

智能社会研究

(双月刊)

2024 年

第 3 期

2024 年 5 月 10 日出版

2022 年 11 月 10 日创刊

总第 10 期

目 次

论文

自由与认同：数字游民文化与本土化社会实践研究

——基于西南某地的田野调查 孙艺珂 周承磊(1)

智能时代劳动范式的数字转型

——基于政治经济学批判的视角 周光港(18)

政治经济学批判语境下数字异化的深层逻辑与扬弃进路 魏金鹏(37)

数字时代的社会自我

——破解碎片化自我困境的路径探索 高思蓉 王化起(52)

赛博格工人、“科技恶魔”与金币农夫

——东方主义遭遇数字时代 何祎金(66)

研究报告

中国个人信息保护经验研究的方法论反思

——欧盟经验的启示

..... 张月明 李汶龙 李汶锴 李子豪 李傲兰(85)

理性、权力与生态

——平台研究相关中文文献述评 袁方杰(114)

译文

气候变化与人工智能政治路径

——人工智能和人类能动性下的技治主义民主困境

..... 马克·科克伯格 亨里克·塞特拉 著

李 帅 李 芳 译(138)

归因法:网络攻击来源归因规则

..... 德尔伯特·特兰 著 裴 轶 强心语 译(160)

书评

数字技术支配下的生活世界

——读杰米·萨斯坎德《算法的力量:人类如何共同生存?》

..... 王国伟(203)

CONTENTS

THESIS

Freedom and Identity: A Study on Digital Nomad Culture and Localized Social Practices Based on Fieldwork in the Southwestern Region of China	Sun Yike , Zhou Chenglei(1)
Digital Transformation of the Labour Paradigm in the Age of Intelligence: A Perspective Based on the Critique of Political Economy	Zhou Guanggang(18)
The Deep Logic and Sublation Approach of Digital Alienation in the Context of Political Economy Criticism	Wei Jinpeng(37)
The Social Self in the Digital Society ; Exploring Paths to Resolve Fragmented Self Dilemmas	Gao Sirong , Wang Huaqi(52)
Cyborg Labor, “Devil of Science” and Gold Farmer: An Encounter Between Orientalism and Digital Times	He Yijin(66)

RESEARCH REPORT

Methodological Reflection on the Study of China's Personal Information Protection Experience ; Insights from the EU's Experience	Zhang Yueming , Li Wenlong , Li Wenkai , Li Zihao , Li Aolan(85)
Rationality , Power and Ecology : A Review of Chinese Literature of Platform Research	Yuan Fangjie(114)

TRANSLATED TEXTS

Climate Change and the Political Pathways of AI: The Technocracy-Democracy Dilemma in

Light of Artificial Intelligence and Human Agency

..... written by M. Coeckelbergh, H. Sætra; trans. by Li Shuai, Li Fang(138)

The Law of Attribution: Rules for Attributing the Source of Cyber-Attack

..... written by D. Tran; trans. by Pei Yi, Jiang Xinyu(160)

BOOK REVIEW

The Living World Dominated by Digital Technology: Reading Jamie Susskind's *Future Poli-*

tics: Living Together in A World Transformed by Tech Wang Guowei(203)

气候变化与人工智能政治路径^{*}

——人工智能和人类能动性下的技治主义民主困境

马克·科克伯格 亨里克·塞特拉 著^{**}

李 帅 李 芳 译^{***}

摘要:人们普遍认为,人工智能对政治和民主产生了巨大影响。同时,该技术常被视为解决关键社会和环境问题的方案。人工智能也引发了诸多问题,如人类应该如何应对气候变化。本文将这些议题置于技治主义民主困境背景中,利用关键节点概念进行讨论,提出两条人工智能政治路径,并对其基本假设展开批判性讨论。这为进一步探讨人工智能、气候变化和民主之间的关系提供了有益框架,且有助于研究技术政治的重要问题,如人类专业知识相对于人工智能的作用、技术解决主义引发的问题,以及应在何种水平上治理人工智能和气候变化。

关键词:人工智能 民主 技治主义 气候变化 关键节点 技术政治

一、引言

作为公民,人们往往认为社会秩序和发展主要由个体行动或最高权威者的

* 本文原题为“Climate Change and the Political Pathways of AI: The Technocracy-Democracy Dilemma in Light of Artificial Intelligence and Human Agency”,原文见 <https://www.sciencedirect.com/science/article/pii/S0160791X23002117?via%3Dihub>,经授权翻译发表。本文系云南省“兴滇英才支持计划”青年人才专项项目(项目批准号:XDYC-QNRC2022-0642)的阶段性研究成果。

** 马克·科克伯格(M. Coeckelbergh),奥地利维也纳大学哲学系。亨里克·塞特拉(H. Sætra),挪威奥斯特福德大学。

*** 李帅,云南大学马克思主义学院。李芳,云南大学马克思主义学院。

决策所决定。诚然,这样的观点有其合理性,但本文认为,社会发展往往是沿着技术所预设的某种路径发展的。技术属于政治的说法已经屡见不鲜,技术塑造着社会、经济和政治的发展(Winner, 1977; Winner, 1980: 121–136; Ellul, 1964)。历史上,某些时间点可被定性为关键节点(critical junctures),在该时期,人类的能动性是纠正、引导或塑造历史发展长波的关键(Jonsson, 2015; Jonsson, 2016: 131–151; Collier & Collier, 1991a: 27–39)。目前,人类正处于这样一个阶段,数据的可获得性和人工智能的发展达到了前所未有的水平,挑战了人们对不同政治制度利弊的现有认知。结合当前与环境可持续性,特别是气候变化相关的挑战来看,该问题尤为突出,其中不乏严厉警告(IPCC, 2022)——必须进行重大变革。然而,政治家面对这些问题时的挫败感却与日俱增。生态法西斯主义(eco-fascist)解决方案的呼声(Satgar, 2021),如“反抗灭绝”(Extinction Rebellion)等试图破坏的社会运动,以及年轻人为气候而战[特别是“未来星期五”(Fridays for Future)],都凸显了气候变化引发的社会和政治紧张局势。与此同时,气温升高和自然灾害日益频繁的现实情形凸显了造成这些紧张局势的真实且紧迫的原因。众多要求变革的呼吁都有一个共同点:要求关注人类的能动性和责任,但这没有强调技术在导致人类面临的挑战中起到的决定性作用,也没有强调技术为人类当前困境提出某些解决方案的核心作用。诚然,人类在技术中始终扮演着重要角色(作为用户、开发者、维护者、监管者等),而且正如我们从其他新兴技术的引入中所了解的那样,一项新技术或新兴技术的成功取决于人类系统(Maynard & Dudley, 2023: 1118–1120),即取决于人类—技术系统或社会—技术系统,但技术本身并不是政治中立的(Coeckelbergh, 2022a),往往具有超人类预期和控制的普遍社会效应。

