

人工智能可能拥有“自我意识”吗？

——基于马克思的存在论之思

万云峰

贵州大学哲学学院，贵州 贵阳

收稿日期：2024年4月20日；录用日期：2024年5月10日；发布日期：2024年5月21日

摘要

人工智能(Artificial Intelligence, 简称AI)是一门研究如何使计算机能够模拟和执行人类智能任务的科学和技术领域。在科学技术飞速发展的今天，人工智能以极大的速度发展和应用到极宽广的领域。由人工智能的飞速发展引发人们对人工智能的一系列思考，人工智能是否可能拥有“自我意识”便是其中激烈讨论的话题。而马克思的存在论对人之为人的本质规定的思考极其深刻，故本文基于马克思的存在论来考察人工智能是否可能拥有“自我意识”。

关键词

人工智能，马克思，存在论，自我意识

Is It Possible for AI to Be “Self-Aware”?

—Based on Marx’s Ontological Thinking

Yunfeng Wan

College of Philosophy, Guizhou University, Guiyang Guizhou

Received: Apr. 20th, 2024; accepted: May 10th, 2024; published: May 21st, 2024

Abstract

Artificial Intelligence (AI) is a field of science and technology that studies how computers can simulate and perform tasks of human intelligence. Today, with the rapid development of science and technology, artificial intelligence is developing and being applied to a wide range of fields at a great speed. The rapid development of artificial intelligence has triggered a series of thoughts about artificial intelligence, and whether artificial intelligence may have “self-awareness” is a topic of heated discussion. Marx’s ontology is extremely profound about the essence of human beings,

so this paper examines whether artificial intelligence may have “self-awareness” based on Marx’s ontology.

Keywords

Artificial Intelligence, Marx, Ontology, Self-Awareness

Copyright © 2024 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

当今的人工智能发展突飞猛进，人们对其远超出人的强劲的能力抱有警惕甚至恐惧的态度。人工智能一旦具有了“自我意识”，其存在对人类来说将可能会是毁灭性的，因为人类不知道这样具有自我意识的人工智能对人类将会采取什么态度，是友好或是敌对？而“绝对强者不需要为伤害申请理由” [1]，无疑拥有自我意识的人工智能对人类来说就是这样的“绝对强者”，因为人类相较于人工智能所具有的那点优越性——即自我意识、理性或思维能力——不存在了。一定程度上反映了这一事实的例子就是：包括特斯拉 CEO 马斯克、苹果联合创始人沃兹尼亚克和人工智能领域顶尖专家、图灵奖得主本吉奥在内的 1000 多名行业高管和专家签署公开信，呼吁暂停开发比 GPT-4 更强大的 AI 系统至少 6 个月，称其“对社会和人类构成潜在风险”。¹人类对人工智能的种种态度表明，对人工智能的反思已经到了避无可避的境地了。这种反思的必要性，甚至决定了人类对人工智能的进一步发展方向。正如学者赵汀阳所认为的那样，“任何智能的危险性都不在其能力，而在于意识” [2]，一个具有强大功能但却不具备自我意识的人工智能是不具备危险性的，其存在充其量只是一个工具。而正是在这里，对人工智能是否可能拥有“自我意识”的讨论，显示出了现实的必要性，因为这关涉到了人类的根本生存大计。

马克思向来是关心人的，马克思的哲学也从来不把注意力放到倒悬在空中的形而上的世界。一百多年前的马克思哲学从来没有因为时间的流逝而褪色，恰恰相反，马克思哲学在当代也显示出了其顽强的生命力，它对我们当代的生活依旧蕴含极强有力的解释力。故本文从马克思的存在论的角度来考察人工智能问题，探讨人工智能是否可能拥有“自我意识”。

2. 自我意识与实践的关系

马克思的存在论的实质究竟是什么呢？恩格斯称马克思的《关于费尔巴哈的提纲》是“包含着新世界观的天才萌芽的第一个文件”，而马克思新世界观的核心、实质和根本就在于“实践”。无疑，其存在论的实质或者说根本重要之点也就在于“实践”。“‘实践’从根本上来说，首先是关于存在与非存在、关于存在的现实性、关于自然界和人的现实存在的基本原则” [2]，既然马克思的“实践”首先就是在存在论高度上谈的，那我们就该首先考察“实践”在存在论上的意义。

