智能社会研究
Journal of Intelligent Society
ISSN 2097-2091
CN 23-1615/C

JIS

第2卷
2023
第 4 期

智能社会研究

第 2 卷
2023
第 4 期

智能社会研究

Journal of Intelligent Society

中华人民共和国工业和信息化部主管　　哈尔滨工程大学主办

中华人民共和国工业和信息化部主管

ZHINENG SHEHUI YANJIU

杂志公众号二维码
官网网址 www.jis.ac.cn

ISSN 2097-2091
9 772097 209239 07>

定价：45.00 元

智能社会研究

（双月刊）

2023 年
第 4 期
2023 年 7 月 10 日出版

2022 年 11 月 10 日创刊

总第 5 期

目　次

智能社会与人口研究专题

研究报告

译　文

书　评

访　谈

CONTENTS

TRANSLATED TEXT

BOOK REVIEW

INTERVIEW

反思人工智能的愿景、神话与未来

——评《人工智能地图集:人工智能的权力、政治和全球代价》

曹立坤　茅泓锴*

摘要:人工智能技术是当前最受关注的新兴技术领域之一,它提高了经济效率,但其生产与运行也可能带来一系列严重的社会问题。《人工智能地图集:人工智能的权力、政治和全球代价》一书从社会科学理论视角出发,批判性地反思了人工智能的环境与资源成本、人工智能对人类工作与劳动的影响、数据与算法中的权力结构,以及人工智能带来的公权力扩张等问题。作者认为,人工智能技术可能在算法与人类之间形成新的权力关系,且它会在多个维度加剧人类社会的不平等,因此有必要限制和监管人工智能技术在特定社会领域中的运用。本文对《人工智能地图集:人工智能的权力、政治和全球代价》的主要思路进行了介绍,并从该书内容出发讨论了技术社群的最新进展和局限性。

关键词:人工智能　技术伦理　技术创新　社会理论

一、人工智能的兴起

近年来,ChatGPT 等大规模语言模型的发布让人工智能技术(artificial intelligence)走到聚光灯下,引起广泛的关注。然而,尽管人工智能似乎颇具颠覆性,它的发展壮大却并不是一夜之间就能完成的:从 1956 年达特茅斯会议开始,人工智能的出现已有几十年历史,其中不乏人们对人工智能技术前景感到怀疑、濒临放弃的时刻。根据美国专利商标局(USPTO)的报告,可被归

* 曹立坤,芝加哥大学社会学系;茅泓锴,芝加哥大学计算社会科学项目(MACSS)。

类为“人工智能”的专利申请数量在2000年前后开始快速增长,其有望在下个时代与电力、信息技术并列成为支撑人类经济活动的通用技术(Brynjolfsson & McAfee, 2014; Crafts, 2021)。人工智能技术改变了商业竞争版图(Krakowski, Luger & Raisch, 2022)、学术范式(Sourati & Evans, forthcoming),甚至复现了人类思考过程(Fei, Lu & Gao et al., 2022)。

人工智能技术蕴含的巨大潜力和利益,让人们对它的未来充满星辰大海一般的想象与期待。但是,正如所有的技术一样,人工智能并非完全中立、价值无涉的——它的诞生和发展都嵌入具体的社会情境中:硬件生产过程依赖于不平等的利益分配模式,算法本身也内含特定的世界观框架,通过商业、政治甚至军事应用完成再生产。与人工智能带来的新功能、新增长相比,它的代价往往更为隐蔽,但对人类社会的影响可能非常深远。凯特·克劳福德(K. Crawford)的《人工智能地图集:人工智能的权力、政治和全球代价》(*Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*,以下简称《人工智能地图集》)深入探讨了这些问题。经过数十年的田野研究与思考,她总结了人工智能可能带来的问题,以及人类可能的应对方案。

二、人工智能的社会、文化与政治代价

在反思“人工智能”的得失利弊之前,首要的问题是:什么是人工智能?如果单纯以算法与模型对人工智能进行概括,那么我们可能陷入对人工智能的片面想象中:它似乎飘浮在无形的虚拟空间(云服务器)里,绝对正确理性,脱离人类的偏见、情绪和利益。克劳福德认为这种想象在两种意义上是错误的(Crawford, 2021: 8,以下仅括注页码):首先,它并不完全是人造的(artificial)。除了虚拟的软件交互,人工智能还依赖于有形的硬件存在,而硬件生产的供应链则依赖于大量的自然资源和人力成本投入,有形世界是虚拟世界的基础。其次,它也不完全智能(intelligent)。训练集常常不完整甚至内含错误,而算法逻辑本身也基于过于简化的认识论框架,这导致人工智能算法不仅无法摆脱人类的认知错误,反而可能强化它。

