文章编号:1000-8934(2024)7-0099-08
DOI:10.19484/j.cnki.1000-8934.2024.07.006

跨文化人工智能体的伦理设计何以可能?

——基于设计师伦理责任意识的培养

王 亮,马紫依

(西安交通大学 马克思主义学院,西安 710049)

摘要:面对当今文化多样化的社会现实,人工智能体的伦理设计已不再只遵循一种文化传统,而必须兼顾多种文化传统。设计师在跨文化人工智能伦理设计中扮演关键角色,可从外部约束和内在修养两方面培养他们的伦理意识:外部约束包括建立跨文化伦理原则、采用文化适应性"警言"等方式,强化设计师的伦理责任;内在修养则涉及提升设计师的文化理解力、道德敏感性和"道德想象力"等,增强他们的跨文化视野和伦理自觉。此外,儒家伦理的丰富资源也为设计师的跨文化道德能力培养提供理论支持。重视设计师跨文化伦理责任意识的培养,不仅有助于突出人类在AI系统开发中的重要性,而且是缩小人工智能伦理理论与实践差距的有效路径。

关键词:跨文化;人工智能体;伦理设计;设计师

中图分类号:N031 **文献标识码**:A

2023年11月,包括中国、美国、英国在内的28个国家以及欧盟共同签署了《布莱切利宣言(The Bletchley Declaration)》,承诺以安全、以人为本、值得信赖和负责任的方式设计、开发、部署和使用人工智能(AI),在这一背景下,跨文化人工智能伦理设计变得尤为关键。近些年,全球已有不少国家和地区已经制定了一系列人工智能伦理原则,并且学术界围绕这些原则(隐私、自主、公正和可解释性等)进行了深入研究,"人们对潜在问题的认识正在快速提高,但人工智能界采取行动降低相关风险的能力仍处于起步阶段"[1]2141。人工智能伦理设计与设计师等工程人员密切相关,设计人员的道德决策偏好直接影响所设计的人工智能体的道德倾向。因此,本文将以设计师为突破口,专注于理论探讨,通过利用各种理论资源,推动人工智能伦理实现从"是什么"到"如何做"之间的转换,旨在为缩小人工智能伦理理论与实践的差距提供启发性思考。

一、设计师的外在伦理约束

1.跨文化伦理原则"指南"

如何保证设计师在设计人工智能体的过程中既考虑到人工智能技术的伦理问题,又能兼顾不同的文化传统呢?首先,可以为设计师提供相关的跨文化伦理原则进行参考。跨文化沟通合作的模式最早来自于跨国公司,为了在全球谋求最大的利益以实现合作共赢,跨国公司往往遵循当地的某些风俗和习惯从而制定相应的道德声明、框架和行为准则。奥斯兰(Osland)认为文化差异是跨文化商贸的最大障碍,其伴随的文化传播过程包含创建、发送、存储和传递信息这四个基本要素,一种文化共同体内部的成员自然可以交接到该文化的内涵,而外部人员最初无法理解某共同体文化,进而也就无法运用语言、实物和非语言行为进行交流。[2]因此,为了

收稿日期:2023-11-28

基金项目:国家社会科学基金后期资助重点项目"社交机器人'拟人化'伦理风险研究"(23FZXA011);陕西省社会科学基金年度项目"'非对称'智能人机交互技术的伦理风险与应对策略研究"(2023C007)。

作者简介:王亮(1985—),湖北黄冈人,哲学博士,西安交通大学马克思主义学院副教授,主要研究方向:马克思主义理论、科技伦理;马紫依(1999—),女,河南周口人,西安交通大学马克思主义学院硕士研究生,主要研究方向:马克思主义理论、科技伦理。

成功地进行跨文化的商业贸易,发送者和接收者必须以相同的方式理解语言、对象和非语言行为,也就是创建一套共同认可的文化伦理原则。目前,这种跨文化伦理原则的制定也被视为是人工智能系统进行正确的道德设计、开发和部署的关键前提。2019年乔宾(Jobin)等人为了进一步推动人工智能伦理的全球化发展,分析了全球84份人工智能伦理原则的指南,并揭示出当前人工智能伦理原则在许多方面出现了全球趋同的趋势。[3]莫利(Morley)等人进一步认为人工智能伦理原则的全球共识为东西方文化背景下"人工智能领域的协同设计"提供了"一个共同基础轮廓",因而具备跨文化人工智能伦理指导性。[1]2145然而,我们仍然可以看到这些跨文化人工智能伦理原则对实际负责设计人工智能算法人员的约束力极为有限,在人工智能的跨文化设计中依然存在许多伦理问题,如种族歧视、文化偏见等。ProPublica的一项调查就发现用于预测犯罪的软件往往表现出对黑人的偏见,黑人被预测为高风险犯罪人群,而白人则被预测为低风险人群。[4]ImageNet曾将一张传统的美国新娘照片贴上"新娘"、"礼服"、"女人"等标签,而将一张北印度新娘照片贴上"表演艺术"标签。[5]莫利等人认为造成这一局面的原因在于,尽管世界范围内有如此之多的人工智能伦理原则"道德指南",但很少有相应的监督措施来"检查这些指南的遵守情况"。[6]