实践，在马克思那里，基本可以和感性的活动或者说对象性的活动做同义替换，原因在于它们都指

¹ 在未来生命研究所(Future of Life Institute)官网上，包括图灵奖得主约书亚·本吉奥(Yoshua Bengio)、特斯拉 CEO 埃隆·马斯克(Elon Musk)、苹果公司联合创始人史蒂夫·沃兹尼亚克(Steve Wozniak)、DeepMind 高级研究科学家扎卡里·肯顿(Zachary Kenton)等在内的数千名 AI 领域企业家、学者、高管发出了一封题为《暂停大型人工智能研究》公开信。这些 AI 专家们在信中强烈呼吁：所有 AI 研究室立刻暂停训练比 GPT-4 更加强大的 AI 系统，为期至少 6 个月，并建议各大企业、机构共同开发一份适用于 AI 研发的安全协议，同时信中还提到各国政府应当在必要的时候介入其中。

“人的现实的存在或现实的人的存在” [2]。而对象性的活动在马克思的语境之中，其所以是对人的存在论高度的规定，最主要的就是因为它揭示了现实的人的存在方式。马克思曾说：“意识在任何时候都只能是被意识到了的存在，而人们的存在就是它们的现实生活过程” [3]。首先，这就是说从人的存在来说本身就是这样的感性的活动，而不是费尔巴哈式的将人的存在视为感性的对象，也就是说，人的存在从一开始就不能视为一种实体般的存在，然后把一种外在于这种实体式的人的活动赋予它，人的存在从一开始就内在地是人的感性的活动。而从这里就可以看出实践关于人的存在方面的问题，人是在对象性的活动中才成其为人，假如人失掉了对象，人不再活动(对象性地)，那人就只是抽象地存在着的人，如此这般存在抽象地存在的人其实也是无。而人的这般的存在方式何以使人有了人的自我意识呢？

通常我们认为，意识是指人类或动物对自身、环境和外部刺激的感知和认知能力，它是一种主观的、个体化的经验，包括知觉、思维、感情和意愿等方面。从这样的对意识的认识和理解我们可以知道，有意识的人或动物面对外部环境的变化或刺激时会相应地做出反应。而自我意识虽然是意识的一种，是意识的一种特殊形式，但却与意识有了本质上的差别，即自我意识是人独有的意识，自我意识是能够意识到自己的独立存在状态的意识，它能够以自我为对象，这也就是为何人类个体能够进行自我认识和自我反思，自我意识使得人类可以思考自己的思维、情感和行为，从而有能力进行自我调节和自我完善。从古至今人们对自我意识有相当程度的思考，如古希腊一座名为德尔菲的神庙就镌刻了“认识你自己”的箴言，这同时也是苏格拉底的哲学主张，再如近代的笛卡尔，他所说的“我思故我在”同样为世人所耳熟能详。但他们所提出的这样的关于自我意识的思考，都存在这样的特征：抛开具体的人的活动来思考人。苏格拉底说认识你自己，开启了西方两千多年来对人的自我的认识历程，其中自我意识是其中重要的一个方面；但哲学家们似乎都对现实的人以及现实的人的活动不感兴趣，直到马克思，才改变了这一状况。

