的交互中体验这个特征?机器人对我们做了什么,取决于它们在我们面前的样子,而不是它们真正在想什么”^⑤。“外观伦理”的独特之处就在于研究基于表象的人机互动,而非人工智能体的自主性、伦理责任等属性。相较而言,“外观伦理”的优势在于,它弥合了因实体与其外观的二分而造成的伦理理论与现实情境脱节的解释力鸿沟。科克尔伯格认为,“属性观假设一个实体只有一个‘正确’的本体状态和意义,与机器人的‘外观’和‘感知’形成对比。那些指责人们行为不‘应该’的人依赖于道德立场的科学,而道德立场的科学假定了实体(例如,机器人,作为一个物自体本身)和实体的外观之间的二分法。但可以想到另一种非二元论的认识论,它拒绝这种二分法,接受一个实体可以以几种方式出现在我们面前,而这些方式都没有先验的本体论或解释学的优先权”^⑥。科克尔伯格对于实体(人工智能体)属性观的探讨将人工智能伦理范式中的“外观伦理”问题直接推向了其理论的源头,即对道德地位的分析。很明显,科克尔伯格不赞成基于实体(人工智能体)属性的道德地位观点,他认为,“我们需要将道德地位置于情境中,而不是在道德地位科学的解剖剧场中把实体作为原子主义的珍玩来研究”^⑦。科克尔伯格的道德地位情境论和他的实体与其外观的非二元论是一致的,在具体的情境之中实体(人工智能体)及其外观直接向人呈现,和人发生相互作用,并且与人是一种天然的关系性存在,正因如此,“一个实体的道德地位是建立在作为认知主体的实体与我们之间的关系的土壤上的”^⑧。在具体的情境中,不存在任何“本体论”的优先性,在分析与人相互作用的实体(人工智能体)属性之前,关系就已经存在。这种生长于情境关系的道德地位观较好地解决了伦理理论与现实情境脱节问题,例如,当人们在探讨智能服务机器人伦理时,“欺骗”是一个很重要的伦理问题,但现实的情况却并非如此,一些老年人或者儿童在现实的人机互动过程中十分投入,完全忽视了“欺骗”问题,同时这种投入的人机互动还对他们产生了积极的效果。“欺骗”的前提是从本体论上首先认定机器人不是情感物,对于传统的伦理学理论而言“欺骗”确实是不道德的,而“外观伦理”理论认为在具体的关系情境中,不存在任何“本体论”的优先性,只有人机互动的关系

① Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

② Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

③ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

④ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

⑤ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

⑥ Coeckelbergh M., The moral standing of machines: Towards a relational and non-Cartesian moral hermeneutics, *Philosophy & technology*, 2014, Vol.27, No.1, pp.61-77.

⑦ Coeckelbergh M., The moral standing of machines: Towards a relational and non-Cartesian moral hermeneutics, pp.61-77.

⑧ Coeckelbergh M., The moral standing of machines: Towards a relational and non-Cartesian moral hermeneutics, pp.61-77.

性和体验性,所以也无所谓“欺骗”,正如科克尔伯格所言:“首先,我对外观(荷兰语: *verschijning*; 德语: *Erscheinung*)的关注旨在引起人们对体验的关注。重点不在于机器人的作用(它的行为),而在于它对我们的作用。”^①如果将“外观伦理”理论应用于更大的情境之中——人类生活——就会形成另一种新的伦理内容,即“人类善的伦理”。

