

基于朱子学动机视角探讨AI绘画的道德困境

姚孟非

中国科学技术大学马克思主义学院, 安徽 合肥

收稿日期: 2024年1月29日; 录用日期: 2024年3月1日; 发布日期: 2024年3月8日

摘要

2023年是AI绘画“泛平台化”的重要发展阶段, 基于算法技术的进步, 人工智能的运行降低了对计算机软硬件的要求, 从而使其用户量激增, 并派生出许多风格各异、功能不同的应用程序。随着AI绘画愈受关注, 其自身所存在的道德问题就更为显著, 用户群体的差异、程序设计的缺陷和绘画行为本身经验和道德需求使远未成熟的数字模型的学习和交互模式暴露出动机短板。在朱熹预设的道德情境下, 一个行为的开始至完成, 动机比实效具有更显著和重要的地位, 人工智能的底层逻辑与用户群体的广泛参与催生出朱子学视域下“学”与“私”的双重困境, 进而导致AI绘画无论是在艺术层面还是在本质上皆不具备道德意义。

关键词

AI绘画, 道德动机, 朱熹

Exploring the Moral Dilemma of AI Painting from the Perspective of Zhu Xi's Motivation

Mengfei Yao

School of Marxism, University of Science and Technology of China, Hefei Anhui

Received: Jan. 29th, 2024; accepted: Mar. 1st, 2024; published: Mar. 8th, 2024

Abstract

2023 is an important stage of development for AI painting's "pan flat platform". Based on the advancement of algorithm technology, the operation of artificial intelligence has reduced the requirements for computer software and hardware, resulting in a surge in its user base and the emergence of many applications with different styles and functions. However, with the increasing attention paid to AI painting, its own moral issues become even more prominent. Differences in user groups,

文章引用: 姚孟非. 基于朱子学动机视角探讨 AI 绘画的道德困境[J]. 哲学进展, 2024, 13(3): 407-412.

DOI: 10.12677/acpp.2024.133061

flaws in program design, and the experience and moral needs of painting behavior itself expose motivational weaknesses in the learning and interaction patterns of immature digital models. In the moral context pre-set by Zhu Xi, motivation plays a more significant and important role than effectiveness in the process of starting and completing an action. AI painting lacks its own moral normative ability, exposing the dual dilemma of “learning” and “private” in daily situations. As a result, it cannot give rise to moral realization both at the artistic level and in essence, and is more likely to lose the user’s recognition of the meaning of the creative form.

Keywords

AI Painting, Moral Motivation, Zhu Xi

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

AI 绘画是基于算法及对抗性网络，通过对艺术作品和设定的理解，再经由数据分析和计算机处理，最终得到类似于“仿象”的“创作”绘画产物。随着时间的推移，AI 创作的模式已经更新迭代，就最新的产出方式探讨其主要原理，即当用户或者开发者对该算法进行一些素材的预设，被投入系统的素材将变成数据或特征被拆分成简单的符号或组成部分存入数据库，这些被拆分的“相”会因为彼此的相似与联系程度被存储在计算机不同类别的库中。当 AI 需要以某种风格或是特征进行绘画时，其指向会使它从不同判定的素材库里抽取有用的资源，再经过拼接等艺术手法的加工，或是重新着色等风格设定的处理，最终使不同的“相”形成一个可供欣赏的、相对和谐的“共相”。也就是一副“以假乱真”的类人工作品的诞生。