那么,如何对人工智能进行公允的评价呢?如果对人工智能的理想化想象只展示了它的一个侧面,那么强调它的生产过程也只是另一个角度的思考,未必比前者更加全面。为了解决这个问题,作者采取了鸟瞰视角对人工智能的发展进行了全面追踪:从硬件的生产到训练集的收集,从人工标注到模型训练,作者观察、记录与人工智能相关的所有社会结构与过程,并将它们联系起来——这正是该书主书名“人工智能地图集”的题中应有之义。作者并不将理论思考限于特定学科,而是积极地与科学学、法学、政治哲学对话(p. 12)。基于实证证据与问题意识,该书的思考可以被总结为四个主题:人工智能的环境与资源成本、人工智能对人类劳动的异化、算法中的知识权力结构以及公权力的扩张与私有化。

(一)人工智能的环境与资源成本

作者的旅途从美国内华达州的克莱顿河谷展开。这里的盐湖是锂的来源,而锂是人工智能硬件的必要原料。在过去的 10 年中,这里的矿物被大量开发,制成寿命极短的电子产品被卖出,再随着电子产品的迭代被废弃。除锂之外,用于人工智能硬件的稀有元素有 17 种之多(p. 33)。

在世界范围内,采矿业本身会带来一些问题:采矿业收入常常引发贫穷地区的暴力争斗,矿区内部的劳动条件也较差,而科技巨头往往不愿意对这些供应链之内的具体问题进行有效监管。更严重和更直接的问题则可能是大量高速采矿对环境造成的影响:矿石加工产生大量废料,而挖掘本身也对自然环境产生了不可逆的破坏,许多自然物种因此灭绝。

技术对环境的影响本身并不是一个全新的主题,1962 年,生物学家雷切尔·卡森(R. Carson)就在《寂静的春天》中讨论了化学农药对生态系统的严重破坏(Carson, 1962)。针对技术的生态成本和可能造成的不平等,许多学者也在环境正义视角下进行过讨论(Bullard, 2000; Ito, 2019)。那么,人工智能的生态效应有什么特殊性?作者强调了两点。首先,由于无休止追求模型计算能力,人工智能产业产生的能源消耗巨大,并且极速增加。目前,全球计算设施的碳排放量已经与运输业高峰时期持平(p. 42),而科技巨头对此的

解决方案是购买更多的碳排放指标。其次,人工智能产业依赖全球供应链。运输本身会产生大量碳排放和有毒物质,集装箱的废弃或遗失也会对海洋产生污染。由于这两点特性,人工智能可能比其他技术产生更严重的环境危害。

(二)人工智能视野下的工作与劳动

与对自然环境的影响相比,人工智能对人类工作和生活的影响更加直观。在学术界与政策研究中,一种观点认为人工智能算法会使得部分技能失去稀缺性,带来失业(Tong, Wu & Evans, 2022)。但是,该书的关注点并不在于人工智能的未来趋势,而在于它对当下劳动过程的塑造。作者认为,人工智能对人类劳动过程的重塑并不是将会来到,而是已经发生在千千万万的工厂与企业中。

在人工智能时代到来之前,许多学者就讨论过技术与劳动过程的关系。马克思在《资本论》中分析了机器对劳动关系的影响,认为机器消除了工人的体力差别,从而使劳动更容易被资本控制(Marx, 1930)。布雷弗曼(H. Braverman)继承了这一思路,认为自动化降低了劳动过程对专业技能的需求,使资本能对劳动进行更标准的控制和监督,从而产生“去技能化”过程(Braverman, 1998)。埃德沃兹(R. Edwards)进一步强调了“数值控制”,认为计算机的引入使得工作流程被完全控制,“技术控制”和“科层制控制”一道成为垄断资本主义体系不可或缺的一环(Edwards, 1980)。

