

# 延续和重构：人工智能伦理研究的多重视角

王雪莹，邵国松

(上海交通大学媒体与传播学院，上海 200240)

**摘要：**人工智能的发展具有两面性，既有技术带来的经济效应，也存在技术带来的伦理风险。关注人工智能的伦理问题，要从其技术逻辑出发，看到大数据、算法和算力之间的关系。人工智能的伦理问题具有复杂性，既要技术发展史看到人工智能的技术延续，从传统伦理学寻求理论的滋养，又要看到人工智能对现实的重构，从应用场景、媒介技术哲学、未来哲学、跨学科等多重视角进行探讨。通过多重视角的人工智能伦理研究，确保科技向善。

**关键词：**人工智能；伦理；研究视角

**中图分类号：**B82-057

**文献标识码：**A

**文章编号：**2096-8418 (2023) 01-0070-08

卢西亚诺·弗洛里迪指出，人类在经历了哥白尼革命、达尔文革命、神经科学革命之后，自我认知已经开始第四个革命，即图灵革命。<sup>[1]</sup> 人类已经进入人工智能时代，而当我们这样讲的时候，我们正在用一种媒介史的角度看待人工智能，因为“我们正在用这一时代的支配性的媒介技术界定这一时代的特征”<sup>[2]</sup>。得益于计算机、大数据、云计算等技术的发展，人工智能技术实现重大突破。一项新技术的出现往往能产生两方面的影响：一方面，人工智能作为一种底层基础技术可以提高生产效率，带来经济效益；另一方面，新技术也会带来隐患。做到对技术经济性关注的同时，也应当关注其风险性的一面。更重要的是，这种风险已经实实在在地发生了，比如大数据“杀熟”<sup>[3]</sup>“回音室效应”和隐私被侵犯<sup>[4]</sup>，等等。出于对风险的担忧，人工智能伦理成为重要的研究领域，也备受新闻传播学者的关注。传播学研究人工智能及其伦理问题的合法性在于我们将人工智能视为媒介。从泛媒介的角度看，媒介与技术基本上可以画上等号，媒介的范畴不仅包括报纸、广播、电视、电影等，更包括服饰、住宅、货币、汽车等。<sup>[5]</sup> 从这个角度看，人工智能当然是媒介，正如麦克卢汉所言：“一切技术都是媒介，一切媒介都是我们自己的外化和延伸。”<sup>[6]</sup> 对人工智能伦理问题的研究正是对人工智能隐忧的回应。本研究试图以人工智能的底层逻辑和伦理研究的发展脉络作为起点，试图梳理出人工智能伦理研究几种可行的视角和研究路径，为其提供一定的创新性思路。

## 一、人工智能的底层逻辑：大数据、算法和算力

1956年的达特茅斯会议一般被认为是人工智能正式诞生的起点。发展至今，人工智能的概念并不明晰。技术层面每一次新的进展都会刷新人类对人工智能的认知。人工智能的本质在于拟人、像人，然而人的智能是什么、人的智能能发展到何种程度尚未可知，因此对人工智能的未来预测也存在局限。目前的人工智能是帮助人类解决难题的、没有自我意识的应用程序，被称为弱人工智能；而在未来可能会出现真正像人、拥有像人类的自我意识的智能，被称为强人工智能。目前我们所探讨的现实中的人工智能多为前者，主要包括专家系统、机器学习、深度学习。此类人工智能主要是以机器学习为核