科克尔伯格的“人类善的伦理”与“外观伦理”一脉相承,他提出,“我建议,我们也应该,尤其是考虑个人机器人可能为人类的福祉做出的潜在贡献,而不是将人机交互的伦理评估局限于与道德权利有关的问题”^②。很明显,“人类善的伦理”的要旨或者目标是指向人类的福祉,而非“局限于与道德权利有关的问题”,它在理论方向上与“外观伦理”保持高度的一致,尽力避免对人工智能伦理进行抽象的道德属性探讨。为了更深入地阐述这一理论主张,科克尔伯格对比了两种不同的人机交互伦理方法,“人类的善是多样的。然而,这些人类的善是什么?为了评估个人机器人需要什么样的方法?以上提出的伦理问题至少可以用两种方法来回答”^③。“一种方法是从人类的善、人类的繁荣、幸福、友谊和爱的某种观念出发”^④,这种方法可以称之为“观念的方法”,科克尔伯格认为它能够给出“更精确和更可行的人类善的定义”,并且能提供“标准清单”,借此可以对机器人进行伦理方面的判断,是一种“技术伦理的创新”^⑤。但同时,科克尔伯格也深刻意识到了“观念的方法”的弊端,他直言不讳,“然而,如果这些关于善的概念和这些清单被视为先验标准,善的先存概念(*pre-conceptions of good*),那么这种方法是有问题的。前文我拒绝了外在主义的道德方法。标准可能是如此笼统或远离人机交互情境,它们并不是非常有用。因此,我们必须通过更强有力的方式将理论与实践联系起来,认真对待新技术的特性及其对我们的影响”^⑥。可以看出,科克尔伯格将“观念的方法”看作是一种“外在主义的道德方法”,认为它是从具体的人机交互情境中抽象出去的、外在的、先验的是将理论与人类生活实践脱离的,因此,“是有问题的”,且“不是非常有用”。那么,对于“人类善的伦理”而言,什么样的方法才是合理的呢?科克尔伯格基于其“关系情境”理论提出了一种“体验和想象的方法”,他认为:“我们需要从人机交互的具体体验和想象开始,然后讨论善在性能清单方面是如何被理解的,而不是从作为先验道德规范的性能清单开始。”^⑦“因此,无论人类善是否可能外在于人类体验,我们必须研究、想象和塑造善可能会出现于其中的人类与机器人共同生活的具体情境。让我们聆听人们的体验,并运用我们的道德想象力,找出是否有可能增进人类繁荣和福祉的人—机共同生活。”^⑧科克尔伯格将人—机共同生活看作是人机交互的最广阔情境,而“人类善的伦理”就是倡导人们利用已有的人机交互体验和“道德想象力”,构建“增进人类繁荣和福祉的人—机共同生活”。

至此,可以看出,科克尔伯格利用其“关系情境主义”理论为人工智能伦理勾勒了一幅极具吸引力的美德伦理式的宏伟画卷。他不仅通过“人类善的伦理”将人工智能伦理范式从“以人工智能体为中心”转换为“以人为中心”,而且将人工智能伦理的研究视域从抽象的概念、属性分析转换为对具体的、体验式的美好生活的探讨。更为重要的是,科克尔伯格十分注重伦理理论对于现实的解释力,强调理论的可用性及其与人类生活实践的关联性,其关于人机共同塑造“美好生活”“人类福祉”等方面的理论阐述对于人工智能伦理范式的转向具有重要的价值导向性。然而就实现这些美好的伦理价值追求的方法方式而言,科克尔伯格却背离了理论与实践应强有力结合的主张,“体验”和“道德想象力”的方法使其

① Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

② Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

③ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

④ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

⑤ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

⑥ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

⑦ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

⑧ Coeckelbergh M., Personal robots, appearance, and human good: A methodological reflection on roboethics, pp.217-221.

伦理范式掉入了认识论窠臼，与其说科克尔伯格的人工智能伦理范式体现的是“人工智能体应该是什么”这一命题，不如说其所表达的是“人类眼中或者期待的人工智能体应该是什么”。要想走出科克尔伯格人工智能伦理范式的理论困境，必须立足于更为宏大、深远的理论视野来分析人工智能与人类伦理道德之间的深刻关系，而马克思主义道德学说的“语境主义”恰恰具备这一理论品质。