作者认为,与之前的劳动控制相比,人工智能技术对劳动过程产生了三点独特的影响。首先,人工智能技术对劳动过程的监控达到史无前例的精度。工人的行动、社交媒体和沟通信息都可以通过设备被记录,并被分解为微观行为数据,这些数据又可以通过人工智能模型被分析和追踪,大大降低了资本控制工人的成本和难度(p. 62)。其次,控制生产过程的程序高度中心化。在许多科技巨头中,中心主机控制其他计算设施与数据库,而这些设施控制了工人的劳动过程和时间。这使得一种中心化的意义秩序被生产出来——设计生产程序的硅谷程序员和工程师大多数是年轻的男性,他们不会考虑个体在家庭与社区中的服务时间,这使得劳动进一步挤压了工人的社会

属性(p. 77)。最后,训练人工智能模型需要大量人类的低技能辅助工作,这类低收入、高重复的工作往往由相对贫困的国家和社会群体完成,一部分工作(如标注仇恨言论和暴力视频)甚至可能给工作者带来永久性精神损伤。在部分情况下,许多号称人工智能完成的工作,可能是人类直接在底层完成的。作者将其总结为“虚假自动化”(fauxtomation),认为其进一步加剧了劳动力异化过程。

(三)数据与算法中的知识权力结构

以上论述大多针对人工智能的社会情境。一种可能的观点是,尽管人工智能可能造成负外部性,但其自身内容仍然具有纯粹的正价值:算法可以为人类提供无偏误、高效率的帮助,从而提升社会的总效益,在功利主义意义上实现社会共赢。该书的第 3—5 章对这种观点做出了回应。作者认为,人工智能算法的数据基础和世界观框架存在着根本性的偏误和扭曲:如果不加以修正,弱势群体可能受到系统性的伤害。

人工智能算法从根本上来说是从已有的大量数据(训练集)中辨识数据的经典模式(pattern recognition),并将数据中存在的模式运用于新的场景。对于决策树这类相对简单的算法而言,人类可以追踪到机器预测的过程,从而看到机器模式识别的原理,对原理进行解释(explanation);但大部分算法的预测过程则难以追踪,更多被用于预测(prediction)。作者的批判正是针对人工智能算法的这几个阶段展开的。在作者看来:首先,算法的基础即数据集并不可靠;其次,算法中蕴含特定的世界观框架,这种框架本身过于武断;最后,算法的难以解释性使得算法的错误不可被追责,在算法与人类之间构造了新的权力关系。

许多计算机科学家使用现存的数据集训练他们的模型,这些数据集也被学界广泛接受。然而,许多数据集的收集过程都存在伦理问题,这也间接影响了科学研究的准确性。有时,学者从公共摄像头中获取人脸数据,而这没有经过这些人的同意;有时,学者从不合适的渠道获取大量数据,用在完全不相关的场景中(p. 102),不恰当的文本迁移导致模型偏误和逻辑扭曲,对模型

用户产生难以预测的影响;有时,学者雇佣低薪劳动力来做数据标注,他们的工作质量完全不被监督,许多人甚至在标签中随意使用侮辱性词语,这些错误标签直到最近才被清除。互联网的兴起带来了大量网络痕迹数据,这些数据使得数据库规模得到了爆炸性增长。但作者认为,这之中蕴含的道德基础是危险的:在人工智能时代,每个人的信息都天然地成为数据,不需要任何关怀、同意或者对风险的考量。

相对而言,算法逻辑本身的问题更加隐蔽,但影响更加深远。分类模型[①]是人工智能时代的主要知识形式,折射出人工智能系统对世界的认知。以福柯的视角观之,正如历史中所有的知识形式一样,算法产生的知识也与权力密不可分:算法的语言规则已经蕴含了特定的世界观,对什么是真实与正常进行定义(Foucault, 1982)。许多人工智能模型诞生于工程需求,初始设置中采纳了关于社会现实的简单假设;这在具体问题上或许有其实用性,但当模型的应用场景扩散到社会生活的方方面面时,这些假设则可能把多面向的社会生活强行套入过于简化的单维度理解。例如,在大多数分类模型中,性别被分为男、女两类,对其他可能存在的性别取向不加考虑;“疯子”“瘾君子”等带有侮辱性的类别存在于被广泛使用的数据集中,并被作为特定类别而为其他变量所预测。由此可以看到人工智能世界观的潜在问题:机器不仅默认种族、职业是天然类别,具有不可更改性,并且认为它们具有统一特征,可以被外表等因素定位。这些都使得人们的刻板印象被不断加深,对弱势群体来说尤为不利。社会生活之所以复杂有趣,就在于其建构性——意义在社会空间中流动,被人们在互动和理解中共同建构起来。然而,人工智能世界观使得这种建构性被挤压殆尽。