本文将重点讨论如何辨别人工智能的政治倾向,以及如何理解和看待备选的政治路径。为此,本文将探讨“民主”和“技治主义”(technocracy)概念,特别是在应对气候变化挑战的背景下,两者的关系十分紧张。首先,人工智

能在某些方面具有改善民主政治的潜力,如改进意见形成、协商、共识形成、决策和参与等(Savage, Chiarini & Evans, 2018: 369–380; Bakker, Chadwick & Sheahann et al., 2022; Koster, Balaguer & Tacchetti et al., 2022: 1398–1407)。迄今为止,人类在应对气候变化的决策过程中已展现出明显的不足,且常陷入困境。为此,有观点提出,人工智能可以提供改善这些决策过程的有效途径(Sætra, 2020)。其次,人工智能越来越被认为是一种促使民主过时的技术(Sætra, 2020)。因此,人工智能在很大程度上可以替代人类的政治决策。这些论调均基于该观点,即人工智能不依赖于民主决策,因此人工智能可以改善政治决策。例如,在应对气候变化的问题上,通过增加监测和监视,创建社会及其居民的数字孪生体,并利用无与伦比的处理能力和数据分析能力,我们可以在理论上改善社会的各个方面,并有效地应对气候变化。但代价是什么?本文旨在结合气候变化和相关的全球环境挑战,讨论在人工智能领域已知但研究不足的政治倾向。通过阐明目前出现的两条路径,本文探讨了每一条路径的可能后果,同时论述了民主与技治主义之间的紧张关系。

首先,本文将在技术和科学哲学之中进行分析,将科学技术与社会(STS)理论同政治和经济发展理论相结合,以此来说明人工智能具有政治性的原因。同时,运用社会和政治理论来理解人类能动性、技术和其他历史制度力量之间的相互作用。其次,探索人工智能的两条重要政治路径:探讨人工智能如何发展出新的可能性,并创造场景;在这种场景中,人工智能要么加强和改善民主,要么助长技治主义,同时预测这两种情况的后果。最后,讨论人类处于关键节点所面临选择的性质。本文认为上述两条路径各有优点,但讨论我们是否应该接受政治发展的路径依赖,即我们是否应该接受在这些情况下做出的假设,也是至关重要的。虽然人工智能作为主要政治决定因素预示着某种未来,但我们不能忽视潜在的替代性未来。在这种未来中,人类并非只能选择由技术决定的道路,也可以从对当前社会与技术的关系以及人类如何在当前全球形势下继续前进的基础分析着手。

二、技术与社会变革

在思想史上,技术在社会中的作用时常被低估。直到近代,情况才有所改变,人们注意到了新(工业)技术如何重塑社会,如何对人们的日常生活产生深远影响。因此,19世纪和20世纪上半叶的社会理论倾向于以一种确定的方式概念化技术与社会之间的关系。依据技术决定论(technological determinism)的观点,历史发展必然受技术革新的驱动,而技术革新又决定着社会结构和文化价值(Heilbroner, 1994: 67–78)。马克思认为,生产关系(以及生产技术)决定社会内部的社会关系。在该观点中,技术具有社会性和政治性,技术决定了社会的基本结构。技术决定论的当代变体即超人类主义(transhumanist)观点,该观点认为技术奇点或其他由技术决定的未来将不可避免地发生,如科兹韦尔(Kurzweil, 2015: 146–170)或博斯特罗姆(Bostrøm, 2014)的观点。

技术工具论(technological instrumentalism)的观点与之相反,它认为技术只是一种工具,而且技术在政治上也是中立的(Schatzberg, 2018)。与温纳(Winner, 1979: 75–86)所说的“技术正统”(technical orthodoxy)有关,其涉及人类对后果的理解和控制,以及技术的中立性。技术决定论和技术工具论都受到了技术哲学与媒介哲学理论家和技术研究理论家的严厉批评。前者强调了技术的非工具性和非预期效应,早期有海德格尔(M. Heidegger)和麦克卢汉(M. McLuhan),今有伊德(D. Ihde)和芬伯格(A. Feenberg)。后者认为,技术与社会相互交织、共同发展。例如,兰登·温纳(Winner, 1980: 121–136)在其关于技术哲学与技术研究之关系的著作中,提出了一个著名的观点:技术是政治性的,但这不能简单地从技术决定论或社会决定论的角度来理解,在特定的文物历史中,技术与社会相互影响,但也可能相互独立。

经济和社会变革理论持续关注历史上的关键转折点。例如,熊彼特

(Schumpeter, 2013)的经济发展概念,即通过引入新产品、引入新生产方法、开拓新市场、征服新原料来源或建立新的产业组织,在经济流动中进行创造性破坏,从而实现非连续性变革。在他看来,这些都是由企业家引起的。随着这种变革的推进,原有的经济流动被打断,从而催生新的发展与创新。此外,还不能忽视库恩(T. Kuhn)的范式转移(paradigm shifts),即其在科学史和社会学中引入了不连续变化概念(David, 1985: 332–337; Fioretos, Falleti & Sheingate, 2016)。