针对当前设计师遵循跨文化伦理原则所面临的问题,莫坎德(Mökander)和弗洛里迪(Floridi)提出可以从"道德审计"治理机制入手,所谓"道德审计"是指一种治理机制,设计人工智能系统的组织和设计师们可以使用该机制来控制或影响人工智能体的相关行为,并且在实践过程中,通过结构化的流程去评估某实体的行为是否符合相关原则或规范。[7]324正如任何一个企业的成功都必须具备相应的基础设施一样,道德审计机制通过对道德层面的基础设施进行审计和监管,推动了人与人之间的良性互动。主要有三种形式去弥合伦理手册和设计师实际实践之间的不足:一是伦理的功能性审计,主要是负责对人工智能体决策背后所体现的伦理原则进行审计,并且允许采用适应特定情境的人工智能治理方法;二是代码审计,通过对人工智能算法、代码的源头进行审计,确保基于原初代码做出的决策的正确性;三是影响审计,通过对人工智能系统输出结果所造成的影响范围、程度进行审计,既可以将结果可视化从而为设计师们做出正确的设计决策提供支持,又可以预视不好的结果,让设计师反推正确的做法从而规避风险。[7]325与此同时,道德审计机构作为独立于人工智能体生产商和消费者的第三方机构,它可以更好地平衡各方面的利益关系,此外,通过将敏感信息悬置于第三方机构,也可以进一步增强生产者和消费者之间的信任。由此可见,通过这三种形式的综合运用,我们就可以在设计前的伦理预置、设计中的代码编写、以及设计结束后的影响后果三方面增强跨文化伦理原则"指南"的应用。当然,道德审计机制在运用的范围和程度上也存在着一定的限制,其中最大的限制在于道德审计机制的"不稳定性"。伦理原则的多样性决定了其不可能成为一个确定的范本,这也就意味着对于道德的审计不是一蹴而就的稳定形态。事实上,即使是同一个人在面对同一个场景时前后两次也可能做出不同的选择,因此道德审计机制应当是一个辩证、持续的过程。

2. 制定工程性团体"誓言"

制定类似如医学上的"希波克拉底誓言"来约束设计师的行为。工程师团体可以制定文化适应性的工程类"誓言",需要解决的问题有三个:一是程序性问题,即"誓言"如何形成。二是"誓言"应该包含哪些具体内容。三是如何在实际操作中保证"誓言"的有效性。

针对第一个问题,我们可以参考"希波克拉底誓言(Hippocratic Oath)"的形成过程。"希波克拉底誓言"由医生希波克拉底创立,后因希腊政权的垮台而消声灭迹,在中世纪被重新发现并于1508年在维滕贝格大学的一个仪式上使用,1948年由世界医学协会(WMA)进行修订即《日内瓦宣言》(Declaration of Geneva)。在第二次世界大战后,WMA开始为全世界的医生制定道德准则并做出承诺:"我不允许考虑年龄、疾病或残疾、信仰、种族血统、性别、国籍、政治派别、种族、性取向、社会地位或任何其他因素来干预我与他人之间的关系职责和我的病人。"[8]至此,举办"希波克拉底"宣誓仪式成为医学生中的"成人礼"。我们可以从"希波克拉底誓言"的形成过程中总结出一些规律,即:(1)由行业内道德高尚且专业的从业者制定;(2)在大学课程的培养过程中去实施,让从业者在学生时代便受到"誓言"的浸润,以此形成一种行业精神;(3)誓言应具备跨文化的共识性理念。基于此,本文认为"人