**基金项目：**国家社会科学基金重大项目“5G时代互联网传播方式的变革和治理对策研究”(19ZDA328)；上海市教委人文社科重大项目“人工智能时代的新闻伦理与法规”(2019-01-07-00-02-E00010)。

**作者简介：**王雪莹，女，博士研究生；邵国松，男，教授，博士。

心, 通常被置于搜索引擎里面, 通过大数据可以自行判断输出结果, 代表性应用是算法推荐。

在提及人工智能时往往会提到大数据和算法, 这是因为大数据、算法与人工智能密不可分。数据量和算法可以分别比作人工智能的燃料和发动机。算法是计算机基于所训练的数据集归纳出的识别逻辑, 好的算法模型可以实现精准的物体和场景识别。数据集的丰富性和大规模性对算法训练尤为重要。<sup>[7]</sup> 人工智能近几年之所以能快速发展, 得益于诸多方面: 一是传感器性能和数量的大幅度提高、移动互联网的爆发使大量用户接入互联网积累了海量数据; 二是计算成本的降低和算力的进步为处理、分析海量数据提供了可能; 三是算法的进步让程序更智能、程序输出结果更精准。通过对大数据、算法和算力的关系分析我们可以发现, 大数据被称作是 21 世纪的“新石油”并不是抽象意义上的比喻, 而是一种现实的客观描述: 数据确实如石油一般具有巨大的经济效益。在某种程度上, 算法已经成为人工智能的三大基石——大数据、算法和算力——中最核心的议题, 甚至可以说“几乎所有的经济运行流程, 或是人们的活动, 都是由算法在幕后重新建立一种秩序”<sup>[8]</sup>。“算法社会的前提是万物数据化, 这也包括人的数据化 (datafication)”<sup>[9]</sup>, 将一切数据化、实现数据流最大化是算法社会内在的逻辑。

回到人工智能技术逻辑本身, 人工智能所面临的伦理问题正是大数据和算法所面临的伦理问题。从人工智能的底层技术看, 探讨人工智能伦理问题可以从大数据、算法和云计算这三个源头考虑。根据对相关研究的梳理, 这一层面的伦理问题主要关注两个方面: 一个是实然的研究, 即人工智能出现以后带来了什么实质性的影响, 尤其是负面后果; 一个是应然的研究, 人工智能作为一种技术, 应该对其建构怎样的伦理。

实然的研究集中于大数据人工智能的数据收集如何引发侵犯个人隐私<sup>[10]</sup>、形成“全景监狱”式的社会监控<sup>[11]</sup>等问题, 以及算法如何带来回音室效应<sup>[12]</sup>、偏见问题<sup>[13]</sup>等。训练算法需要借助大数据, 因此大数据的质量尤为重要, 基于“garbage in, garbage out”的原则, 如果原始数据有问题, 那么算法建模运行后的结果一定有问题。算力是帮助算法运行的基础, 比起算法模型的准确性和大数据来源的精准性, 算力并不是核心要素。因为即使算力空前发展, 如果大数据和算法建模存在问题, 算力的强大只会让我们更快地得出错误的结果。所以, 算力是作为硬件为软件服务的。大数据之“大”, 在于其体量大, 因为需要云存储和云计算。借助于“云”的隐喻, 云计算看起来像是没有实体, 但是实际上云计算背后是物质性实体, 需要建立数据储存中心。从这个维度上说, 云计算带来的电子污染、隐私、安全、结构性失业的问题必不可免。云计算的问题, 也是大数据的问题。<sup>[14]</sup>

应然的研究多是规范性研究, 关注对人工智能的治理, 旨在提出伦理准则。<sup>[15]</sup> 提出怎样的伦理准则依然与人工智能的底层逻辑相关。人工智能如何表现关键在于我们需要人工智能实现什么样的目标——实现的目标是人类的目标, 而不是人工智能自己的目标。目标指导算法结构及其运算的目的。因此, 制定什么样的目标、如何制定目标成为关键。一方面, 从功能论的角度看, 人工智能的目标一定是实现效用最大化。但是实现的效用是谁的效用? 当人工智能掌握在大公司手中, 实现的一定是商业公司的目标。价值涉入和偏好并不能兼顾到社会上所有的利益群体: 比如普通大众。而多元价值、多元利益共存的世界, 要求利益各群体都投入到议题协商中。无论是政府、企业、学界, 还是公众, 都应该明晰人工智能的发展对个人意味着什么。

媒介的逻辑非常重要, 因此需要解开算法黑箱, 了解人工智能的运作逻辑和基本原理, 明晰使用人工智能带来的结果, 帮助个体识别人工智能、警惕人工智能。媒介素养的概念自报纸、广播、电视等传统媒体时代就被提出了, 其核心意涵在于如何提高公众识别信息的能力。基于人工智能的新媒体成为公众主要的信息来源, 媒介素养的内容应随之转换。