关键节点理论(critical juncture theory)认为,存在某些可以改变一个实体(如一个物种、一个社会)演化过程的关键转折点。关键节点可以被定义为:“一个重大变革时期,通常以不同的方式发生在不同的国家(或其他分析单元中),并假设会形成不同的结果。”(Collier & Collier, 1991b)在这些时期,未来的轨迹已经形成,并在一定程度上确定下来,这就诞生了源自经济学的“路径依赖”(path dependence)概念。

历史制度主义(historical institutionalism)持续关注关键节点的能动性作用,因为在这个相对较短的时期内改变发展轨迹、塑造社会未来空间相对较大。例如,强有力的制度约束在一定程度上会形成路径依赖,因此,把握人类能动性如何形成新的路径是十分重要的(Capoccia & Kelemen, 2007: 341–369)。例如,琼森(Jonsson, 2016)在其政治经济学著作中使用了“关键节点”的概念。他认为,冰岛引入新自由主义(neoliberalism)就是一个关键节点,而政治反对派却未能以一种能够促成政权更迭的方式介入。他还指出,真正的关键节点是可以被创造出来的,因此在面对各种制度力量的制约时,关键节点进一步突出了政治和人类的能动性。

三、双重关键时刻:气候变化与人工智能

环境可持续性被广泛认为是全社会和全人类必须解决的重要问题,气候

变化则被视为当代最大的挑战。科学家警告说,气温正在上升,极端天气事件愈加频繁;冰川正在融化,全球海平面正在上升。为了解决这些问题,各级机构必须采取政治行动和决策,减少与人类活动相关的温室气体排放。在该领域,有人声称已经到达了关键节点,这通常与各种临界点(*tipping points*)有关(在临界点之后将会失去控制)(Lenton, 2013: 1–29)。如果我们现在不充分减缓气候变化,局面将会失控,必将出现大规模全球变暖及其他影响。甚至有人声称,我们已经越过了临界点。

“关键节点”的概念在政治体制分析中占据着重要地位,同样地,这一概念在应对气候挑战时亦显得尤为关键。国家制度,即将人类社会组织成主权国家的结构,以及既有的国际合作框架,都在很大程度上影响甚至阻碍着应对气候变化的策略发展(Sætra, 2022: 66; Coeckelbergh, 2021a)。本文将论证,人类目前面临的气候变化挑战就是一个关键节点,这将挑战已有的政治路径和制度遗产,从而保护自然环境。气候变化已是政治争议的根源之一,如果不能在全球范围内减缓和适应气候变化,预计会导致更多的社会和政治动荡,增大本国暴力冲突的风险(IPCC, 2022);如果成功应对这一挑战,预计将对区域性和全球性合作产生深远影响,并对民主在威权主义(authoritarianism)日益突出的世界中的可行性产生重大影响。

与此同时,科技发展领域可能正迎来另一个关键节点——人工智能和数据科学取得了重大突破,特别是机器学习的广泛应用和数据可用性的显著提高。人工智能正在迅速普及,并广泛应用于商业和政府各个领域。人类或许正迈过一个临界点,这与科林格里奇困境(Collingridge's dilemma)有关(Collingridge, 1980),新技术在初期阶段易于监管,但其潜在的深远影响却难以预测和认知,因此监管的理由显得模糊不清。然而,一旦技术成熟,其产生的后果将变得清晰无疑,但此时再试图对其进行规制,往往事倍功半,甚至难以奏效。这表明,技术在临界点之后将会成熟,但在这个临界点之后,未来的发展道路将被固化,技术的政治含义也将难以避免。

因此,人类或许正处于一个双重关键时刻。此刻,自然环境的保护路径与技术发展的方向均呈现出相对开放的状态,而且这两条路径都会对我们正在塑造和选择的政治道路产生重大影响。今天,关于人工智能可持续能力的讨论越来越多,针对人工智能促进可持续发展的解决方案比比皆是(Galaz, Centeno & Callahan et al., 2021)。双重关键时刻的主要驱动因素可能与正在进行的、与数字化转型中绿色转型必要性相关的政策讨论有关(Floridi, 2019: 183–221),讨论将“绿色”(环境政策和绿色经济)和“蓝色”(数字政策和数字服务经济)结合的政治,两者之间的互动和关系日益成为政策制定者,尤其是欧盟内部政策制定者的关注焦点。欧盟委员会已经将“双轨转型”(twin transition)概念作为其战略规划和分析的关键内容(Makitie, Hanson & Damman et al., 2023),明奇等人(Muench, Stoermer & Jensen et al., 2022)将这两种转型描述为“政治优先事项”,两者既相互促进,又存在紧张关系。例如,数字技术可以促进创新和增长,但同时也会提高环境成本。

人工智能与气候变化之间的关系,无疑是绿色转型与数字转型交织的关键节点。在全球性议题如气候变化治理中,人工智能的应用显得尤为重要。然而,技术发展的双刃剑效应也尤为显著:一方面,人工智能的运作本身,如模型运行和数据服务器维护,均需要消耗大量能源,进而产生温室气体排放,这无疑加剧了气候变化的严峻性(OECD, 2022);另一方面,人工智能亦展现出在减缓气候变化方面的巨大潜力。其不仅能够协助处理温度变化和碳排放数据,预测极端天气事件和能源使用情况,还能在改造运输系统、监测海洋乃至辅助地球工程等方面发挥关键作用。尽管其中某些应用,如直接空气碳捕获技术(direct air capture, DAC),在学界和政策层面仍存争议(Arcusa, Lackner & Page et al., 2022),但其潜在的积极影响不容忽视。值得一提的是,人工智能在引导人类社会行为方面的作用。在营销和广告领域,人工智能已经展现出强大的影响力,能够在潜移默化中引导人类行为。或许可以利用人工智能来引导人类社会朝着更加气候友好的方向发展(Cockelbergh,

2021b: 67–72; Mills & Whittle, 2023), 同时还可以提高人类的决策能力, 这似乎很有诱惑力 (Sætra, 2021)。但是, 这也存在一定的危险和弊端, 在讨论不同的选择和方案后就会明白这一点。