工智能设计师的誓言”的形成也应当至少具备以上三个要素。

要解决第二个问题，跨文化伦理誓言的内容应该涵盖跨文化可通约性的伦理共识。目前所发布的许多“宣言”都十分重视这一点。例如2011年英国工程与物理科学研究委员会（EPRSC）和艺术与人文研究委员会（AHRC）就为人工智能体设计师们发布了一套供参考的伦理原则，以确保人工智能体的相关生产过程能够参考一套共同的价值观和伦理原则。[9]然而，我们不得不承认不同地区文化之间确实存在着许多不可通约性。比如欧洲对人工智能伦理通常采取怀疑论和义务论态度；美国人工智能伦理的出发点一般是功利主义；而东亚国家一般遵循儒家伦理文化。[10]尽管如此，不同文化地区的管理者们最关心的问题还是如何降低人工智能技术应用的风险，包括伦理道德风险。2019年北京智源人工智能研究院联合多所高校发布《人工智能北京共识》，其中就人工智能的设计应当降低道德风险、合乎伦理做了明确规定：“人工智能的研发应采用符合伦理的设计方法以使得系统可信，包括但不限于：使系统尽可能公正，减少系统中的歧视与偏见；提高系统透明性，增强系统可解释度、可预测性，使系统可追溯、可核查、可问责等。”[11]保罗·特里奥洛（Paul Triollo）在分析国际上比较有影响力的人工智能共识后也指出，“人工智能广泛应用的障碍显然是极难平衡的隐私问题”，但也面临着“AI应造福全人类”、“公益原则”等广泛共识。[12]因此，我们可以从另一个角度思考“誓言”的跨文化伦理共识问题，即，如何设计“誓言”内容，从而使其成为降低伦理风险的工具？首先，我们需要明确人工智能体可能造成的两种伦理风险：一是由于客观上对环境等外在因素判断失误所造成的伦理风险，这属于技术上的问题。二是因为其主观故意的“恶行”所导致的伦理风险。“誓言”所包含的内容应当努力降低这两种风险的存在。针对第一种风险，“誓言”中应当包含如何提高设计师总体工作质量的内容，即，设计师应当对自己可以在跨文化情况下保持其专业性做出承诺。针对第二种风险，“誓言”中的内容应当有利于设计师与道德框架联系起来，从而降低其内在不道德的风险。

就第三个问题而言，目前有许多的较为成熟的工程师宣言已经出台。比如2016年IEEE（电气和电子工程师协会）所做全球倡议报告指出：全球倡议的使命是“确保参与自主和智能系统设计和开发的每个利益相关者都接受教育、培训并被授权优先考虑道德因素，以便这些技术为造福人类而进步”[13]3。其鼓励设计师在人工智能体的设计和研发过程中优先思考人类道德伦理问题，并且提出应当遵循人类利益、责任、透明性和教育等四原则。[13]6 2018年在天津举办的雷克大会也发布了《人工智能创新发展道德伦理宣言》，其中，第三章“人工智能具体接触人员的道德伦理要求”专门针对人工智能体设计者和使用者提出了道德要求，同时也明确指出只要可以直接操纵或影响人工智能系统的人员都应当遵循相关原则。[14]然而，正如“希波克拉底誓言”所面临的实际效用有限的问题，在实际操作中，也有人对“誓言”的有效性提出质疑。罗伯特·马丁在希腊敏捷峰会上就指出：很多程序员对他在2015年所做的《程序员誓言》很感兴趣，但大多数人认为其是荒谬和愚蠢的。[15]基于此，我们可以从两个方面去增强“誓言”的实用性：第一，通过宣誓仪式来提高设计师内心的成就感。正如每个人在结婚时都要进行庄重的婚前宣誓一样，人工智能体设计师可以通过举办一些庄重且有意义道德宣誓仪式来获得内心的充实感，从而增强其对于该行业的认同感和归属感，将誓言做到内化于心。第二，可以将誓言作为人工智能相关协会的入会门槛。例如IEEC为了确保其行业的规范性，为其软件开发人员建立了相应的认证规则，并协同ACM联合组建了软件工程方面的道德规范原则，而不符合或者违背相关道德规范原则的开发人员将不被允许成为该协会会员。[16]

二、设计师内在跨文化道德能力的养成

1. 设计师的人机交互道德体验

设计师并非是哲学家，人工智能体的设计者通常关注的是智能系统能做什么，以及程序如何实现的问题。设计师通常利用已经预设好的伦理标准来思考人工智能伦理问题，这种“自上而下”的模式有很多弊端：（1）设计师很难做到对伦理标准的合理判断；（2）书面化的伦理标准也很难应付复杂多变的现实情境；（3）伦理标准本身的正确性也有待考证。因此，设计师需要一种“自下而上”的模式来重新思考

人工智能伦理设计问题。蒙泰亚努(Munteanu)等人在2015年做了有关人机交互情境伦理的试验,结果发现既定的伦理原则往往与进行定性实地研究的现实情况不相符合,并通过四个实际案例说明了人机交互情境能够增强人工智能体设计师的伦理设计能力。[17]科克尔伯格(Mark Coeckelbergh)也认为,要结合文化差异考察人工智能伦理问题,他倡导人们利用已有的人机交互"道德体验"和"道德想象力",构建"增进人类繁荣和福祉的人-机共同生活"。[18]因此,本文认为人机交互道德体验是一种"自下而上"的模式,在这一模式下,设计师可以通过自身的人机交互体验或者通过观察测试者的人机交互体验来增加自身对相关伦理原则的理解和运用,从而推动自身内在道德能力的养成。