动机是朱子学中的一个重要概念，也是朱熹进行道德评价的核心标准。在其理学家的立场中，朱熹认为道德的价值应与动机密切联系，在他的道德预设下，其所继承并完善的义理学说和天理系统排斥私欲，试图教导人们通过“格物致知”进行学习，继而达到理想中完美的道德境界。这同人工智能的发展路径显然存在偏差。一个无法被忽略的事实是，人工智能自面世以来就饱受非议，尽管其带来无数利益和超越其本身领域的价值，但正如生物医疗领域的基因技术的进步引发了伦理矛盾，科技的迭代一定会伴随着道德问题的到来。AI 绘画自然不能摆脱该魔咒，如果说早期的小众性或是偏重“发明”的过程让研发过程更为理性和冰冷，那么自 2022 年末，AI 创作这一话题开始大面积在常规互联网媒体、互联网平台之间流通，大量非软件技术人员参与到话题的讨论，使能够提供 AI 绘画这一功能的软件广为传播，AI 作画的发展流变已经从“发明”期过渡到了“使用”期。由于市场的急速膨胀，公共算法已经不能满足各开发商的胃口，为了迎合客户，更是为了独特算法提升竞争力，开发人员需要研制独特的学习模型，自然更加欢迎普通用户对程序智能度的调教作为协助。于是，大量的使用者以协助的身份涌入市场，作用于 AI “仿相”的开发环节和体验环节，AI 创作不再是工作室或者专业人员的特权，故而其道德矛盾更为显著，在网络上引起怀疑与抵制的浪潮。不少反对人工智能介入创作的用户认为，AI 绘画存在天然道德缺陷，然而其中症结在何处却未有定论。若对该问题引入朱子学动机视角研判，通过朱熹道德预设情境之中为“学”与去“私”的进路，我们便可自 AI 绘画本身的运作原理与对 AI 绘画用户群体的分析两方面，探讨 AI 绘画在当前发展过程中呈现出的道德困境。

2. 对朱子道德动机的解释

朱熹曾以“可见其立心之本在于功利”反驳陈亮提出的“正欲揽金银铜铁镕作一器，要以适用为主耳”[1]，是故以朱熹的道德标准判断，动机是优先于结果被考虑的东西，即使最终得到的效果并不理想，但假如动机是好的，那么就可以评价整个行为是道德的。“如意志完全丧失了实现其意图的能力，所剩下的只是善良，它仍然如一颗宝石，自身就发射着耀眼的光芒，自身之内就具有价值。”[2]与同样重视“善良意志”的西方哲学家康德类似，朱子的二元论是价值与事实的二元论，是道德哲学意义上的“应该”和“是”的二元论，道德作为一种道德规范，赋予事物的“实然”以“应然”的意义[3]。在一个行动中，道德动机的存在具有绝对不可替代的意义，朱子以为，“心”主宰道德的产生，“心，主宰之谓也”，即作为道德律的“绝对命令”，是道德的先验、客观、必然的原因，又由于心与理一，故心主宰身体的活动，也发用情的流行，而道德意志就是先验的仁心[4]，所以在时间上，动机是意和情流行的开始，也是心的官能所展现的开端，而在意义上，对动机的判断影响到对整个行为的评价，不仅是事前评估值不值得、应不应该的标准，也是事后判断有没有效、正确与否的规则。要使一件事作为成果而非恶果，在性质上，出于赋予实效以价值的目的，对动机的修正和转化也影响着行为的道德意义。

3. 从运作原理探讨 AI 绘画之“学”的困境

2023 年是 AI 绘画发展过程中最为枝繁叶茂的一年，实际上，将人工智能应用于绘画并非罕事，早在上世纪，机械臂 AARON 的出现就已经昭示着 AI 绘画的可行性。该机械臂由一定的程序进行引导，并在开发者的控制下进行绘画创作。不同于早期从形式上模仿人类在画纸上进行创作的过程，由于算法的精进和开发者的思路转换，从主动收集进步到被动数据处理，新型的 AI 绘画软件所采用的学习型数字 AI，作为内核，其表现终于更加趋向众人想象之中的人工智能。基于 AI 绘画软件本身而言，学习型 AI 是模拟人类知识经验的积累以提高自身算法精确程度的模型。在目前软件所生成的图像表现中，作品的生成多半来源于对海量的无版权、无授权图像的破坏重组。算法将代表美术的特征：色彩、笔触、线条等等艺术构成，解构并转换为可供计算机读取的数据代码，待用户的指令唤起后再将之提取、分配，以排布的方式重现出一副成型画作。那么这就意味着素材如何被拆解、分类再重组是 AI 本身应当注重的问题。目前大多数 AI 会依据用户上传的图片进行模仿，但仅靠使用者的培养是不够的，学习型模型需要大量的“临摹”素材以供学习和参照，可以说，样本的丰富度直接影响其运算过程的智慧程度。所以，实际上 AI 会长期不间断地在互联网上进行图片的“吞噬”，即使在人类画师休息的时候，作为电子画师的它们正在无休止将一切可以为己所用的作品收入囊中。这和 AI 智能语音、AI 棋手的原理是一致的，学习型模型必须通过这种方式进行算法的成长。对于它们而言，一切深刻复杂的艺术作品都会变成简单的、可供描述或是指向的数据或符号，经由算法处理和加工变成素材，等待着使用者的字、词、句将这些被打散的碎片从分类归置的库中取出，再糅合拼接出一副符合逻辑的画作。