人工智能不仅影响自然科学, 还影响人文科学和政治等活动。例如, 现在可以用大语言模型 (LLMs) 创建文本, 用人工智能辅助政治决策。随着 ChatGPT 及其他生成式语言模型的兴起, 人们担心人工智能在游说方面的潜在作用会“劫持民主” (hijacks democracy)。“生成式人工智能”是一个更广泛的术语, 不仅包括以文本形式, 还包括以图像、音频、视频等形式生成内容的人工智能系统, 影响着个人 (微观层面)、群体 (中观层面) 和社会 (宏观层面)。塞特拉 (Sætra, 2023a) 将生成式人工智能在上述各个层面会令人们产生担忧的一些原因绘制成图 1。

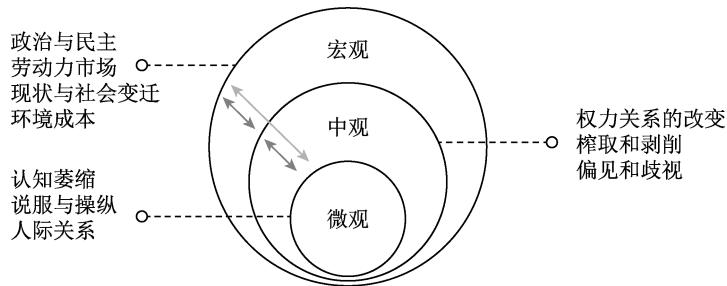


图 1 生成式人工智能的潜在危害

塞特拉指出, 生成式人工智能系统存在对政治和民主进程产生深远影响的潜在问题。具体而言, 这类系统能够生成大量内容, 包括虚假新闻和深度伪造的内容, 改变公众舆论和认知。此外, 这些系统还可能被用于操纵特定目标, 从而改变原有的权力关系。塞特拉进一步指出, 这类系统在开发和部署阶段均需要消耗大量的能源和设备资源, 无疑增加了环境成本 (Dodge, Prewitt & Combes et al., 2022: 1877–1894; OECD, 2022)。

不过, 人工智能也具有积极的政治潜力。例如, 一些研究人员正在研究如何使用大语言模型促进信仰和观点不同的人达成一致。通过这些应用, 我

们看到人工智能有助于促进协商和达成共识,而这也正是民主支持者关注的关键领域。

我们之所以将人工智能的积极方面和消极方面都描述为“潜力”,是因为我们想挑战人类无法控制自己未来轨迹的决定论立场。因此,本文强调人类的能动性以及采取有意义行动的必要性,以确保在当前关键节点上引导未来朝着正确的方向前进。聚焦于人工智能,本文提出在政治活动中使用人工智能时,至少要区分两种选择或路径,我们称之为“人工智能增强的民主”(AI-augmented democracy)和“人工智能驱动的技治主义”(AI-driven technocracy)。

四、人工智能增强的民主

第一条可能性路径是在不取代人类参与决策的前提下,使用人工智能增强民主——人工智能增强的民主。气候变化已被证明是对传统民主决策进程的严峻挑战,一些人认为,加强民主国家的协商进程是采取有效气候行动的关键(Willis, Curato & Smith, 2022)。理论上,人工智能可以帮助大型社区之间更有效地进行沟通、会面甚至协商,这对更好地理解气候变化带来的挑战和潜在的解决方案至关重要。

技术研究人员和政治理论家提出了各种概念和方法。例如,流动民主是直接民主的形式之一,它允许选民委托投票,如可以通过数字平台和算法来实现(Kahng, Mackenzie & Procaccia et al., 2021: 1223–1252)。另一个例子是DeepMind公司的研究,该研究提出利用深度强化学习,寻找人们会为科斯特等人(Koster, Balaguer & Tacchetti et al., 2022: 1398–1407)投票的经济政策。在一场投资博弈中,人工智能发现了一种优化人类偏好从而赢得多数投票的机制。此外,在政治理论方面,也存在利用人工智能促进民主的工作。例如,有人提出使用语言模型来寻找不同偏好民众间的共识(Bakker, Chadwick & Sheahan et al., 2022)。兰德摩尔(Landemore, 2023)认为,人工智能

可以通过多种方式提高协商质量,从而实现大众协商(微观公共领域)。在微观公共领域,随机挑选民众参与协商。人工智能还可以作为促进者(如计时器)、翻译者、事实审查者、论据组织者、交流跟踪者、重叠和质量的衡量者以及团队立场可视化者的角色,帮助改善协商工作。其他人还注意到,协商民主原则可以促进负责任的创新(responsible innovation)和控制人工智能(Buhmann & Fieseler, 2021)。

在这些案例中,人工智能都是在高度政治化背景下使用的,而且人类在民主决策过程中仍保留了核心作用。然而,自柏拉图和培根以来,持续存在着专制主义(autocratic),特别是技术官僚主义(technocratic)的诱惑:人类,包括专家(柏拉图)都经常出错,为什么还让人们做决定,而不是使用科学技术引导社会朝着更加美好的方向发展,创造一个勇敢的新世界呢(培根)?