如何通过人机交互道德体验推动设计师内在道德能力的养成呢?设计师可以通过与自己设计的人工智能体之间的情境交互测试获得体验。这种人机交互情境体验可以增加设计师对产品的理解,并更好地领悟相关的伦理原则。梅可勒(Mekler)和霍恩拜克(Hornbæk)设计了一个与产品交互的意义框架并指出:"通过与产品进行交互式体验,可以帮助人们提高对产品意义的理解以及进行有价值的计算,从而有助于体验者福祉的实现。"[19]然而,由于体验感的主观能动性、偏好性,在提倡设计师进行产品自测的基础上,还需要调查、了解更多其他用户的人机交互体验。这就要求设计师应认真观察不同客户或同事在人机交互过程中对其设计产品的态度和行为,在理解的基础上形成与他们基本一致的价值观和行为逻辑。

此外,为保证人机交互道德体验的跨文化特性,设计师可以选择不同文化背景的测试者。罗伯逊(Robertson)指出,在面对"道德困境"时,个体如何做决策与自身的文化水平有关。[20]测试者的文化价值观非常影响人机交互实践的测试结果。比如,我们可以参考霍夫斯泰德提出的四种文化价值维度来挑选测试者,包括:个人主义、不确定性规避、男子气概和权力距离。[21]依据这四种维度,设计师可以有目的的挑选不同文化维度的测试者:(1)测试者偏向于个人主义还是集体主义?(2)测试者对于不确定性是否容易感到威胁?(3)测试者偏向于男子气概还是女性气质?(4)测试者更接受大权力距离还是小权力距离?通过对测试者的挑选,人机交互道德实践的跨文化特性将得到进一步保障。此外,还有直接依据国家、地区进行的测试者分类,最典型的例子如阿瓦德等学者所做的全球范围内无人驾驶汽车道德决策偏好实验。[22]

2. 设计师的道德想象力

韦斯特鲁姆(Westrum)认为,具有开放文化的组织通过积极鼓励员工们利用自己的想象力去探究各种产品的潜在问题的方式,可以有效地验证正在开发的系统的好坏,并防止坏的后果发生。[23]道德想象力指的是一个人能够意识到自己行为所包含的道德含义,并且能够通过构造相关的道德情境和创造道德替代方案以克服出现的问题的能力。莫伯格(Moberg)和考德威尔(Caldwell)将道德想象力分为三个子认识系统:一是道德敏感性,即能够对不同情境中所涉及的道德伦理原则、意义进行敏锐捕捉和识别判断的能力;二是观点截取能力,即能够超越自身角色界定,并站在一切利益相关者的立场上去考虑不同行为可能造成的影响和后果的能力;三是创造性想象的能力,即能够脱离具体情境去发散思考不同的结果,甚至创造出全新可能的能力。[24]归根到底,道德想象力不仅包括产生实用性想法的能力,还包括形成关于何为善、何为道德的想法的能力,并且可以将最符合实际情况的道德想法付诸行动,最终为他人服务的能力。

那么,为什么要充分发展这种道德想象力呢?首先,对于设计师而言,特定情境下运用何种道德伦理往往不是一个选择性的问题,而是一个动态的预判过程。这就需要设计师具备一定的道德想象能力,设计师能够通过预测后果来辅助自己进行正确的人工智能体设计。正如杜威指出,想象可以看作是一场戏剧排练,人们创造性地探索和排练不同的行动方案,而预想的可能性结果和对他人的影响将指导人们做出相应的道德决策。[25]其次,增强设计师的道德想象力可以进一步拓宽设计师的视野,减轻其"微观视觉"带来的弊端。在工程类团体组织中,"微观视觉"通常指专注于所从事的领域,而对更广泛的社会问题则"视而不见"。戴维斯(Davis)指出,造成恶的结果的人,并不一定是因为其意志力比较薄弱或者是本身带有恶的念头,而是因为其视野过于狭窄,无法预见更多的后果。[26]最后,道德想象力的发展也显示着一个人的道德成熟度。一个人是否在道德上成熟取决于他是否能够想象自己的行为如何影响他人,是否能够共情并理解他人,以及是否在必要时能够设想其他行动方案。维

贝克(Peter – Paul Verbeek)指出可以从“道德想象力”上对设计人员提出责任要求,他认为,“当设计人员试图想象他们所设计的技术在用户行为中可能扮演的中介角色时,他们可以将预期反馈到设计过程中”[27]。

如何培养设计师的道德想象力呢?雷斯特(Rest)设计了一个道德推理模型,并认为道德行为的产生依赖于四个要素:道德敏感性、道德判断、道德动机以及执行,并且只有当这四个部分都出现的时候,才能形成一个道德行为。[28]根据雷斯特提出的道德推理模型,我们可以得出增强设计师道德想象力的四个步骤:第一步,设计师应当扩大其道德共同体的范围,通过观察他人不同的道德伦理想法和观点来增强自己的道德敏感性,并进一步发现设计过程中所存在的道德问题;第二步,设计师们应当根据所面临的实际情况不断地提高自己的道德判断能力,进而增强自己的道德想象能力;第三步,设计师需要参照自己的道德判断进一步反思自己的道德动机;第四步,设计师基于自己的道德意图做出相应的道德行为。