总而言之, 看到技术的逻辑, 明确思考人工智能的伦理应从哪些维度着手。同时, 也应该看到定位人工智能目标的重要性, 呼唤社会各方提高参与公共事务的意识。

## 二、延续：并非仅仅是人工智能才有伦理问题

伦理研究由来已久，伦理道德是一套调节人与人之间社会关系的价值系统，包括内在的价值理想和外在的行为规范。<sup>[16]</sup>“伦理学就是要去证明，当我们追问何为正确的行为时，什么是正确的答案；就是要对为什么认为我们关于人之品格或行为之道德的表述是对的或错的给出理由。”<sup>[17]</sup>伦理学是关于人的伦理学，当伦理的概念应用于人工智能，追问的是人工智能如何表现才是正确的行为。能给人工智能伦理提供指导作用的理论非常多，比如中外具有代表性的适度伦理原则：亚里士多德的中间之道、孔子的中庸之道，还有康德的绝对律令、密尔的效用原则、罗尔斯的无知之幕、诺丁思的关怀伦理学。伦理意味着我们人类如何理解他人和这个世界，以及对他人和世界是什么样的期待。因此，伦理研究从本质上就有规范研究的倾向，即这个世界“应当”是什么样子的判断。从宏观角度说，伦理是社会得以运行的原则；从中观角度说，不同的职业系统有不同的伦理原则（比如医学伦理、新闻传播伦理等）；从微观角度说，人与人之间的日常交往也需要伦理道德的规范。伦理研究在于追求“善”，让生活更美好。因此，伦理问题当然不是人工智能出现后人类所面对的独有的问题，而是在伦理学脉络上逐渐具体化的体现。

人工智能作为一种技术，其伦理问题可以被看作是技术伦理的一种；人工智能应用于实践，也就有了相应的职业伦理，比如机器人写作的伦理问题、算法推荐的伦理问题等。采用何种技术，决定了我们如何看待世界，也会改变我们与外部世界的关系。大数据和人工智能是一种新技术，但是，所谓新型技术也只是某一阶段的“新”。人类从第一阶段的农业社会到第二阶段的工业社会，再到第三阶段的信息社会，阶段之间跃升所花的时间越来越短。关照现实固然有意义，互联网出现后我们研究互联网，社交媒体出现后我们研究社交媒体，人工智能出现后我们研究人工智能，区块链出现后我们研究区块链……但是这种追“新”，需要问一个问题：然后呢？“赶时髦”的研究反而“乱花渐欲迷人眼”，让人抓不到问题的实质。孙玮提到：“大量的所谓新媒体研究，只不过是在旧有范式中填充入新媒体材料，是不断地在用新材料验证旧理论。”<sup>[18]</sup>

为了解决这个问题，我们可以从“媒介”的角度看人工智能。虽然强调媒介可能容易陷入媒介决定论的危险，但是，它却是一种非常有效的看待事物的方法。尤其是当我们从媒介技术发展史的角度探讨相关问题时，这种有效性更加明显。我们所要探讨的“大数据”“人工智能”等新技术，它们虽然从时间线上看是最新出现的，但是事实上它们也不过是整体媒介技术的一种具体表现形式罢了。将人工智能置于媒介发展史的长河中看，当有更新的技术出现后，所谓新技术又变成旧技术了。当我们从这样的视角切入探讨大数据、人工智能和社会各行各业的关系时，就不再是孤立的，而是从一种整体观的路径出发。这种分析视角使我们探讨问题更加深入、更加具有广泛性。

回溯伦理学的研究脉络，人工智能被划分为拥有自我意识的“强人工智能”和无自我意识的“弱人工智能”，二者的伦理问题并不能在同一层次的逻辑上探讨。对于强人工智能，我们可以沿袭机器伦理的思路，先确定其主体性问题，再探讨其伦理问题。但是，强人工智能从技术上是否能实现，目前暂时还未知。虽然需要未雨绸缪，但是提出“可能存在”的伦理问题需要有一个立得住的前提，即“可能存在伦理问题”的人工智能从技术上来说是可实现的。否则，人类所谈论的人工智能的伦理问题只能是人的臆想。对于弱人工智能，因其是工具，沿袭技术伦理的思路，将其看作是“客体”，探讨其伦理问题就自然而然了。将人工智能看作一种技术，人工智能伦理不再是孤立的，比如技术伦理本身就已经是一个非常成熟的概念了。因此，人工智能伦理问题是新现象背后的“老问题”。

就伦理学而言，依然存在很多悬而未决的问题。比如，伦理学追求“善”，那么实现这一目标是目的导向还是结果导向呢？当做出一种行为是出于“善”的目的但是却产生了坏的结果，那么这是可接



受的吗? 当做出一种行为尽管是出于“非善”的目的但是却产生了好的结果, 那么这是可接受的吗? 当类似的重大议题没有解决时, 研究人工智能伦理需要继续推进伦理学的研究向前, 从总纲领的角度为人工智能伦理研究提供理论基础。