三、唯物史观视野下的人工智能伦理

加拿大著名哲学家凯·尼尔森(Kai Nielsen)立足于唯物史观提出了马克思主义道德学说的“语境主义”理论，这不仅开阔了马克思主义伦理学学理基础的研究，而且也拓展了伦理学理论的广度和深度。尼尔森是如何从唯物史观的丰富资源中萃取“道德精华”的呢？首先，尼尔森对“语境主义”进行了准确的说明，他认为：“语境主义（当它被运用于道德时）是这样—个命题，即，道德要求几乎毫无例外地随语境而发生较大程度的改变。语境主义并不是那种认为对或错、善或恶是由一个人的态度和承诺或一个人所愿意接受的任何可普遍化命令来决定的主张；毋宁说，对或错、善或恶，在相当大的程度上是由客观情境决定的，在这些情境中，人们可以发现他们自身。”^①在此，尼尔森表达出了三点看法：第一，在道德学说的“语境主义”理论中，道德不是抽象的、永恒的存在，而是随情境变化发生改变；第二，变化着的情境是“客观情境”而非取决于“态度”“承诺”等主观因素的情境；第三，“在这些情境中，人们可以发现他们自身”。前两点是尼尔森“语境主义”与唯物史观结合的关键，历史唯物主义强调人类历史发展规律的客观性，它体现了更为宏大的客观情境。正因如此，尼尔森认为：“历史唯物主义自身创造了这样一种从道德层面观察事物的方式，在重要的意义上，这种观察方式是语境主义的，但依然绝非相对主义的。”^②不可否认，“语境主义”的提出为历史唯物主义中的道德理论合理性提供了有力的支撑。然而，道德合理性不等于道德本身，“语境主义”的道德内容正是隐藏于上述的第三个观点之中，即人在客观情境中发现自我。在人类宏大的历史发展进程中，发现自我的过程就是人的自由实现的过程。

“作为人的存在形态，自由既有其本体论的意义，又包含价值论的内涵。从价值论的角度看，自由具体展开为二个维度：在类或社会的层面，它以人向自身的复归为其深刻内容，在个体层面，它所指向的，则是真、善、美统一的存在境界；二者统一于走向自由王国的历史过程”^③。而这也被尼尔森敏锐地捕捉到了，并将其作为评判道德进步的标准，“即，人有能力理性地评价道德规定并相信道德进步。随着生产力的发展，它们为幸福、人的兴旺，特别是为更多的人拥有更大的生活控制权进而为更多的人享有更大的自主性，开启了越来越多的可能”^④。在这里，尼尔森并没有直观地提到人的自由，而是重点阐述了唯物史观视野下人的“自主性”程度和广度的拓展问题。然而，自主性、自由与道德伦理又是紧密相关的，“正因为我们在自由之中，我们的生活与行动在任何时候都不会有‘非如此不可’的必然性处境，或者说，我们的生活或行为在任何时候都不会被锁定在一个既定的唯一轨道上；相反，总还有其他可能性呈现在我们眼前。在这个意义上，自由使我们永远处在能够自主决断的处境之中，自由使我们永远面临着自主决断的任务。所以，对于人这种存在者，才面临着应当或不应当的问题，也才会提出应当或不应当的要求，并因此才会有各种伦理价值。在这个意义上，我们可以说，自由使一切伦理价值成为可能而构成了一切伦理价值的基础”^⑤。

其实，作为德国古典哲学的代表人物，康德和黑格尔都曾经深入研究过自由与伦理之间的关系。“康德把道德基础从经验外在对象物转移到先验主体人的意志上来，以意志自律原则批判了经验的他律原则，