当然,在培养设计师道德想象力的过程中也会遇到一系列困扰:(1)我们如何获取未发生事件以及自身之外事件的主观体验;(2)道德想象可能不足以使行为合乎道德自主,例如在遵循道德推理的过程中,人工智能体设计师的行为可能会违背他们个人的直接利益或者其他部分人的利益,这需要很大的道德力量;(3)道德想象力的发展可能会进一步固化人们的某种道德观念,正如前文提到的,受“微观视觉”影响,人们很难想象自己认知以外的情境,而在狭窄视野下,过度发展想象力,反而会一进步增强人们已形成的“刻板印象”。

在设计师道德想象力的培养上,如何解决这些问题呢?第一,设计师需要对人工智能体所造成的道德影响保持高敏感度,需要想象不同的行动选择及其会造成的后果和影响。而我们获取他人主观体验的关键一步在于培养我们对细节的感知力,以及要学会随着细节变动而随之调整相关举措。默多克(Murdoch)指出:这种感知力很大程度上取决于一个人习惯于将自己的注意力放在哪里,如果一个人的注意力被感知到的威胁线索所吸引,那么道德知觉就会缩小到自我保护的范围。[29]因此,设计师们为了更好地增强自己的道德想象能力,需要有意识地使自己的注意力突破自身角色局限,并拓展至整个社会关系网中。第二,道德想象能力不仅仅指头脑中对各种情境后果的预演和思考,更包含着将信念和理想化的目标付诸实际的勇气和决心。有学者曾作出“道德想象的神经生物学根源”假定,在这个假定根源中,个体可以利用前额叶皮层进行自我调节,并通过“自由意志”防止有害行为,参与反思抽象。[30]32这种将想象付诸实际的勇气和努力不仅仅依靠设计师的自主性,更需要运用前面章节所提出的一系列外部约束。第三,针对在“微观视角”下的道德想象力可能会加剧设计师的刻板印象这一问题,有学者认为,“协调”往往会在处理自主性和社区等多元价值观方面起着重要的作用,并指出,“情感和理性的协调、人的主观思考能力和无意识的被动适应能力的协调都是十分重要的”[30]33。设计师需要进一步发展协调自身与社会、道德情感与认知的能力,通过这种协调能力,实现自身“微观视角”的突破,从而进一步提高自己的道德想象能力,发展自己的多元价值观。

此外,设计师道德想象力的培养与其文化理解力的提升是正相关的。木村刚(Takeshi Kimura)认为机器人工程师只有熟悉了不同的社会、文化和伦理内涵,才能在产品设计过程中自觉地考虑伦理问题。[31]设计师可以通过短期和长期两种模式进一步提升自己的文化理解力。短期模式包括:(1)设计师需要定期接受“全球公民培训”,这种培训的课程应当包括沟通方式、商务礼仪等内容。埃文斯(Evans)等人认为,全球公民教育应当进行诸如“世界意识”的身份和会员资格、全球背景下的权利和责任、信仰和价值观的多样性、重要的公民素养能力、管理和理解冲突、公平和社会正义等方面的教育。[32](2)要尽可能增加设计师团队组成人员的多样性,使其能够完全代表相关用户群中身份群体的多样性和交叉性。而之所以要增加设计师团队的组成人员,是因为设计师团队属于“嵌入式用户”(员工和用户双重身份)。[33]他们可以为人工智能体设计提供更多的创新思路,既兼顾公司所要求的的市场效益,又能根据自身情况充分理解某些隐形歧视或者由于文化障碍所导致的不道德问题,从而解决人工智能体在设计过程中所面临的跨文化伦理问题。而长期模式主要是指对设计师实施跨文化教育从而培养其跨文化好奇心,减轻其以自我为中心的态度。例如欧洲委员会举办的 LIAM(成年移民语言融合)项目通过一系列开放性措施帮助成员国制定基于欧洲委员会共同价值观的包容性语

言政策,增强各成员国对他国文化的理解与包容。[34]此外,立足于长期模式,还应当培养设计师们开放、主动思考的态度,并增强其反思批判的能力,以便其学会更积极地理解他国文化,从而正确地处理和应付所遇到的跨文化伦理问题。