### 三、重构: 人工智能伦理研究的多重视角

人工智能带来的变革一触即发, 互联网行业之外的其他行业也会被颠覆, 旧的系统有可能面临崩溃。既然人工智能的出现带来的问题如此之大, 人工智能伦理研究应从哪几个方向入手呢? 根据前述对人工智能伦理现有研究的梳理发现, 现有研究往往聚焦于现状、问题、对策。这种研究具有一定的价值, 但是问题在于, 如果研究视角过于狭窄, 缺乏宏观视角统筹, 研究就无法归类、无法与上一层级的学术传统进行对话、无法站在巨人的肩膀上将理论再往前推进一步, 同时也无法看清人工智能伦理的本质, 直击问题的痛点, 给出切实可行的对策建议。为了突破目前研究的局限性, 本文依然延续人工智能作为媒介的逻辑, 提出四种研究路径。

#### (一) 将人工智能置于应用场景中进行探讨

目前大部分关于人工智能伦理问题的探讨多是从人工智能的应用出发, 从实践中探究人工智能的伦理问题。人工智能的原理就是借助于海量数据, 通过算法, 帮助实现智能。人工智能内在的应有之义就是获得更多的数据, 因此互联网公司想方设法收集用户的信息。从个人的角度讲, 人有保护自己的隐私、保护个人数据的需要。在人工智能大数据时代就形成了“数据收集”和“数据保护”的悖论。<sup>[19]</sup> 从隐私到用户数据, 互联网的发展改变了隐私的内涵和外延, 甚至模糊了公共领域与私人领域之间的界限。要解决这个悖论, 实际上要解决的是公共利益与私人利益边界的问题。

生活中的人工智能主要用于医疗(智能诊断)、交通(自动驾驶)、智能家居等。除此之外, 还有对社会排斥、社会公正、不平等问题的关注。人工智能的发展必然会挑战既有的人类价值, 促使人类去重新思考人类的基本属性与伦常关系。甚至还有对超级智能是否会“失控”, 反过来取代、控制人类的隐忧。<sup>[20]</sup> 从应用中探讨人工智能的伦理问题, 无疑是明智的。具体问题具体分析, 可以保证谈论的问题具体到实处, 进而制定行之有效的对策。在寻找解决之道时, 主要体现在以下三个方面: 第一, 从理论层面和规范层面提出各种伦理原则以用于实践时, 必须考量其现实可能性并寻求折中的解决方案; 第二, 应摆脱未来学家简单的乐观主义与悲观主义立场, 从具体问题入手强化人的控制作用与建设性参与; 第三, 要从人机协作和人机共生而不是人机对立的角度探讨负责任的人工智能的可能。<sup>[21]</sup> 但是问题在于, 提出对策应当是自然而然、水到渠成的, 探讨人工智能需要这种原则, 探讨其他的媒介技术就不需要这种原则了吗? 也就是说, 提出的对策必然要看到人工智能伦理问题所面临的普遍性和特殊性。如此, 才能真正理解人工智能伦理的内涵所在。

既然人工智能实现一定的目标是靠算法设计操纵的, 那么解决伦理的问题也可以从源代码的设计上规避伦理风险。比如, 通过效用函数或决策树的设计解决人工智能的伦理困境, 实现利益最大化。<sup>[22]</sup> 但是事实上, 这种做法从技术上可行, 在现实生活中往往很难实现。这就是“应然”和“实然”的区别。这背后依然是商业主义与人文主义的冲突。人工智能的开发和利用往往是大公司先行, 商业公司首要的目标是盈利。在进行人工智能应用时, 希望人工智能可以带来市场和利润。算法是有价值取向的, 这种价值取向来自于算法执行的目的。在使用人工智能时, 首先要制定执行目标, 根据不同的执行目标进行算法的设置以实现相应目标。因此, 如果公司的目标是盈利, 那么算法的执行自然就是要获取更多的点击量、购买量。这也就解释为什么从算法设置这一步就能解决算法偏见, 以及信息流通过程中所带来的信息茧房问题而商业公司却毫不作为——将“善”的伦理嵌入算法结构并不能为公司带来利益。这一点与人工智能的技术逻辑相一致。因此, 为了解决商业主义与人文主义的冲突, 依