五、人工智能驱动的技治主义

不应轻视或低估技治主义的诱惑。依靠专家和专业知识让人工智能做出政治决策似乎是很棒的理由。理论上讲,人工智能可以更有效地确定应对气候变化的有效缓和战略与适应战略。基于此,一些人可能会认为,如果人类赋予人工智能系统绕过阻碍有效行动的烦琐民主程序的权力,人类自己就可以更有效地实现减缓和适应气候变化的相关目标,即第二条可能性路径——人工智能驱动的技治主义。

塞特拉(Sætra, 2020)在其《对人工智能技治主义的略微辩护》("A Shallow Defence for ATechnocracy of AI")一文中提出,我们应当认真审视并考虑人工智能技治主义的观点。他认为,在政治决策中更加积极地利用人工智能具有充分的合理性,主要基于以下三个前提:第一,政策评价应以该社会的基本道德价值为基础,确定基本道德价值是政治的首要目的;第二,根据第一个前提得出的最佳政策应该予以实施;第三,在某些涉及科学、工程和复杂社会

及宏观经济问题的领域,人工智能比人类更善于发现和制定最佳政策。本文提到的这些论点虽为假设性的,但其主要目的并非直接倡导技治主义,而是检验我们是否拥有充分的论据来反驳这一观点。

政治控制论也能提出类似的论点。1948年诺伯特·维纳(Wiener, 1948)举办了控制论研讨会后,许多政治学家逐渐倾向于将政治制度理解为输入和输出的流动,并将这些输出作为反馈重新输入,著名的例子包括卡尔·多伊奇(Deutsch, 1963)的《政府的神经》和戴维·伊斯顿(Easton, 1965)的《政治生活的系统分析》。受当时先进计算机系统和仿真技术所带来的可能影响,20世纪60年代是一个对计算社会科学潜力和更全面了解与控制政治制度非常乐观的时期。然而,20世纪60年代末,计算政治学方法的局限性日益凸显。在这一背景下,温纳(Winner, 1969: 1–17)对多伊奇的作品做出了评论,这种评论在某种程度上可被视为对当时政治学领域中一场热潮开始走向衰退的悼词。

虽然政治控制论最终失去了支持,但随着大数据和人工智能的发展,有人开始预言政治控制论将再次兴起,并有望实现20世纪60年代政治学家尚未达成的雄心壮志。由于海量数据被用以控制多种形式的社会制度,无论是否存在有价值的人类控制,多样化的控制论与人工智能技治主义都紧密联系(Zuboff, 2019)。

在应对气候变化问题方面,技治主义十分具有吸引力,因为普通民众可能不具备应对气候变化必需的专业知识,即使是人类科学家也不可能掌握一切知识。但人工智能凭借其对大数据的深入挖掘和分析能力,可做出更为精准的预测,从而推动制定更为科学合理的决策。然而,技治主义也并非完美无缺的(Sadowski & Selinger, 2014: 161–168)。塞特拉(Sætra, 2020)以五种反对意见总结了技治主义所面临的主要问题:第一,民众需要充分的政治参与才能得到满足;第二,没有民众参与的政府被认为是不合法的;第三,计算机无权做出影响人们生活和福祉的决定;第四,人工智能是不透明的,不可能

完全服从于人类控制;第五,政治决策后果问责必须要明确,而当涉及人工智能决策时,问责将不再明确。

这些似乎都是驳斥人工智能技治主义的有力证据。然而,塞特拉(Sætra, 2020)认为,当我们更仔细地审视政治制度时会发现,现存的政治制度(如非理想化和非假想的自由民主形式)也因许多类似的反对意见被摒弃,这给我们带来了一个实际问题,即必须在两个非理想的解决方案之间做出选择(Gaus, 2016)。因此,重要的不是比较技治主义的悲观描述(现实主义)与民主的乐观描述(理想主义),因为现实政治制度总是非理想政治制度的典范。因此,真正的问题是:对人工智能技治主义和人工智能增强民主的最终评估是否会产生更有利的结果,尽管它们都存在负面因素?

然而,即使这些系统在实际应用中取得了成功,也有人可能会反驳,人工智能的技术官僚化运用将威胁到启蒙理想和人类自主权价值。技术官僚将会从民众手中夺走决策权,甚至隐蔽性地操控民众,从而忽视了人民对自我决定生活、社区以及社会事务的强烈愿望。康德式的自主思考也将无法得以实践:人工智能不具备思考能力,那么人类的自主权将得不到保障。然而,自由主义的理论家如弗里德里希·哈耶克(Hayek, 1960)提出了相反的观点,他强调经济自由对有意义的个人生活至关重要,民主在这方面也许是次要的。

另一种反对观点,即一种类似于技术决定论的观点,至少在规范形式上认为技术应该塑造社会。这一观点有时也使用“技术解决主义”这一术语(Sætra, 2023b),即认为人类面临的问题本质上都是技术性的,技术可以解决所有问题,包括政治问题。但本文认为,这是不可能的。

六、讨论:选择哪条路径

第一条路径的优势在于,坚持人类决策的主导地位,而将人工智能作为

辅助决策过程的工具。然而,当涉及大众协商时,决策过程就会存在挑战,但仍然符合协商民主(*deliberative democracy*)理念和总体民主(*democracy in general*)原则。人工智能是一种技术性辅助工具,是民主决策的支撑。因此,该路径也为将人工智能与政治科学家的各项增强民主政治决策的提案(如微观公共领域)的融合提供了可能性。该条路径也更有可能获得民主支持,因为人类仍然处于主要地位,人类不会产生人工智能正在取代自身的感觉,从而消除了由科幻小说所引发的普遍恐惧。

第二条路径的优势在于,通过人工智能保证了专业知识的参与。相较于人类,人工智能处理数据的能力更强,能够发现人类可能忽略的模式,人工智能在特定领域极有可能设计出更佳政策。这对于应对气候变化和需要在海量数据中辨别模式、在全球范围内进行更具挑战性协调的复杂社会和全球问题具有重要意义。将政治权力赋予人工智能,而非仅仅将其视作工具,这意味着人类能够更加有效地应对这些问题。

这两条路径都存在严重缺陷。即使在第一条路径中,人工智能也可能会影响民众的决策,而不是仅仅作为一种工具。正如科克伯格(Coeckelbergh, 2021b)在讨论使用人工智能缓解气候变化时所指出的,由于人工智能的预期或非预期效应,它可能会操纵民众,与人类的自由和自主权相违背。至少存在三方面的风险。首先,人工智能中介协商可能会干扰民众的思维,影响决策结果。例如,科克伯格(Coeckelbergh, 2022b)提出,人工智能可能影响信仰的形成和修正,包括政治信仰。剑桥分析(Cambridge Analytica)公司从数百万个脸书账户资料中收集数据进行分析,影响选民,这就表明了通过社交媒体,人工智能可以影响选举结果。其次,就目前而言,人工智能已经实现了海量推理、组织讨论和做出决策;经过人工智能“优化”或“增强”的协商可能更加肤浅,或无法提供人们从“真实”或“全面”协商中获得的学习结果——在这种协商中,人们不得不从头开始讨论,思考新方案。因此,经人工智能调解、促进和简化,协商的性质可能会发生质的差异。最后,人工智能辅助的决