三、儒家伦理对设计师跨文化道德能力养成的启示

相较于西方伦理注重道德抽象知识、原则,中国儒家伦理更注重道德经验和实践,并且这些都是根植于自身在家庭、社会关系中所承担的重要角色,即从所承担的社会角色中找到道德行动的"指南"。[35]223齐景公向孔子请教为政之道时,孔子用一句话概括了不同身份角色应当承担的义务:"齐景公问政于孔子。孔子对曰:'君君,臣臣,父父,子子。'"(《论语·颜渊》)华人学者刘纪璐认为,"社会角色不只是社会任务(social assignment);它还是道德任务(moral assignment)。"[36]君尽君道,臣尽臣道,父尽父道,子尽子道。这里面不仅包含了不同身份角色所对应的职责,也包含了其应有的道德责任,比如君王关爱体恤百姓,大臣忠心辅佐君王,父母抚养教育孩子,孩子要孝敬双亲。设计师不仅是一个职业,也是一个重要社会角色,其能够通过设计的技术产品影响社会,设计师的道德责任至少包含两个层次:一是对客户和用户负责;二是承担社会道德责任。前者是从具体的人工智能产品用途出发,需要设计师遵循职业道德或行为准则,以避免人工智能伦理风险的发生;后者是基于技术对社会的塑造性、系统性影响,设计师要主动肩负起责任,将道德或价值纳入设计过程中,比如将隐私、安全、人的自主性等理念贯彻人工智能产品设计全过程,最终实现人工智能产品对个人、社会或环境的整体性目标。[37]13-16

然而,"以对道德负责的方式进行设计是一个进化的过程,我们不能一概而论地试图按部就班地遵循预先确定的规则,因为环境在变,人在变,整个系统也在进化"[37]16。作为设计师该如何应对这一动态的变化性呢?在儒家文化中"学"既是人的道德养成路径,也是在动态的变化中寻找"恰当性"的重要方法。在《论语·阳货篇》中记载了一段孔子与仲由的对话:"子曰:'由也,女闻六言六蔽矣乎?'对曰:'未也。''居。吾语女。好仁不好学,其蔽也愚;好知不好学,其蔽也荡;好信不好学,其蔽也贼;好直不好学,其蔽也绞;好勇不好学,其蔽也乱;好刚不好学,其蔽也狂。'"可见,在孔子眼中"学"不仅是一种知识的掌握,更是要结合具体的情境来学习并践行道德行为,不懂得道德行为的"恰当性"就会走向另一种极端。设计师在面对动态变化且是"跨文化"的道德情境时,也要"学",具体来说,(1)设计师需要细致入微地观察不同文化的特点以及在其背景下人们倾向于做出的道德决策,以更全面地了解各种文化的价值取向和偏好;(2)设计师需要积极参与跨文化的交流,倾听不同文化背景的人对人工智能体伦理原则嵌入的看法和反馈,拓宽自己的设计视野;(3)设计师要在观察的基础上对不同情况下的不同道德决策进行批判性反思和评估,能够更好地理解自己的伦理取向;(4)设计师通过实践整合过往的经验,逐步发展自身的道德敏感性,能够更加敏锐地察觉潜在的伦理问题,进而可以在跨文化的设计环境中做出恰当的伦理决策。

此外,儒家所提出的"恕"、"和"等伦理思想本身就蕴含着跨文化包容性设计理念。何谓"恕"?子曰:"其'恕'乎!己所不欲,勿施于人。"(《论语·卫灵公》)这样一种"道德金律"至少可以从两个方面为设计师跨文化道德能力养成带来启示:一是尊重差异。杜维明认为,"恕道首先是尊重他者。有了尊重才能承认差异,才能够互相学习和交流"[38]。不同文化之间必然存在差异,如果不尊重彼此的差异对话很难展开。设计师需要在设计过程中有意识地考虑不同文化和价值观的差异,并学会尊重差异,在制定人工智能体的行为准则和决策模型时兼顾跨文化道德差异性。二是表达社会的共同诉求。考虑跨文化道德设计时不仅要考虑彼此的差异,还要看到不同文化中的道德共性。"恕"正是凭借着"对人的内在的没有强制性的表达与诉求",而"可能成为整个社会的表达与诉求",因此,可以"作为我们社会每一角色所遵守的'最低伦理准则'"。[39]另外,儒家还倡导"和而不同",它体现了差异性与共性之间的辩证统一。"和"不是机械地通过"削弱"差异性的"和合",而"是一个创造性与丰富性结果——差异性被协调为导致出一种最佳状态。"[35]223-249就跨文化人工智能伦理设计来说,这种"最佳状态"是不同文化、文明互鉴的结果。因此,设计师应当通过学习、理解多元文化,培养自

身对不同文化背景下伦理、价值观的宽容态度,最终设计出更具包容性的人工智能产品,这些产品不仅能够更好地满足全球用户的需求,更在跨文化伦理适应和社会互动方面展现出更高的成熟度。