然要向“多元协商”的策略靠拢：当公众意识到算法、大数据的应用影响到自身的生活并且产生了负面影响，公众必然要采取抵抗行动，通过行动倒逼商业公司进行算法的改良。比如剑桥分析数据泄露事件引发了用户#delete Facebook#的行动，倒逼 Facebook 采取更强的隐私保护策略。<sup>[23]</sup> 公众的行动之所以尤为重要，是因为只有公众的抵抗行动影响到企业的利益时，企业才会采取改善的措施。除此之外，各方应成立伦理审查委员会，从专业的角度提出人工智能发展和应用应遵循的原则。正如学者段伟文所称：“鉴于当前的人工智能发展还远未达到这一或许会出现的未来情境，面向人工智能时代的价值校准与伦理调适，应与未来学家的悲观主义和乐观主义预见保持一定距离，而更多地诉诸人工智能当下的真实发展程度与可预见的未来的可能性。唯其如此，才能系统深入地慎思、明辨、审度与应对人工智能的伦理挑战。”<sup>[21]</sup>

## （二）将人工智能置于媒介技术哲学更大的背景下探讨

对媒介技术的研究并不稀奇，从哈罗德·伊尼斯、马歇尔·麦克卢汉以来的媒介技术学派，以及后继的以约书亚·梅罗维茨为代表的媒介环境学派，都看到了媒介的重要性。每种新媒介的出现，一定是给当时的社会同时带来新的机遇和恐慌。因此，当我们把媒介的研究聚焦到人工智能上，虽然“媒介形式”是新的，但是其可能面对的隐忧和风险并不是新的。而对媒介的研究，核心关怀在于看到“媒介与人和社会”的关系。这种关系，并非是单向的，而是双向的。一方面，我们要看到政治、经济、文化和社会对技术发展的形塑；另一方面，是要看到技术的反向作用。人工智能不仅作为工具为人所使用，更重要的是人工智能技术结构所具有的内在规则对人的思想和行为的形塑。因此，就会存在这样的疑问：到底是人在利用人工智能还是人工智能在利用人？

一种技术塑造一种新的生活方式，麦克卢汉通过“媒介即讯息”的描述指出，最重要的不是媒介传递的信息内容，而在于媒介形式自身。麦克卢汉还称，媒介即人的延伸，延伸意味着“截除”。也就是说，技术在一定程度上延伸了人的某种功能，但是却会损害人的另一种功能。比如文字的发明帮助人记忆的同时，也使得人的记忆能力变弱，人开始过度依赖体外化的媒体。人工智能将人的延伸演化到极致——因为它延伸的是人脑。人工智能的出现，使得人们在信息爆炸时代高效地获取需要的信息，但是同时也让人失去了自主性——过度依赖算法推荐的信息做出决策。因此，人们总是假装在选择，实际上顺从了算法的决定。

随着现代科学技术的快速发展，正如凯瑟琳·海勒所言，人类已成为“后人类”。人与机器的联系越来越密切，依赖越来越深，差异越来越小。将人工智能放置于媒介技术哲学背景下考察有一个好处是可以看到人工智能出现的脉络，进而解决如何分辨人工智能伦理的一般性和特殊性的问题。

对于技术的理解有两种理论可以参考：工具理论（instrumental theory）和实体理论（substantive theory）<sup>[24]</sup>。从工具理论的角度看，一定会得出技术是中立的，技术是善是恶关键是人如何使用它。但是从实体理论的角度看，一切技术都有价值偏向性。即使是最简单的弱人工智能，因为其依靠算法，所以会包含价值导向：无论是有意识的企业价值导向，还是无意识的程序员的价值导向。所以，绝不能从中立性的角度看待人工智能。除此之外，人能在多大程度上控制人创造出的东西？技术是否有自主性？

回答这些问题，我们需要借助媒介技术哲学思考问题的方式，回到媒介环境学派学者那里看媒介发展的历史，看各种媒介是如何嵌入社会、给人和社会带来影响的；回到海德格尔、基特勒、齐泽克等学者那里寻找理解媒介的灵感；回到雅克·埃吕尔、兰登·温纳等学者那里寻找思考技术政治的方法。技术不断使自身更加“技术化”<sup>[25]</sup>，自主技术具有“自我生成、自我维持和自我规划”的能力<sup>[26]</sup>。技术作为一种人造物（artifact）存在对非物的牵制性作用基本上已经达成了共识<sup>[27]</sup>。伦理学研究又以其鲜明的价值取向为特征，而这一点正与媒介技术哲学的研究取向相关：从媒介技术哲学的立场上来

说, 技术从来都不是中立的。因此, 当技术出现了明显的偏向性, 就存在内在的伦理问题。人工智能底层的算法逻辑, 比以往任何一种媒介更趋显自主性。媒介哲学的反思潮流, 是面对当前数字场景传统研究思路已然捉襟见肘的应然之举。<sup>[28]</sup>