马克思所提的实践或者对象性活动是对人的存在的本质规定。人正是在对象性活动中，人的自我意识才得以形成，物之为物才得以生成。马克思说：“当现实的、肉体的、站在坚实的呈圆形的地球上呼出和吸入一切自然力的人通过自己的外化把自己现实的、对象性的本质力量设定为异己的对象时，设定并不是主体；它是对象性的本质力量的主体性，因此这些本质力量的活动也必须是对象性的活动” [4]。马克思的这段话表明，正是在人的本质力量的对象化活动过程中，人才获得了“对象性的本质力量的主体性”，什么是人的主体性呢？主体性就是人在面对客体时所拥有的那种自在自为的本质力量的那种特性，动物就不具备主体性，它自身就是环境的一部分。而主体性的根基在哪里呢？在自我意识。由对象性活动，人获得了或者说生成了自己得以超越自然意识的那个主体性，生成了自我意识。而对象呢？回答是对象也是在对象性活动过程中得以生成的。在马克思之前的哲学家对对象的物的理解，大体上可以分为三种，分别是自然物、机械物和观念物[5]。第一，古希腊的哲学家倾向于把万物的本原归结为某种自然物质，用以说明整个世界的生成和发展，此时人们理解的对象或物是一种自然物；第二，近代机械唯物主义认为人的活动遵循自然因果的规律，人的精神归根结底也只是物性，这就是说，他们以同样的方式对待人和物，人和物没有区别，他们把一切都视为机械物；第三，唯心主义主要将物理解为观念的物，黑格尔哲学是这方面的渊薮，这就是马克思所批判黑格尔的地方，马克思说：“自我意识的外化设定物性。因为人 = 自我意识，所以人的外化的、对象性的本质即物性(对他来说是对象的那个东西，而且只有对他来说是本质的对象并因而是他的对象性的本质的东西，才是他的真正的对象。既然被当作主体的不是现实的人本身，因而也不是自然——人是人的自然——而只是人的抽象，即自我意识，所以物性只能是外化的自我意识) = 外化的自我意识，而物性是由这种外化设定的” ([4]: p. 208)。而到了马克思，他颠覆了传统哲学的对象的理解，他把物理解为与人相联系的对象，他说的“设定并不是主体，它是对象性的本质力量的主体性，因此这些本质力量的活动也必定是对象性的活动” [4]。表明，事物只

有在主体存在时，作为客体的事物才是存在的，否则便是一种非存在，而主体只有在设定客体时，主体才称其为主体。

综上所述，人在进行对象性活动即实践时才获得了其主体性，而主体性所由的主体的自我意识同样也是在这个过程中生成的。

3. 人工智能的“活动”与实践的关系

前面讨论了自我意识与实践的关系，认为自我意识是在实践过程中形成起来的。那么，仿照人类行为模式而建立的人工智能的“活动”(活动一词可能引起误会，仿佛人工智能也如同人一般，但为了讨论的方便，在这里我们把人工智能的运行也用活动来指称)与人的实践之间有什么关系呢？下面我们将简单探讨一下人工智能的工作原理或者说人工智能的“活动”和基本情况。

人工智能的背景可以追溯到上世纪 50 年代，当时科学家们开始探索如何让机器模拟人类的智能行为。然而，由于计算机处理能力的限制以及缺乏足够的数据和算法，人工智能的发展进展缓慢。随着计算机技术和算法的进步，尤其是机器学习和深度学习的兴起，人工智能开始迎来爆发式的发展。机器学习使得计算机能够通过数据学习和改进性能，而深度学习则基于神经网络模型实现了更高级别的模式识别和抽象能力。这些技术的发展推动了人工智能在各个领域的广泛应用，如自然语言处理、计算机视觉、语音识别等。人工智能的定义也在不断演变。现代人工智能强调计算机系统能够模仿人类智能的各个方面，包括感知、学习、推理和决策。人工智能的目标是使计算机具备智能的能力，能够自主地解决复杂问题，并与人类进行自然和智能的交互。

人工智能的基本活动原理是通过模仿人类智能的思维和行为方式，利用计算机系统进行信息处理和决策。人工智能的基本原理涵盖了数据获取与处理、机器学习、深度学习、自然语言处理以及推理与决策等方面。这些原理的结合与应用使得人工智能可以自动化和智能化地处理和分析数据，学习和改进性能，理解和生成自然语言，并做出推理和决策。