四、结 论

随着人工智能技术的跨文化发展,因不符合当地文化传统而导致智能歧视、算法偏见的伦理事件时有发生。本文从设计师伦理责任意识培养的角度出发,认为设计师是直接影响跨文化人工智能伦理设计的关键要素之一。培养设计师的伦理责任意识可以从外在伦理原则和内在道德能力两个方面入手,具体而言,外在约束可以通过构建跨文化伦理原则、文化适应性工程类"誓言"推动强化设计师伦理责任的落实。而设计师内在道德能力的提升可以通过培养设计师的文化理解力、道德敏感性、"道德想象力"等途径来实现,这些途径从内部保证了设计师的跨文化视野、伦理自觉以及伦理道德责任。此外,儒家伦理基于生活实践、身份角色的道德养成路径,以及"恕"、"和"等包容性伦理思想为设计师跨文化道德能力养成提供了丰富的理论资源。总之,通过强调设计师在跨文化人工智能伦理设计过程中的重要作用,进一步突显了人类在人工智能系统开发中的重要性。然而,随着 ChatGPT、Sora 等 AI 大模型的快速迭代,人工智能体的智能自主化程度明显提升,它们自身也可以通过伦理算法、机器学习等方式来自主学习人类的道德伦理,这对我们进一步深入思考跨文化人工智能伦理设计提出了新挑战。

参考文献

[1] Morley J, Floridi L, Kinsey L, et al. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices[J]. *Science and Engineering Ethics*, 2020, 26(4): 2141 - 2168.

[2] Osland, Gregory E. Doing Business in China: A Framework for Cross - Cultural Understanding[J]. *Marketing Intelligence & Planning*, 1990, 8(4): 4 - 14.

[3] Jobin A, Ienca M, Vayena E. The Global Landscape of AI Ethics Guidelines[J]. *Nature Machine Intelligence*, 2019, 1(9): 389 - 399.

[4] Angwin J, Larson J, Mattu S. Machine Bias[EB/OL]. [2016 - 05 - 23](2022 - 12 - 25). https://www. propublica. org/article/machine - bias - risk - assessments - in - criminal - sentencing.

[5] Zou J, Schiebinger L. AI can be Sexist and Racist - It's Time to Make It Fair[J]. *Nature*, 2018, 559(7714): 324 - 326.

[6] Morley J, Kinsey L, Elhalal A, et al. Operationalising AI Ethics: Barriers, Enablers and Next Steps[J]. *AI & SOCIETY*, 2021: 1 - 13.

[7] Mökander J, Floridi L. Ethics - based Auditing to Develop Trustworthy AI[J]. *Minds and Machines*, 2021, 31(2): 323 - 327.

[8] Oxtoby K. Is the Hippocratic Oath still Relevant to Practising Doctors Today? [EB/OL]. [2016 - 11 - 04](2022 - 12 - 25). https://www. bmj. com/content/355/bmj. i6629. full

[9] Ashrafian H. AIonAI: A Humanitarian Law of Artificial Intelligence and Robotics[J]. *Science and Engineering Ethics*, 2015, 21: 29 - 40.

[10] Fung P, Etienne H. Confucius, Cyberpunk and Mr. Science: Comparing AI Ethics Principles between China and The EU[J]. *AI and Ethics*, 2023, 3(2): 505 - 511.

[11] 智能研究院. 人工智能北京共识[EB/OL]. [2019 - 05 - 25](2022 - 12 - 25). https://www. baai. ac. cn/portal/article/index/type/center_result/id/110. html.

[12] Triolo P. 建立全球人工智能伦理规范的挑战[EB/OL]. [2018 - 10 - 30](2021 - 04 - 05). http://m. sohu. com/a/272210929_468736.

[13] IEEE standards institute. Ethically Aligned Design v2 — Overview [EB/OL]. [2017 - 12 - 12](2022 - 12 - 25). https://standards. ieee. org/industry - connections/ec/ead - v1/.

[14] 人工智能产业创新联盟.《人工智能创新发展道德伦理宣言》助力产业健康发展[J]. 机器人产业, 2018(4): 4.

[15] Ben Linders. Oath for Programmers[EB/OL]. [2017 - 09 - 16](2022 - 12 - 25). https://www. infoq. com/news/2017/09/oath - programmers/.

[16] Gotterbarn D, Miller K, Rogerson S. Computer Society and ACM Approve Software Engineering Code of Ethics[J]. *Computer*, 1999, 32(10): 84 - 88.

[17] Munteanu C, Molyneaux H, Moncur W, et al. Situational Ethics: Re - thinking Approaches to Formal Ethics Requirements for Human - Computer Interaction[J]. *ACM*, 2015. (4): 105 - 114.

[18] Coeckelbergh M. Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics[J]. *International Journal of Social Robotics*, 2009, 1(3): 217 - 221.

[19] Mekler E D, Hornbæk K. A Framework for The Experience of Meaning in Human - Computer Interaction[C]//*Proceedings of The* 2019 *CHI Conference on Human Factors in Computing Systems.* 2019: 1 - 15.

[20] Robertson C J, Crittenden W F, Brady M K, et al. Situational Ethics across Borders: A Multicultural Examination[J]. *Journal of Business Ethics*, 2002, 38: 328.