### (三) 将强人工智能置于未来哲学的语境中探讨

对强人工智能的探讨最初往往来源于科幻小说。而对于人工智能尚未利用但是未来有可能涉足的领域, 学者往往采用“可能性”来推断这种伦理问题。对强人工智能伦理的探讨首先要解决从技术上讲强人工智能能否实现的问题。卡鲁姆·蔡斯在其著作《人工智能革命: 超级智能时代的人类命运》中探讨了这一问题: 我们能造出来吗? 如果能, 什么时候能造出来? 它安全吗? 作者称, 第一个问题是最接近答案的, 这个答案是“只要人类没有在此之前灭绝, 有可能”。而就第二个问题, 作者认为, 我们确实不知道人类能否制造一个大脑或者有意识的机器出来。<sup>[29]</sup>

这是作者的一种看法, 当然, 也有学者认为可能实现。比如乐观主义者雷·库兹韦尔认为: “只要拥有真实大脑的详细数据, 我们就能模拟出生物学意义上的大脑。”<sup>[30]</sup> 既然学界对此有争议, 那么本研究就不能轻易下结论。做出“只要人类没有灭亡, 依靠进化的能力一定能实现”的推论并不具有现实意义。强人工智能作为一种是否能实现还未知的事物, 只能从源头上加以防范。强人工智能的开发是否应该成为科学研究的禁区?

世界上第一个拥有公民身份的机器人——索菲亚——虽然并不能被看作是强人工智能机器人。但是她的出现让我们思考: 人工智能机器人是否具有人格、主体性? 什么是机器人的权利? 这些暂时都无法回答。但是这些风险和挑战涉及伦理、法律和社会等方方面面, 需要政府、学界、业界以及公众共同应对。作为工具使用的弱人工智能的伦理问题关键在于使用主体, 工具是否作恶全在于人如何使用它, 应该从制度上建立起弱人工智能的使用标准与规范。而强人工智能只能从源头上加以防范。借助未来哲学的语境, 对人工智能发展的可能性进行预测, 可预测性是人工智能, 尤其是强人工智能的伦理原则之一。<sup>[31]</sup>

### (四) 将人工智能置于跨学科的背景下探讨

如今的世界是具有复杂性的世界, 任何单个人、单个学科的力量无法应对高度复杂的问题。因此, 我们需要跨学科的研究视角。“跨学科研究是回答问题、解决问题或处理问题的进程, 这些问题太宽泛、太复杂, 靠单门学科不足以解决; 它以学科为依托, 以整合其见解、构建更全面认识为目的。”<sup>[32]</sup> 也就是说, 跨学科研究是以研究问题为导向, 突破学科之间的藩篱, 发挥各学科的优势来回答研究问题。因此跨学科仅仅是解决问题的手段和路径。并非所有问题或课题都适用于跨学科研究, 不能为了跨学科而跨学科, 而应该坚持问题导向。当研究问题和课题是较为复杂的、尚未解决的社会需求或议题、单个学科并不能全面处理或解决该问题、问题的重要见解或理论由两门及以上学科提供,<sup>[32](85)</sup> 那么自然而然就“跨”了学科。

为什么在人工智能时代需要强调跨学科? 首先, 因为人工智能的议题过于复杂, 而同时学科分化、专业化、碎片化的程度之高, 依靠单个学科的力量无法解决。“现在社会的技术知识是如此高度专业化和分散, 以至于大多数人仅仅能掌握一个微小的部分。”<sup>[26](23)</sup> 其次, 人工智能应用于实践, 技术人员更倾向于关注其是否达到预定的目标和效果, 忽略了伦理问题。一项对照研究发现, 道德准则的有效性几乎为零, 它们并不会改变技术界专业人员的行为。<sup>[33]</sup> 与之相对应的, 关注伦理问题的学者可能不懂技术, 文科学者对技术保持天然的警惕, 对维护人类安全、幸福的大宪章问题有天然的敏感性。而由于不懂人工智能运作的背后原理, 尤其是算法黑箱的问题, 文科学者在探讨伦理问题时并不能直击人工智能的软肋。因此, 在这样的背景下重提跨学科研究, 发挥各学科的优势, 达到协同创新的目的。