故此，从学习进路来看，传统儒家的视域下学习的概念并非简单的知识增加或经验积累，而是对个人境界的修养兼顾知识层次的提升。在朱子看来，宇宙本体和道德本体具有天然的关联，作为形上本体的理落实在个人层面，就是要以心去显化“性即理也”的性。性理同一故而心理同一，心理的一贯由是才符合“心包万理”的先验假设。周敦颐在其《太极图说》中提到：“惟人也得其秀而最灵”[5]是因为人得以形神具备、五性感动，所以各一本体之性。而数据程序则不能在运算中获得理一本而万殊的效果，其在诞生之初就没有被赋予道德完全体的责任，所以不具备道德先天的心器，便也不具备心的官能，无法承担道德义务，也不能形成道德自律。在道德行动的动力机制上，理论上而言，道德行动的动力是由行动者的道德动机提供的，在朱熹看来，道德行动者因具认知与意志双重道德动机状态，从而能够强烈地激发其做出恰当的道德行动[6]。所以具备认知为前提的动机尤为重要，同时出自学习主体的认知能力

和独立意志是能够端正动机的前提，“大凡道理皆是我自有之物，非从外得。”^[7]但学习型模型的进步显然不能依靠闭合逻辑，代码的开源与无筛选用户群体致使其道德自律丧失，不得不妥协于道德他律。此外，在朱熹的认识论中有一个知名的“工夫”，即格物致知。不过在格物这一途径上，朱子有两个非常鲜明的限定，一者在于，格物致知的主体是动机持有者本身，《大学》有载“修己治人”，即“欲修其身者，先正其心。欲正其心者，先诚其意。欲诚其意者，先致其知。致知在格物。”^[8]正心诚意、格物致知，其价值导向和指归都是道德主体自身，而并不存在其他客体。其二则是朱子的格物致知论对于知识积累的量变质变存在非常鲜明的解释路径，而经验是学的重要构成，朱子以为“上而无极、太极，下而至于一草、一木、一昆虫之微，亦各有理。”^[7]，要穷理，就要格天下之物，没有分别，没有限制。那么在 AI 创作过程中，无论是程序本性不具备涵养道德动机的前提，还是人工智能的学习过程同朱子格物致知的理想状态完全不符，二者俱表现出完全偏差于朱子的为学路线的情况。

此外，从学习对象来看，随着 AI 绘画的流行程度和覆盖广度，在越来越多的人使用的情况下，就会有越来越多的“素材碎片”出现在库中。其中不乏许多画作的水印被一并作为文字素材。这些水印在以往是作为绘者的落款，且在一定程度上起到以遮盖画作主体的方式保护版权的作用，但对于 AI 来说这种手段几无效果，故而很多对原创性极为重视且对于 AI 绘画的不告而取行为有深度恶感的原创画师，正在集中智慧试图发明更为容易但有效的、针对 AI 搜索和读取的防盗水印，那么 AI 想要更专注的训练，就必须克服这一新的问题，破解新的水印或是图片格式以优化自己的识图能力。这样的迭代带来的后果有二，第一，如果仅以 AI 创作这一学习方式界定，其行为的后果竟然是非其程序本身在自行成长，而是因其学习的对象的进步引起了主客体的颠倒，并在第一阶段相当增长了客体的经验，以此推进了本体的被动学习；第二，这种进步是有限制的，经验的范围被限制在防盗与破解的命题下，所以针对性的学习和训练，最终带来的是开发者动机的偏离，起因即同样在于人工智能自原理上不具备道德规范能力，故其行为举止不为动机所自发，反而为外部条件所限。法国心理学家梅洛·庞蒂认为，对于具身认知的本质解释是身体、知觉和世界是一个统一的整体，人的知觉存在于身体的主体之中^[9]。所以，学习型 AI 模型不能使知觉与主体从二元分裂中解脱出来，从而无法具备身心一元化的具身道德。