更严重的是,人工智能不仅预测职业等社会属性,也预测情绪、性格等不可见的因素。尽管许多心理学家提出人类的表情和外在动作是在具体情境中生成的,无法作为独立的数据输入模型而产生准确的预测结果,但科技界还是在不断推动算法在这些领域的应用。单一算法在不同情境中的准确适

① 在机器学习中,分类(classification)和回归(regression)是两类基本问题。在这里,作者重点关注了分类问题,但类似的观点同样可以在回归问题上适用。

用,需要社会规律的普遍性作为前提。然而目前阶段的社会科学,与自然科学具有可比性的普遍性是难以想象的。在技术企业商业利益的驱动下,人工智能的用途被无限扩张,却可能被套用在量化认识论尚不成熟的软科学领域,从而伤害人类个体和社会。

(四)公权力的扩张与私有化

作者对人工智能的最后一点批判指向了政府与国家。从 1950 年开始,美国军方就是人工智能技术和互联网技术的主要赞助方之一,并积极将其应用于国家安全和军事领域。美国前国防部长阿什顿·卡特(A. Carter)更是积极推动美国军方和硅谷建立合作,并将其称为"第三次抵消战略"——用压倒性技术优势弥补军事实力上的不足之处,从而取得军事领先地位。

人工智能在军事上的应用非常广泛,如用图像识别技术和分类算法从人群中识别潜在的恐怖分子,并通过无人机将其定点清除。虽然与军方的合作引发了部分科技公司对技术作恶的担忧,但大多数美国公司对于技术和军事的结合还是喜闻乐见、积极参与的。在与政府的合作中,人工智能的两个潜在问题可能恶化。首先,民用分类算法的错误只会加剧不平等,但军用分类算法的错误却可能导致误杀目标,以及其他不可逆的严重后果。其次,公司在与军方的合作中能够发展出大量数据和监控工具。在与军方的合同结束后,公司常常将这些数据与技术卖给政府机构、警方甚至私人公司,将这些军事手段应用于民事(civic)事务。一方面,这导致政府权力扩张,警务(policing)越来越多地依赖算法监控,精细追踪普通民众的行动。另一方面,这也存在公共监控私有化问题。由于美国业界的技术水平常常高于政府机构,许多美国政府机构开始向私人公司购买监控数据和算法系统(p. 200),政府的一部分职能被私人公司所取代。公权与私权的结合不仅使得政府的行为更容易逃避行政追责,也为私人公司收集巨量公众数据用于牟利大开方便之门。在这样的背景下,电子监控如何影响社会管理,已经成为美国社会学界近年来很重要的议题(Buechi, Festic & Latzer, 2022; Burrell & Fourcade, 2021)。

在技术公司和政府的合作中,许多法律和伦理问题很难通过目前的法律框架得到妥善解决。例如,分类算法可能内含对弱势群体的歧视,这种歧视通过公权力的执行被放大;训练集的系统性缺陷导致特定群体的利益被剥夺。由于许多人工智能算法仍在黑箱之中,由此而来的公共损失无法被完全追责(accountable)。作者由此提问:人工智能技术是否已经足够成熟,能够分享公权力的一部分,从根本上干预人类生活?

三、技术嵌入性和人类未来

在《大转型》中,卡尔·波兰尼(Polanyi, 2001)首次提出"嵌入性"概念。他认为,经济制度与市场结构并非独立自存的实体,而是嵌入在社会制度与文化规范之中的。该书的视角与波兰尼有异曲同工之妙:正如市场并非空中楼阁一样,人工智能技术也有它的社会、政治、文化与经济土壤。唯有将这些因素都考虑在内,才能看到人工智能技术的全貌。

如何理解该书的理论关切?我们认为有两点。

首先,该书作者认为人工智能技术是非中立的,并且可能系统地加深社会不平等。近年来,关于技术非中立性的大规模实证研究越来越多,例如科宁(R. Koning)等人发现,女性发明的专利更有可能解决与女性需求相关、服务女性的技术问题(Koning, Samila & Ferguson, 2021)。尽管技术的影响很大程度上取决于使用者的意图,但技术赖以生存的生态系统、发明过程、文化逻辑,本身就已经赋予技术以先天属性和成长方向。在作者看来,人工智能的生产过程依赖于对落后国和弱势群体的自然资源、人力和数据的索取,而作为算法基础的数据集对弱势群体也存在着误解和偏见。这使得人工智能影响下的政治经济体系先天不平等,且比旧技术中的偏见更难以定位、修正和追责。