人工智能的应用场景极其广泛。无论是医疗行业中的疾病诊断和个性化治疗、金融行业中的风险管理和欺诈检测、制造行业中的智能生产和预测维护、交通行业中的自动驾驶和交通管理、教育行业中的个性化学习和智能辅导，还是零售行业中的智能推荐和无人店铺，以及能源行业、农业行业、娱乐行业和安全与监控行业，人工智能都在不同领域展现出了巨大的应用潜力。普通人对人工智能的第一反应实质上就是对人工智能取代他们岗位的担忧，由此进一步引发了对人工智能的其他思考，如人工智能是否拥有以及是否可能拥有自我意识等问题。

从人工智能的发展来看，其更多的还是依靠对人的模仿，比如对人的神经网络的模仿建造的模型的让其有了更高级别的模式识别和抽象能力；可以看到，现代人工智能更进一步，甚至能够深入学习和自我迭代。我们不禁要问：如此智能的人工智能，当与其对话时甚至察觉不出交谈的对象是人工智能时，这样的人工智能拥有自我意识吗？尽管人工智能是通过模仿人类实现其现状，但我们通过比较其活动与人类实践的底层逻辑就可以得到答案。

人类的实践，从马克思存在论的角度看，是对象性的活动。首先，这里的“活动”，不是德国古典哲学家们所说的活动，不是康德所说的“纯粹的活动”，不是费希特的“活动本身”，也不是谢林的“无限活动”，更不是黑格尔的“自我活动”，上述所提及的“活动”都是自我意识本身的活动，而在马克思，现实的活动不是这样的主体内部的活动(自我意识本身的运动)，马克思所言的是“对象性的本质力量的主体性”，也就是对象性的活动。其次，这里的“对象性”或者说对象性原理来自费尔巴哈的哲学。在费尔巴哈看来，存在的“感性”才是真正现实的，对象性使主体的现实性得到了保证。对象性原理主要体现在费尔巴哈的这两个命题之中：第一，没有了对象，主体就成了无；第二，主体必然与其发生关

系的那个对象，不外是这个主体固有而又客观的本质。而人工智能的活动，其“活动”首先有赖于人类事先输入的程序，在此基础上，运行人类输入的指令，这是最重要也是最关键的一点，那就是人工智能不具备自主性，人工智能没有可供自身遵守的自己的“意愿”。与人的对象性的活动相比，人工智能的“活动”不是其基本的存在规定，人工智能的存在是人类赋予的，属于“物”的范畴，即便它是貌似“聪明”一点的物，但一定意义上连“智能”都算不上。与人类活动的对象性相比，人工智能活动的对象(如其检索的数据或者编写的代码)也是人类事先就已经发明出来的，人工智能在其中不会遇到感性障碍。

4. 人工智能的“活动”不能使其产生自我意识

人的实践从根本上来讲即本质性方面是一种关联了人与人的对象的活动，人所以被认为是人，人的对象所以被认为是人的对象，就在于两者是在人的实践活动中被这种实践活动先行征用为关联项，由此人与人的活动才获得其本质界定。一方面，既然人的实践才是两者获得各自本质界定的终极根据，那么人在其中是逻辑在先的，即没有人，便没有这样的把人与人的对象关联起来的实践活动，便没有把人“设定”为人使其获得主体性的对象性活动，也就没有把物“设定”为对象的对象性活动。另一方面，人在实践活动中所获得的对象即前文所言的“物”的对象可以分为自然界、人类社会和人的思维，在这里人工智能的“活动”与人的实践的最大差异就是：人能够以自身为实践对象，人在人的生命过程中能够观照自身，即能够反思人类社会和人自身的思维或精神的病症；反观人工智能，它只能在人类为其构建的数据库中即已有的信息中畅游，它不能够突破这个“已有的”疆界而从无到有开辟出一个新的世界。人可以认识到自己是人，但人工智能未能认识到自己是“人工智能”。