“虽然大部分学院的学科障碍仍然阻遏了观念的交换和信息的整合, 但是要求各学科或多学科合并



的呼声现在非常高涨，几乎已经成为一种潮流了。”<sup>[34]</sup> 人工智能伦理研究当然也应该促进文科与理工科的交叉融合，促进科技向善。传播学天然地对媒介持有敏感性，传播学学者总是先人一步地讨论与媒介相关的新现象、新问题，然而由于缺乏成熟和公认的理论传统，随着研究的深入传播，研究最后总是被收编到其他学科的研究路径之下，失去了自身学科的发言权。<sup>[35]</sup> 而从媒介的角度理解人工智能，关注人工智能的伦理问题可以激发人工智能伦理研究的活力，从传播学的视角为人工智能伦理这一交叉领域做出自己的贡献。

## 四、余论：迈向未来的人工智能研究

海德格尔和埃吕尔都从技术实体论的角度出发，主张技术不是价值无涉的，“它承载着一定的价值内涵并因此构成我们难以摆脱和改变的社会环境和生活方式”<sup>[36]</sup>。关注人工智能伦理，关注的是人工智能是如何嵌入到社会结构中，对人和社会产生影响的。而人工智能的发展也符合梅卡夫法则：使用某种技术的人数越多，产生的效益越是呈倍数级增长。正如研究技术的人忽视了关注人工智能的伦理问题一样，研究伦理学的学者也应该反思是否夸大了技术的负面效应？比如每次新的技术出来都有人看到技术的负面效应：文字的出现让人丧失主动的记忆能力、印刷术的出现使得大量“机器复制的艺术作品”丧失了“灵韵”、电视的发明让人担心“娱乐至死”，互联网和人工智能出现后自然也会有新的问题。在对媒介的批评中，印刷媒体、广播电视在互联网的冲击下似乎已经退居二线。反观其负面效应，似乎也没有学者预测的那样严重。但是问题在于，当学者放弃了对技术的批评，商业主义必然占据控制技术发展的主导地位，多元协商的力量不再平衡，那么我们是否能由人工智能媒介时代“安全”过渡到下一个更新的媒介时代，就不可知了。学者作为多元利益主体的一分子，需要做出自己的行动。

“海德格尔已经说得很明白：你批判什么东西，不等于要逃避它。恰恰相反，你要直面它、拥抱它。”<sup>[37]</sup> 关注人工智能伦理，不是限制人工智能的发展。而是在发展的同时，直面人工智能技术发展可能带来的风险。伦理学最终回答的还是如何让人类生活得更幸福的问题。因此，了解人工智能的技术逻辑，重回伦理学研究的起点，通过应用场景、媒介技术哲学、未来哲学和跨学科的多重视角，也许能够在人工智能技术广泛应用的今天，在伦理研究的问题上提供更有价值的贡献。

### 参考文献：

- [1] [意] 卢西亚诺·弗洛里迪. 第四次革命：人工智能如何重塑人类现实 [M]. 王文革，译. 杭州：浙江人民出版社，2016：107.
- [2] 胡翌霖. 媒介史强纲领：媒介环境学的哲学解读 [M]. 北京：商务印书馆，2019：290.
- [3] 张爱军，王芳. “大数据杀熟”的政治安全风险 [J]. 未来传播，2021，28（2）：46-51+124-125.
- [4] 潘红霞. 智媒时代智能信息推荐算法的缺陷及正向重构 [J]. 未来传播，2020，27（5）：36-41.
- [5] [加] 马歇尔·麦克卢汉. 理解媒介：论人的延伸 [M]. 何道宽，译. 南京：译林出版社，2011，目录：2.
- [6] [美] 林文刚. 媒介环境学：思想沿革与多维视野 [M]. 何道宽，译. 北京：中国大百科全书出版社，2019：281.
- [7] 曹婷. 人工智能行业研究报告 [EB/OL]（2017-07-21）[2021-07-05]. <http://www.199it.com/archives/565872.html>.
- [8] 徐恪，李沁. 算法统治世界：智能经济的隐形秩序 [M]. 北京：清华大学出版社，2017，序言：10.
- [9] 彭兰. 生存、认知、关系：算法将如何改变我们 [J]. 新闻界，2021（3）：45-53.
- [10] 邵国松，黄琪. 人工智能中的隐私保护问题 [J]. 现代传播（中国传媒大学学报），2017，39（12）：1-5.
- [11] 匡文波. 智能算法推荐技术的逻辑理路、伦理问题及规制方略 [J]. 深圳大学学报（人文社会科学版），2021，38（1）：144-151.
- [12] [美] 伊莱·帕里泽. 过滤泡：互联网对我们的隐私操纵 [M]. 方师师，杨媛，译. 北京：中国人民大学出版社，2020：8.
- [13] Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Confer-

ence on fairness, accountability and transparency, 77-91.