其次,作者认为人工智能具有政治性,它以前所未有的方式改变了不同社会群体之间的权力分布。在政治经济学意义上,以硅谷为代表的技术行业垄断了人工智能的生产与使用,诞生出新技术阶级。技术精英、商业资本与

传统公权力的结合,将弱势群体置于更严厉的监管和控制之下。从微观政治角度出发,人类的行为甚至思想都可以被精密追踪,权力第一次真正通过毛细血管渗透到社会有机体的每个角落,重塑了资本与工人、政府与民众、军方与民事机构之间的社会关系。

四、技术社群的回应与局限

该书提及的问题,技术社群内部并非毫无觉察,近年来许多相关研究热点就是围绕该书提及的问题展开的。那么,针对这些问题,人工智能科学家做出了哪些努力,又有哪些局限性?很大程度上,该书提到的问题在今天仍然存在,技术社群面临的挑战为社会科学提供了新的思路和机会。

第一,针对人工智能的自然和人力成本问题,技术社群有一个重要研究方向——算法与配套设计的降本增效。这一类的尝试包括稀疏化或者修剪神经网络模型,寻找性能相仿的子网络,尝试不同的模型训练方式,在压缩网络上进行推理,在不过多影响计算精度的基础上降低计算位数,减少分布式训练时的通信,寻找更高效的硬件加速器,乃至更新云服务器架构等,目的都是降低存储和计算消耗(Caulfield, Chung & Putnam et al., 2016; Frankle & Carbin, 2018; Han, Liu & Mao et al., 2016; Micikevicius, Stosic & Burgess et al., 2022; Morcos, Yu & Paganini et al., 2019; Sattler, Wiedemann & Müller et al., 2019)。然而,需要指出的是,许多这样的尝试都是出于降低成本和对使用者负责的考量,主要立足点是提升经济效益。这虽与减少人工智能的环境和资源成本的愿景有重合,但因为少有对环境负责的直接考量,其带来的正面效益往往零散而有限。近年来,技术社群也开始思考人工智能的环境可持续性(Patterson, Gonzalez & Le et al., 2021; Patterson, Gonzalez & Hölzle et al., 2022; Wu, Raghavendra & Gupta et al.,2022),但关注点主要在人工智能的碳足迹上,且也仅是刚刚起步,如何评估并解决人工智能对环境的总体负面影响依然任重道远。

第二,关于人工智能和人类劳动过程的关系问题,技术社群最初对此关

注并不多,但近年来也开始涉猎此领域。例如,作为技术性的代表之一,OpenAI 的一支研究队伍近期就评估了人工智能对劳动者的正面效应。研究人员在调研后发现,大型语言模型对不同工资水平的工作者都有积极影响,能加快任务完成速度,从而助推经济增长(Eloundou, Manning & Mishkin et al., 2023)。不过,技术社群内部也有不同声音。从就业角度来看,有学者认为人工智能可能带来失业,会在继生产工具和材料丧失后,再让劳动本身丧失从而导致人的异化(Wogu, Olu-Owolabi & Assibong et al., 2017)。此外,近期也有学者研究了生成式人工智能对网络上技术问答社群的影响(Rio-Chanona, Laurentsyeva & Wachs, 2023),揭示了问答人工智能可能带来的个人图书馆私有化和知识存储中心化的挑战。虽然网络社群内的技术问答并非工作场所内的互动,但是由于场景的相似性,可以预见,使用类似于人工智能的辅助劳动也存在凯特·克劳福德所说的中心化的意义秩序固化的风险。人工智能和人类劳动过程的关系问题,依然需要研究者进行广泛而深入的探索。