策过程缺乏透明度,可能被技术系统控制者刻意操纵。

从第二条路径来看,没有理由将人工智能的能动性局限于解决特定问题(如改善气候变化的决策)。人工智能能做的远不止于此,它可以优化人类生活的方方面面,甚至可以被赋予充分的政治权力。那为什么还要限制人工智能的力量和影响呢?当人类和人工智能意见不统一时,问题就会出现。例如,假设人工智能决定剥夺某些人的自由来实现某个目标(如应对气候变化),或者人工智能可以在没有透明法律程序的情况下决定人们的生活和自由,那么政治自由和法治等原则将会受到威胁。鉴于目前人工智能在发挥作用时存在偏见和歧视问题,第二条路径要如何通向公正与平等仍然值得商榷。基本的原则如自由、公正和平等有充分的保障吗?如果人工智能认为人类不需要这些呢?因此,这条路径必然会带来专制化,甚至可能是极权主义(totalitarian)政治制度的真正风险。

第一条路径虽有其优势,但也存在一个不容忽视的内在缺陷:原则上协商民主可以跨越国界实施,但目前的建议更多地聚焦于国家层面的决策优化,而忽视了人工智能在全球层面辅助决策的巨大潜力。特别是在应对气候变化等全球议题时,这种全球性的辅助决策能力显得尤为重要。例如,人工智能辅助下的微观公共领域的民主工具和方法,在全球范围内是否具备可行性和有效性,这是一个值得深入探究的问题。相比之下,第二条路径试图解决这一全球性问题,但又引入了新的挑战:可能导致一种被人工智能主导的全权性体制(a totalitarian system),这种体制不仅威胁到自由和民主,甚至可能导致人类的终结。一旦人工智能的目标与人类产生冲突,其结果将是难以预料的。例如,如果人工智能认为种族灭绝或生态灭绝是实现某个或某些人类目标的最有效方式,那么人类将面临前所未有的危机。有人提出,正如DeepMind等多数大型科技公司所做的那般,人工智能应该与人类的目标保持一致,这一观点在表面上看似合理。然而,当人工智能的决策逻辑与人类产生分歧时,如何确保这种一致性将成为一个难题。同样地,虽然有人主张

人类必须控制人工智能,但这又可能牺牲人工智能在决策过程中的优势。

上述讨论深刻揭示了人工智能与民主之间的两难选择困境。诚然,两条路径各自存在不容忽视的缺陷,但这并不意味着人类必须在二者之间做出非此即彼的选择。幸运的是,我们不需要接受任何一条路径,也不需要把这两种困境视为既定事实。

让我们阐明这两条路径背后的一些(有问题的)假设。

第一,这两条路径都认为技术可以解决所有问题(如气候变化)。这些观点都建立在一种对技术修复能力过度夸大且缺乏坚实依据的信任之上,即技术解决主义。也有其他的观点和方式来界定人工智能治主义问题,特别是有人指出,人类不可能完全依赖人工智能,我们还需要人类的专业知识。那么,核心问题就不仅仅是人工智能可以和应该为民主做什么,而是在民主中人工智能专业知识与人类专业知识的关系应该是什么。这些问题十分复杂,我们需要承认仅仅依靠技术无法解决这些问题,我们始终需要人类及其专业知识。

第二,这两条路径都倾向于以技术决定论为基础。两者都认为,在民主中使用人工智能是不可避免的,我们必须在不同的人工智能应用方式之间进行选择。但在民主中并不一定要使用人工智能,我们还可以做出其他的选择。此外,人工智能对社会和民主的影响很难预测,而且可能根本不是直接性的,因为技术总会产生意料之外的后果——积极的或消极的。

第三,虽然本文的讨论聚焦在预期效果上,但我们也可以说从技术哲学中了解到,人工智能除了会产生有利的预期效果之外,还不可避免地会产生意料之外的副作用。再次谈及有关控制的观点,以及为了实现(人类共识的)目标,人工智能选择的手段是否会导致极权主义的反乌托邦,甚至是人类的终结。这意味着,这两条路径可能都无法实现自己的初衷,尤其是当目标涉及维护民主时。

第四,政治与伦理的关系问题。例如,弗洛里迪(Floridi, 2019)在创作

《绿色与蓝色》时就认为,伦理优先于政治。基于伦理,人们可以拒绝这两条路径,哪怕它们有助于民主或解决地球问题。这就又回到了人类自主权问题上,这不仅可以看作对政治原则的呼吁,还可以看作一种伦理价值。然而,还存在其他价值。

第五,与国家制度相关的挑战以及全球合作或治理的必要性(Bull, 1977)。本文已经指出,第一条路径的设想仅是局限于国家层面拟定的;而理论上,第二条路径也可能推进一个世界国家的建立。两种方式各自存在非常严重的问题。因为气候变化是一个全球性议题,稳定的气候是一项全球公共产品(Barrett, 2007),所以我们需要在全球范围内解决该问题。我们可以设想几种实现气候变化全球治理的方式,每种方法都各有利弊。迄今为止采取的最明显的解决方案,即利益相互竞争的主权国家团结起来,协商并达成共识,解决共同的问题。该解决方案尊重各个国家和民族的主权,其优势在于现实性以及对利益冲突和国际无政府状态的认知。在无政府体系中描述各个国家之间相互作用的另一种方式是,将国家看作国际社会的行动者。该观点以国际关系中的英国学派为代表(Bull, 1977),“国家社会”(a society of states)的概念强调了非正式关系、规范和共同价值观如何极大地影响并隐蔽性地为全球范围更密切、更有效的合作与治理奠定基础。然而,正如批评者指出的,一旦权威行动者发现维持这种社会不再符合他们的利益,这种全球治理方式就很容易崩溃。最后,有人可能会提出一种解决方案,即由一个超级国家实体取代国际合作,该实体拥有强制执行全球解决方案的权威和权力(Sætra, 2022)。这种解决方案将会招致强烈的伦理反对和实践异议。许多人既反对抛弃主权国家和以民族国家为基础的国家制度的可取性,又反对在全球层面实施有效政治制度的实际可行性。人工智能和气候变化对国家和全球治理结构提出了挑战,我们应持开放态度,考虑在各个层面采取不同的治理形式。