- [14] [加] 文森特·莫斯科. 云端: 动荡世界中的大数据 [M]. 杨睿, 陈如歌, 译. 北京: 中国人民大学出版社, 2017: 11-12.
- [15] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P. & Dignum, V., et al. (2018). Ai4people—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28 (4): 689-707.
- [16] 金东寒. 秩序的重构: 人工智能与人类社会 [M]. 上海: 上海大学出版社, 2017: 39.
- [17] [美] G. E. 摩尔. 伦理学原理 [M]. 陈德中, 译. 北京: 商务印书馆, 2017: 1.
- [18] 孙玮. 从新媒介通达新传播: 基于技术哲学的传播研究思考 [J]. 暨南学报 (哲学社会科学版), 2016, 38 (1): 66-75+131.
- [19] 田新玲, 黄芝晓. “公共数据开放”与“个人隐私保护”的悖论 [J]. 新闻大学, 2014 (6): 55-61.
- [20] 孙伟平. 关于人工智能的价值反思 [J]. 哲学研究, 2017 (10): 120-126.
- [21] 段伟文. 人工智能时代的价值审度与伦理调适 [J]. 中国人民大学学报, 2017 (6): 98-108.
- [22] De Swarte, T., Boufous, O. & Escalle, P. (2019). Artificial intelligence, ethics and human values: The cases of military drones and companion robots. *Artificial Life and Robotics*, 24 (3): 291-296.
- [23] Brown, A. J. (2020). “Should I stay or should I leave?”: Exploring (dis) continued Facebook use after the Cambridge analytica scandal. *Social Media+ Society*, 6 (1): 2056305120913884.
- [24] [美] 安德鲁·芬伯格. 技术批判理论 [M]. 韩连庆, 曹观法, 译. 北京: 北京大学出版社, 2005, 前言: 3.
- [25] [美] 阿尔文·托夫勒. 未来的冲击 [M]. 黄明坚, 译. 北京: 中信出版社, 2018: 20.
- [26] [美] 兰登·温纳. 自主性技术: 作为政治思想主题的失控技术 [M]. 杨海燕, 译. 北京: 北京大学出版社, 2014: 203.
- [27] [美] 唐·伊德. 让事物“说话”: 后现象学与技术科学 [M]. 韩连庆, 译. 北京: 北京大学出版社, 2008: 47.
- [28] 骆世查. 反还原论与媒介哲学: 哈曼的“物向本体论”初探 [J]. 新闻大学, 2020 (4): 36-52+120.
- [29] [英] 卡鲁姆·蔡斯. 人工智能革命: 超级智能时代的人类命运 [M]. 张尧然, 译. 北京: 机械工业出版社, 2017: 63.
- [30] [美] 雷·库兹韦尔. 人工智能的未来: 揭示人类思维的奥秘 [M]. 盛杨燕, 译. 杭州: 浙江人民出版社, 2016: 125.
- [31] Bostrom, N. & Yudkowsky, E. (2014). The ethics of artificial intelligence. In Frankish, K. & Ramsey, W. M. (eds.). *The Cambridge handbook of artificial intelligence*. Cambridge: Cambridge University Press, 316-334.
- [32] [美] 艾伦·雷普克. 如何进行跨学科研究 [M]. 傅存良, 译. 北京: 北京大学出版社, 2016: 17.
- [33] McNamara, A., Smith, J. & Murphy-Hill, E. (2018). Does ACM’s code of ethics change ethical decision making in software development? In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 729-733.
- [34] [美] 阿尔文·托夫勒. 第三次浪潮 [M]. 黄明坚, 译. 北京: 中信出版社, 2018: 313.
- [35] 胡翼青, 张婧妍. 重新发现“媒介”: 学科视角的建构与知识型转变——2018年中国传播研究综述 [J]. 编辑之友, 2019 (2): 39-45.
- [36] 刘同舫. 技术的当代哲学视野 [M]. 北京: 人民出版社, 2017: 44.
- [37] [美] 戴维·J. 贡克尔, 保罗·A. 泰勒. 海德格尔论媒介 [M]. 吴江, 译. 北京: 中国传媒大学出版社, 2019: 25.

[责任编辑: 谢薇娜]