第三,数据与算法中的知识权力结构问题,被技术社群总结为人工智能技术的“毒性”(toxicity)问题:模型的系统性偏误可能对社会产生破坏与伤害,而这种负面影响对弱势群体尤甚。技术社群对此有许多研究,并倾注了许多精力和资金尝试解决。首先,针对训练集内可能存在的偏见和错误,技术社群尝试将模型和用户目标对齐(alignment),建立基于人类反馈的强化学习(RLHF)(Christiano, Leike & Brown et al., 2017),把人类的偏好和正确的价值观作为奖励信号,对模型进行调整(Liu, 2023; Ouyang, Wu & Jiang et al., 2022)。其次,针对不准确的训练集,技术社群积极寻找、整合高质量训练集,甚至在学术文本训练集上进行训练(Taylor, Kardas & Cucurull et al., 2022)。最后,技术社群一直努力使人工智能摆脱对弱势群体的歧视,甚至具有帮助弱势群体的愿景。技术社群中有许多人致力于发现模型中的歧视,消除模型的偏见,甚至提出如何系统性地评估人工智能风险或者促进社会公平(Gabriel, 2022; Shah, Varma & Kumar et al., 2022; Shevlane, Farquhar & Garfinkel et al., 2023)。然而,上述努力也存在一些问题:基于人类反馈的强

化学习很大程度上依赖于第三国家劳工的廉价劳动,不断阅读有害内容可能造成情感剥削;类似于学术文本的高质量训练集,相比于常用互联网数据集存在数量级上的差异;目前人们对社会机制的理解还非常有限,其限制了人们主动干预社会过程的能力。当下对人工智能广泛而深刻的应用,本意是帮助弱势群体,但其施行不当可能会产生更严重的非预期后果(Eubanks, 2018)。人工智能如何有效嵌入社会环境中,减轻对弱势群体的伤害,依然需要诸多探索。

第四,针对公权力与私权力结合带来的技术垄断,技术社群的软件开源举措可能是一个行之有效的办法。对政府部门而言,也许开源难以从根本上阻止公权力扩张的脚步,但却有助于对人工智能进行审查,并对可能带来的公共损失进行追责。首先,开源意味着公开(publicity)。政府采用人工智能进行管理,将应用流程和算法细节置于公众目光审视之下,一旦最后的结果出现了损害民众利益的情况,便可进行责任追溯,向基于人工智能的社会公平靠近(Gabriel, 2022)。其次,在定位潜在问题后,开源允许技术社群成员有效应对算法带来的公共损失,例如通过算法补丁清除性别歧视(Bhardwaj, Majumder & Poria, 2021),从而实现更加公平的人工智能。最后,对私人公司而言,开源有助于解决个体和企业之间的算法权力不对等问题,从而制约公权私有化。例如,Hugging Face 是目前全球最大的人工智能开源社区,被认为是对抗人工智能垄断的前沿阵地——用户可以自由上传、下载、使用、修改模型,不必受大公司的约束,也不必把自己的数据交由他人管理。虽然开源有助于应对公权和私权结合的问题,但是其也并非毫无缺点。近年来,人工智能催生出诸多乱象,如有好事者合成虚假色情图片、制造和传播虚假信息等,这在一定程度上可以归因于此。同时,越来越高昂的人工智能训练和维护成本也使得许多“开源”名不副实。即使如 OpenAI 这样最初把自己定位为非营利性开源人工智能的组织,现如今的举措也有点名不副实。因此,开源也存在着基于现实考量的阻力。如何平衡公权力与私权力、自由与规则,可能成为社会政策中的一个难点,甚至需要建立新的治理框架予以应对(Zuiderwijk, Chen & Salem, 2021)。

从近年的发展来看,技术社群对凯特·克劳福德所提出的人工智能问题有许多回应,然而他们的出发点大多是解决具体技术问题中的缺陷,缺乏对社会过程的整体洞察。因此,社会科学在人工智能时代依然重要,甚至有史无前例的迫切性:对不平等的生产和再生产的理解将帮助技术社群做出有效的干预,对弱势群体进行赋权(empowerment);对人与技术关系的理解将帮助技术社群重新思考资源、人力投入与效益产出的性价比,并将无法避免的自然与人力损耗真正有效地转换为人类的根本福祉;对意义建构过程的理解将帮助技术社群重新思考预测方法的边界条件,在固化的分类体系中为流动的意义保留存在空间。从这个意义上讲,社会科学将是人类文明与工程技术之间的枢纽与桥梁。凯特·克劳福德的《人工智能地图集:人工智能的权力、政治和全球代价》正是在这样的背景下实现了其理论意义:尽管具体问题可能具有时效性,但该书关于环境正义、劳工权益、意义构建和权力边界的思考,为理解人工智能时代的技术创新提供了完整的学术视角和理论工具。

参考文献

Bhardwaj, R., N. Majumder & S. Poria 2021, "Investigating Gender Bias in Bert." *Cognitive Computation* 13(4).