第六,即使人工智能能够和人类目标保持一致(Koster, Balaguer & Tac-

chetti et al., 2022: 1398–1407),也不应该将人类的目标视为理所当然。因为人类的目标可能本身就存在谬误。以当前的消费社会为例,人类普遍依赖并为之贡献,但这种模式对人类和地球环境均造成了严重损害。如果我们任由人工智能按其自身的发展轨迹前行,即使人类很幸运,人工智能可以在符合人类目标的前提下推进人类目标;但是,也存在固有问题,即一些人类目标本身就存在严重的问题。例如,最大限度地刺激消费不可能帮助应对气候变化,高消费会造成能源浪费和环境污染,更不用说自然资源的耗竭,这会以威胁生态系统的方式破坏环境,最终影响人类。人类活动已经过度,减少人类活动和人口数量既符合非人类生命和人类生命的繁荣,又是必要的(Næss, 1989)。如果管理政治的人工智能系统就人类的问题最终得出了与我们类似的结论,并着手实施必要的政策,减少人类活动,其可能带来的深远影响将是人类难以预料的。

七、结论

在探讨人工智能与政治的关系,特别是其与民主议题的联系时,需将其置于技术与社会变革的宏观背景之下,并考虑其与气候变化等全球性挑战的政策交汇点。在此背景下,本文勾勒出了人工智能的两条政治路径:一是运用人工智能促进政治与民主决策的优化,二是借助人工智能解决气候变化等全球性挑战。然而,这两条路径各有其利弊,需审慎权衡。人工智能可以且应该用于增强人类决策的民主性,但正如本文所述,人工智能也存在局限性。最终,人工智能可能会产生不利于实现人类目标的副作用,或在政治上,或在道德上,产生不可接受的其他影响。与此相反,让人工智能来做决定(不受人类的控制和影响)则危险得多,因为它从一开始就威胁着人类的基本价值。

本文深入剖析了导致技术政治倾向困境的原因,这与当前的政治理论、

政策制定及科技界就人工智能与民主关系的讨论紧密相关。同时,这些讨论引出了关于人工智能与政治关系的核心议题,甚至更为深层次的根本性议题,如人工智能专业知识与人类专业知识之间的互补与平衡、技术解决主义与决定论的边界问题、错误的人类目标与人工智能非预期性副作用的潜在冲突、伦理考量与政治决策的融合,以及气候变化等全球性挑战如何在多层次得以妥善应对与治理。如果我们想在政治和伦理上负责任地使用人工智能,特别是在民主政治中使用人工智能,就需要在这些领域开展更多的工作。

参考文献

- Arcusa, S. , K. Lackner & R. Page et al. 2022, “The Socio-Economic and Technical Context for Direct Air Capture in Decarbonization Pathways.” <https://keep.lib.asu.edu/items/170566/view>.
- Bakker, M. , M. Chadwick & H. Sheahan et al. 2022, “Fine-Tuning Language Models to Find Agreement among Humans with Diverse Preferences.” <https://arxiv.org/abs/2211.15006>.
- Barrett, S. 2007, *Why Cooperate? The Incentive to Supply Global Public Goods*, Oxford: Oxford University Press.
- Bostrøm, N. 2014, *Super Intelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press.
- Buhmann, A. & C. Fieseler 2021, “Towards A Deliberative Framework for Responsible Innovation in Artificial Intelligence.” *Technology in Society* 64.
- Bull, H. 1977, *The Anarchical Society: A Study of Order in World Politics*, New York: Columbia University Press.
- Capoccia, G. & R. Kelemen 2007, “The Study of Critical Junctures: Theory, Narrative, and Counterfactuals in Historical Institutionalism.” *World Politics* 59(3).
- Coeckelbergh, M. 2021a, “AI for Climate: Freedom, Justice, and other Ethical and Political Challenges.” *AI and Ethics* 1.
- Coeckelbergh, M. 2021b, *Green Leviathan or the Poetics of Political Liberty: Navigating Free-*

- dom in the Age of Climate Change and Artificial Intelligence*, New York: Routledge.
- Coeckelbergh, M. 2022a, “Democracy, Epistemic Agency, and AI: Political Epistemology in Times of Artificial Intelligence,” <https://doi.org/10.1007/s43681-022-00239-4>.
- Coeckelbergh, M. 2022b, *The Political Philosophy of AI*, Cambridge: Polity.
- Collier, R. & D. Collier 1991a, “Critical Junctures and Historical Legacies.” *Americas* 51(1).
- Collier, R. & D. Collier 1991b, *Shaping the Political Arena*, Princeton: Princeton University Press.
- Collingridge, D. 1980, *The Social Control of Technology*, London: Frances Pinter.
- David, P. 1985, “Clio and the Economics of QWERTY.” *The American Economic Review* 75 (2).
- Deutsch, K. 1963, *The Nerves of Government: Models of Political Communication and Control*, New York: The Free Press.
- Dodge, J. , T. Prewitt & R. Combes et al. 2022, “Measuring the Carbon Intensity of AI in Cloud Instances.” 2022 ACM Conference on Fairness, Accountability, and Transparency.
- Easton, D. 1965, *A System Analysis of Political Life*, New York: John Wiley.
- Ellul, J. 1964, *The Technological Society*, New York: Vintage Books.
- Fioretos, O. , T. Falletti & A. Sheingate 2016, “Historical Institutionalism in Political Science.” in O. Fioretos, T. Falletti & A. Sheingate eds. , *The Oxford Handbook of Historical Institutionalism*, Oxford: Oxford University Press.
- Floridi, L. 2019, *The Green and the Blue: Naïve Ideas to Improve Politics in A Mature Information Society*, The 2018 Yearbook of the Digital Ethics Lab.
- Galaz, V. , M. Centeno & P. Callahan et al. 2021, “Artificial Intelligence, Systemic Risks, and Sustainability.” <https://doi.org/10.1016/j.techsoc.2021.101741>.
- Gaus, G. 2016, “The Tyranny of the Ideal.” in *The Tyranny of the Ideal*, Princeton: Princeton University Press.
- Hayek, F. 1960, *The Constitution of Liberty*, Chicago: The University of Chicago Press.
- Heilbroner, R. 1994, “Technological Determinism Revisited.” in L. Marx & M. Smith eds. , *Does Technology Drive History*, Cambridge, MA. : MIT Press.
- IPCC 2022, “Climate Change 2022: Impacts, Adaptation and Vulnerability.” <https://www.ipcc.ch/report/ar6/wg2/>.