① [加]凯·尼尔森：《马克思主义与道德观念——道德、意识形态与历史唯物主义》，李义天译，北京：人民出版社，2014年，第1页。

② [加]凯·尼尔森：《马克思主义与道德观念——道德、意识形态与历史唯物主义》，第159页。

③ 杨国荣：《自由的形上意蕴》，《文史哲》2004年第6期。

④ [加]凯·尼尔森：《马克思主义与道德观念——道德、意识形态与历史唯物主义》，第159页。

⑤ 黄裕生：《论自由与伦理价值》，《清华大学学报（哲学社会科学版）》2016年第3期。

在伦理学上进行了一场所谓‘哥白尼式革命’，建立起以人的意志自由为中心的主体伦理思想体系”^①。黑格尔在《法哲学原理》中直接强调，“伦理是自由的理念。它是活的善，这活的善在自我意识中具有它的知识和意志，通过自我意识的行动而达到它的现实性；另一方面自我意识在伦理性的存在中具有它的绝对基础和起推动作用的目的。因此，伦理就是成为现存世界和自我意识本性的那种自由的概念”^②。可以看出，康德和黑格尔都从主体意识、意志等主观方面来界定伦理，强调主体对道德律令、伦理秩序有发自内心的期盼和认同，并且自觉、自愿遵从，这些都深刻体现了主体意志之自由。恩格斯也认为：“如果不谈谈所谓自由意志、人的责任、必然和自由的关系等问题，就不能很好地讨论道德和法的问题。”^③然而要注意的是，唯物主义和唯心主义在关于自由的伦理观立场上有本质的区别，马克思认为：“从前的一切唯物主义——包括费尔巴哈的唯物主义——的主要缺点是：对对象、现实、感性，只是从客体的或者直观的形式去理解，而不是把它们当作人的感性活动，当作实践去理解，不是从主体方面去理解。因此，结果竟是这样，和唯物主义相反，唯心主义却发展了能动的方面，但只是抽象地发展了，因为唯心主义当然是不知道现实的、感性的活动本身的。”^④唯心主义过于夸大主体的能动性，而唯物主义在肯定主体的能动性的同时强调了实践对于主体的重要意义。唯物主义的立场恰恰是“语境主义”道德学说的理论出发点：在人类历史的发展进程中，道德的进步体现为，人在现实的实践中自由的实现程度。这里的自由是人对于自身的解放，人向自身本质的复归的自由，也正因此，它体现为一种“解放的道德”^⑤。

如果对尼尔森的“语境主义”再做进一步探讨就会提出质疑：当“自由的实现”成为道德进步的标准时，“自由”本身是否会成为脱离实践的抽象原则？事实上，为了更好地说明客观情境与道德之间的关系，尼尔森预设了“道德公理（moral truism）”的存在，但他同时也表明了其与先验真理、道德原则等之间的区别，他认为，“在没有什么争议性的意义上，道德公理也很琐碎，它与永恒原则或道德真理（假定我们知道其内涵）的差别在于，它不是先验真理（分析的或综合的，如果具体说来就是这些），也不是自明地确定的道德真理或道德原则”^⑥。“虽然它们在理论或原则上可能消失，但事实上不会，而且没有理由认为，它们会在任何心理或社会层面上有所改变的环境中消失。”^⑦可以看出，尽管“道德公理”没有理论的抽象形式，但它却在实践中永存。尼尔森通过这种方式将道德的根基深深地扎入了客观情境之中，结合唯物史观的视角，“人的自由是善”这一命题就是隐藏于人类历史实践活动之中的“道德公理”，它也因此成为道德判断的标尺。马克思曾经将人的发展过程分为了三个大的历史阶段：“人的依赖关系”阶段、“物的依赖关系”阶段、“自由个性”阶段，其中“自由个性”的阶段是建立在“共同的社会生产能力”之上的。^⑧所以也就不难理解为什么尼尔森将生产力的发展看作是道德进步的动力。

科技是第一生产力，人工智能从本质上来看是一种新兴的科学技术，它引领并推动着第四次工业革命。马克思指出：“自然科学却通过工业日益在实践上进入人的生活，改造人的生活，并为人的解放做准备。”^⑨人工智能所引领的第四次工业革命分别从物质财富、自由时间、主体性等方面，为人的解放、人的自由的实现作着充分的准备，从而也为人工智能伦理的“语境主义”进路奠定了坚实的