4. 从对用户群体的分析中探讨 AI 绘画之“私”的困境

近几个月来，随着网络社交媒体的不断传播和国内诸多类似于网易、腾讯等知名度较高的厂商的带头造势，AI 绘画逐渐脱离专业的软件支持，变成一种嵌入式的交互体验，被放置在诸多应用程序的选项界面。而针对这些厂商的研发思路，他们的目标群体由于创作模型的风格和功能倾向的不同，也产生了较大的区别。由于不同的功能划分，不同的用户群体在使用过程中的需求倾向和体验深度不同，所以其对工具的精度要求和成果预期原本也有所差别，这就意味着每个人在进行绘画前一定有一个判断，即这样的作品是出自于数据和算法，并不出于我本人完全的原创性设计，我的创作思路的最终成型依赖于该软件的程序输出。从而当引入动机视角时，我们会发觉在使用者层面已经先入为主地产生了自然的动机立场与道德判断。朱熹对于动机的重视完全体现在他的思想之中，在他的学说体系下，理具有不可动摇的形上地位，他以为：“至于天下之物，则必各有所以然之故，与其所当然之则，所谓理也。”^[10]所以，在朱熹的思维里“所当然”和“所以然”是分离的，但同样都是需要关注的部分。朱熹又在后文提到，“所当然”的事情或有“不容己”者，那么在这时就应该去求其“所以然”。按说理应当具备万殊之义，那么“不容己”是如何产生的，朱子以为，是因为“心”和“理”产生偏离的原因。“理便在心之中，心包蓄不住，随事而发。”^[11]事物引起心产生活动，心作为道德动机的缘起，再通过引导行动落实心的官能。“动机在实质的层面表现为行动的动力因。”^[12]心所蕴含的先验认知、情感显发和欲望生成，都意味着其具有道德判断的本领，同时也具有规范的作用和实现的能力。在朱子的概念体系中，“发于智

识念虑处皆是情”、“志和意都属情”，情正为心所统，那么，基于使用者而言，在参与交互的行为中，不可能没有心的参与，然而在 AI 创作过程中，本应为心所掌握的道德能力却似突然消失了，究其原因，或可以归于“私”的作用。

在朱熹的定义中，“涉于人为”就是人心发用之私的开始。而相比较“公”，私正是具有特殊性的，独属于某一位所有的一种情绪、一种立场。“私”是针对 AI 绘画的使用者的评判，从使用动机上看，用户在使用该工具之前先有自身的计量，他们预先将所需所感输入到系统中，已率先存在意愿的偏向。那么，想要不涉及人为，就需要“无心”“无情”，并不是说用虚无的状态应接，而是在行为举止中不存在自为的计较。亚里士多德说：“我们不能对产生欲望的东西说好说坏。”[13]，也就是说，在平衡未遭到打破的时候，人性不能被称为善或恶，而朱熹也秉承了性善说的观点，私的生发处，就在于心的发用。然而，在关于私的探讨中，时常将私和欲联系在一起，因为私本身就包含着“耳目口鼻之欲”的定义。不过有关于欲本身的优劣，其评价是根据对欲望属性的裁定，“然人不能无私欲之累”，也就是说其实私欲是不可以简单地从人的想法中刨除的，只在于私欲本身的性质不同，它们受到的待遇便也不同。合理的、符合生理需求的欲望是人的本能，也是天理流行使然，不能被归于恶的存在，但是在生活中要“克去己私”是为了克服什么，朱子的答案是，要尽力去除那些超出天理的部分，即过度的欲望。过度的欲望会引导人生发恶行，所以，他为恶寻找的生发处就是超过天理的私欲，即在于生命中那些非必须的但又能够引起人巨大满足感的行为。回归到绘画创作的行为上来，朱子对于艺术的定义也存在不偏于私的观点，他称赞王羲之“不与法缚，不求法脱，真所谓一一从自己胸襟流出者。”[14]他认为“不与法缚，不求法脱”是艺术的体现，也是不为法度束缚，但也不偏离于法度的一个对于“中”的践行。美的起点和终点是自符合伦理的道德动机始，至感性情感与理性认知达到和谐统一而终。在这一点上，纯粹的道德动机不仅在整个行为展开时是避免人为过分的匡正，也是最终行为结束效果形成时达到理想状态下艺术和理性平衡相合的必备条件。所以，在创作过程开始之前就保持中正之义，这对于使用 AI 模型这一算法的用户来说几乎是不可成的命题，一者，整个创作并非简单的使用个人技巧的简单能动形式，工具本身作为相对独立的运行存在，虽为意识倾向所影响，但本不为用户自身素质决定；二者，极强的目的性替代了原始艺术创造的单纯动机，并为部分依靠软件盈利的用户群体而言，创作作品已不是需要用道德评判的行为，而是牵扯实际利益的谋生手段。因而用户的广泛参与实际已造成了“私”的加入，即不同群体在使用人工智能软件生成作品时，本应为其心所统摄的道德前因已被目的性所取代，最终造成了能够规范动机的中正之“理”的溃败。