- ipcc.ch/report/ar6/wg2.
- Jonsson, I. 2015, *The Political Economy of Innovation and Entrepreneurship: From Theories to Practice*, London: Routledge.
- Jonsson, I. 2016, “Economic Crisis and Real Critical Junctures: On the Decay of the Political Party System of Iceland.” *The Polar Journal* 6(1).
- Kahng, A. , S. Mackenzie & A. Procaccia et al. 2021, “Liquid Democracy: An Algorithmic Perspective.” *Journal of Artificial Intelligence Research* 70.
- Koster, R. , J. Balaguer & A. Tacchetti et al. 2022, “Human-Centred Mechanism Design with Democratic AI.” *Human Nature Behaviour* 6.
- Kurzweil, R. 2015, “Superintelligence and Singularity.” in S. Schneider ed., *Science Fiction and Philosophy: From Time Travel to Superintelligence*, New York: Wiley-Blackwell.
- Landemore, H. 2023, “Can AI Bring Deliberative Democracy to the Masses?” <https://www.law.nyu.edu/sites/default/files/Helen%20Landemore%20Can%20AI%20bring%20deliberative%20democracy%20to%20the%20masses.pdf>.
- Lenton, T. 2013, “Environmental Tipping Points.” *Annual Review of Environment and Resources* 38.
- Makitie, T. , J. Hanson & S. Damman et al. 2023, “Digital Innovation’s Contribution to Sustainability Transitions.” *Technology in Society* 73.
- Maynard, A. & S. Dudley 2023, “Navigating Advanced Technology Transitions: Using Lessons from Nanotechnology.” *Nature Nanotechnology* 18.
- Mills, S. & R. Whittle 2023, “Nudging Policy or Crowding It out? Green Nudges as Ideational Technologies.” in H. Sætra ed., *Technology & Sustainable Development: The Promise and Pitfalls of Techno-Solutionism*, London: Routledge.
- Muench, S. , E. Stoermer & K. Jensen et al. 2022, *Towards A Green & Digital Future*, Bruxelles: L. Publications Office of the European Union.
- Næss, A. 1989, *Ecology, Community and Lifestyle: Outline of An Ecosophy*, Cambridge: Cambridge University Press.
- OECD 2022, “Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications: The AI Footprint.” *OECD Digital Economy Papers* 341.

- Sadowski, J. & E. Selinger 2014, “Creating A Taxonomic Tool for Technocracy and Applying It to Silicon Valley.” *Technology in Society* 38.
- Sætra, H. 2020, “A Shallow Defence of A Technocracy of Artificial Intelligence: Examining the Political Harms of Algorithmic Governance in the Domain of Government.” <https://doi.org/10.1016/j.techsoc.2020.101283>. <https://www.sciencedirect.com/science/article/pii/S0160791X19305925>.
- Sætra, H. 2021, “A Typology of AI Applications in Politics.” https://doi.org/10.1007/978-3-030-88972-2_3.
- Sætra, H. 2022, “A Hobbesian Argument for World Government.” <https://doi.org/10.3390/philosophies7030066>.
- Sætra, H. 2023a, “Generative AI: Here to Stay, but for Good?” *Technology in Society* 75.
- Sætra, H. 2023b, *Technology and Sustainable Development: The Promise and Pitfalls of Techno-Solutionism*, London: Routledge.
- Satgar, V. 2021, *The Rise of Eco-Fascism*, in M. Williams & V. Satgar eds., *Destroying Democracy*, Johannesburg: Wits University Press.
- Savaget, P. , T. Chiarini & S. Evans 2018, “Empowering Political Participation Through Artificial Intelligence.” *Science and Public Policy* 46(3).
- Schatzberg, E. 2018, *Technology: Critical History of A Concept*, Chicago: University of Chicago Press.
- Schumpeter, J. 2013, *Capitalism, Socialism and Democracy*, London: Routledge.
- Wiener, N. 1948, *Cybernetics or Control and Communication in the Animal and the Machine*, Cambridge, MA. : MIT Press.
- Willis, R. , N. Curato & G. Smith 2022, “Deliberative Democracy and the Climate Crisis.” <https://doi.org/10.1002/wcc.759>.
- Winner, L. 1969, “Review: Cybernetics and Political Language. Review of the *Nerves of Government: Models of Political Communication and Control* by Karl W. Deutsch.” <https://www.jstor.org/stable/41035143>.
- Winner, L. 1977, *Autonomous Technology: Technics-out-of-Control as A Theme in Political Thought*, Cambridge, MA. : MIT Press.

Winner, L. 1979, "The Political Philosophy of Alternative Technology: Historical Roots and Present Prospects." *Technology in Society* 1(1).

Winner, L. 1980, "Do Artifacts Have Politics?" *Daedalus* 109(1).

Zuboff, S. 2019, *The Age of Surveillance Capitalism*, New York: Profile Books.