① 程惠莲：《康德伦理学的主体性思想及其影响》，《湖北大学学报（哲学社会科学版）》1993年第2期。

② [德]黑格尔：《法哲学原理》，范扬、张企泰译，北京：商务印书馆，1961年，第164页。

③ 《马克思恩格斯选集》（第3卷），中共中央马克思恩格斯列宁斯大林著作编译局译，北京：人民出版社，1972年，第152—153页。

④ 《马克思恩格斯选集》（第1卷），中共中央马克思恩格斯列宁斯大林著作编译局译，北京：人民出版社，1995年，第58页。

⑤ [英]史蒂文·卢克斯：《马克思主义与道德》，袁聚录译，北京：高等教育出版社，2009年，第27页。

⑥ [加]凯·尼尔森：《马克思主义与道德观念——道德、意识形态与历史唯物主义》，第14页。

⑦ [加]凯·尼尔森：《马克思主义与道德观念——道德、意识形态与历史唯物主义》，第15页。

⑧ 《马克思恩格斯全集》（第46卷·上），中共中央马克思恩格斯列宁斯大林著作编译局译，北京：人民出版社，1979年，第104页。

⑨ 《马克思恩格斯文集》（第1卷），中共中央马克思恩格斯列宁斯大林著作编译局译，北京：人民出版社，2009年，第193页。

现实基础。人工智能技术渗透到各行各业，深刻地改变着人的生产方式和生活方式，为人类带来日益丰富的产品和服务，为社会创造出巨大的物质财富。人工智能在家居、教育、医疗、交通、物流、金融、社交、传播、装备制造等领域显现出巨大的应用潜力，就近期来看，“IDC 统计数据显示，人工智能市场规模在 2024 年将达到 470 多亿美元”^①。此外，人工智能的应用使得劳动生产率大幅度提高，这为人类节省出非常多的“自由时间”。马克思认为时间对人的存在和发展意义重大，他指出，“时间实际上是人的积极存在，它不仅是人的生命的尺度，而且是人的发展的空间”^②。并且强调，“自由时间”是“从事较高级活动的时间”，“个人得到充分发展的时间”。^③那么什么样的活动是“较高级活动”？人工智能如何使人有大量从事较高级活动的自由时间，进而让人自身得到充分发展呢？福克斯（Christian Fuchs）曾经按照不同的劳动（工作）对象对人类所从事的劳动进行了简单的划分，他认为：“农业和采掘劳动以自然为对象，工业劳动以人造自然为对象，信息劳动以思想和人的主体性为对象。”^④很明显，从目前人类发展的历史阶段来看，在这三种劳动中，“信息劳动”可以称得上高级劳动形式，而人工智能则是信息劳动的高级阶段，它是人类对脑信息、感官信息乃至心理信息、情感信息等进行模拟的产物或过程。“尽管目前的人工智能距离所设想的超级人工智能还十分遥远，其在局部领域展示出的强大能力和应用潜力已经足够令人瞠目结舌。IBM 开发的 Watson 系统与医生在提供肺癌、结肠癌和直肠癌治疗建议方面一致性比例分别高达 96%、81% 和 93%，最近中国本土开发人工智能神经影像系统判别病例的准确率超过了人类医学专家约 20 个百分点。谷歌和微软旗下的语音识别技术准确率都已经达到 95% 左右，接近人耳的极限。微软提供的智能语音语义分析不仅能够对于单一语言有效识别，更能够进行分析和转化，自动翻译成世界范围内 60 种语言。”^⑤不难看出，人工智能通过对人的感官能力、大脑思维等各种复杂信息处理过程的模拟，使人从复杂的脑力劳动中得以解放，从而使人有更多的时间来从事更为高级的脑力劳动——创造，而创造力本身就是人类智能最核心的能力，世界著名心理学家霍华德·加德纳（Howard Gardner）曾经强调：“智能是一种计算能力——即处理特定信息的能力，这种能力源自人类生物的和心理的本能。尽管老鼠、鸟类和计算机也具有这种能力，但是人类具有的智能，是一种解决问题或创造产品的能力。”^⑥但同时也要注意，人的创造能力不仅仅局限于智能领域，人作为一种主体性的存在，具有复杂而综合的本质特性，智能的发展只是为主体创造能力的提升拓展了空间，这种创造不仅仅是霍华德所说的“创造产品的能力”，而是进行自我创造的能力。“而人工智能时代的来临，可能会更加鼓励我们选择区别于此两者的第三种方法，即对自我的创造。也就是说，不仅仅是‘重估一切价值’，更是重估一切赖以重估的存在本身。人不仅要突破‘话语关系’的束缚，还要突破肉体生命的束缚，从而彻底地掌握自己”^⑦。归根到底，人的创造活动是要最终实现人的自由，由此反观之，致力于人的自由的实现的人的创造活动才是真正意义上的“较高级活动”，而人工智能的应用和发展使人的创造活动得以深化，进而也促进了人的自由的实现和人的充分发展。正是在这个意义上，与人的自由紧密相关的马克思主义道德学说的“语境主义”赋予了人工智能伦理全新的研究范式和价值内涵。