Braverman, H. 1998, *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*, New York: Monthly Review Press.

Brynjolfsson, E. & A. McAfee 2014, *The Second Machine Age: Work, Progress, and Prosperity in A Time of Brilliant Technologies*, New York: W. W. Norton & Company.

Buechi, M., N. Festic & M. Latzer 2022, "The Chilling Effects of Digital Dataveillance: A Theoretical Model and An Empirical Research Agenda." *Big Data & Society* 9(1).

Bullard, R. 2000, *Dumping in Dixie: Race, Class, and Environmental Quality*, Boulder: Routledge.

Burrell, J. & M. Fourcade 2021, "The Society of Algorithms." *Annual Review of Sociology* 47.

Carson, R. 1962, *Silent Spring*, Boston: Houghton Mifflin Harcourt.

Caulfield, A., E. Chung & A. Putnam et al. 2016, "A Cloud-Scale Acceleration Architec-

ture.” 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO).

Christiano, P., J. Leike & T. Brown et al. 2017, *Deep Reinforcement Learning from Human Preferences* 30.

Crafts, N. 2021, “Artificial Intelligence as A General-Purpose Technology: An Historical Perspective.” *Oxford Review of Economic Policy* 37.

Crawford, K. 2021, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence,* New Haven: Yale University Press.

Edwards, R. 1980, *Contested Terrain*, New York: Basic Books.

Eloundou, T., S. Manning & P. Mishkin et al. 2023, “Gpts Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.” arXiv preprint arXiv: 2303.10130.

Eubanks, V. 2018, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York: St. Martin's Publishing Group.

Fei, N., Z. Lu & Y. Gao et al. 2022, “Towards Artificial General Intelligence Via A Multimodal Foundation Model.” *Nature Communications* 13(1).

Foucault, M. 1982, *The Archaeology of Knowledge: And the Discourse on Language*, New York: Vintage.

Frankle, J. & M. Carbin 2018, “The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks.” International Conference on Learning Representations.

Gabriel, I. 2022, “Toward A Theory of Justice for Artificial Intelligence.” *Daedalus* 151(2).

Han, S., X. Liu & H. Mao et al. 2016, “EIE: Efficient Inference Engine on Compressed Deep Neural Network.” *ACM SIGARCH Computer Architecture News* 44(3).

Ito, J. 2019, *Resisting Reduction: Designing Our Complex Future with Machines*, Cambridge, Ma.: MIT Press.

Koning, R., S. Samila & J.-P. Ferguson 2021, “Who Do We Invent for? Patents by Women Focus More on Women's Health, but Few Women Get to Invent.” *Science* 372(6548).

Krakowski, S., J. Luger & S. Raisch 2022, “Artificial Intelligence and the Changing Sources of Competitive Advantage.” *Strategic Management Journal*, https://doi.org/10.1002/

smj. 3387.

Liu, G. 2023, "Perspectives on the Social Impacts of Reinforcement Learning with Human Feedback." arXiv https://doi.org/10.48550/arXiv.2303.02891.

Marx, K. 1930, *Capital*, Vol. 1, London: Dent.

Micikevicius, P., D. Stosic & N. Burgess et al. 2022, "FP8 Formats for Deep Learning." arXiv https://doi.org/10.48550/arXiv.2209.05433.

Morcos, A., H. Yu & M. Paganini et al. 2019, "One Ticket to Win Them All: Generalizing Lottery Ticket Initializations Across Datasets and Optimizers." Proceedings of the 33rd International Conference on Neural Information Processing Systems.

Ouyang, L., J. Wu & X. Jiang et al. 2022, "Training Language Models to Follow Instructions with Human Feedback." *Advances in Neural Information Processing Systems* 35.

Patterson, D., J. Gonzalez & Q. Le et al. 2021, "Carbon Emissions and Large Neural Network Training." ArXiv https://www.semanticscholar.org/paper/Carbon-Emissions-and-Large-Neural-Network-Training-Patterson-Gonzalez/79b8ef3905a42b771248719495a2117271906445.

Patterson, D., J. Gonzalez & U. Hölzle et al. 2022, "The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink." *Computer* 55(7).