[21] Lonner W J, Berry J W, Hofstede G H. *Culture's Consequences: International Differences in Work - Related Values*[M]. New York:

Sage Publications, 1984: 65, 110, 148, 176.
[22] Awad E, Dsouza S, Kim R, et al. The Moral Machine Experiment [J]. *Nature*, 2018, 563(7729): 59.
[23] Westrum R. Cultures with Requisite Imagination[C]//Wise J A, Hopkin V D, Stager P. (eds.) *Verification and Validation of Complex Systems: Human Factors Issues*. Heidelberg: Springer, 1993: 401-416.
[24] Moberg D, Caldwell D F. An Exploratory Investigation of The Effect of Ethical Culture in Activating Moral Imagination[J]. *J Bus Ethics*, 2007, 73: 193-204.
[25] [美]斯蒂文·费什米尔. 杜威与道德想象力: 伦理学中的实用主义[M]. 徐鹏, 马如俊, 译注. 北京: 北京大学出版社, 2010: 104.
[26] Davis M. Explaining Wrongdoing[J]. *Journal of Social Philosophy*, 1989(20): 74-90.
[27] Verbeek P P. *Moralizing Technology: Understanding and Designing The Morality of Things*[M]. Chicago: University of Chicago Press, 2011: 99.
[28] Narvaez D, Rest J. The Four Components of Acting Morally[J]. *Moral Behavior and Moral Development: An Introduction*, 1995, 1(1): 385-400.
[29] Murdoch I. *The Sovereignty of Good*[M]. London: Routledge, 1970: 28-29.
[30] Narvaez D, Mrkva K. The Development of Moral Imagination[J]. *The Ethics of Creativity*, 2014: 25-45.
[31] Kimura T. Roboethical Arguments and Applied Ethics: Being A Good Citizen[J]. *Cybernics: Fusion of Human, Machine and Information Systems*, 2014: 289-298.
[32] Evans M, Ingram L, MacDonald A, et al. Mapping The Global Dimension of Citizenship Education in Canada: The Complex Interplay between Theory, Practice, and Context[J]. *Citizenship Teaching & Learning*, 2009, 5(2): 16-34.
[33] Schweisfurth T G, Herstatt C. How Internal Users Contribute to Corporate Product Innovation: The Case of Embedded Users[J]. *R & D Management*, 2016, 46(S1): 107-126.
[34] Council of Europe. Reference Guide on Literacy and Second Language Learning for The Linguistic Integration of Adult Migrants (LASLLIAM)[EB/OL]. [2022-06-30](2022-12-25). https://www.coe.int/en/web/lang-migrants.
[35] 安乐哲. 儒家"角色伦理"基本阐释概念[EB/OL]. 孟巍隆, 译. [2022-11-29](2024-01-16). https://www.chinakongzi.org/rxmj/anlezhe/202211/t20221129_554868.htm.
[36] Liu J. Confucian Robotic Ethics[EB/OL]. [2021-11-01](2024-01-16). https://www.researchgate.net/profile/Jeeloo_Liu/publication/319391008_Confucian_Robotic_Ethics/links/61a4342af1d624457171ec56/Confucian-Robotic-Ethics.pdf.
[37] Fiore E. Ethics of Technology and Design Ethics in Socio-Technical Systems: Investigating The Role of The Designer[J]. *FormAkademisk*, 2020, 13(1): 1-19.
[38] 杜维明. 儒家的恕道是文明对话的基础[J]. 人民论坛, 2013(36): 76-77.
[39] 刘火. "恕"的当代意义: 兼议《儒家角色伦理学》[J]. 文史杂志, 2022(1): 92-96.

How can Ethical Design of Transcultural AI be Possible?: Based on the Cultivation of Designer's Ethical Responsibility Consciousness

WANG Liang, MA Zi-yi
(College of Marxism, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: In light of the contemporary culturally diverse social landscape, the ethical design of artificial intelligence (AI) can no longer be bound by a single cultural tradition, but must acknowledge and incorporate multiple cultural traditions. Designers play a key role in the ethical design of transcultural AI, which can be cultivated from the perspectives of external constraints and internal moral cultivation: External constraints include establishing transcultural ethical principles and adopting culturally adaptive "pledges" to strengthen designers' ethical responsibilities. Internal moral cultivation involves enhancing designers' cultural understanding, moral sensitivity, and "moral imagination", thereby strengthening their transcultural perspective and ethical awareness. Moreover, the rich resources of Confucian ethics provide theoretical support for cultivating designers' transcultural moral capabilities. Emphasizing the cultivation of designers' transcultural ethical responsibility awareness not only helps to highlight the importance of humans in AI system development, but also is an effective path to bridging the gap between AI ethics theory and practice.

Key words: transcultural; artificial intelligence; ethical design; designer

(本文责任编辑:崔伟奇 郑 泉)