5. 结语

不能忽视的是，朱熹所秉持的道德先验论这一起点是基于人类人格而设定的。当前尽管很多科学家已经解码了人类的生物基因和普遍的进化进程，而意识本体却是一个难以解开的谜团，即使人工智能产生了类似的人类意识，这个计算机意识将仍然是人类意识的延伸。如是当下的人工智能仅仅是一种技术，其操作还是完全依赖于意志主体，因此人工智能技术的出现并未消解主体，也没有消解艺术创作媒体的意义[15]。回到作品本身，艺术审美和创作是概念抽象和客观经验的融合，需要感官的加工，也需要对认知进行解构和重组。在这个层面上，程序的学习基于机械式的模仿，其行为本质无流行的发生可言。《宗镜录》卷五五释影像尘云“影者流类义，像者相似义，即所变相分是本质之流类，又与本质相似，故名影像。”影像，即是影子的映像，虽然看起来是有，相似于本有对象，但既经根识之变现，便非认识对象本质尘之实体，也非全同于本质尘的原样，也就是“仿象”的存在含义[16]。人工智能画作亦可以称作人类艺术成就的“影”，由是对于 AI 绘作的交互完成过程而言，道德评价的对象应该是使用者、开发者、程序本身还是最终形成的作品，需要针对不同的算法偏好进而做出不同的判断。当面对动机的道德性欠

缺时，AI 绘画程序的学习板块应该更注重算法的动机自为，以形成先机的规范和全期的持守效用，从而减少种种因素影响道德自律的自然形成，如何达到这种预期并解决已形成的道德困境，这是今后人工智能的开发者和使用者均需要考虑的现实问题。

致 谢

感谢导师耐心的指导、友人的帮助及所有本文引用材料之著者。

参考文献

- [1] 朱熹. 朱子全书(第二十一卷) [M]. 上海: 上海古籍出版社, 2010: 1589.
- [2] 康德著. 道德形而上学原理[M]. 苗力田, 译. 上海: 上海人民出版社, 2012: 43.
- [3] 张汝伦. 朱陈之辩再思考[J]. 复旦学报(社会科学版), 2012(3): 58-72.
- [4] 彭长程. 朱子心论研究[D]: [硕士学位论文]. 南昌: 江西师范大学, 2021: 43.
- [5] 黄宗羲, 全祖望. 宋元学案: 第一册[M]. 北京: 中华书局, 1986: 497-498.
- [6] 姚勇. 朱熹论“克己复礼”与“主敬行恕”的关系——基于道德行动动力机制的阐释[J]. 南昌大学学报(人文社会科学版), 2021, 52(3): 13-22.
- [7] 黎靖德. 朱子语类[M]. 北京: 中华书局, 1986: 295-382.
- [8] 朱熹. 四书章句集注[M]. 北京: 中华书局, 2012: 3-4.
- [9] 梅洛·庞蒂, 著. 知觉现象学[M]. 姜志辉, 译. 北京: 商务印书馆, 2001: 86-97.
- [10] 朱熹. 朱子全书(第六卷) [M]. 上海: 上海古籍出版社, 2010: 512.
- [11] 朱熹. 朱子全书(第十四卷) [M]. 上海: 上海古籍出版社, 2010: 219.
- [12] 杨国荣. 人类行动与实践智慧[M]. 北京: 生活·读书·新知三联书店, 2013: 89.
- [13] 亚里士多德. 伦理学[M]. 北京: 中国社会科学出版社, 1990: 32.
- [14] 崔尔平. 历代书法论文选续编[M]. 上海: 上海书画出版社, 2003: 144-147.
- [15] 吴明澍. 德勒兹欲望理论视野下的 AI 绘画[D]: [硕士学位论文]. 重庆: 四川美术学院, 2020: 37.
- [16] 佛日. 真实论[J]. 法音, 1993(7): 3-15+47.