Polanyi, K. 2001, *The Great Transformation: The Political and Economic Origins of Our Time*, Boston: Beacon Press.

Rio-Chanona, M., N. Laurentsyeva & J. Wachs 2023, "Are Large Language Models A Threat to Digital Public Goods? Evidence from Activity on Stack Overflow." arXiv preprint arXiv:2307.07367.

Sattler, F., S. Wiedemann & K.-R. Müller et al. 2019, "Sparse Binary Compression: Towards Distributed Deep Learning with minimal Communication." 2019 International Joint Conference on Neural Networks.

Shah, R., V. Varma & R. Kumar et al. 2022, "Goal Misgeneralization: Why Correct Specifications Aren't Enough For Correct Goals." arXiv https://doi.org/10.48550/arXiv.2210.01790.

Shevlane, T., S. Farquhar & B. Garfinkel et al. 2023, "Model Evaluation for Extreme Risks." arXiv. https://doi.org/10.48550/arXiv.2305.15324.

Sourati, J., & J. Evans forthcoming, "Accelerating Science with Human Versus Alien Artificial Intelligences." *Nature Human Behaviour*, https://jsourati.github.io/assets/pdf/Writing_sample.pdf.

Taylor, R., M. Kardas & G. Cucurull et al. 2022, "Galactica: A Large Language Model for Science." arXiv. https://doi.org/10.48550/arXiv.2211.09085.

Tong, D., L. Wu & J. Evans 2022, "Low-Skilled Occupations Face the Highest Re-skilling Pressure." arXiv. https://doi.org/10.48550/arXiv.2101.11505.

Wogu, I., F. Olu-Owolabi & P. Assibong et al. 2017, "Artificial Intelligence, Alienation and Ontological Problems of Other Minds: A Critical Investigation into the Future of Man and Machines." 2017 International Conference on Computing Networking and Informatics.

Wu, C.-J., R. Raghavendra & U. Gupta et al. 2022, "Sustainable AI: Environmental Implications, Challenges and Opportunities." *Proceedings of Machine Learning and Systems* 4.

Zuiderwijk, A., Y.-C. Chen & F. Salem 2021, "Implications of the Use of Artificial Intelligence in Public Governance: A Systematic Literature Review and A Research Agenda." *Government Information Quarterly* 38(3).

智能社会研究
Journal of Intelligent Society

主 管 单 位：中华人民共和国工业和信息化部
主 办 单 位：哈尔滨工程大学
出 版 单 位：哈尔滨工程大学出版社
地　　址：哈尔滨市南岗区南通大街 145 号

国际标准连续出版物号：
ISSN 2097-2091
国内统一连续出版物号：
CN 23-1615/C

印刷单位：哈尔滨理想印刷有限公司
创刊年份：2022 年
出版日期：2023 年 7 月 10 日
发行单位：哈尔滨市邮局
订 阅 处：全国各地邮电局
邮发代号：14-375
发行范围：公开发行
定　　价：45.00 元

投稿指南

本刊面向海内外学者征稿，欢迎社会科学及交叉学科的专家学者惠赐稿件。请在来稿首页写明文章标题、作者简介（姓名、工作单位全称、联系电话、详细通信地址、电邮地址等）。文稿需完整，包括标题（中英文）、作者姓名、作者单位、摘要（300 字左右）、关键词（3—5 个）、正文、参考文献等。所投稿件如受基金资助，请在标题上加脚注说明，包括项目全称和项目批准号。来稿请以中文撰写。

稿件采用他人成说的，须在文中以括注方式说明出处，并在篇末列出参考文献；作者自己的注释均作为当页脚注。中外文参考文献分开列出，中文文献在前，外文文献在后，并按音序排列。中文文献参照中文社会学权威期刊格式，外文文献参照APA格式。来稿中的图表要清晰，符合出版质量要求，必要时可单独提供图表压缩包文件。

稿件格式请参考杂志官网（http://www.jis.ac.cn）“下载中心”中的稿件模板。

投稿方式：请登录杂志官网投稿系统（http://www.jis.ac.cn）进行投稿。

编辑部联系方式

地　址：黑龙江省哈尔滨市南岗区南通大街 145 号哈尔滨工程大学主楼北楼 N301 室，《智能社会研究》编辑部
邮　编：150001
电　话：0451-82588881
E-mail：mailtojis@163.com

著作权使用说明