四、结 语

① 张彦坤、刘锋：《全球人工智能发展动态浅析》，《现代电信科技》2017年第1期。

② 《马克思恩格斯全集》（第47卷），中共中央马克思恩格斯列宁斯大林著作编译局译，北京：人民出版社，1979年，第532页。

③ 《马克思恩格斯全集》（第46卷·下），中共中央马克思恩格斯列宁斯大林著作编译局译，北京：人民出版社，1980年，第225—226页。

④ Fuchs C., *Digital Labour and Karl Marx*, New York: Routledge, 2014, p.247.

⑤ 俞建施：《未来基石——人工智能的社会角色与伦理》，北京：中国发展研究基金会、微软公司，2018年，第1—37页。

⑥ [美]霍华德·加德纳：《多元智能新视野》，沈致隆译，北京：中国人民大学出版社，2008年，第7页。

⑦ 高奇琦：《人工智能、人的解放与理想社会的实现》，《上海师范大学学报（哲学社会科学版）》2018年第1期。

尽管科克尔伯格所提出的人工智能伦理范式在尽力避免伦理的抽象化，在论述的过程中也使用了“实践”的概念，但他过于夸大主观作用，他将人机的互动实践置于心理感受的体验之上，将人机互动的美好情境置于主观的道德想象之中。与此不同，马克思主义道德学说的“语境主义”将道德的进步根植于人类具体的感性的历史实践活动之中，从客观的人类历史发展规律出发，消除了主观的伦理道德臆想。因此，新的人工智能伦理范式的出发点不再是人机互动的心理体验，而是唯物史观视野中的人的自由实现；新的人工智能伦理范式的道德推动力也不再是人机互动的道德想象，而是具体而现实的人类实践。同时不可否认，科克尔伯格的“人类善的伦理”为人工智能伦理探索了美德伦理进路，使得人们能更加关注人工智能与“美好生活”的紧密关系。然而，对于有无限发展可能的人而言，真正的“美好生活”不是止于人和机器之间和谐、动人的情境，而是人的本质、人的自由的实现。正如胡尔卡（Thomas Hurka）所言：“这一道德理论始于对人类美好生活或真正的理想生活的描述。它以独特的方式描绘了这种生活。某些构成人的本质或者是人最根本的特性使人成为人自身。美好生活使这些特性发展到一个新的高度或者体现着人的本质的核心。不同版本的（道德）理论可能对相关特性是什么有不同意见，因此对美好生活的内容持不同意见。但他们有着共同的基本观点，即人的本质的实现过程是善。”^①可以看出，基于马克思主义道德学说的“语境主义”的人工智能伦理的另一理论优势就是将人的根本目的——人的本质、人的自由的实现——看作是伦理道德的最终归宿，同时赋予了“美好生活”最深刻的价值内涵。

From “Relational Situationism” to “Contextualism” of Marxist Moral Theory: The Turn of Paradigm Foundation of Artificial Intelligence Ethics

WANG Liang

(School of Marxism, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Artificial intelligence (AI) ethics serves for the prosperity and well-being of human beings. However, there are different theoretical propositions: “Theory of agent center” maintains that agents should conform to human ethical norms, and it is necessary to conduct ethical design and supervision on them; “Relational situationism” holds that human beings are the center of AI ethics, not agents, and AI ethics should be for human good. “Contextualism” of Marxist moral theory links productive forces, morality and human’s freedom together from the height historical materialism and thus lays the foundation of the ethical paradigm of AI from a new height.

Keywords: Relational Situationism, Contextualism, AI Ethics

[责任编辑：谢雨佟]

^① Hurka T., *Perfectionism*, New York: Oxford University Press, 1993, p.16.