诚然，人工智能在其现实性的能力或曰功能上远远地超越了人类，人工智能功能性地超越人类的领域实在数不胜数。人眼的探测范围仅仅局限在可见光，而给人工智能赋予外在设备，那这样的“机器人”的“视觉”就不再局限于可见光，甚至连紫外线、无线电波、红外线、微波以及全频谱波段之电磁波都“看”得到，这直接的结果就是在人眼不可及的视界实现成像，如此以至于，人工智能对世界的认识扩展了，而可见光之外的视界是人之眼远远无法抵达的彼岸。对声音之感知、认知，人工智能其实也远超越人类，于人而言，我们只能在一定的波段内识别自己熟悉的声音，并且范围只能是我们已经掌握的语言，如母语和已习得的外国语；而人工智能则由于语言库的存在而能够实现对人的几乎所有语言的识别，人工智能在声音、语言方面的强势之处就表现在对自然语言之处理，人工智能实现了与全球各民族之交互，尽管是机械地做到这一点，人工智能甚至可以实现“人与机器”、“机器和机器”之间的交互，当然，人毫无疑问能借助人工智能之中介实现与各不同语言之间民族的交流，自然语言的交互能够让机器沟通无障碍。

即便人工智能对于人来说拥有种种优势，但我们仍然不能说人工智能的“活动”已经是一种实践，因为它不能摆脱人类为其准备好的熟悉的东西即人为键入的数据、代码以及各种外在实体设备。人工智能可以依靠这些东西做到某些单个人永远无法完成的工作，甚至于我们可以极大胆地说人工智能的“活动”的完成就是其所积攒的“经验”，因为确有事实表明了人工智能的深度学习为其后面的“活动”所做的准备工作。但正如学者赵汀阳认为，“经验能够发展出意识，但不足以发展出自我意识，意识需要实现客观化，即内在性映射为外在性或外在性映射为内在性”[6]。笔者认为，人工智能的活动从本质上还不能脱离器物的范畴，其活动尚不能脱离人类的输入而独立完成，这就意味着，人工智能无法实现从无到有的过程。因为，没有自主性的人工智能，其底层逻辑还是人类智力的体现，就仍然是马克思所说的“人的对象性的本质力量”人工智能方式，而不是其他。如果连人工智能类似于动物式的活动都无法完成，其几乎难以完成意识的诞生，更遑论自我意识的实现，因为人工智能在其现有的存在条件或存在方式下无法获得主体性的存在。这就是赵汀阳教授所说的“主体性始终是一个创造者而不仅仅是认识

者;主体性的要义不仅在于认识自己,更在于创造自己;从根本上说,主体性不是一个知识论概念,而是一个存在论概念” [6]。

5. 结语

对人类和人工智能之关系的思考和探索,从人工智能技术诞生之日起就没有停过。著名物理学家霍金曾说过,成功创造了人工智能,很可能是人类文明史上最伟大的事件。但如果我们不能学会如何避免风险,那么我们会把自己置于绝境;尼采也曾经说当你凝视深渊的时候,深渊也在凝视着你。所以,对于人工智能是否可能拥有自我意识这件事在不可掉以轻心,哪怕只是在一种极微小的可能性上。但从人工智能与人的类比(这种类比是不可避免的,毕竟人工智能从存在之初就是依靠模仿人类而得以成型)来看,从马克思的存在论角度来观人工智能,人工智能在目前的存在方式中不太可能产生自我意识。一旦人工智能到了“想要”生成“自我意识”的时候,那也就是人工智能的冬天的迫近。

参考文献

- [1] 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(1): 1-8.
- [2] 吴晓明. 马克思实践学说的本体论意义[J]. 南京社会科学, 1996(8): 25-38.
- [3] 马克思, 恩格斯. 马克思恩格斯选集(第1卷) [M]. 北京: 人民出版社, 2012: 152.
- [4] 马克思, 恩格斯. 马克思恩格斯选集(第1卷) [M]. 北京: 人民出版社, 1995: 208-209.
- [5] 陈艳波, 王奕霖. “物”之为历史性的“对象”——论马克思对“物”的分析[J]. 东岳论丛, 2022(7): 95-105.
- [6] 赵汀阳. 替人工智能着想[J]. 哲学动态, 2023(7): 41-45+127.